4. INNER PRODUCTS AND ORTHOGONALITY

In making the definition of a vector space, we generalized the linear structure (addition and scalar multiplication) of $\mathbb{R}^2$ an $\mathbb{R}^3$. We ignored other important features, such as the notions of length and angle. These ideas are embedded in the concept we now investigate, inner products.

4.1 Inner product spaces

We begin with a definition.

Definition 4.1

Let V be a real (or complex) vector space. Suppose that to each pair of vectors $\mathbf{u},\mathbf{v} \in V$ there is assigned a real number denoted by $\langle \mathbf{u},\mathbf{v} \rangle$. This function is called an *inner product* on V if it satisfies the following axioms:

(I1) (Linearity in the first entry) $\langle a\mathbf{u}_1 + b\mathbf{u}_2, \mathbf{v} \rangle = a\langle \mathbf{u}_1, \mathbf{v} \rangle + b\langle \mathbf{u}_2, \mathbf{v} \rangle$.

(I2) (Symmetry) $\langle \mathbf{u},\mathbf{v} \rangle = \langle \mathbf{v},\mathbf{u} \rangle$.

(I3) (Positive definiteness) $\langle \mathbf{u},\mathbf{u} \rangle \geq 0$; and $\langle \mathbf{u},\mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

The vector space V with an inner product is called an *inner product space*.

Using the linearity axiom (I1) and the symmetry axiom (I2), we obtain

$$\langle \mathbf{u}, a\mathbf{v}_1 + b\mathbf{v}_2 \rangle = \langle a\mathbf{v}_1 + b\mathbf{v}_2, \mathbf{u} \rangle = a\langle \mathbf{v}_1, \mathbf{u} \rangle + b\langle \mathbf{v}_2, \mathbf{u} \rangle = a\langle \mathbf{u}, \mathbf{v}_1 \rangle + b\langle \mathbf{u}, \mathbf{v}_2 \rangle.$$

That is, the inner product function is also linear in its second entry. By induction, we obtain

$$\langle a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + \mathbf{a}_r\mathbf{u}_r, \mathbf{v} \rangle = a_1\langle \mathbf{u}_1, \mathbf{v} \rangle + a_2\langle \mathbf{u}_2, \mathbf{v} \rangle + \cdots + a_r\langle \mathbf{u}_r, \mathbf{v} \rangle$$

and

$$\langle \mathbf{u}, b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + \mathbf{b}_s\mathbf{v}_s \rangle = b_1\langle \mathbf{u}, \mathbf{v}_1 \rangle + b_2\langle \mathbf{u}, \mathbf{v}_2 \rangle + \cdots + b_s\langle \mathbf{u}, \mathbf{v}_s \rangle.$$

Combining these two properties yields the following general formula:

$$\left\langle \sum_{i=1}^{r} a_i\mathbf{u}_i, \sum_{j=1}^{s} b_j\mathbf{v}_j \right\rangle = \sum_{i=1}^{r}\sum_{j=1}^{s} a_i b_j \langle \mathbf{u}_i, \mathbf{v}_j \rangle$$

$$= \begin{bmatrix} a_1 & a_2 & \dots & a_r \end{bmatrix} \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{v}_1 \rangle & \langle \mathbf{u}_1, \mathbf{v}_2 \rangle & \dots & \langle \mathbf{u}_1, \mathbf{v}_s \rangle \\ \langle \mathbf{u}_2, \mathbf{v}_1 \rangle & \langle \mathbf{u}_2, \mathbf{v}_2 \rangle & \dots & \langle \mathbf{u}_2, \mathbf{v}_s \rangle \\ \dots & \dots & \dots & \dots \\ \langle \mathbf{u}_r, \mathbf{v}_1 \rangle & \langle \mathbf{u}_r, \mathbf{v}_2 \rangle & \dots & \langle \mathbf{u}_r, \mathbf{v}_s \rangle \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{bmatrix}. \tag{4.1}$$

By axiom (I3), $\langle \mathbf{u}, \mathbf{u} \rangle$ is a non negative number and hence its positive real square root exists. We use the notation

$$\| \mathbf{u} \| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

This non negative real number $\| \mathbf{u} \|$ is called the *norm* of $\mathbf{u}$. The relation $\| \mathbf{u} \|^2 = \langle \mathbf{u}, \mathbf{u} \rangle$ will be used frequently.

Example 4.1

Consider the vector space $\mathbb{R}^n$. The function

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\top \mathbf{v}$$

defines an inner product on $\mathbb{R}^n$. Notice that this is the usual *dot product* $\mathbf{u} \cdot \mathbf{v}$. The norm $\| \mathbf{u} \|$ of a vector in this space is as follows:

$$\| \mathbf{u} \| = \sqrt{\mathbf{u}^\top \mathbf{v}}$$

Although there are many ways to define an inner product on $\mathbb{R}^n$, we will often use the inner product in $\mathbb{R}^n$ defined by the dot product, unless otherwise stated or implied; we will call it the *usual inner product on* $\mathbb{R}^n$.

Example 4.2

Let V be the vector space of real continuous functions on the interval $a \le t \le b$. Then the following is an inner product on V:

$$\langle f, g \rangle = \int_a^b f(t) g(t) \, dt,$$

where $f(t)$ and $g(t)$ are any continuous functions on $[a, b]$.

Another inner product on V is:

$$\langle f, g \rangle = \int_a^b f(t) \, g(t) \, w(t) \, dt,$$

where $w(t)$ is a given continuous function which is positive on the interval $a \le t \le b$. In this case $w(t)$ is called a *weight function* for the inner product.

Example 4.3

Let V denote the vector space of $m \times n$ matrices over $\mathbb{R}$. The following is an inner product on V:

$$\langle A,B\rangle = Tr\left(B^{\top} A\right)$$

where Tr stands form the trace, the sum of the diagonal elements. If $A = \left(a_{ij}\right)$ and $B = \left(b_{ij}\right)$, then

$$\langle A,B\rangle = \mathrm{Tr}\left(B \quad A\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}$$

the sum of the products of the corresponding entries. In particular,

$$\| A \|^2 = \langle A, A\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2$$

the sum of the squares of all the elements of A.

4.2 Gram matrices of an inner product

Let V be an inner product space and let $\mathcal{B} = \left\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\right\}$ be a basis for V. Since any vectors $\mathbf{u}, \mathbf{v} \in V$ can be written as a linear combination of the basis using their coordinates as weights, we can write the inner product using (4.1) as follows

$$\langle \mathbf{u}, \mathbf{v}\rangle = [\mathbf{u}]_{\mathcal{B}}^{\top} \begin{bmatrix} \langle \mathbf{b}_1, \mathbf{b}_1\rangle & \langle \mathbf{b}_1, \mathbf{b}_2\rangle & \cdots & \langle \mathbf{b}_1, \mathbf{b}_n\rangle \\ \langle \mathbf{b}_2, \mathbf{b}_1\rangle & \langle \mathbf{b}_2, \mathbf{b}_2\rangle & \cdots & \langle \mathbf{b}_2, \mathbf{b}_n\rangle \\ \cdots & \cdots & \cdots & \cdots \\ \langle \mathbf{b}_n, \mathbf{b}_1\rangle & \langle \mathbf{b}_n, \mathbf{b}_2\rangle & \cdots & \langle \mathbf{b}_n, \mathbf{b}_n\rangle \end{bmatrix} [\mathbf{v}]_{\mathcal{B}}.$$

The $n \times n$ matrix

$$G^{\mathcal{B}} = \begin{bmatrix} \langle \mathbf{b}_1, \mathbf{b}_1\rangle & \langle \mathbf{b}_1, \mathbf{b}_2\rangle & \cdots & \langle \mathbf{b}_1, \mathbf{b}_n\rangle \\ \langle \mathbf{b}_2, \mathbf{b}_1\rangle & \langle \mathbf{b}_2, \mathbf{b}_2\rangle & \cdots & \langle \mathbf{b}_2, \mathbf{b}_n\rangle \\ \cdots & \cdots & \cdots & \cdots \\ \langle \mathbf{b}_n, \mathbf{b}_1\rangle & \langle \mathbf{b}_n, \mathbf{b}_2\rangle & \cdots & \langle \mathbf{b}_n, \mathbf{b}_n\rangle \end{bmatrix}$$

is called the *Gram matrix if the inner product relative to the basis* $\mathcal{B}$. Notice that by axioms (I2) and (I3), the Gram matrix must satisfy

$$\left(G^{\mathcal{B}}\right)^{\top} = G^{\mathcal{B}} \quad \text{and} \quad 0 \leq \mathbf{x}^{\top} G^{\mathcal{B}} \mathbf{x}, \quad \text{forall} \mathbf{x} \in \mathbb{R}^n.$$

Example 4.4

Recall that if $\mathcal{E}_n$ is the canonical basis in $\mathbb{R}^n$, then for every vector $\mathbf{u} \in \mathbb{R}^n$, we have $[\mathbf{u}]_{\mathcal{E}_n} = \mathbf{u}$. Moreover, the usual inner product is given by $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\top} \mathbf{v}$, which can be written as

$$\langle \mathbf{u}, \mathbf{v} \rangle = [\mathbf{u}]_{\mathcal{E}_n}^{\top} I_n [\mathbf{v}]_{\mathcal{E}_n}.$$

So we conclude that the Gram matrix of the usual inner product with respect to the canonical basis is the identity matrix I_n, that is, $G^{\mathcal{E}_n} = I_n$. This is easy to verify using the definition and is left to the reader as a straightforward exercise.

