10 Orthogonality

10.1 Orthogonal subspaces

In the plane $\mathbf{R}^2$ we think of the coordinate axes as being orthogonal (perpendicular) to each other. We can express this in terms of vectors by saying that every vector in one axis is orthogonal to every vector in the other.

Let V be an inner product space and let S and T be subsets of V. We say that S and T are **orthogonal**, written $S \perp T$, if every vector in S is orthogonal to every vector in T:

$$S \perp T \iff \mathbf{s} \perp \mathbf{t} \text{ for all } \mathbf{s} \in S, \mathbf{t} \in T$$

10.1.1 Example In $\mathbf{R}^3$ let S be the x_1x_2-plane and let T be the x_3-axis. Show that $S \perp T$.

Solution Let $\mathbf{s} \in S$ and let $\mathbf{t} \in T$. We can write $\mathbf{s} = [x_1, x_2, 0]^T$ and $\mathbf{t} = [0, 0, x_3]^T$ so that

$$\langle \mathbf{s}, \mathbf{t} \rangle = \mathbf{s}^T\mathbf{t} = \begin{bmatrix} x_1 & x_2 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} = 0 .$$

Therefore, $\mathbf{s} \perp \mathbf{t}$ and we conclude that $S \perp T$. □

10.1.2 Example In $\mathbf{R}^3$ let S be the x_1x_2-plane and let T be the $x_1x_3$plane. Is it true that $S \perp T$?

Solution Although the planes S and T appear to be perpendicular in the informal sense, they are not so in the technical sense. For instance, the vector $\mathbf{e}_1 = [1, 0, 0]^T$ is in both S and T, yet

$$\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = \mathbf{e}_1^T\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1 \neq 0 ,$$

so $\mathbf{e}_1 \not\perp \mathbf{e}_1$. Viewing the first $\mathbf{e}_1$ in S and the second $\mathbf{e}_1$ in T, we see that $S \not\perp T$.

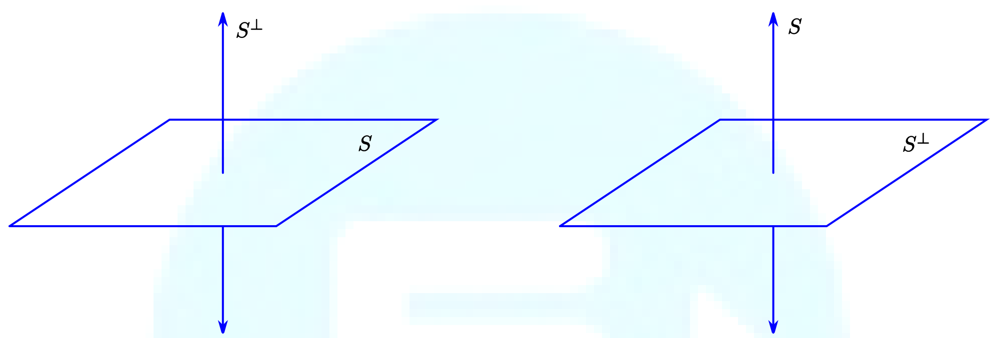
□

Orthogonal complement.

Let S be a subspace of V. The **orthogonal complement** of S (in V), written $S^{\perp}$, is the set of all vectors in V that are orthogonal to every vector in S:

$$S^{\perp} = \{\mathbf{v} \in V \mid \mathbf{v} \perp \mathbf{s} \text{ for all } \mathbf{s} \in S\}.$$

For instance, in $\mathbf{R}^3$ the orthogonal complement of the x_1x_2-plane is the x_3-axis. On the other hand, the orthogonal complement of the x_3-axis is the x_1x_2-plane.



It follows using the inner product axioms that if S is a subspace of V, then so is its orthogonal complement $S^{\perp}$.

The next theorem says that in order to show that a vector is orthogonal to a subspace it is enough to check that it is orthogonal to each vector in a spanning set for the subspace.

Theorem. *Let* $\{\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_n\}$ *be a set vectors in* V *and let* $S = \text{Span}\{\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_n\}$. *A vector* $\mathbf{v}$ *in* V *is in* $S^{\perp}$ *if and only if* $\mathbf{v} \perp \mathbf{b}_i$ *for each* i.

Proof. For simplicity of notation, we prove only the special case when $n = 2$ so that $S = \text{Span}\{\mathbf{b}_1,\mathbf{b}_2\}$. If $\mathbf{v}$ is in $S^{\perp}$, then it is orthogonal to *every* vector in S, so, in particular, $\mathbf{v} \perp \mathbf{b}_1$ and $\mathbf{v} \perp \mathbf{b}_2$. Now assume that $\mathbf{v} \in V$ and $\mathbf{v} \perp \mathbf{b}_1$ and $\mathbf{v} \perp \mathbf{b}_2$. If $\mathbf{s} \in S$, then we can write $\mathbf{s} = \alpha\mathbf{b}_1 + \alpha\mathbf{b}_2$ for some $\alpha_1,\alpha_2 \in \mathbf{R}$.

