# MATH 2101 LINEAR ALGEBRA I, FALL SEMESTER 2023

## 1. Inner product spaces

Key concepts in this section:

- What is an inner product on a vector space?
- Examples of an inner product on a vector space
- Orthonormal basis and Gram-Schmidt process
- Dimension formula for orthogonal complements
- Application: Least square apporximation (if time permits)

### 1.1. **Inner product.**

**Definition 1.1.** Let $V$ be a vector space. An inner product $\langle , \rangle$ on $V$ is a function that assigns a scalar in $\mathbb{R}$ to every pair $(x, y) \in V \times V$ satisfying the following properties:

(a) For any $x, y, z \in V$, $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.
(b) For any $x, y \in V$ and $c \in \mathbb{R}$, $\langle cx, y\rangle = c\langle x, y\rangle$.
(c) For any $x, y \in V$, $\langle x, y\rangle = \langle y, x\rangle$.
(d) For any $x \in V$, $\langle x, x\rangle > 0$ if $x \neq 0$.

A vector space equipeed with an inner product is called an *inner product space*, and we usually denote the inner product by $\langle ., .\rangle$.

We first deduce some immediate consequences:

**Proposition 1.2.** *Let $V$ be an inner product space. Then*

(a) *For any $x, y, z \in V$, $\langle z, x + y\rangle = \langle z, x\rangle + \langle z, y\rangle$.*
(b) *For any $x, y \in V$ and $c \in \mathbb{R}$, $\langle x, cy\rangle = c\langle x, y\rangle$.*
(c) *For any $x \in V$, $\langle 0, x\rangle = \langle x, 0\rangle = 0$.*
(d) *For any $x \in V$, $\langle x, x\rangle = 0$ if and only if $x = 0$.*

*Proof.* The key ideas for (a) and (b) are to transfer the problem to the one in Definition 1.1. For (a),

$$\langle z, x + y\rangle = \langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle = \langle z, x\rangle + \langle z, x\rangle$$

For (b), $\langle x, cy\rangle = \langle cy, x\rangle = c\langle y, x\rangle = c\langle x, y\rangle$. For (c), we use that $0 = 0y$ for any $y \in V$. Hence, we have:

$$\langle 0, x\rangle = \langle 0y, x\rangle = 0\langle y, x\rangle = 0.$$

For (d), if $x \neq 0$, then $\langle x, x\rangle > 0$ by Definition 1.1(d). If $x = 0$, it follows from part (c). $\square$

**Example 1.3.** Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, define

$$\langle x, y\rangle = x_1y_1 + \dots + x_ny_n.$$

Check $\langle , \rangle$ defines an inner product on $\mathbb{R}^n$. This is called a **standard inner product** on $\mathbb{R}^n$ (which we may also sometimes use for column vectors rather than row vectors). When $n = 2$ and $n = 3$, it coincides with the dot product between two vectors.

*Sol:* Let $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ and $z = (z_1, \dots, z_n)$. Let $c \in \mathbb{R}$.
Check (a): $\langle x + y, z\rangle = (x_1 + y_1)z_1 + \dots + (x_n + y_n)z_n = (x_1z_1 + \dots + x_nz_n) + (y_1z_1 + \dots + y_nz_n) = \langle x, z\rangle + \langle y, z\rangle$.
Checking (b) and (c) are straightforward.
Check (d): $\langle x, x\rangle = x_1^2 + \dots + x_n^2 > 0$ for $x \neq 0$.

**Exercise 1.4.** Let $A$ be an $n \times m$-matrix. Let $x \in \mathbb{R}^m$ be a column vector and let $y \in \mathbb{R}^n$ be a column vector. Then,

$$\langle Ax, y\rangle = \langle x, A^Ty\rangle.$$

*Sol:* Recall that $\langle x, y\rangle = x^Ty$ for any $x, y \in \mathbb{R}^m$. Then

$$\langle Ax, y\rangle = (Ax)^Ty = x^TA^Ty = x^T(A^Ty) = \langle x, A^Ty\rangle.$$

**Example 1.5.** Let $\langle , \rangle$ be an inner product on a vector space $V$. Let $r$ be a *positive* real number. Define a new pairing:

$$\langle , \rangle' : V \times V \to \mathbb{R}$$

given by $\langle x, y\rangle' = r\langle x, y\rangle$. Then $\langle , \rangle'$ also defines an inner product.

**Example 1.6.** Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $(y_1, \dots, y_n) \in \mathbb{R}^n$. Let $a_1, \dots, a_n > 0$. Define

$$\langle x, y\rangle = a_1x_1y_1 + \dots + a_nx_ny_n.$$

This defines another inner product on $\mathbb{R}^n$.