Example 4.5

Consider the usual inner product in $\mathbb{R}^3$ and the basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$, where

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{b}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then,

$$\begin{aligned} \langle \mathbf{b}_1, \mathbf{b}_1 \rangle = 1, \quad \langle \mathbf{b}_1, \mathbf{b}_2 \rangle = 1, \quad \langle \mathbf{b}_1, \mathbf{b}_3 \rangle = 1, \\ \langle \mathbf{b}_2, \mathbf{b}_2 \rangle = 2, \quad \langle \mathbf{b}_2, \mathbf{b}_3 \rangle = 2, \quad \langle \mathbf{b}_3, \mathbf{b}_3 \rangle = 2. \end{aligned}$$

So the Gram matrix of the usual inner product relative to $\mathcal{B}$ is

$$G^{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

Notice that the Gram matrix of the usual inner product in $\mathbb{R}^3$ relative to the canonical basis is I_3. In fact, the Gram matrix of an inner product is, in general, different for each choice of basis, that is, the Gram matrix of an inner product depends on the basis. This is true for any inner product space.

Example 4.6

Consider the vector space $\mathbb{P}_2$ endowed with the inner product defined by

$$\langle f,g\rangle = \int_0^1 f(t)\,g(t)\,dt\,.$$

For the standard basis $\mathcal{S}_2 = \{1,t,t^2\}$, we have

$$\langle 1,1\rangle = \int_0^1 dt = 1,\quad \langle 1,t\rangle = \int_0^1 t\,dt = \frac{1}{2},\quad \langle 1,t^2\rangle = \int_0^1 t^2 dt = \frac{1}{3}$$

$$\langle t,t\rangle = \int_0^1 t^2 dt = \frac{1}{3},\quad \langle t,t^2\rangle = \int_0^1 t^3 dt = \frac{1}{4},\quad \langle t^2,t^2\rangle = \int_0^1 t^4 dt = \frac{1}{5}\,.$$

Then the Gram matrix of the inner product relative to the standard basis is

$$G^{\mathcal{S}_2} = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}.$$

We can use the Gram matrix to compute the inner product between to elements of $\mathbb{P}_2$. If $p(t) = 1 - t$ and $q(t) = 1 - t + t^2$, then

$$\langle p(t),q(t)\rangle = [p(t)]_{\mathcal{S}_2}^{\top} G^{\mathcal{S}_2} [q(t)]_{\mathcal{S}_2} = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \frac{5}{12}$$

The definition of an inner product is independent of the choice of basis. Nevertheless, the Gram matrices of an inner product relative to an old basis $\mathcal{B}$ and a new basis $\mathcal{C}$ are, in general, not equal; that is, $G^{\mathcal{B}} \neq G^{\mathcal{C}}$. So it is sensible to conclude that the two Gram matrices are related in some way. We explore this relation here.

First, using the Gram matrix $G^{\mathcal{B}}$, we can write

$$\langle \mathbf{u},\mathbf{v}\rangle = [\mathbf{u}]_{\mathcal{B}}^{\top} G^{\mathcal{B}} [\mathbf{v}]_{\mathcal{B}}.$$

Recall that for any vector $\mathbf{x} \in V$, the coordinate vector $[\mathbf{x}]_{\mathcal{B}}$ is related to the coordinate vector $[\mathbf{x}]_{\mathcal{C}}$ by means of the change of basis matrix $P_{\mathcal{B}}^{\mathcal{C}}$:

$$[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}}^{\mathcal{C}} [\mathbf{x}]_{\mathcal{C}}.$$

Therefore, we can use the change of basis matrix to change the coordinate vectors relative to the old basis to coordinate vectors relative to the new basis:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \left(P_{\mathcal{B}}^{\mathcal{C}} [\mathbf{u}]_{\mathcal{C}} \right)^{\top} G^{\mathcal{B}} \left(P_{\mathcal{B}}^{\mathcal{C}} [\mathbf{v}]_{\mathcal{C}} \right) = [\mathbf{u}]_{\mathcal{C}}^{\top} \left(\left(P_{\mathcal{B}}^{\mathcal{C}} \right)^{\top} G^{\mathcal{B}} P_{\mathcal{B}}^{\mathcal{C}} \right) [\mathbf{v}]_{\mathcal{C}}.$$

Comparing the last equation with

$$\langle \mathbf{u}, \mathbf{v} \rangle = [\mathbf{u}]_{\mathcal{C}}^{\top} G^{\mathcal{C}} [\mathbf{v}]_{\mathcal{C}},$$

we conclude that

$$G^{\mathcal{C}} = \left(P_{\mathcal{B}}^{\mathcal{C}} \right)^{\top} G^{\mathcal{B}} P_{\mathcal{B}}^{\mathcal{C}}.$$

Example 4.7

Consider the usual inner product in $\mathbb{R}^3$. Its Gram matrices relative to two different bases were computed in Example 4.4 and Example 4.5. The Gram matrix relative to the canonical basis is $G^{\mathcal{E}_3} = I_3$, where as the Gram matrix relative to the basis $\mathcal{B}$ given in Example 4.5 is

$$G^{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

The matrix $G^{\mathcal{B}}$ can be obtained from $G^{\mathcal{E}_3} = I_3$ by means of the change of basis matrix

$$P_{\mathcal{E}_3}^{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Indeed, we compute

$$\left(P_{\mathcal{E}_3}^{\mathcal{B}} \right)^{\top} G^{\mathcal{E}_3} P_{\mathcal{E}_3}^{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} = G^{\mathcal{B}}.$$

4.3 Cauchy-Schwarz inequality

The following inequality is called the *Cauchy-Schwarz inequality*; it is used in many branches of mathematics.

Theorem 4.1

For any vectors $\mathbf{u},\mathbf{v} \in V$,

$$|\langle \mathbf{u},\mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|.$$

This inequality is an equality if and only if one of $\mathbf{u},\mathbf{v}$ is a scalar multiple of the other.

Proof. Let $\mathbf{u},\mathbf{v} \in V$. If $\mathbf{v} = \mathbf{0}$, then both sides of the inequality are equal to 0 and the desired result holds. Thus we can assume that $\mathbf{v} \neq \mathbf{0}$.