So using properties of the inner product we get

$$
\begin{aligned}
\langle \mathbf{v}, \mathbf{s} \rangle &= \langle \mathbf{v}, \alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2 \rangle && \text{(ii), (iii), and (iv)} \\
&= \alpha_1 \langle \mathbf{v}, \mathbf{b}_1 \rangle + \alpha_2 \langle \mathbf{v}, \mathbf{b}_2 \rangle && \mathbf{v} \perp \mathbf{b}_1 \text{ and } \mathbf{v} \perp \mathbf{b}_2 \\
&= \alpha_1 \cdot 0 + \alpha_2 \cdot 0 \\
&= 0.
\end{aligned}
$$

Therefore, $\mathbf{v} \perp \mathbf{s}$. This shows that **v** is orthogonal to every vector in S so that $\mathbf{v} \in S^{\perp}$. ☐

Theorem. *If* **A** *is a matrix, then the orthogonal complement of the row space of* **A** *is the null space of* **A***:* $(\text{Row}\mathbf{A})^{\perp} = \text{Null}\mathbf{A}$.

Proof. Let **A** be an $m \times n$ matrix and let $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m$ be the rows of **A** written as columns so that $\text{Row}\mathbf{A} = \text{Span}\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m\}$. Let **x** be a vector in $\mathbf{R}^n$. The product **Ax** is the $m \times 1$ matrix obtained by taking the dot products of **x** with the rows of **A**:

$$
\mathbf{Ax} = \begin{bmatrix} \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_m^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{b}_1^T \mathbf{x} \\ \vdots \\ \mathbf{b}_m^T \mathbf{x} \end{bmatrix}
$$

so, in view of the previous theorem, **x** is in $(\text{Row}\mathbf{A})^{\perp}$ if and only if $\mathbf{b}_i^T \mathbf{x} = 0$ for each i and this holds if and only if **x** is in Null**A**. ☐

10.1.3 Example Let S be the subspace of $\mathbf{R}^4$ spanned by the vectors $[1,0,-2,1]^T$ and $[0,1,3,-2]^T$. Find a basis for $S^{\perp}$. *Solution* We first note that $S = \text{Row}\mathbf{A}$, where

$$
\mathbf{A} = \begin{bmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 3 & -2 \end{bmatrix}.
$$

According to the theorem, $S^{\perp} = (\text{Row}\mathbf{A})^{\perp} = \text{Null}\mathbf{A}$ so we need only find a basis for the null space of A. The augmented matrix we write to solve $\mathbf{Ax} = \mathbf{0}$ is already in reduced row echelon form (RREF):

$$
\left[\begin{array}{cccc|c} 1 & 0 & -2 & 1 & 0 \\ 0 & 1 & 3 & -2 & 0 \end{array}\right]
$$

so $\text{Null}\mathbf{A} = \{[2t - s, -3t + 2s, t, s]^T \mid t, s \in \mathbf{R}\}$. We have

$$
\begin{bmatrix} 2t - s \\ -3t + 2s \\ t \\ s \end{bmatrix} = t \begin{bmatrix} 2 \\ -3 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix},
$$

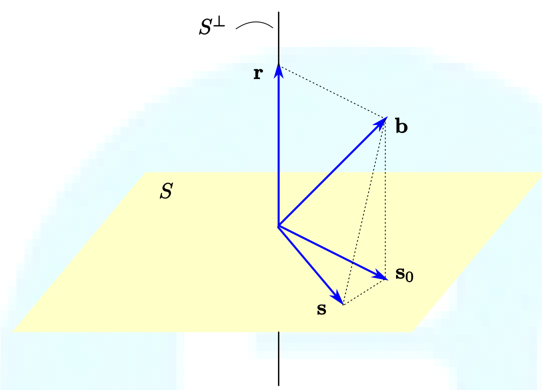
and therefore $\{[2, -3, 1, 0]^T, [-1, 2, 0, 1]^T\}$ is a basis for Null**A** (and hence $S^{\perp}$).

□

The next theorem generalizes the notion that the shortest distance from a point to a plane is achieved by dropping a perpendicular.

Theorem. *Let S be a subspace of V, let $\mathbf{b}$ be a vector in V and assume that $\mathbf{b} = \mathbf{s}_0 + \mathbf{r}$ with $\mathbf{s}_0 \in S$ and $\mathbf{r} \in S^{\perp}$. For every $\mathbf{s} \in S$ we have*

$$\mathrm{dist}(\mathbf{b}, \mathbf{s}_0) \le \mathrm{dist}(\mathbf{b}, \mathbf{s}).$$