We also have some non-examples:

**Example 1.7.** For $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$, define

$$\langle (x_1, x_2), (y_1, y_2)\rangle = x_1y_1 - x_2y_2.$$

Determine if $\langle ., .\rangle$ defines an inner product.
*Sol:* Note that $\langle (0, 1), (0, 1)\rangle = -1 < 0$ which violates Definition 1.1(d).

We also have an inner product on other type of examples:

**Example 1.8.** Let $a, b \in \mathbb{R}$. Let $V$ be the vector space of integrable functions from $[a, b]$ to $\mathbb{R}$. For $f, g \in V$, define

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx.$$

Then $\langle , \rangle$ defines an inner product on $V$.

**Exercise 1.9.** Let $T : V \to W$ be an injective linear transformation. Let $\langle ., .\rangle'$ be an inner product on $W$. For $x, y \in V$, define

$$\langle x, y\rangle = \langle T(x), T(y)\rangle'.$$

Show that $\langle ., .\rangle$ is an inner product.
*Sol:* For $x, y, z \in V$, define

$$\langle x+y, z\rangle = \langle T(x+y), T(z)\rangle' = \langle T(x)+T(y), T(z)\rangle' = \langle T(x), T(z)\rangle'+\langle T(y), T(z)\rangle' = \langle x, z\rangle+\langle y, z\rangle.$$

This checks (a).

For $x, y \in V$ and $c \in \mathbb{R}$, define

$$\langle cx, y\rangle = \langle T(cx), T(y)\rangle' = \langle cT(x), T(y)\rangle' = c\langle T(x), T(y)\rangle' = c\langle x, y\rangle.$$

This checks (b).

For $x, y \in V$,

$$\langle x, y\rangle = \langle T(x), T(y)\rangle' = \langle T(y), T(x)\rangle' = \langle y, x\rangle.$$

This checks (c).

For $x \in V$ and $x \neq 0$, $T(x) \neq 0$ since $T$ is injective. Then $\langle x, x\rangle = \langle T(x), T(x)\rangle' = 0$. This checks (d).

**Exercise 1.10.** What happens for Exercise 1.9 if one drops the condition that $T$ is injective?
*Sol:* In such case, $\langle ., .\rangle'$ is not necessarily an inner product. This is because if $T$ is not injective, one may find a non-zero $x$ such that $T(x) = 0$. In such case, $\langle x, x\rangle = \langle T(x), T(x)\rangle' = 0$. This violates Definition 1.1(d).

We also have a more interesting example:

**Example 1.11.** Let $M_{n\times n}$ be the vector space of $n \times n$ matrix. The trace of the matrix $A$ is defined as:

$$\mathrm{tr}(A) = A_{11} + A_{22} + \ldots + A_{nn}.$$

For $A, B \in M_{n\times n}$, define

$$\langle A, B\rangle = \mathrm{tr}(A^T B).$$

Show that $\langle .,. \rangle$ is an inner product.
*Sol:* To show this, one considers

$$(A^T B)_{ii} = A_{1i}B_{1i} + A_{2i}B_{2i} + \ldots + A_{ni}B_{ni}.$$

Then,

$$\langle A, B \rangle = \sum_{i,j=1}^{n} A_{ij}B_{ij}.$$

One may check that $\langle .,. \rangle$ is an inner product from definition. Here, we give a better explanation.

We now give another way to do that via an invertible linear transformation. Define $T : M_{n\times n} \to \mathbb{R}^{n\times n}$ given by:

$$T(A) = (A_{11}, A_{12}, \ldots, A_{1n}, A_{21}, \ldots, A_{2n}, \ldots, A_{n1}, \ldots, A_{nn}).$$

Indeed, $T$ is an isomorphism (recall the definition from the last chapter!).

Let $\langle .,. \rangle'$ be the standard inner product on $\mathbb{R}^{n\times n}$. Then one has, for any $A, B \in M_{2\times 2}$,

$$\langle A, B \rangle = \langle T(A), T(B) \rangle'.$$

We have checked in Exercise 1.9 that $\langle .,. \rangle$ is an inner product.

We now use the concept of inner products to generalize the length:

**Definition 1.12.** Let $V$ be an inner product space. For $x$, define $||x|| = \sqrt{\langle x, x \rangle}$, which is called the *length* of $x$.

**Example 1.13.** Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then

$$||x|| = \sqrt{x_1^2 + \ldots + x_n^2}.$$

**Exercise 1.14.** Let $V$ be an inner product space. Let $T : V \to V$ be a linear transformation. Suppose $||T(x)|| = ||x||$ for all $x$. Show that $T$ is injective.
*Sol:* Suppose $T(x) = 0$. Then $||x|| = ||T(x)|| = ||0|| = 0$. This implies that $x = 0$ by Proposition 1.2(d). Hence, $T$ is injective.