By axiom (I3), for all $\lambda \in \mathbb{R}$, we get:

$$\langle \mathbf{u} - \lambda\mathbf{v}, \mathbf{u} - \lambda\mathbf{v}\rangle \geq 0.$$

Consequently,

$$\|\mathbf{u}\|^2 - 2\lambda\langle \mathbf{u},\mathbf{v}\rangle + \lambda^2 \|\mathbf{v}\|^2 \geq 0.$$

Taking $\lambda = \dfrac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{v}\|^2}$, we obtain

$$\|\mathbf{u}\|^2 - 2\frac{\langle \mathbf{u},\mathbf{v}\rangle^2}{\|\mathbf{v}\|^2} + \frac{\langle \mathbf{u},\mathbf{v}\rangle^2}{\|\mathbf{v}\|^2} \geq 0.$$

That is,

$$\|\mathbf{u}\|^2 - \frac{\langle \mathbf{u},\mathbf{v}\rangle^2}{\|\mathbf{v}\|^2} \geq 0.$$

Multiplying both sides by $\|\mathbf{v}\|^2$,

$$\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - \langle \mathbf{u},\mathbf{v}\rangle^2 \geq 0s,$$

which implies

$$\langle \mathbf{u},\mathbf{v}\rangle^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2 .$$

Taking the positive square root of both sides, we finally get

$$|\langle \mathbf{u},\mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|.$$

Observe that if $\mathbf{u} = \lambda \mathbf{v}$, then

$$\langle \mathbf{u} - \lambda \mathbf{v}, \mathbf{u} - \lambda \mathbf{v} \rangle = 0.$$

In this case, each step of the proof is valid with $\leq$ and $\geq$ replaced with equality signs. Moreover, each step (with equality replacing inequalities) is reversible. Therefore, the Cauchy-Schwarz inequality is an equality if and only if $\mathbf{u}$ or $\mathbf{v}$ is a scalar multiple of the other. ■

Example 4.8

Consider any real numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$. Then the Cauchy-Schwarz inequality applied to the usual inner product in $\mathbb{R}^n$ implies

$$\left(a_1 b_1 + a_2 b_2 + \cdots + a_n b_n\right)^2 \leq \left(a_1^2 + a_2^2 + \cdots + a_n^2\right)\left(b_1^2 + b_2^2 + \cdots + b_n^2\right).$$

Example 4.9

Let f and g be any real continuous functions defined on the interval $a \leq t \leq b$. Then, applying the Cauchy-Schwarz inequality to the first inner product introduced in Example 4.2, we obtain

$$\left(\int_a^b f(t) g(t) dt\right)^2 \leq \left(\int_a^b f(t)^2 dt\right)\left(\int_a^b g(t)^2 dt\right).$$

The next theorem gives basic properties of the norm; the proof of the third property requires the Cauchy-Schwarz inequality.

Theorem 4.2

Let V be an inner product space. Then the norm in V satisfies the following properties:

(N1) $\| \mathbf{v} \| \geq 0$; and $\| \mathbf{v} \| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

(N2) $\| k\mathbf{v} \| = |k| \| \mathbf{v} \|$.

(N3) (Triangle inequality) $\| \mathbf{u} + \mathbf{v} \| \leq \| \mathbf{u} \| + \| \mathbf{v} \|$; this inequality is an equality if and only if one of $\mathbf{u}, \mathbf{v}$ is a non negative multiple of the other.

Proof. Property (N1) is a direct consequence of axiom (I3). Property (N2) is a direct consequence of the definition of norm and the linearity of the inner product in both entries:

$$\| k\mathbf{v} \| = \sqrt{\langle k\mathbf{v}, k\mathbf{v} \rangle} = \sqrt{k^2 \langle \mathbf{v}, \mathbf{v} \rangle} = |k| \| \mathbf{v} \|.$$

Most of our effort is in proving property (N3). For $\mathbf{u}, \mathbf{v} \in V$,

$$\begin{aligned} \| \mathbf{u} + \mathbf{v} \|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\ &= \| \mathbf{u} \|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \| \mathbf{v} \|^2 \\ &\leq \| \mathbf{u} \|^2 + 2 |\langle \mathbf{u}, \mathbf{v} \rangle| + \| \mathbf{v} \|^2 \\ &\leq \| \mathbf{u} \|^2 + 2 \| \mathbf{u} \| \| \mathbf{v} \| + \| \mathbf{v} \|^2 \quad (\text{byCauchy} - \text{Schwartz}) \\ &= (\| \mathbf{u} \| + \| \mathbf{v} \|)^2 \end{aligned}$$

Taking square roots of both sides of the inequality above gives the triangle inequality. The proof above shows that the triangle inequality is an equality if and only if $\langle \mathbf{u}, \mathbf{v} \rangle = \| \mathbf{u} \| \| \mathbf{v} \|$. If one $\mathbf{u}, \mathbf{v}$ is a non negative multiple of the other, then the above equality holds. Conversely, if the above equality holds, then the condition for the Cauchy-Schwarz inequality implies that one of $\mathbf{u}, \mathbf{v}$ must be a scalar multiple of the other. If $\mathbf{u} = \lambda \mathbf{v}$, then bye (N2), $\langle \mathbf{u}, \mathbf{v} \rangle = \| \mathbf{u} \| \| \mathbf{v} \|$ implies $\lambda = |\lambda|$, which forces the scalar in question to be non negative, as desired. ■

If $\| \mathbf{u} \| = 1$, or, equivalently, if $\langle \mathbf{u}, \mathbf{u} \rangle = 1$, then $\mathbf{u}$ is called a *unit vector* and is said to be *normalized*. Every non zero vector $\mathbf{v} \in V$ can be multiplied by the reciprocal of its norm to obtain a unit vector

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{\| \mathbf{v} \|}$$

which is a positive multiple of **v**. This process is called *normalizing* **v**.

The non negative real number

$$d(\mathbf{u}, \mathbf{v}) = \| \mathbf{u} - \mathbf{v} \|$$

is called the *distance* between $\mathbf{u}$ and $\mathbf{v}$. For any non zero vectors $\mathbf{u}, \mathbf{v} \in V$, the angle between $\mathbf{u}$ and $\mathbf{v}$ is defined to be the angle θ such that $0 \leq \theta \leq \pi$ and

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\| \mathbf{u} \| \| \mathbf{v} \|}.$$

By the Cauchy-Schwartz inequality, $-1 \leq \cos\theta \leq 1$ and so the angle always exists and is unique.

4.4 Orthogonality

Let V be an inner product space. The vectors $\mathbf{u},\mathbf{v} \in V$ are said to be *orthogonal* and $\mathbf{u}$ is said to be *orthogonal* to $\mathbf{v}$ if

$$\langle \mathbf{u},\mathbf{v} \rangle = 0.$$

The relation is clearly symmetric; that is, if $\mathbf{u}$ is orthogonal to $\mathbf{v}$, then $\langle \mathbf{v},\mathbf{u} \rangle = 0$ and so $\mathbf{v}$ is orthogonal to $\mathbf{u}$. We note that $\mathbf{0} \in V$ is orthogonal to every vector $\mathbf{v} \in V$ for

$$\langle \mathbf{0},\mathbf{v} \rangle = \langle 0\mathbf{v},\mathbf{v} \rangle = 0\langle \mathbf{v},\mathbf{v} \rangle = 0.$$

Conversely, if $\mathbf{u}$ is orthogonal to every $\mathbf{v} \in V$, then $\langle \mathbf{u},\mathbf{u} \rangle = 0$ and hence $\mathbf{u} = \mathbf{0}$ by axiom (I3). Observe that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\cos\theta = 0$ where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$, and this is true if and only if $\theta = \pi / 2$.

Example 4.10

Consider an arbitrary vector $\mathbf{u} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}^\top$ in $\mathbb{R}^n$. Then a vector $\mathbf{v} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^\top$ is orthogonal to $\mathbf{u}$ if

$$\langle \mathbf{u},\mathbf{v} \rangle = a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0.$$

In other words, $\mathbf{v}$ is orthogonal to $\mathbf{u}$ if $\mathbf{v}$ satisfies a homogeneous equation whose coefficients are the entries of $\mathbf{u}$.

Example 4.11

Suppose that we want a non zero vector which is orthogonal to $\mathbf{v}_1 = \begin{bmatrix} 1 & 35 \end{bmatrix}^\top$ and $\mathbf{v}_2 = \begin{bmatrix} 0 & 1 & 4 \end{bmatrix}^\top$ in $\mathbb{R}^3$. Let $\mathbf{w} = \begin{bmatrix} x & y & z \end{bmatrix}^\top$. We want

$$0 = \left\langle \mathbf{v}_1,\mathbf{w} \right\rangle = x + 3y + 5z \quad \text{and} \quad 0 = \left\langle \mathbf{v}_2,\mathbf{w} \right\rangle = y + 4z.$$

Thus we obtain the homogeneous system

$$\begin{aligned} x + 3y + 5z &= 0, \\ y + 4z &= 0. \end{aligned}$$

with general solution

$$\mathbf{w} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \lambda \begin{bmatrix} 7 \\ -4 \\ 1 \end{bmatrix}, \quad \lambda \in \mathbb{R}.$$

Then, $\mathbf{w}$ with $\lambda \neq 0$ is orthogonal to $\mathbf{v}_1$ and $\mathbf{v}_2$.