Proof. Let $\mathbf{s} \in S$. We have

$$\|\mathbf{b} - \mathbf{s}\|^2 = \|(\mathbf{b} - \mathbf{s}_0) + (\mathbf{s}_0 - \mathbf{s})\|^2$$

$$= \|\mathbf{b} - \mathbf{s}_0\|^2 + \|\mathbf{s}_0 - \mathbf{s}\|^2 \qquad \text{Pythagorean theorem}$$

$$\ge \|\mathbf{b} - \mathbf{s}_0\|^2$$

(note that the Pythagorean theorem applies since $\mathbf{b}-\mathbf{s}_0 = \mathbf{r} \in S^{\perp}$ and $\mathbf{s}_0-\mathbf{s} \in S$ so that $(\mathbf{b} - \mathbf{s}_0) \perp (\mathbf{s}_0 - \mathbf{s})$). Therefore,

$$\mathrm{dist}(\mathbf{b},\mathbf{s}_0) = \|\mathbf{b} - \mathbf{s}_0\| \le \|\mathbf{b} - \mathbf{s}\| = \mathrm{dist}(\mathbf{b},\mathbf{s}).$$

□

10.2 Least squares

Suppose that the matrix equation $\mathbf{Ax} = \mathbf{b}$ has no solution. In terms of distance, this means that $\mathrm{dist}(\mathbf{b},\mathbf{Ax})$ is never zero, no matter what $\mathbf{x}$ is.

Instead of leaving the equation unsolved, it is sometimes useful to find an $\mathbf{x}_0$ that is as close to being a solution as possible, that is, for which the distance from **b** to $\mathbf{Ax}_0$ is less than or equal to the distance from **b** to **Ax** for every other **x**. This is called a "least squares solution."

A least squares solution can be used, for instance, to find the line that best fits some data points (see Example 10.2.1).

> Least squares.
>
> Let **A** be an $m \times n$ matrix and let **b** be a vector in $\mathbf{R}^n$. If $\mathbf{x} = \mathbf{x}_0$ is a solution to
>
> $$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}.$$
>
> then, for every $\mathbf{x} \in \mathbf{R}^n$,
>
> $$\text{dist}(\mathbf{b},\mathbf{Ax}_0) \le \text{dist}(\mathbf{b},\mathbf{Ax}).$$
>
> Such an $\mathbf{x}_0$ is called a **least squares solution** to the equation $\mathbf{Ax} = \mathbf{b}$.

Proof. Let $\mathbf{x}_0 \in \mathbf{R}^n$ and assume that $\mathbf{A}^T\mathbf{Ax}_0 = \mathbf{A}^T\mathbf{b}$. Letting $\mathbf{s}_0 = \mathbf{Ax}_0$, we have

$$\mathbf{A}^T(\mathbf{b} - \mathbf{s}_0) = \mathbf{A}^T(\mathbf{b} - \mathbf{Ax}_0) = \mathbf{0},$$

so that

$$\begin{aligned} \mathbf{b} - \mathbf{s}_0 &\in \text{Null}\mathbf{A}^T \\ &= (\text{Row}\mathbf{A}^T)^\perp \qquad \text{(Theorem in Section 10.1)} \\ &= (\text{Col}\mathbf{A})^\perp \\ &= S^\perp, \end{aligned}$$

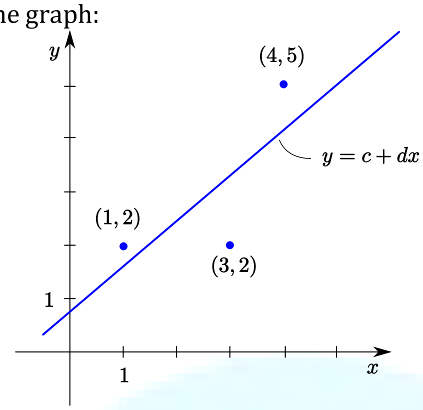
where $S = \{\mathbf{Ax} | \mathbf{x} \in \mathbf{R}^n\}$ (the product **Ax** can be interpreted as the linear combination of the columns of **A** with the entries of **x** as scalar factors, so S is the set of all linear combinations of the columns of **A**, which is Col**A**). This shows that the vector $\mathbf{r} = \mathbf{b} - \mathbf{s}_0$ is in $S_\perp$, so that $\mathbf{b} = \mathbf{s}_0 + \mathbf{r}$ with $\mathbf{s}_0 \in S$ and $\mathbf{r} \in S_\perp$. By the last theorem of Section 10.1, for every $\mathbf{x} \in \mathbf{R}^n$, we have

$$\text{dist}(\mathbf{b},\mathbf{Ax}_0) = \text{dist}(\mathbf{b},\mathbf{s}_0) \le \text{dist}(\mathbf{b},\mathbf{Ax}).$$

□

10.2.1 Example Use a least squares solution to find a line that best fits the data points (1,2), (3,2), and (4,5).

Solution Here is the graph:



If we write the desired line as $c+dx = y$, then ideally the line would go through all three points giving the system

$c + d = 2$ c
$+ 3d = 2$ c
$+ 4d = 5,$

which can be written as the matrix equation $\mathbf{Ax} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}.$$