We have some more properties of the inner product

**Theorem 1.15.** *Let $V$ be an inner product space. Then, for all $x, y \in V$ and $c \in \mathbb{R}$,*

(a) $||cx|| = |c|||x||$;
(b) $||x|| = 0$ *if and only if* $x = 0$;
(c) *(Cauchy-Schwartz inequality)* $\langle x, y \rangle \leq ||x||||y||$ *for any* $x, y \in V$;
(d) *(Triangle inequality)* $||x + y|| \leq ||x|| + ||y||$.

*Proof.* For (a),

$$\begin{aligned} ||cx|| &= \sqrt{\langle cx, cx\rangle} \\ &= \sqrt{c^2\langle x, x\rangle} \\ &= |c|\sqrt{\langle x, x\rangle} \\ &= |c|||x|| \end{aligned}$$

For (b), if $x = 0$, by Proposition 1.2, $\langle x, x\rangle = 0$ and so $||x|| = \sqrt{\langle x, x\rangle} = 0$.

If $||x|| = 0$, $\sqrt{\langle x, x\rangle} = 0$ and $\langle x, x\rangle = 0$ and so by Proposition 1.2, $x = 0$.

For (c), we consider

$$0 \leq \langle x - cy, x - cy\rangle = \langle x, x\rangle - c\langle x, y\rangle - c\langle y, x\rangle + c^2\langle y, y\rangle.$$

Let $c = \frac{\langle x,y\rangle}{||y||^2}$. Then, the inequality can be simplified as:

$$0 \leq ||x||^2 - \frac{\langle x, y\rangle^2}{||y||^2}.$$

Then $\langle x, y\rangle^2 \leq ||x||^2 \cdot ||y||^2$. Thus, $\langle x, y\rangle \leq ||x|| \cdot ||y||$.

For (d),

$$\begin{aligned} ||x + y||^2 &= \langle x + y, x + y\rangle \\ &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\ &= ||x||^2 + 2\langle x, y\rangle + ||y||^2 \\ &\geq ||x||^2 + 2||x||||y|| + ||y||^2 \\ &= (||x|| + ||y||)^2. \end{aligned}$$

□

**Exercise 1.16.** Interpret geometrically the Cauchy-Schwartz inequality for the standard inner product for $\mathbb{R}^2$.

**Exercise 1.17.** Let $A$ be an $n \times n$ matrix. Then nullity$(A^T A)$ = nullity$(A)$ and rank$(A^T A)$ = rank$(A)$.

## 1.2. **Orthonormal set.**

**Definition 1.18.** Let $V$ be an inner product vector space.

- For $x, y \in V$, we say that $x$ and $y$ are orthogonal if $\langle x, y\rangle 0$.
- A subset $S$ of $V$ is orthogonal if for any two *distinct* $x, y \in V$, $\langle x, y\rangle = 0$.
- A vector $v \in V$ is said to be an *unit* if $||v|| = 1$.
- A subset $S$ of $V$ is said to be orthonormal if $S$ is orthogonal and any vector in $S$ is an unit.

**Example 1.19.** The vectors $\mathbf{e}_i$ are units in $\mathbb{R}^n$. The set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal in $\mathbb{R}^n$.

**Example 1.20.** The set $\{2\mathbf{e}_1, 3\mathbf{e}_2, \ldots, (n+1)\mathbf{e}_n\}$ is orthogonal, but not orthonormal in $\mathbb{R}^n$.

**Example 1.21.** The set

$$S = \left\{ \frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(1,-1) \right\}$$

is an orthonormal set in $\mathbb{R}^2$.

**Exercise 1.22.** Let $v$ be a non-zero vector in $V$. Then $\frac{1}{||v||}v$ is always a unit vector. We shall say that the vector $\frac{1}{||v||}v$ is obtained by normalizing $v$.
*Sol:* $||\frac{1}{||v||}v||2 = \langle \frac{1}{||v||}v, \frac{1}{||v||}v \rangle = \frac{1}{||v||^2}\langle v, v\rangle = \frac{1}{\langle v,v\rangle}\langle v, v\rangle = 1$. Hence, $||\frac{1}{||v||}v|| = 1$.

**Example 1.23.** We normalize the vector $(2,3)$ to obtain the unit vector $\frac{1}{\sqrt{13}}(2,3)$.

**Exercise 1.24.** Let $v$ be an unit vector in $V$. Show that $-v$ is also an unit vector.
*Sol:* $||-v|| = |-1|||v|| = 1$.