Let S be a subset of an inner product space V. The *orthogonal complement* of S, denoted by $S^{\perp}$ (read "S perp") consists of those vectors in V which are orthogonal to every vector $\mathbf{u} \in S$:

$$S^{\perp} = \{\mathbf{v} \in V :\ \mathbf{v},\mathbf{u}\ = 0 \text{ for every } \mathbf{u} \in S\}.$$

We show that $S^{\perp}$ is a subspace of V. Clearly $\mathbf{0} \in S^{\perp}$ since $\mathbf{0}$ is orthogonal to every vector in V. Now suppose that $\mathbf{u},\mathbf{w} \in S^{\perp}$. Then, for any scalars a and b and any vector $\mathbf{u} \in S$, we have

$$\langle a\mathbf{v} + b\mathbf{w},\mathbf{u}\rangle = a\langle \mathbf{v},\mathbf{u}\rangle + b\langle \mathbf{w},\mathbf{u}\rangle = a0 + b0 = 0.$$

Thus $a\mathbf{v} + b\mathbf{w} \in S^{\perp}$ and therefore $S^{\perp}$ is a subspace of V.

Example 4.12

Suppose we want to find a basis for the subspace S = Span$\{\mathbf{v}\}$ in $\mathbb{R}^3$ where $\mathbf{u} = [13-4]^{\top}$. Note that $S^{\perp}$ consists of all vectors $\mathbf{w}$ = $[x\ y\ z]^{\top}$ such that

$$\langle \mathbf{w},\mathbf{u}\rangle = x + 3y - 4z = 0.$$

The general solution of this homogeneous equation is

$$\mathbf{w} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \lambda \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} + \mu \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \quad \lambda, \mu \in \mathbb{R}.$$

Therefore, a basis for $S^{\perp}$ is

$$\left\{ \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

The next theorem is a basic result in linear algebra.

Theorem 4.3

Let S = Span$\{\mathbf{v}\}$ where $\mathbf{v} \in V$ and $\mathbf{v} \neq \mathbf{0}$. Then for every $\mathbf{x} \in V$ there is a unique $\mathbf{u} \in S$ and a unique $\mathbf{w} \in S^{\perp}$ such that

$$\mathbf{x} = \mathbf{u} + \mathbf{w}.$$

Proof. Let λ be a scalar. Then

$$\mathbf{x} = \lambda\mathbf{v} + (\mathbf{x} - \lambda\mathbf{v}).$$

Notice that $\lambda \in S$. Thus we need to choose λ so that $\mathbf{v}$ is orthogonal to $(\mathbf{x} - \lambda\mathbf{v})$. In other words, we want

$$0 = \langle \mathbf{x} - \lambda\mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{x}, \mathbf{v} \rangle - \lambda \| \mathbf{v} \|^2 .$$

The equation above shows that we should choose λ to $\langle \mathbf{x}, \mathbf{v} \rangle / \| \mathbf{v} \|^2$. Making this choice of λ, we can write

$$\mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{v} \rangle}{\| \mathbf{v} \|^2}\mathbf{v} + \left(\mathbf{x} - \frac{\langle \mathbf{x}, \mathbf{v} \rangle}{\| \mathbf{v} \|^2}\mathbf{v} \right).$$

■

Example 4.13

Consider the vectors $\mathbf{x} = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top$ and $\mathbf{v} = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$ in $\mathbb{R}^2$ with the usual inner product. Then $\langle \mathbf{x}, \mathbf{v} \rangle = 1, \| \mathbf{v} \|^2 = 2$, and

$$\mathbf{w} = \mathbf{x} - \frac{\langle \mathbf{x}, \mathbf{v} \rangle}{\| \mathbf{v} \|^2}\mathbf{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

Then $\mathbf{x}$ can be written uniquely as

$$\mathbf{x} = \frac{1}{2}\mathbf{v} + \mathbf{w},$$

where $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

Example 4.14

Consider the vector space $\mathbb{P}_2$ endowed with the inner product defined by

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt.$$

Relative to the standard basis $\mathcal{S}_2 = \left\{ 1, t, t^2 \right\}$, we have the Gram matrix

$$G^{\mathcal{S}_2} = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}.$$

If $p(t) = 1 - t$ and $q(t) = 1 - t + t^2$, then

$$\langle p(t), q(t) \rangle = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \frac{5}{12}$$

and

$$\langle q(t), q(t) \rangle = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \frac{7}{10}$$

Then

$$p(t) = \frac{25}{42} q(t) + \left(p(t) - \frac{25}{42} q(t) \right) = \frac{25}{42} q(t) + \left(\frac{17}{42} - \frac{17}{42} t - \frac{25}{42} t^2 \right)$$

where

$$\left\langle q(t), \frac{17}{42} - \frac{17}{42} t - \frac{25}{42} t^2 \right\rangle = \begin{bmatrix} \frac{17}{42} & -\frac{17}{42} & -\frac{25}{42} \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = 0.$$

4.5 Orthogonal sets and bases

A set S of vectors in V is called *orthogonal* if each pair of vectors in S are orthogonal, and S is called *orthonormal* if S is orthogonal and each vector in S has unit norm. In other words, $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ is orthogonal if

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0 \quad \text{for} \quad i \neq j$$

and is *orthonormal* if

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{i,j} = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$$

Normalizing an orthogonal set S refers to the process of multiplying each vector in S by the reciprocal of its norm in order to transform S into an orthonormal set of vectors. The following theorems apply.

Theorem 4.4

Suppose that S is an orthogonal set of non zero vectors. Then S is linearly independent.

Proof. Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ be an orthogonal set. Suppose that the scalars $a_1, a_2, \ldots, a_r$ satisfy

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_r\mathbf{u}_r = \mathbf{0}.$$

Computing the inner product of both sides of the above equation with $\mathbf{u}_i$ for $i = 1, 2, \ldots, r$, by the orthogonality of S we obtain

$$\begin{aligned} 0 &= \langle \mathbf{u}_i, \mathbf{0} \rangle = \langle \mathbf{u}_i, a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_r\mathbf{u}_r \rangle \\ &= a_1 \langle \mathbf{u}_i, \mathbf{u}_1 \rangle + \cdots + a_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle + \cdots + a_r \langle \mathbf{u}_i, \mathbf{u}_r \rangle = a_i \|\mathbf{u}_i\|^2, \quad \text{for } i = 1, 2, \ldots, r. \end{aligned}$$

This implies that $a_i = 0$ for $i = 1, 2, \ldots, r$ is the unique solution of the homogeneous equation. Therefore S is a linearly independent set. ■

Theorem 4.5: Pythagoras

Suppose that $S = \{\mathbf{u}, \mathbf{v}\}$ is an orthogonal set of two vectors. Then

$$\| \mathbf{u} + \mathbf{v} \|^2 = \| \mathbf{u} \|^2 + \| \mathbf{v} \|^2 .$$

Proof. Suppose that $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Then

$$\| \mathbf{u} + \mathbf{v} \|^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle = \| \mathbf{u} \|^2 + \| \mathbf{v} \|^2$$

which gives the desired result. ■

Example 4.15

Consider the canonical basis $\mathcal{E}_3 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ in $\mathbb{R}^3$ with the usual inner product. It is clear that

$$\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = \langle \mathbf{e}_1, \mathbf{e}_2 \rangle = \langle \mathbf{e}_2, \mathbf{e}_3 \rangle = 0 \quad \text{and} \quad \langle \mathbf{e}_1, \mathbf{e}_1 \rangle = \langle \mathbf{e}_2, \mathbf{e}_2 \rangle = \langle \mathbf{e}_3, \mathbf{e}_3 \rangle = 1.$$

Thus $\mathcal{E}_3$ is an orthonormal basis of $\mathbb{R}^3$. More generally, the canonical basis in $\mathbb{R}^n$ with the usual inner product is orthonormal for every n.

Example 4.16

Consider the vector space $\mathbb{P}_2$ endowed with the inner product defined by

$$\langle f, g \rangle = \int_0^1 f(t)\, g(t)\, dt.$$

Relative to the standard basis $\mathcal{S}_2 = \{1, t, t^2\}$, we have the Gram matrix

$$G^{\mathcal{S}_2} = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}.$$

Let $S = \{1, 1-2t, 2-12t+12t^2\}$ is an orthogonal set since

$$\langle 1, 1-2t \rangle = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} = 0$$

$$\langle 1, 2-12t+12t^2 \rangle = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \begin{bmatrix} 2 \\ -12 \\ 12 \end{bmatrix} = 0$$

$$\langle 1-2t, 2-12t+12t^2 \rangle = \begin{bmatrix} 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \begin{bmatrix} 2 \\ -12 \\ 12 \end{bmatrix} = 0$$

The most useful property of orthogonal bases is that the coordinates of any vector relative to an orthogonal basis can be computed directly one by one, without the need to solve a system of linear equations.