The least squares solution is obtained by solving the equation $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$. We have

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 26 \end{bmatrix}$$

and

$$\mathbf{A}^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 9 \\ 28 \end{bmatrix},$$

so the matrix equation has corresponding augmented matrix

8 | 2628) 3∼ 21 3 | $\begin{array}{cc|c} 3 & 8 & 9 \\ 0 & 14 & 12 \end{array}$

89 | −8

7

∼ $\begin{array}{cc|c} 3 & 8 & 9 \\ 0 & 7 & 6 \end{array}$ −8)

1 $\begin{array}{cc|c} 21 & 0 & 15 \\ 0 & 7 & 6 \end{array}$

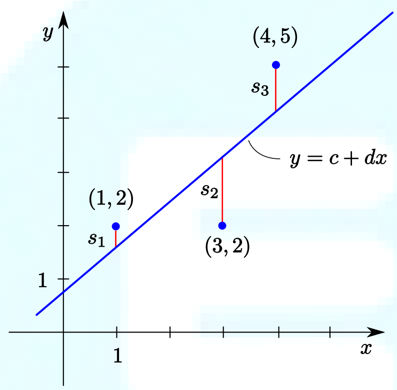
∼ 21 $\begin{array}{cc|c} 1 & 0 & \frac{5}{7} \\ 0 & 1 & \frac{6}{7} \end{array}$ 17

Therefore, $c = \frac{5}{7}$ and $d = \frac{6}{7}$ and the best fitting line is $y = \frac{5}{7} + \frac{6}{7}x$, which is the line shown in the graph.

(We could tell in advance that the matrix equation **Ax** = **b** has no solution since the points are not collinear. In general, however, it is not necessary to first check that an equation has no solution before applying the least squares method since, if it has a solution, then that is what the method will produce.)

□

We can use the last example to explain the terminology "least squares." The method gives a line that minimizes the sum of the squares of the vertical distances from the points to the lines. This sum of squares is $s_1^2 + s_2^2 + s_3^2$, with s_1, s_2, and s_3 as indicated:



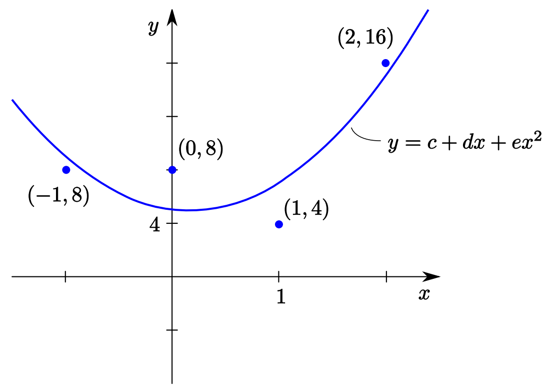
We have

$$\begin{aligned} s_1^2 + s_2^2 + s_3^2 &= \left(2 - (c + d)\right)^2 + \left(2 - (c + 3d)\right)^2 + \left(5 - (c + 4d)\right)^2 \\ &= \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2 \\ &= \operatorname{dist}(\mathbf{b}, \mathbf{A}\mathbf{x})^2 \end{aligned}$$

A least squares solution $\mathbf{x} = [c,d]^T$ makes dist(**b**,**Ax**) as small as possible and therefore makes $s_1^2 + s_2^2 + s_3^2$ as small as possible as well.

10.2.2 Example Use a least squares solution to find a parabola that best fits the data points (−1,8), (0,8), (1,4), and (2,16).

Solution Here is the graph:



Ideally, a parabola would go through all four points, so if we write its equation as $c + dx + ex^2 = y$, then

$c - d + e = 8\ c$

$= 8$

$c + d + e = 4$

$c + 2d + 4e = 16,$

which can be written as the matrix equation $\mathbf{Ax} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ d \\ e \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 8 \\ 8 \\ 4 \\ 16 \end{bmatrix}.$$

The least squares solution is obtained by solving the equation $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$. We have

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix}$$

and

8

□ $_T$1 1 □ 1 1□ □ □36□

8

A b = □−11 00 1124□□□□ 4 □□□ = □2876□,

16

so the matrix equation has corresponding augmented matrix

  4 2 636 $\frac{1}{2}$ 2 1 3 − 3 18 1

2

 2 6 828 $2\frac{1}{1}2$~ 13 34 94 1438 −2

6 8 1876 1

2 1 3 18

~ 00 −55 −95 −2210

2 1 3 18

~ 00 −05 −45 −1210 −$5\frac{1}{1}4$

2 1 3 18

~ 00 10 11 23 − 3 −1

2 1 0 9

~ 00 01 10 −31 −1

2 0 0 10 $\frac{1}{2}$

~ 00 10 01 −31

1 0 0 5

~ 00 01 01 −31

yTherefore,= 5 − x + 3c = 5x^2, which is the parabola shown in the graph., d = −1, and e = 3, and the best fitting parabola has equation