**Exercise 1.25.** Describe how to obtain an orthonormal set from an orthogonal set in an inner product space $V$.
*Sol:* Let $S = \{v_1, \ldots, v_r\}$ be an orthogonal set in $V$. Then, by normalization, we obtain an orthonormal set:

$$\left\{ \frac{v_1}{||v_1||}, \ldots, \frac{v_r}{||v_r||} \right\}.$$

One main property of an orthogonal set is the following:

**Theorem 1.26.** *Let $S$ be an orthogonal set of non-zero vectors in an inner product space $V$. Then $S$ is linearly independent.*

*Proof.* Let $S = \{v_1, \ldots, v_r\}$. Let $a_1, \ldots, a_r \in \mathbb{R}$ such that

$$a_1 v_1 + \ldots + a_r v_r = 0.$$

Then

$$\langle a_1 v_1 + \ldots + a_r v_r, v_i \rangle = \langle 0, v_i \rangle = 0.$$

Hence, $a_1\langle v_1, v_i\rangle + \ldots + a_r\langle v_r, v_i\rangle = 0$. Since $\langle v_j, v_i\rangle = 0$ for $i \neq j$. It leaves:

$$a_i\langle v_i, v_i\rangle = 0.$$

Since $v_i \neq 0$, we then have $\langle v_i, v_i\rangle \neq 0$ and so $a_i$ must be 0. This shows all $a_i = 0$. $\square$

**Example 1.27.** Let

$$v_1 = (1, 2, -1), v_2 = (2, -1, 0), v_3 = (1, 2, 5).$$

Show, using the inner product that $v_1, v_2, v_3$ are linearly independent. *Sol:* Then, one checks:

$$\langle v_1, v_2 \rangle = (1)(2) + (2)(-1) = 0$$
$$\langle v_1, v_3 \rangle = (1)(1) + (2)(2) + (5)(-1) = 0$$
$$\langle v_2, v_3 \rangle = (2)(1) + (-1)(2) = 0.$$

Hence, $\{v_1, v_2, v_3\}$ is a set of linearly independent vectors.

**Remark 1.28.** The converse of Theorem 1.26 is not true. For example. $(1, 1)^T$ and $(0, 1)$ are linearly independent in $\mathbb{R}^2$, but they do not form an orthogonal set.

**Exercise 1.29.** Let $V$ be an inner product space. Suppose $\dim(V) = n$. Let $S$ be an orthogonal set in $V$. Show that the number of vectors in $S$ is less than or equal to $n$.

Theorem 1.26 can deal with some more elegant example that cannot be dealth with by solving a system of linear equations:

**Example 1.30.** Let $V$ be the set of integrable functions from $[0, 2\pi]$ to $\mathbb{R}$. Show that $\{\sin x, \sin(2x), \sin(3x), \ldots, \sin(2023x)\}$ forms a set of linearly independent vectors. *Sol:* It is a trigonometric question that

$$\int_0^{2\pi} \sin(nx) \sin(mx) dx = 0$$

if $n \neq m$. Hence, by Theorem 1.26, the set is linearly independent in $V$.

**Theorem 1.31.** *Let $V$ be an inner product space. Let $S = \{v_1, \ldots, v_r\}$ be an orthogonal subset of $V$ consisting of non-zero vectors. For any $y \in \mathrm{span}(S)$,*

$$y = \frac{\langle v_1, y \rangle}{||v_1||^2} v_1 + \ldots + \frac{\langle v_r, y \rangle}{||v_r||^2} v_r.$$

We remark that Theorem 1.26 implies that $S$ forms a basis for $\mathrm{span}(S)$. Thus, for any $y \in \mathrm{span}(S)$,

$$y = a_1 v_1 + \ldots + a_r v_r$$

for some unique $a_1, \ldots, a_r$. Theorem 1.31 is to give an explicit description for those $a_1, \ldots, a_r$.

*Proof of Theorem 1.31.* Since $y \in \mathrm{span}(S)$,

$$y = a_1 v_1 + \ldots + a_r v_r.$$

Then,

$$\langle v_i, y \rangle = \langle v_i, a_1 v_1 + \ldots + a_r v_r \rangle = a_1 \langle v_i, v_1 \rangle + \ldots + a_r \langle v_i, v_r \rangle.$$

Then $\langle v_i, y\rangle = a_i\langle v_i, v_i\rangle$ and so

$$a_i = \frac{\langle v_i, y\rangle}{||v_i||^2}$$

for all $i$. □

**Example 1.32.** Show that $\{(3,2),(2,-3)\}$ is an orthogonal set. Express $(\sqrt{2},-1)$ in terms of a linear combination of the two orthonormal vectors.
*Sol:* Let $v_1 = (3,2)$ and $v_2 = (-2,3)$. Then

$$\langle v_1, v_2\rangle = (3)(2) + (-2)(3) = 0.$$

Let $v = (\sqrt{2},-1)$. Now we compute

$$\langle v, (3,2)\rangle = 3\sqrt{2} - 2$$

and $\langle v, (2,-3)\rangle = 2\sqrt{2} + 3$. Moreover,

$$||(3,2)|| = \sqrt{3^2+2^2} = \sqrt{13}, \quad ||(2,-3)|| = \sqrt{2^2+(-3)^2} = \sqrt{13}.$$

Then

$$v = \frac{3\sqrt{2}-2}{13}(3,2) + \frac{2\sqrt{2}+3}{13}(-2,3).$$

### 1.3. **Orthonormal basis and Gram-Schmidt process.**

**Definition 1.33.** Let $V$ be an inner product space. A subset $S$ of $V$ is said to be an *orthonormal basis* for $V$ if $S$ is an ordered basis and is also orthonormal.