Theorem 4.6

Suppose that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthogonal basis for V. Then, for any $\mathbf{v} \in V$,

$$\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1 + \frac{\langle \mathbf{v}, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle}\mathbf{u}_2 + \cdots + \frac{\langle \mathbf{v}, \mathbf{u}_n \rangle}{\langle \mathbf{u}_n, \mathbf{u}_n \rangle}\mathbf{u}_n.$$

Proof. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis for V, any vector $\mathbf{v} \in V$ can be written as a linear combination of the basis; that is,

$$\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n$$

for suitable scalars $a_1, a_2, \ldots, a_n$. Using the orthogonality of the basis, we compute

$$\begin{aligned} \langle \mathbf{v}, \mathbf{u}_i \rangle &= \langle a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n, \mathbf{u}_i \rangle \\ &= a_1 \langle \mathbf{u}_1, \mathbf{u}_i \rangle + \cdots + a_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle + \cdots + a_n \langle \mathbf{u}_n, \mathbf{u}_i \rangle = a_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle, \quad \text{for } i = 1, 2, \ldots, n. \end{aligned}$$

This implies that $a_i = \langle \mathbf{v},\mathbf{u}_i \rangle / \langle \mathbf{u}_i,\mathbf{u}_i \rangle$ for $i = 1,2,\ldots,n$, which is the desired result. ■

The above scalars

$$a_i = \frac{\langle \mathbf{v},\mathbf{u}_i \rangle}{\langle \mathbf{u}_i,\mathbf{u}_i \rangle} = \frac{\langle \mathbf{v},\mathbf{u}_i \rangle}{\|\mathbf{u}_i\|^2}, \quad \text{for } i = 1,2,\ldots,n,$$

are called the *Fourier coefficients* of $\mathbf{v}$ with respect to $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$.

Consider a non zero vector $\mathbf{w}$ in an inner product space V. For any $\mathbf{v} \in V$, we have showed that

$$\lambda = \frac{\langle \mathbf{v},\mathbf{w} \rangle}{\langle \mathbf{w},\mathbf{w} \rangle} = \frac{\langle \mathbf{v},\mathbf{w} \rangle}{\| \mathbf{w} \|^2}$$

is the unique scalar such that $\hat{\mathbf{v}} = \mathbf{v} - \lambda \mathbf{w}$ is orthogonal to $\mathbf{w}$. The *projection* of $\mathbf{v}$ along $\mathbf{w}$, denoted by $proj_{\mathbf{w}}\mathbf{v}$, is defined as

$$\text{proj}_{\mathbf{w}}\mathbf{v} = \frac{\langle \mathbf{v},\mathbf{w} \rangle}{\langle \mathbf{w},\mathbf{w} \rangle}\mathbf{w} = \frac{\langle \mathbf{v},\mathbf{w} \rangle}{\| \mathbf{w} \|^2}\mathbf{w}.$$

The above notion is generalized as follows.

Theorem 4.7

Suppose that $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_r$ form an orthogonal set of non zero vectors in V. Let $\mathbf{v}$ be any vector in V. Define

$$\hat{\mathbf{v}} = \mathbf{v} - \left(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_r\mathbf{u}_r\right)$$

where

$$c_1 = \frac{\langle \mathbf{v},\mathbf{u}_1 \rangle}{\langle \mathbf{u}_1,\mathbf{u}_1 \rangle}, \quad c_2 = \frac{\langle \mathbf{v},\mathbf{u}_2 \rangle}{\langle \mathbf{u}_2,\mathbf{u}_2 \rangle}, \quad \ldots, \quad c_r = \frac{\langle \mathbf{v},\mathbf{u}_r \rangle}{\langle \mathbf{u}_r,\mathbf{u}_r \rangle}.$$

Then $\hat{\mathbf{v}}$ is orthogonal to $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_r$.

Proof. We verify that $\hat{\mathbf{v}}$ is orthogonal to $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_r$. For $i = 1,2,\ldots,r$, compute using the orthogonality of $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_r$,

$$\begin{aligned}\langle \hat{\mathbf{v}}, \mathbf{u}_i \rangle &= \langle \mathbf{v} - (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_r\mathbf{u}_r), \mathbf{u}_i \rangle \\ &= \langle \mathbf{v}, \mathbf{u}_i \rangle - c_1 \langle \mathbf{u}_1, \mathbf{u}_i \rangle - c_2 \langle \mathbf{u}_2, \mathbf{u}_i \rangle - \cdots - c_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle - \cdots - c_r \langle \mathbf{u}_r, \mathbf{u}_i \rangle \\ &= \langle \mathbf{v}, \mathbf{u}_i \rangle - c_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle \\ &= \langle \mathbf{v}, \mathbf{u}_i \rangle - \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \langle \mathbf{u}_i, \mathbf{u}_i \rangle = \langle \mathbf{v}, \mathbf{u}_i \rangle - \langle \mathbf{v}, \mathbf{u}_i \rangle = 0.\end{aligned}$$

Therefore $\hat{\mathbf{v}}$ is orthogonal to $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$. ■

Notice that each c_i in the theorem above is the Fourier coefficient of $\mathbf{v}$ along the corresponding vector $\mathbf{u}_i$.

The notion of the projection of a vector $\mathbf{v} \in V$ along a subspace W of V is defined as follows. If $W = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$ where the $\mathbf{w}_i$ form an orthogonal set, then the *projection* of $\mathbf{v}$ along W, denoted by $proj_W \mathbf{v}$, is defined as

$$\text{proj}_W \mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle} \mathbf{w}_1 + \frac{\langle \mathbf{v}, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle} \mathbf{w}_2 + \cdots + \frac{\langle \mathbf{v}, \mathbf{w}_r \rangle}{\langle \mathbf{w}_r, \mathbf{w}_r \rangle} \mathbf{w}_r.$$

4.6 Gram-Schmidt orthogonalization

Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of an inner product space V. One can use this to construct an orthogonal basis $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ of V as follows. Set

$$\begin{aligned}&\mathbf{w}_1 = \mathbf{v}_1, \\ &\mathbf{w}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{w}_1} \mathbf{v}_2, \\ &\mathbf{w}_3 = \mathbf{v}_3 - \text{proj}_{\mathbf{w}_1} \mathbf{v}_3 - \text{proj}_{\mathbf{w}_2} \mathbf{v}_3, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\mathbf{w}_n = \mathbf{v}_n - \text{proj}_{\mathbf{w}_1} \mathbf{v}_n - \text{proj}_{\mathbf{w}_2} \mathbf{v}_n - \cdots - \text{proj}_{\mathbf{w}_{n-1}} \mathbf{v}_n.\end{aligned}$$

In other words, for $k = 2, 3, \ldots, n$, we define

$$\mathbf{w}_k = \mathbf{v}_k - proj_{W_{k-1}} \mathbf{v}_k, \quad \text{where} \quad W_{k-1} = Span\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{k-1}\}.$$

By Theorem 4.7, each $\mathbf{w}_k$ is orthogonal to the preceding $\mathbf{w}$'s. Thus, $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ forms an orthogonal basis for V as claimed. Normalizing will then yield an orthonormal basis for V.

The above construction is known as the *Gram-Schmidt orthogonalization process*. Each vector $\mathbf{w}_k$ is a linear combination of $\mathbf{v}_k$ and the preceding

w's. Hence, it can easily be shown, by induction, that each $\mathbf{w}_k$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. Because taking multiples of vectors does not affect orthogonality, it may be simpler in hand calculations to clear fractions in any new $\mathbf{w}_k$, by multiplying $\mathbf{w}_k$ by an appropriate scalar, before obtaining the next $\mathbf{w}_{k+1}$.

Suppose $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ are linearly independent, and so they form a basis for

$$U = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}.$$

Applying Gram-Schmidt orthogonalization to the u's yields an orthogonal basis for U. The following theorems use the above algorithm and remarks.

Theorem 4.8

Suppose that $S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$ is an orthogonal basis for a subspace W of a vector space V. Then one may extend S to an orthogonal basis for V; that is, one may find vectors $\mathbf{w}_{r+1}, \ldots, \mathbf{w}_n$ such that $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ is an orthogonal basis for V.