□

10.3 Gram-Schmidt process

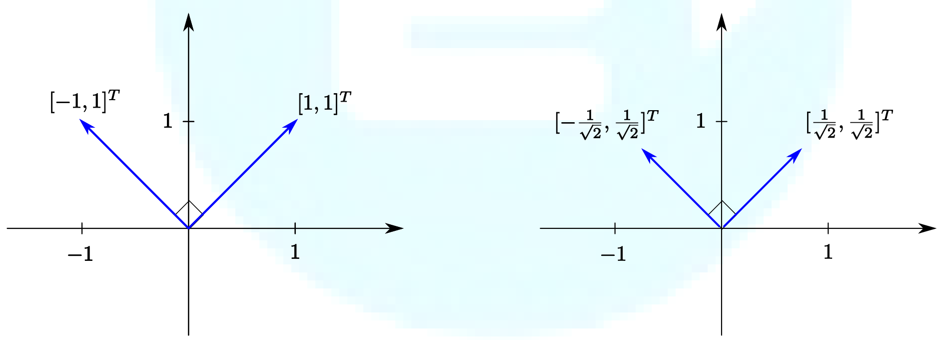
The standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ in $\mathbf{R}^n$ have two useful properties:

- they are pairwise orthogonal (any two are orthogonal),
- each is a unit vector (a vector of norm one).

If we have a basis for an inner product space V that does not already have these properties, then we can change it into a basis that does by using the Gram-Schmidt process.

> Orthonormal set.
>
> Let $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_s\}$ be a set of vectors in the inner product space V. The set is **orthogonal** if $\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0$ for all $i \neq j$ (the vectors are pairwise orthogonal).
>
> The set is **orthonormal** if it is orthogonal and each vector is a unit vector.

Any orthogonal set of nonzero vectors can be changed into an orthonormal set by dividing each vector by its norm. For instance, $\{[1,1]^T, [-1,1]^T\}$ is an orthogonal set in $\mathbf{R}^2$ and each of these vectors has norm $\sqrt{2}$, so $\{[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T, [-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T\}$ is an orthonormal set:

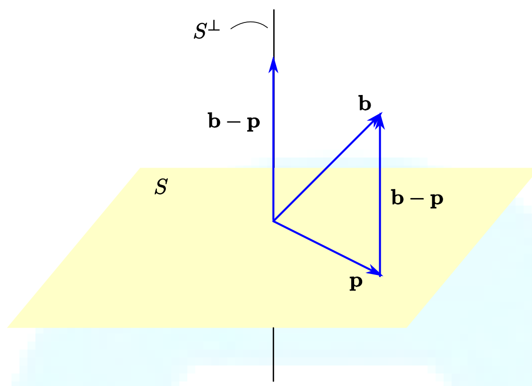


Let V be an inner product space. The Gram-Schmidt process requires a formula for the projection of a vector on a subspace S.

Theorem. *Let S be a subspace of V and let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_s\}$ be an orthonormal basis for S. Let $\mathbf{b}$ be a vector in V and let*

$$\mathbf{p} = \sum_{i=1}^{s} \langle \mathbf{b}, \mathbf{u}_i \rangle \mathbf{u}_i.$$

Then $\mathbf{p} \in S$ and $\mathbf{b} - \mathbf{p} \in S^{\perp}$:



The vector $\mathbf{p}$ is the projection *of $\mathbf{b}$ on S.*

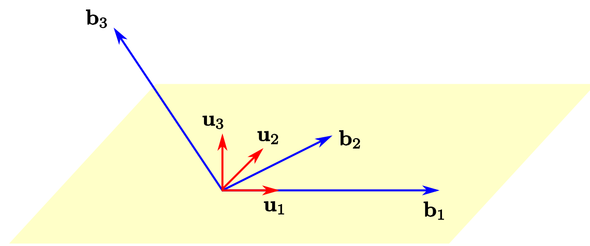
Proof. We prove only the special case $s = 2$ so that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is a basis for S and $\mathbf{p} = \langle \mathbf{b}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{b}, \mathbf{u}_2 \rangle \mathbf{u}_2$.

Since $\mathbf{p}$ is a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$, it is in S. We have

$$\begin{aligned}
\langle \mathbf{b} - \mathbf{p}, \mathbf{u}_1 \rangle &= \langle \mathbf{b} - (\langle \mathbf{b}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{b}, \mathbf{u}_2 \rangle \mathbf{u}_2), \mathbf{u}_1 \rangle \\
&= \langle \mathbf{b}, \mathbf{u}_1 \rangle - \langle \mathbf{b}, \mathbf{u}_1 \rangle \langle \mathbf{u}_1, \mathbf{u}_1 \rangle - \langle \mathbf{b}, \mathbf{u}_2 \rangle \langle \mathbf{u}_2, \mathbf{u}_1 \rangle && \text{(iii) and (iv)} \\
&= \langle \mathbf{b}, \mathbf{u}_1 \rangle - \langle \mathbf{b}, \mathbf{u}_1 \rangle (1) - \langle \mathbf{b}, \mathbf{u}_1 \rangle (0) && \mathbf{u}_1 \text{ is unit vector and } \mathbf{u}_1 \perp \mathbf{u}_2 \\
&= 0,
\end{aligned}$$