**Example 1.34.** The standard ordered basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.

**Example 1.35.** The set

$$\left\{\frac{1}{\sqrt{10}}(1,3), \frac{1}{\sqrt{10}}(-3,1)\right\}$$

is another orthonormal basis for $\mathbb{R}^2$.

We have the following result to find an orthonormal basis for an inner product space.

**Theorem 1.36.** *Let $V$ be an inner product space and let $S = \{w_1, \ldots, w_n\}$ be a basis for $V$. Let $v_1 = w_1$ and, for $2 \le j \le n$, let*

$$v_i = w_i - \sum_{j=1}^{i-1} \frac{\langle w_i, v_j\rangle}{||v_j||^2} v_j.$$

*Then, $\{v_1, \ldots, v_k\}$ is a set of orthogonal vectors which forms a basis for $V$.*

A consequence is the following existence result for an orthonormal basis:

**Corollary 1.37.** *Let $V$ be an inner product space. There always exists an orthonormal basis for $V$.*

*Proof.* There always exists a basis for $V$. Then, by Theorem 1.36, one has a basis which is orthogonal to $V$. Now one normalizes each vector in the basis, which gives an orthonormal basis. $\square$

*Proof of Theorem 1.36.* Let $S' = \{v_1, \ldots, v_n\}$. We first show that $S'$ forms a basis for $V$.

- Spanning condition: We can rewrite
$$w_i = v_i + \sum_{j=1}^{i-1} \frac{\langle w_i, v_j \rangle}{||v_j||^2} v_j$$
i.e. written as a linear combinations of $v_1, \ldots, v_n$. Thus, $w_i \in \operatorname{span}(S')$. Hence, $V = \operatorname{span}(\{w_1, \ldots, w_n\}) \subset \operatorname{span}(S')$. The inclusion $\operatorname{span}(S') \subset V$ is clear. Thus $\operatorname{span}(S') = V$.
- Linearly independence: We consider the equation:
$$b_1 v_1 + \ldots + b_n v_n = 0$$
If we rewrite each $v_i$ into $w_i$ in the above equation, we have:
$$b_1' w_1 + \ldots + b_{n-1}' w_{n-1} + b_n w_n = 0$$
for some $b_1', \ldots, b_{n-1}'$ in $\mathbb{R}$. This implies that $b_n = 0$ by using the linear independence of $w_1, \ldots, w_n$. This reduces the equation to the form:
$$b_1 v_1 + \ldots + b_{n-1} v_{n-1} = 0.$$
Inductively, we can show $b_{n-1} = \ldots = b_1 = 0$.
- Orthonormal: We assume $l < k$
$$\begin{aligned} \langle v_k, v_l \rangle &= \langle w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{||v_j||^2} v_j, v_l \rangle \\ &= \langle w_k, v_l \rangle - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{||v_j||^2} \langle v_j, v_l \rangle \\ &= \langle w_k, v_l \rangle - \frac{\langle w_k, v_l \rangle}{||v_l||^2} \langle v_l, v_l \rangle \\ &= 0 \end{aligned}$$
where the second equality follows the property of the inner product, the third equality follows from an induction that $\langle v_j, v_l \rangle = 0$ for $j \neq l$, and the last equality follows from $\langle v_l, v_l \rangle = ||v_l||^2$.

**Definition 1.38.** The orthonormal basis obtained in Theorem 1.36 is called an *Gram-Schmidt orthonormal basis* and the process to get that is called the *Gram-Schmidt orthogonalization process.*

**Remark 1.39.** For any vector subspace of $\mathbb{R}^n$, let

$$T : V \to \mathbb{R}^n$$

be an injective linear transformation given by $T(v) = v$. Let $\langle , \rangle$ be the standard inner product on $\mathbb{R}^n$. Then, by Exercise 1.9, $\langle , \rangle$ also defines an inner product for $V$.

For any subspace of $\mathbb{R}^n$, *we shall use this standard inner product unless specified.*

**Example 1.40.** Find an orthonormal basis for the vector subspace

$$\text{span}(\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \right\}).$$

*Sol:* Let $w_1 = (1, 0, -1)^T$ and $w_2 = (2, 1, -1)^T$. We carry out the Gran-Schmidt orthogonalization process: Set $v_1 = w_1$. Then we find:

$$\begin{aligned} v_2 &= w_2 - \frac{\langle v_1, w_2 \rangle}{||v_1||^2} v_1 \\ &= (2, 1, -1)^T - \frac{3}{2}(1, 0, -1)^T \\ &= \begin{pmatrix} \frac{1}{2} \\ 1 \\ \frac{1}{2} \end{pmatrix} \end{aligned}$$

One then normalizes to get the orthonormal set:

$$\left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\}.$$

**Exercise 1.41.** Draw vectors $w_1, w_2, v_1, v_2$ of Example 1.40 in $\mathbb{R}^3$ to visualize the orthogonalization process.