Proof. Since S is a linearly independent set in V, we can extend S to a basis of V. That is, we can find vectors $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n$ such that $\{\mathbf{w}_1, \ldots, \mathbf{w}_r, \mathbf{u}_{r+1}, \ldots \mathbf{u}_n\}$ is a basis for V. For $k = r+1, \ldots, n$, we define

$$\mathbf{w}_k = \mathbf{u}_k - \text{proj}_{W_{k-1}} \mathbf{u}_k, \quad \text{where} \quad W_{k-1} = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{k-1}\}.$$

This is just the Gram-Schmidt orthogonalization process applied to $\{\mathbf{w}_1, \ldots, \mathbf{w}_r, \mathbf{u}_{r+1}, \ldots \mathbf{u}_n\}$, which yields an orthogonal basis for V. ■

Theorem 4.9

Let W be a subspace of an inner product space V. Then each vector $\mathbf{v} \in V$ can be expressed uniquely in the form

$$\mathbf{v} = \mathbf{w} + \widehat{\mathbf{w}}, \quad \mathbf{w} \in W, \quad \widehat{\mathbf{w}} \in W^{\perp}$$

Proof. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ be a basis for W. Applying the Gram-Schmidt orthogonalization process, we obtain an orthogonal basis for $W : S = \{\mathbf{w}_1, \ldots, \mathbf{w}_r\}$. Using S, for every $\mathbf{v} \in V$, we can write

$$\mathbf{v} = \text{proj}_W \mathbf{v} + (\mathbf{v} - \text{proj}_W \mathbf{v})$$

Clearly, $\mathrm{proj}_W\mathbf{v} \in W$. By Theorem 4.7, $\left(\mathbf{v} - \mathrm{proj}_W\mathbf{v}\right)$ is orthogonal to every vector in S and, thus, to every vector in W. Therefore, $\left(\mathbf{v} - \mathrm{proj}_W\mathbf{v}\right) \in W^{\perp}$.

Suppose that we can write

$$\mathbf{v} = \mathbf{w} + \widehat{\mathbf{w}}, \quad \mathbf{w} \in W, \quad \widehat{\mathbf{w}} \in W^{\perp}$$

Then subtracting the above expression from $\mathbf{v} = \mathrm{proj}_W\mathbf{v} + \left(\mathbf{v} - \mathrm{proj}_W\mathbf{v}\right)$, we get

$$\mathbf{0} = \left(\mathrm{proj}_W\mathbf{v} - \mathbf{w}\right) + \left(\mathbf{v} - \mathrm{proj}_W\mathbf{v} - \widehat{\mathbf{w}}\right).$$

Observe that $\left(\mathrm{proj}_W\mathbf{v} - \mathbf{w}\right) \in W$ and $\left(\mathbf{v} - \mathrm{proj}_W\mathbf{v} - \widehat{\mathbf{w}}\right) \in W^{\perp}$. Therefore,

$$\begin{aligned} 0 = \left\langle \mathbf{0}, \mathrm{proj}_W\mathbf{v} - \mathbf{w} \right\rangle &= \left\langle \left(\mathrm{proj}_W\mathbf{v} - \mathbf{w}\right) + \left(\mathbf{v} - \mathrm{proj}_W\mathbf{v} - \widehat{\mathbf{w}}\right), \mathrm{proj}_W\mathbf{v} - \mathbf{w} \right\rangle \\ &= \left\langle \mathrm{proj}_W\mathbf{v} - \mathbf{w}, \mathrm{proj}_W\mathbf{v} - \mathbf{w} \right\rangle. \end{aligned}$$

This implies that $proj_W\mathbf{v} - \mathbf{w} = \mathbf{0}$ or, equivalently, $\mathbf{w} = proj_W\mathbf{v}$. Similarly,

$$\begin{aligned} 0 = \left\langle \mathbf{0}, \mathbf{v} - \mathrm{proj}_W\mathbf{v} - \widehat{\mathbf{w}} \right\rangle &= \left\langle \left(\mathrm{proj}_W\mathbf{v} - \mathbf{w}\right) + \left(\mathbf{v} - \mathrm{proj}_W\mathbf{v} - \widehat{\mathbf{w}}\right), \mathbf{v} - \mathrm{proj}_W\mathbf{v} - \widehat{\mathbf{w}} \right\rangle \\ &= \left\langle \mathbf{v} - \mathrm{proj}_W\mathbf{v} - \widehat{\mathbf{w}}, \mathbf{v} - \mathrm{proj}_W\mathbf{v} - \widehat{\mathbf{w}} \right\rangle. \end{aligned}$$

Therefore, $\mathbf{v} - \mathrm{proj}_W\mathbf{v} - \widehat{\mathbf{w}} = \mathbf{0}$ or, equivalently, $\widehat{\mathbf{w}} = \mathbf{v} - \mathrm{proj}_W\mathbf{v}$. This proves that each $\mathbf{v} \in V$ is expressed uniquely as $\mathbf{v} = \mathrm{proj}_W\mathbf{v} + \left(\mathbf{v} - \mathrm{proj}_W\mathbf{v}\right)$. ■

Example 4.17

Apply the Gram-Schmidt orthogonalization process to find an orthogonal basis for the subspace U of $\mathbb{R}^4$ spanned by

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 5 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ -3 \\ -4 \\ -2 \end{bmatrix}.$$

Therefore, we first set $\mathbf{w}_1 = \mathbf{v}_1$. Then we set

$$\mathbf{w}_2 = \mathbf{v}_2 - \mathrm{proj}_{\mathbf{w}_1}\mathbf{v}_2 = \mathbf{v}_2 - \frac{12}{4}\mathbf{w}_1 = \begin{bmatrix} -2 \\ -1 \\ 1 \\ 2 \end{bmatrix}.$$

Finally, compute

$$\mathbf{v}_3 - \text{proj}_{\mathbf{w}_1}\mathbf{v}_3 - \text{proj}_{\mathbf{w}_2}\mathbf{v}_3 = \mathbf{v}_3 - \frac{(-8)}{4}\mathbf{w}_1 - \frac{(-7)}{10}\mathbf{w}_2 = \begin{bmatrix} \frac{8}{5} \\ -\frac{17}{10} \\ -\frac{13}{10} \\ \frac{7}{5} \end{bmatrix}.$$

Multiplying the last vector by 10 to clear fractions, we set

$$\mathbf{w}_3 = \begin{bmatrix} 16 \\ -17 \\ -13 \\ 14 \end{bmatrix}.$$

Thus, $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ form an orthogonal basis for U.

Example 4.18

Let V be the vector space of polynomials with inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(t)\, g(t)\, dt$$

Apply the Gram-Schmidt orthogonalization process to $\{1, t, t^2, t^3\}$ to find an orthogonal basis for $\mathbf{P}_3$.

Here we use the fact that, for $r + s = n$,

$$\left\langle t^r, t^s \right\rangle = \int_{-1}^{1} t^n dt = \left.\frac{t^{n+1}}{n+1}\right|_{-1}^{1} = \begin{cases} 2/(n+1), & \text{when } n \text{ is even} \\ 0, & \text{when } n \text{ is odd} \end{cases}$$

We compute

$$f_0(t) = 1$$
$$f_1(t) = t - proj_{f_0} t = t - 0 = t$$
$$f_2(t) = t^2 - proj_{f_0} t^2 - proj_{f_1} t^2 = t^2 - \frac{1}{3}$$
$$f_3(t) = t^3 - proj_{f_0} t^3 - proj_{f_1} t^3 - proj_{f_2} t^3 = t^3 - \frac{3}{5} t.$$

Thus, $\left\{1, t, t^2 - \frac{1}{3}, t^3 - \frac{3}{5}t\right\}$ is the required orthogonal basis.

4.7 Orthogonal projections and minimization

Suppose that U is a subspace of an inner product space V. Then each vector $\mathbf{v} \in V$ can be written uniquely in the form

$$\mathbf{v} = \mathbf{u} + \mathbf{w},$$

where $\mathbf{u} \in U$ and $\mathbf{w} \in U^{\perp}$. We use this decomposition to define a linear mapping on V, denoted by $proj_U$, called the *orthogonal projection* of V along U. For $\mathbf{v} \in V$, we define $proj_U\mathbf{v}$ to be the vector $\mathbf{u}$ in the decomposition above.

The orthogonal projection along a subspace satisfies the following properties.

Theorem 4.10

Let U is a subspace of an inner product space V. Then,

1. range $\mathrm{proj}_U = U$
2. ker $\mathrm{proj}_U = U^{\perp}$
3. $\mathbf{v} - \mathrm{proj}_U\mathbf{v} \in U^{\perp}$ for every $\mathbf{v} \in V$.
4. $\mathrm{proj}_U^2 = \mathrm{proj}_U$
5. $\|\mathrm{proj}_U\mathbf{v}\| \leqslant \|\mathbf{v}\|$ for every $\mathbf{v} \in V$.