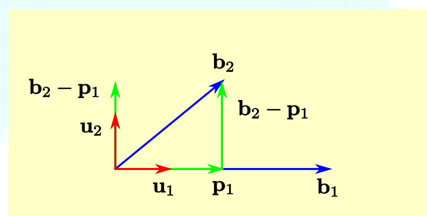
so $\mathbf{b} - \mathbf{p}$ is orthogonal to $\mathbf{u}_1$. Similarly, $\mathbf{b} - \mathbf{p}$ is orthogonal to $\mathbf{u}_2$. Since $S = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, it follows from the theorem in Section 10.1 that $\mathbf{b} - \mathbf{p}$ is in $S^{\perp}$. □

We now describe the Gram-Schmidt process in the case of three initial vectors. Suppose that $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is a basis for an inner product space V. The GramSchmidt process uses these vectors to produce an *orthonormal* basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for V:



Here are the first two steps of the process (the view is looking straight down at the plane spanned by $\mathbf{b}_1$ and $\mathbf{b}_2$ and we write $\frac{\mathbf{v}}{\|\mathbf{v}\|}$ to mean $\frac{1}{\|\mathbf{v}\|}\mathbf{v}$):

$$\mathbf{u}_1 = \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|}$$



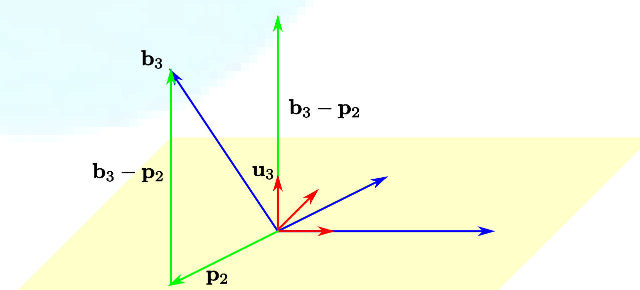
$$\mathbf{u}_2 = \frac{\mathbf{b}_2 - \mathbf{p}_1}{\|\mathbf{b}_2 - \mathbf{p}_1\|}$$

where $\mathbf{p}_1 = \langle \mathbf{b}_2, \mathbf{u}_1 \rangle \mathbf{u}_1$

Here is the third step of the process:

$$\mathbf{u}_3 = \frac{\mathbf{b}_3 - \mathbf{p}_2}{\|\mathbf{b}_3 - \mathbf{p}_2\|}$$

where $\mathbf{p}_2 = \langle \mathbf{b}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{b}_3, \mathbf{u}_2 \rangle \mathbf{u}_2$



The general statement is as follows:

Gram-Schmidt process.

Let $\{\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_n\}$ be a basis for the inner product space V.

Define vectors $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n$ recursively by

$$\mathbf{u}_1 = \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|}$$

$$\mathbf{u}_k = \frac{\mathbf{b}_k - \mathbf{p}_{k-1}}{\|\mathbf{b}_k - \mathbf{p}_{k-1}\|} \quad \text{where } \mathbf{p}_{k-1} = \sum_{i=1}^{k-1} \langle \mathbf{b}_k, \mathbf{u}_i \rangle \mathbf{u}_i \quad (k > 1)$$

Then $\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n\}$ is an orthonormal basis for V. Moreover, $\mathrm{Span}\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k\} = \mathrm{Span}\{\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_k\}$ for each k.

We will not give a detailed proof. However, we note that for each k the vector $\mathbf{b}_k - \mathbf{p}_{k-1}$ is in the orthogonal complement of $\mathrm{Span}\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_{k-1}\}$ by the preceding theorem so that $\mathbf{u}_k \perp \mathbf{u}_i$ for $1 \le i < k$.

10.3.1 Example Let $\mathbf{b}_1 = [1,2,2,4]^T$, $\mathbf{b}_2 = [-2,0,-4,0]^T$, and $\mathbf{b}_3 = [-1,1,2,0]^T$, and let S be the span of these vectors. Apply the Gram-Schmidt process to $\{\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3\}$ to obtain an orthonormal basis $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ for S.

Solution Since the given vectors are in $\mathbf{R}^4$, the inner product is the dot product and the norm of a vector $\mathbf{x}$ is given by the formula

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + x_3^2 + x_4^2}.$$

We use the formula $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ to simplify computations.