**Example 1.42.** Let $P_2(\mathbb{R})$ be the vector space of polynomials of degree at most 2. Find an orthonormal basis for $P_2(\mathbb{R})$ using the inner product in Example 1.8 with $a = 0$ and $b = 1$.
*Sol:* Let $f_1(x) = 1$, $f_2(x) = x$ and $f_3(x) = x^2$. Then, by Gram-Schmidt orthogonalization process, we have:

$$w_1(x) = f_1(x)$$
$$w_2(x) = f_2(x) - \frac{\langle f_1, f_2 \rangle}{||f_1||^2} f_1(x) = x - \frac{1}{2}$$

$$w_3(x) = f_3(x) - \frac{\langle w_1, f_3\rangle}{||w_1||^2} w_1(x) - \frac{\langle w_2, f_3\rangle}{||w_2||^2} w_2(x) = x^2 - \frac{1}{3} - \frac{1}{4}x$$

We then normalize to obtain an orthonormal basis:

$$w_1(x), 12w_2(x), \frac{w_3(x)}{||w_3||}.$$

## 1.4. **Orthogonal complements.**

**Definition 1.43.** Let $S$ be a non-empty subset of an inner product space $V$. Define

$$S^\perp = \{x \in V : \langle x, y\rangle = 0 \text{ for all } y \in S\ \}.$$

Literally, $S^\perp$ is the subset of $V$ containing all the vectors orthogonal to all vectors in $S$. We call it the **orthogonal complement** of $S$.

**Example 1.44.** Let $S = \{0\}$. Then $S^\perp = V$ by Proposition 1.2(c).

**Example 1.45.** Let $S = V$. Then $S^\perp = \{0\}$ since if $v \neq 0$, then $\langle v, v\rangle \neq 0$.

**Example 1.46.** Let $S = \{\mathbf{e}_2\} \subset \mathbb{R}^3$. Then $S^\perp = \{(x, 0, z) : x, z \in \mathbb{R}\}$ is the $xz$-plane.

**Exercise 1.47.** Let $S$ be a subset for $V$. Show that $S^\perp = \text{span}(S)^\perp$.
*Sol:* We first show $\text{span}(S)^\perp \subset S^\perp$. Let $x \in \text{span}(S)^\perp$. Then $\langle x, y\rangle = 0$ for all $y \in \text{span}(S)$. In particular, $\langle x, y'\rangle = 0$ for all $y' \in S$ since $S \subset \text{span}(S)$. This shows that $\text{span}(S)^\perp \subset S^\perp$.

We next show that $S^\perp \subset \text{span}(S)^\perp$. Let $v \in S^\perp$. Write $S = \{w_1, \ldots, w_n\}$. For any $y \in \text{span}(S)$,

$$y = a_1 w_1 + \ldots + a_n w_n$$

for some $a_1, \ldots, a_n$. Then

$$\langle v, y\rangle = \langle v, a_1 w_1 + \ldots + a_n w_n\rangle = a_1\langle v, w_1\rangle + \ldots + a_n\langle v, w_n\rangle = 0,$$

where the last equality follows from $\langle v, w_i\rangle = 0$ for all $i$ since $v \in S^\perp$. This shows $S^\perp \subset \text{span}(S)^\perp$.

**Theorem 1.48.** *Let $S$ be a non-empty subset of $V$. Then $S^\perp$ is a vector subspace of $V$.*

*Proof.*
- Check addition: Let $x, y \in S^\perp$. For any $z \in S$,

$$\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle = 0.$$

- Check scalar multiplication: Let $x \in S^\perp$ and let $c \in \mathbb{R}$. For any $z \in S$,

$$\langle cx, z\rangle = c\langle x, z\rangle = 0.$$

Hence, $cx \in S^\perp$. By the definition, we have that $S^\perp$ is a vector subspace of $V$. □

Another property is on the dimension:

**Theorem 1.49.** *Let $V$ be an inner product space. Let $W$ be a vector subspace of $V$. Then*

$$\dim(W) + \dim(W^{\perp}) = \dim(V).$$

*Proof.* We briefly sketch the proof. Let $n = \dim(V)$. Let $\{v_1, \ldots, v_r\}$ be an orthonormal basis for $W$. We find vectors $w_{r+1}, \ldots, w_n$ such that $\{v_1, \ldots, v_r, w_{r+1}, \ldots, w_n\}$ forms a basis for $V$. Then by carrying out the Gram-Schmidt process, we obtain an orthonormal basis $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$. Then one checks that $v_{r+1}, \ldots, v_n$ are in $W^{\perp}$ and form a basis for $W^{\perp}$. □

**Exercise 1.50.** Investigate the proof of Theorem 1.49 and describe a way to find an orthonormal basis for $W^{\perp}$ if we know an orthonormal basis for $W$.