Proof. 1. If $\mathbf{u} \in U$, then $\mathbf{u} = \mathrm{proj}_U\mathbf{u}$. Therefore, $\mathbf{u} \in$ range proj_U. Conversely, if $\mathbf{u} \in$ range proj_U, then by definition, $\mathbf{u} \in U$. Altogether, we have $U \subseteq$ range proj_U and range $\mathrm{proj}_U \subseteq U$, which implies range $\mathrm{proj}_U = U$.

2. If $\mathbf{u} \in U^{\perp}$, then it is expressed uniquely as $\mathbf{u} = \mathbf{0} + \mathbf{u}$ since it has no component in U other than the zero vector. Then $\mathrm{proj}_U\mathbf{u} = \mathbf{0}$ which means that $\mathbf{u} \in$ ker proj_U. Conversely, if $\mathbf{u} \in$ ker proj_U then $\mathrm{proj}_U\mathbf{u} = \mathbf{0}$. But $\mathbf{u}$ is expressed uniquely as $\mathbf{u} = \mathbf{0} + \mathbf{w}$ where $\mathbf{w} \in U^{\perp}$. It follows that $\mathbf{u} \in U^{\perp}$. Altogether, we have $U^{\perp} \subseteq$ ker proj_U and ker $\mathrm{proj}_U \subseteq U^{\perp}$, which implies ker $\mathrm{proj}_U = U^{\perp}$.

3. Since $\mathbf{v}$ can be expressed as $\mathbf{v} = \mathbf{u} + \mathbf{w}$ where $\mathbf{w} \in U^{\perp}$ and, by definition, $\mathbf{u} = \mathrm{proj}_U\mathbf{v}$. It follows that $\mathbf{w} = \mathbf{v} - \mathrm{proj}_U\mathbf{v}$ and, thus, $\mathbf{v} - \mathrm{proj}_U\mathbf{v} \in U^{\perp}$.

4. Consider any vector $\mathbf{v} \in V$. Then, by the first and second statements above, it can be written as $\mathbf{v} = \mathrm{proj}_U\mathbf{v} + \mathbf{w}$ where $\mathbf{w} \in \ker \mathrm{proj}_U$. Therefore, applying the orthogonal projection to both sides of this equation, we get

$$\mathrm{proj}_U\mathbf{v} = \mathrm{proj}_U^2\mathbf{v} + \mathrm{proj}_U\mathbf{w} = \mathrm{proj}_U^2\mathbf{v}$$

But this is true for any $\mathbf{v} \in V$. Then, $\mathrm{proj}_U^2 = \mathrm{proj}_U$.

5. For every $\mathbf{v} \in V$, by Pythagoras,

$$\left\|\text{proj}_U\mathbf{v}\right\|^2 \leqslant \left\|\text{proj}_U\mathbf{v}\right\|^2 + \left\|\mathbf{v} - \text{proj}_U\mathbf{v}\right\|^2 = \| \mathbf{v} \|^2$$

Taking the positive square root of both sides, we get $\left\|\text{proj}_U\mathbf{v}\right\| \leqslant \| \mathbf{v} \|$.

■

By the definition of orthogonal projection, if $U = \text{Span}\left\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\right\}$ where the $\mathbf{u}_i$ form an orthogonal basis, then

$$\text{proj}_U\mathbf{v} = \frac{\left\langle \mathbf{v}, \mathbf{u}_1 \right\rangle}{\left\langle \mathbf{u}_1, \mathbf{u}_1 \right\rangle}\mathbf{u}_1 + \frac{\left\langle \mathbf{v}, \mathbf{u}_2 \right\rangle}{\left\langle \mathbf{u}_2, \mathbf{u}_2 \right\rangle}\mathbf{u}_2 + \cdots + \frac{\left\langle \mathbf{v}, \mathbf{u}_r \right\rangle}{\left\langle \mathbf{u}_r, \mathbf{u}_r \right\rangle}\mathbf{u}_r.$$

The following problem often arises: given a subspace U of V and a vector $\mathbf{v} \in V$, find a vector $\mathbf{u} \in U$ such that $\| \mathbf{v} - \mathbf{u} \|$ is as small as possible. The next theorem shows that this minimization problem is solved by taking $\mathbf{u} = \text{proj}_U\mathbf{v}$.

Theorem 4.11

Suppose that U is a subspace of V and $\mathbf{v} \in V$. Then

$$\left\|\mathbf{v} - \text{proj}_U\mathbf{v}\right\| \leq \| \mathbf{v} - \mathbf{u} \|$$

for every $\mathbf{u} \in U$. Furthermore, if $\mathbf{u} \in U$ and the inequality above is an equality, then $\mathbf{u} =$ projv.

Proof. Suppose that $\mathbf{u} \in U$. Then

$$\begin{aligned} \left\|\mathbf{v} - \text{proj}_U\mathbf{v}\right\|^2 &\leq \left\|\mathbf{v} - \text{proj}_U\mathbf{v}\right\|^2 + \left\|\text{proj}_U\mathbf{v} - \mathbf{u}\right\|^2 \\ &= \left\|\left(\mathbf{v} - \text{proj}_U\mathbf{v}\right) + \left(\text{proj}_U\mathbf{v} - \mathbf{u}\right)\right\|^2 \\ &= \| \mathbf{v} - \mathbf{u} \|^2, \end{aligned}$$

where the first equality comes from Pythagoras, which applies because $\left(\mathbf{v} - \text{proj}_U\mathbf{v}\right) \in U^\perp$ and $\left(\text{proj}_U\mathbf{v} - \mathbf{u}\right) \in U$. Taking the square roots gives the desired inequality, which is an equality if and only if

$$\left\|\mathbf{v} - \text{proj}_U\mathbf{v}\right\|^2 = \left\|\mathbf{v} - \text{proj}_U\mathbf{v}\right\|^2 + \left\|\text{proj}_U\mathbf{v} - \mathbf{u}\right\|^2.$$

This happens if and only if $\left\|\text{proj}_U\mathbf{v} - \mathbf{u}\right\| = 0$, which happens if and only if $\mathbf{u} =$ projv. ■

Example 4.19

Let V be the vector space of polynomials with inner product

$$\langle f,g\rangle = \int_{-1}^{1} f(t)g(t)dt$$

Then, $\{1,t\}$ is an orthogonal basis of $\mathbf{P}_1$. The polynomial $q(t) \in \mathbf{P}_1$ such that $\left\|t^{2k} - q(t)\right\|$ is as small as possible is

$$q(t) = \text{proj}_{\mathbf{P}_1} t^{2k} = \frac{1}{2k+1},$$

and the polynomial $r(t) \in \mathbf{P}_1$ such that $\left\|t^{2k+1} - r(t)\right\|$ is as small as possible is

$$r(t) = \text{proj}_{\mathbf{P}_1} t^{2k+1} = \frac{3t}{2k+3}.$$

4.8 Least-squares problems

Inconsistent systems of equations $A\mathbf{x} = \mathbf{b}$ arise often in applications. When a solution is demanded and none exists, the best one can do is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$. Think of $A\mathbf{x}$ as an *approximation* to $\mathbf{b}$. The smaller the distance between $\mathbf{b}$ and $A\mathbf{x}$, given by $\|\mathbf{b} - A\mathbf{x}\|$, the better the approximation. The *general least-squares problem* is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible.

Definition 4.2

If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a *least-squares solution* of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

The most important aspect of the least-squares problem is that no matter what $\mathbf{x}$ we select, the vector $A\mathbf{x}$ will necessarily be in the column space, Col$A$. So we seek an $\mathbf{x}$ that makes $A\mathbf{x}$ the closest point in Col$A$ to $\mathbf{b}$. Of course, if $\mathbf{b}$ happens to be in Col$A$, then $\mathbf{b}$ *is* $A\mathbf{x}$ for some $\mathbf{x}$, and such an $\mathbf{x}$ is a "least-squares solution."