First,

$$\mathbf{u}_1 = \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|} = \frac{[1,2,2,4]^T}{\|[1,2,2,4]^T\|} = \tfrac{1}{5}[1,2,2,4]^T.$$

Next,

$$\mathbf{p}_1 = \langle \mathbf{b}_2, \mathbf{u}_1 \rangle \mathbf{u}_1 = \left\langle [-2,0,-4,0]^T, \tfrac{1}{5}[1,2,2,4]^T \right\rangle \mathbf{u}_1 = -2\mathbf{u}_1 = -\tfrac{2}{5}[1,2,2,4]^T \text{ and}$$

$$\mathbf{b}_2 - \mathbf{p}_1 = [-2,0,-4,0]^T + \tfrac{2}{5}[1,2,2,4]^T = \tfrac{4}{5}[-2,1,-4,2]^T,$$

so

$$\mathbf{u}_2 = \frac{\mathbf{b}_2 - \mathbf{p}_1}{\|\mathbf{b}_2 - \mathbf{p}_1\|} = \frac{\frac{4}{5}[-2,1,-4,2]^T}{\|\frac{4}{5}[-2,1,-4,2]^T\|} = \tfrac{1}{5}[-2,1,-4,2]^T.$$

Finally,

$$\begin{aligned}\mathbf{p}_2 &= \langle \mathbf{b}_3, \mathbf{u}_1\rangle \mathbf{u}_1 + \langle \mathbf{b}_3, \mathbf{u}_2\rangle \mathbf{u}_2 \\ &= \langle [-1,1,2,0]^T, \tfrac{1}{5}[1,2,2,4]^T\rangle \mathbf{u}_1 + \langle [-1,1,2,0]^T, \tfrac{1}{5}[-2,1,-4,2]^T\rangle \mathbf{u}_2 \\ &= \mathbf{u}_1 - \mathbf{u}_2 = \tfrac{1}{5}[1,2,2,4]^T - \tfrac{1}{5}[-2,1,-4,2]^T = \tfrac{1}{5}[3,1,6,2]^T\end{aligned}$$

and

$$\mathbf{b}_3 - \mathbf{p}_2 = [-1,1,2,0]^T - \tfrac{1}{5}[3,1,6,2]^T = \tfrac{2}{5}[-4,2,2,-1]^T,$$

so

$$\mathbf{u}_3 = \frac{\mathbf{b}_3 - \mathbf{p}_2}{\|\mathbf{b}_3 - \mathbf{p}_2\|} = \frac{\frac{2}{5}[-4,2,2,-1]^T}{\|\frac{2}{5}[-4,2,2,-1]^T\|} = \tfrac{1}{5}[-4,2,2,-1]^T.$$

Therefore, $\{\tfrac{1}{5}[1,2,2,4]^T,\ \tfrac{1}{5}[-2,1,-4,2]^T,\ \tfrac{1}{5}[-4,2,2,-1]^T\}$ is an orthonormal basis for *S*. □

10.3.2 Example Let *S* be the span of the functions 1, *x*, and x^2 in $\mathbf{C}_{[0,1]}$. Apply the Gram-Schmidt process to $\{1, x, x^2\}$ to obtain an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for *S*.

Solution The inner product in $\mathbf{C}_{[0,1]}$ is given by $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$. Let $\mathbf{b}_1 = 1$, $\mathbf{b}_2 = x$, and $\mathbf{b}_3 = x^2$.

First,

$$\|\mathbf{b}_1\| = \sqrt{\langle \mathbf{b}_1, \mathbf{b}_1\rangle} = \sqrt{\langle 1, 1\rangle} = \sqrt{\int_0^1 1\,dx} = 1,$$

so

$$\mathbf{u}_1 = \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|} = 1.$$

Next,

$$\mathbf{p}_1 = \langle \mathbf{b}_2, \mathbf{u}_1\rangle \mathbf{u}_1 = \langle x, 1\rangle \mathbf{u}_1 = \left(\int_0^1 x\,dx\right)\mathbf{u}_1 = \tfrac{1}{2}\mathbf{u}_1 = \tfrac{1}{2}$$

and

$$\mathbf{b}_2 - \mathbf{p}_1 = x - \tfrac{1}{2},$$

so

$$\mathbf{u}_2 = \frac{\mathbf{b}_2 - \mathbf{p}_1}{\|\mathbf{b}_2 - \mathbf{p}_1\|} = \frac{x - \frac{1}{2}}{\|x - \frac{1}{2}\|} = \frac{x - \frac{1}{2}}{\sqrt{\int_0^1 (x - \frac{1}{2})^2\,dx}} = 2\sqrt{3}(x - \tfrac{1}{2}) = \sqrt{3}(2x - 1).$$