*Sol:* We start with an orthonormal basis $\{v_1, \ldots, v_r\}$ for $W$.

- Then we complete to a basis $\{v_1, \ldots, v_r, w_{r+1}, \ldots, w_n\}$ for $V$.
- We carry out the Gram-Schmidt process to obtain an orthonormal basis $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ for $V$.
- Finally, $\{v_{r+1}, \ldots, v_n\}$ forms a basis for $W^{\perp}$.

**Example 1.51.** Let $W = \mathrm{span}(\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\})$. Find $\dim(W^{\perp})$ and find an orthonormal basis for $W^{\perp}$.

*Sol:* By the dimension formula (Theorem 1.49), $\dim(W^{\perp}) = 3 - 1 = 2$.

We now find an orthonormal basis for $W^{\perp}$. First we find a basis

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

We now use the Gram-Schmidt process to find the vectors:

$$v_1 = (1, 1, 1)^T,$$
$$v_2 = (0, 1, 0)^T - \frac{1}{3}(1, 1, 1)^T = (-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3})^T.$$
$$v_3 = (0, 0, 1)^T - \frac{1}{3}(1, 1, 1)^T - \frac{-1}{6}(-1, 2, -1)^T = (-\frac{1}{2}, 0, \frac{1}{2})^T.$$

Then $\left\{(-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3})^T, (-\frac{1}{2}, 0, \frac{1}{2})^T\right\}$ forms an orthonormal basis for $W^{\perp}$.

1.5. **Orthogonal projections.** One importance of the orthogonal complement is the following result:

**Theorem 1.52.** *Let $V$ be an inner product space. Let $W$ be a vector subspace of $V$. Then, for any $v \in V$, there exists unique $x \in W$ and $y \in W^{\perp}$ such that $v = x + y$. The vector $x$ is called the* **orthogonal** *projection of $v$ on $W$.*

Before going into the proof of Theorem 1.52, we try to understand the theorem in some examples:

**Example 1.53.** Let $W = \text{span}(\{\mathbf{e}_1, \mathbf{e}_3\})$. Then $W^\perp = \text{span}(\{\mathbf{e}_2\})$. For $v = (2,3,4)$, we have:

$$v = (2,0,4) + (0,3,0)$$

for $(2,0,4) \in W$ and $(0,3,0) \in W^\perp$. Draw a picture to visualize that.

**Example 1.54.** Let $W = \text{span}(\{(1,-2)\})$ in $\mathbb{R}^2$. Then $W^\perp = \text{span}(\{(2,1)\})$. For example,

$$\langle (3,4), (1,-2) \rangle = -5$$

and so the orthogonal projection of $(3,4)$ on $W$ is

$$\frac{-5}{5}(1,-2) = (-1,2)$$

(One may also use the formula in Corollary 1.55 below.) One can write $(3,4) = (-1,2) + (4,2)$ with $(4,2) \in W^\perp$.

*Proof of Theorem 1.52.* Let $\{v_1, \ldots, v_r\}$ be an orthonormal basis for $W$. Define

$$x = \langle v, v_1 \rangle v_1 + \ldots + \langle v, v_r \rangle v_r.$$

By definitions, $x \in W$.

Now let $y = v - x$. For any $1 \leq i \leq r$,

$$\begin{aligned}
\langle y, v_i \rangle &= \langle v - x, v_i \rangle \\
&= \langle v, v_i \rangle - \langle x, v_i \rangle \\
&= \langle v, v_i \rangle - \langle v, v_1 \rangle \langle v_1, v_i \rangle - \ldots - \langle v_r, v_i \rangle \langle v_r, v_i \rangle \\
&= \langle v, v_i \rangle - \langle v, v_i \rangle \langle v_i, v_i \rangle \\
&= 0
\end{aligned}$$

This shows that $y \in W^\perp$.

The uniqueness part is left as an exercise. □

**Corollary 1.55.** *We use the notations in Corollary 1.52. Let $\{v_1, \ldots, v_r\}$ be an orthonormal basis for $W$. Then such $x$ is given by*

$$x = \langle v, v_1 \rangle v_1 + \ldots + \langle v, v_r \rangle v_r.$$

**Corollary 1.56.** *(Minimizing the length via orthogonal projection) Let $W$ be a subspace of $V$. For $v \in V$, write $v = x + y$ as in Theorem 1.52. Then $x$ is the unique vector in $W$ such that*

$$||v - x'|| \geq ||v - x||$$

*for any $x' \in W$. In other words, $x$ is the unique vector in $W$ that minimizes the length $||v-x||$.*