Finally,

$$\mathbf{p}_2 = \langle \mathbf{b}_3, \mathbf{u}_1\rangle \mathbf{u}_1 + \langle \mathbf{b}_3, \mathbf{u}_2\rangle \mathbf{u}_2 = \langle x^2, 1\rangle \mathbf{u}_1 + \langle x^2, \sqrt{3}(2x-1)\rangle \mathbf{u}_2$$

$$= \left(\int_0^1 x^2\,dx\right)\mathbf{u}_1 + \left(\sqrt{3}\int_0^1 x^2(2x-1)\,dx\right)\mathbf{u}_2 = \tfrac{1}{3}\mathbf{u}_1 + \tfrac{\sqrt{3}}{6}\mathbf{u}_2$$

$$= \tfrac{1}{3}(1) + \tfrac{\sqrt{3}}{6}\left(\sqrt{3}(2x-1)\right) = x - \tfrac{1}{6}$$

and

$$\mathbf{b}_3 - \mathbf{p}_2 = x^2 - (x - \tfrac{1}{6}) = x^2 - x + \tfrac{1}{6},$$

so $\mathbf{u}_3 = \dfrac{\mathbf{b}_3 - \mathbf{p}_2}{\|\mathbf{b}_3 - \mathbf{p}_2\|} = \dfrac{x^2 - x + \frac{1}{6}}{\sqrt{\int_0^1 (x^2 - x + \frac{1}{6})^2\,dx}}$

$$= 6\sqrt{5}(x^2 - x + \tfrac{1}{6}) = \sqrt{5}(6x^2 - 6x + 1).$$

Therefore $\{1,\ \sqrt{3}(2x-1),\ \sqrt{5}(6x^2-6x+1)\}$ is an orthonormal basis for S. □

10–Exercises

10–1 Let $\mathbf{x}_1 = [-1,0,1]^T$, $\mathbf{x}_2 = [0,1,-1]^T$, and $\mathbf{x}_3 = [1,1,1]^T$. Show that $S \perp T$, where $S = \mathrm{Span}\{\mathbf{x}_1,\mathbf{x}_2\}$ and $T = \mathrm{Span}\{\mathbf{x}_3\}$.

10–2 Let S be the subspace of $\mathbf{P}_3$ spanned by x and x^2. Find $S^\perp$, where the inner product on $\mathbf{P}_3$ is as in Section 9.5 with $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$.

Hint: By the first theorem of Section 10.1 a polynomial $p = ax^2 + bx + c$ is in $S^\perp$ if and only if it is orthogonal to both x and x^2.

10–3 Let S be the subspace of $\mathbf{R}^4$ spanned by the vectors $[1,3,-4,0]^T$ and $[-2,-6,8,1]^T$.

(a) Find a basis for $S^{\perp}$.

(b) Verify that the basis vectors you found in (a) are orthogonal to the given vectors.

10–4 Use a least squares solution to find a line that best fits the data points (1,2), (2,4), and (3,3).

10–5 Use a least squares solution to find a curve of the form $y = a\cos x + b\sin x$ that best fits the data points (0,1), (π/2,2), and (π,0).

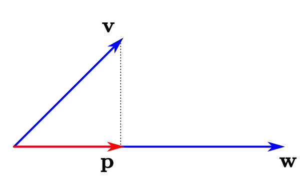
Answer: $y = \frac{1}{2}\cos x + 2\sin x$.

10–6 Let $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ be an orthonormal set in an inner product space V. Prove that $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ are linearly independent.

Hint: Suppose that $\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 = \mathbf{0}$. Apply h—,$\mathbf{u}_1$i to both sides of this equation and use properties of the inner product to conclude that $\alpha_1 = 0$. Note that similarly $\alpha_2 = 0$ and $\alpha_3 = 0$. You may use the fact that h$\mathbf{0}$,$\mathbf{v}$i = 0 for every vector $\mathbf{v} \in V$.

10–7 Let $\mathbf{v}$ and $\mathbf{w}$ be vectors in an inner product space V and assume that $\mathbf{w}$ is nonzero. Let

$$\mathbf{p} = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\langle \mathbf{w}, \mathbf{w} \rangle}\mathbf{w}$$



The vector $\mathbf{p}$ is the **projection of v on w.**

(a) Show that $(\mathbf{v} - \mathbf{p}) \perp \mathbf{w}$.

(b) Find $\mathbf{p}$ when $\mathbf{v} = [1,2]^T$ and $\mathbf{w} = [3,1]^T$ and sketch.

10–8 Let $\mathbf{b}_1 = [0,0,-1,1]^T$, $\mathbf{b}_2 = [1,0,0,1]^T$ and $\mathbf{b}_3 = [1,0,-1,0]^T$, and let S be the span of these vectors in $\mathbf{R}^4$. Apply the Gram-Schmidt process to $\{\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3\}$ to obtain an orthonormal basis $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ for S.

Answer: $\{\frac{1}{\sqrt{2}}[0,0,-1,1]^T,\ \frac{1}{\sqrt{6}}[2,0,1,1]^T,\ \frac{1}{\sqrt{3}}[1,0,-1,-1]^T\}$.

10–9 Let S be the span of the functions x, $x+1$, and x^2-1 in $\mathbf{C}_{[0,1]}$. Apply the Gram-Schmidt process to $\{x, x+1, x^2-1\}$ to obtain an orthonormal basis $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ for S.

Answer: $\{\sqrt{3}x,\ -3x+2,\ \sqrt{5}(6x^2-6x+1)\}$.