*Proof.*

$$\begin{aligned}||v-x'||^2 &= ||y+x-x'||^2\\ &= ||y||^2+||x-x'||^2\\ &\geq ||y||^2\\ &= ||v-x||^2\end{aligned}$$

For the uniqueness, we need the above inequality to become an equality and so $||x'-x||=0$. Hence, $x'$ must be $x$ to acheive the minimum. ☐

**Example 1.57.** Let $W=\text{span}(\{\mathbf{e}_2,\mathbf{e}_3\})$ be a subspace of $\mathbb{R}^4$. Let $v=(1,2,3,4)^T$. Find a vector $x\in W$ such that the length $||v-x||$ is minimized.
*Sol:* The orthogonal projection $x'$ of $v$ is

$$x'=\langle \mathbf{e}_2,v\rangle \mathbf{e}_2+\langle \mathbf{e}_3,v\rangle \mathbf{e}_3=(0,2,3,0)^T,$$

which minimizes the length $||x'-v||$ by Corollary 1.56.

**Example 1.58.** Let $W=\{(x,y,z):x+y-z=0\}$, which is a 2-dimensional subspace in $\mathbb{R}^3$. Find the point in $W$ which has the shortest distance with the point $(1,2,1)$.
*Sol:* We first find an orthonormal basis for $W$. Note that $\{(1,0,1),(1,-1,0)\}$ forms a basis for $W$. Let $v_1=(1,0,1)$ and let

$$v_2=(1,-1,0)-\frac{\langle (1,0,1),(1,-1,0)\rangle}{||(1,0,1)||^2}(1,0,1)=(1,-1,0)-\frac{1}{2}(1,0,1)=(\frac{1}{2},-1,-\frac{1}{2}).$$

Since those vectors are obtained by the Gram-Schmidt process, $\{v_1,v_2\}$ is orthogonal.

We now normalize the vectors to obtain an orthonormal basis formed by two vectors

$$v_1'=\frac{1}{\sqrt{2}}(1,0,1),\quad v_2'=\frac{1}{\sqrt{6}}(1,-2,-1)$$

The the point is given by:

$$\langle (1,2,1),v_1'\rangle v_1'+\langle (1,2,1),v_2'\rangle v_2'=(1,0,1)-\frac{2}{3}(1,-2,-1)=(\frac{1}{3},\frac{4}{3},\frac{5}{3})$$

in $W$.

1.6. **Least square approximation.** We shall now describe a practical question. Suppose we have some data points

$$(t_1, y_1), \ldots, (t_k, y_k).$$

We want to find a line of the equation $y = ct + d$ to approximate those data, that is,

$$y_i \approx ct_i + d.$$

Let

$$A = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \\ t_k & 1 \end{pmatrix}, \quad x = \begin{pmatrix} c \\ d \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix}.$$

We can rewrite into a matrix form:

$$Ax \approx y.$$

To make sense of the approximation, one may find $x_0 \in \mathbb{R}^2$ such that the length $||Ax_0 - y||$ is the minimum. The resulting line is called a **least squares line**.

**Theorem 1.59.** *There exists $x_0 \in \mathbb{R}^2$ such that $||Ax_0 - y||$ is the smallest and $A^T A x_0 = A^T y$.*

*Proof.* Let $W = \text{col}(A)$. Then Theorem 1.52 implies that we can find $x' \in W$ so that $||x' - y||$ is the smallest. Write $x' = Ax_0$ for some $x_0 \in \mathbb{R}^2$. Now we have $||Ax_0 - y||$.

It remains to show another equality. We also have that $Ax_0 - y \in W^\perp$ and so

$$\langle Ax_0 - y, Ax \rangle = 0$$

for all $x \in \mathbb{R}^2$. We use Exercise 1.4 to get:

$$\langle A^T(Ax_0 - y), x \rangle = 0$$

for all $x$. Hence, $A^T(Ax_0 - y) = 0$ and so $A^T A x_0 = A^T y$. □

**Example 1.60.** Find the least square line for the points/data $(2, 3), (1, 5), (3, 7), (4, 2)$.
*Sol:* Let

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}, y = \begin{pmatrix} 3 \\ 5 \\ 7 \\ 2 \end{pmatrix}.$$

Then $A^T A = \begin{pmatrix} 30 & 10 \\ 10 & 1 \end{pmatrix}$ and $A^T y = \begin{pmatrix} 40 \\ 17 \end{pmatrix}$. Then, we can find $x_0$ by solving

$$(A^T A)x_0 = A^T y.$$

Hence,

$$x_0 = (A^T A)^{-1}(A^T y) = \frac{1}{-70}\begin{pmatrix} 1 & -10 \\ -10 & 30 \end{pmatrix}\begin{pmatrix} 40 \\ 17 \end{pmatrix} = \frac{1}{-7}\begin{pmatrix} -13 \\ 11 \end{pmatrix}$$