Given A and $\mathbf{b}$ as above, let

$$\hat{\mathbf{b}} = \text{proj}_{\text{Col}\,A}\mathbf{b}$$

where the orthogonal projection of $\mathbf{b}$ along Col$A$ is defined with respect to a given inner product. Because $\hat{\mathbf{b}}$ is in the column space of A, the equation $A\mathbf{x} = \mathbf{b}$ *is* consistent, and there is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}}$$

Since $\hat{\mathbf{b}}$ is the closest point in $ColA$ to $\mathbf{b}$, a vector $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\hat{\mathbf{x}}$ satisfies $A\mathbf{x} = \mathbf{b}$, which may have many solutions if the equation has free variables.

Suppose that $\hat{\mathbf{x}}$ satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. The projection $\hat{\mathbf{b}}$ has the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to $ColA$, so $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A. If $\mathbf{a}_j$ is any column of A, then $\left\langle \mathbf{a}_j, \mathbf{b} - A\hat{\mathbf{x}} \right\rangle = 0$, and $\mathbf{a}_j^\top G(\mathbf{b} - A\hat{\mathbf{x}}) = 0$, where G is the Gram matrix relative to an appropriate basis. Since $\mathbf{a}_j^\top$ is a row of $A^\top$,

$$A^\top G(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$$

These calculations show that each least-squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the equation

$$A^\top GA\hat{\mathbf{x}} = A^\top G\mathbf{b}$$

This matrix equation represents a system of equations called the *normal equations* for $A\mathbf{x} = \mathbf{b}$.

Theorem 4.12

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the non empty set of solutions of the normal equations $A^\top GA\hat{\mathbf{x}} = A^\top G\mathbf{b}$.

Proof. As shown above, the set of least-squares solutions is non empty and each least-squares solution $\hat{\mathbf{x}}$ satisfies the normal equations. Conversely, suppose that $\hat{\mathbf{x}}$ satisfies $A^\top GA\hat{\mathbf{x}} = A^\top G\mathbf{b}$. Then $\hat{\mathbf{x}}$ satisfies $A^\top G(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$, which shows that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the

rows of $A^\top$ and hence is orthogonal to the columns of A. Since the columns of A span $ColA$, the vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to all of $ColA$. Hence the equation

$$\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}})$$

is the decomposition of $\mathbf{b}$ into the sum of a vector in $ColA$ and a vector in $(ColA)^\perp$. By the uniqueness of the orthogonal decomposition, $A\hat{\mathbf{x}}$ must be the orthogonal projection of $\mathbf{b}$ along $ColA$. That is, $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ and $\hat{\mathbf{x}}$ is a least-squares solution.

Example 4.20

Find the least-squares solution with respect to the usual inner product in $\mathbb{R}^2$ of the inconsistent system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

In this case, we may assume that everything is expressed relative to the canonical basis, thus the Gram matrix is the identity matrix I_2, so the normal equations read $A^\top A\mathbf{x} = A^\top \mathbf{b}$. To use the normal equations, we compute:

$$A^\top A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^\top \mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

The equation $A^\top A\mathbf{x} = A^\top \mathbf{b}$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

Row operations can be used to solve this system, but since $A^\top A$ is invertible and 2×2, it is probably faster to compute

$$\left(A^\top A\right)^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then to solve $A^\top A\mathbf{x} = A^\top \mathbf{b}$ as

$$\hat{\mathbf{x}} = \left(A^\top A\right)^{-1} A^\top \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

In many calculations $A^\top A$ is invertible, but this is not always the case as in the following example.

Example 4.21

Find a least-squares solution relative to the usual inner product of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$$

As in the previous example, the Gram matrix of the inner product relative to the standard basis is the identity matrix. Compute

$$A^\top A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^\top \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}.$$

The augmented matrix for $A^\top A\mathbf{x} = A^\top \mathbf{b}$ is

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

So the general least-squares solution of $\mathbf{x} = \mathbf{b}$ has the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

4.9 Orthogonal diagonalization

Let V be an inner product space. A *symmetric* linear mapping $T : V \to V$ relative to an inner product is a linear mapping such that $\langle T(\mathbf{u}), \mathbf{v}\rangle = \langle \mathbf{u}, T(\mathbf{v})\rangle$ for all $\mathbf{u}, \mathbf{v} \in V$.

Theorem 4.13

Suppose that $T : V \to V$ is a symmetric linear mapping. If $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is an orthonormal basis for V and $M = (m_{j,k})$ is the matrix representation of T relative to $\mathcal{B}$, then $M^\top = M$.

Proof. Recall that we obtain the k-th column of M by writing $T(\mathbf{b}_k)$ as a linear combination of the basis $\mathcal{B}$; the scalars used in this linear combination then become the k-th column of M. Since $\mathcal{B}$ is orthonormal, we know how to write $T(\mathbf{b}_k)$ as a linear combination of $\mathcal{B}$:

$$T(\mathbf{b}_k) = \langle T(\mathbf{b}_k), \mathbf{b}_1\rangle \mathbf{b}_1 + \cdots + \langle T(\mathbf{b}_k), \mathbf{b}_n\rangle \mathbf{b}_n$$

Thus $m_{j,k} = \langle T(\mathbf{b}_k), \mathbf{b}_j\rangle$. Using the symmetry of T and the symmetry of the inner product, we get

$$m_{j,k} = \langle T(\mathbf{b}_k), \mathbf{b}_j\rangle = \langle \mathbf{b}_k, T(\mathbf{b}_j)\rangle = \langle T(\mathbf{b}_j), \mathbf{b}_k\rangle = m_{k,j}$$

Therefore, $M = (m_{j,k}) = (m_{k,j}) = M^\top$. ■

One of the most important properties of symmetric linear mappings is stated in the following theorem.

Theorem 4.14

If $T : V \to V$ is a symmetric linear mapping, then any two eigenvectors from different eigenspaces are orthogonal.

Proof. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors that correspond to distinct eigenvalues, say, λ_1 and λ_2. To show that $\langle \mathbf{v}_1, \mathbf{v}_2\rangle = 0$, compute

$$\lambda_1\langle \mathbf{v}_1, \mathbf{v}_2\rangle = \langle \lambda_1\mathbf{v}_1, \mathbf{v}_2\rangle = \langle T(\mathbf{v}_1), \mathbf{v}_2\rangle = \langle \mathbf{v}_1, T(\mathbf{v}_2)\rangle = \langle \mathbf{v}_1, \lambda_2\mathbf{v}_2\rangle = \lambda_2\langle \mathbf{v}_1, \mathbf{v}_2\rangle$$

Hence $(\lambda_1 - \lambda_2)\langle \mathbf{v}_1, \mathbf{v}_2\rangle = 0$. But $\lambda_1 - \lambda_2 \neq 0$ so $\langle \mathbf{v}_1, \mathbf{v}_2\rangle = 0$. ■

A linear mapping $T : V \to V$ is said to be orthogonally diagonalizable if V has an orthonormal basis consisting of eigenvectors of T. The proof of the following theorem is quite involved and is ommitted.

Theorem 4.15

Let V be an inner product space. A linear mapping $T: V \to V$ is orthogonally diagonalizable if and only if T is symmetric relative to the inner product.

This theorem is rather amazing, because the work in the previous chapter would suggest that it is usually impossible to tell when a linear mapping is diagonalizable. But this is not the case for symmetric matrices.

The next example treats a linear mapping whose eigenvalues are not all distinct.

Example 4.22

Orthogonally diagonalize the linear mapping $T: \mathbb{R}^3 \to \mathbb{R}^3$ defined relative to the canonical basis by

$$T(\mathbf{x}) = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix} \mathbf{x}.$$

The usual calculations produce the characteristic equation

$$0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2).$$

A basis for the eigenspace associated with $\lambda_1 = 7$ is

$$\left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} \right\}$$

and a basis for $\lambda_2 = -2$ is

$$\left\{ \mathbf{v}_3 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix} \right\}$$

Although $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, they are not orthogonal. Applying the Gram-Schmidt orthogonalization process and normalization, we obtain the following orthonormal basis for the eigenspace associated with λ_1:

$$\left\{ \mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix} \right\}.$$

An orthonormal basis for the eigenspace associated with λ_2:

$$\left\{ \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix} \right\}.$$

Hence, $\mathcal{C} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis for $\mathbb{R}^3$ and the matrix representation of T relative to $\mathcal{C}$ is

$$\begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

An interesting observation is that if $P = [\mathbf{u}_1 \ldots \mathbf{u}_n]$ is a square matrix whose columns satisfy $\mathbf{u}_i^\top \mathbf{u}_j = 0$ for $i \neq j$ and $\mathbf{u}_j^\top \mathbf{u}_j = 1$ (as in the example above), then $P^\top P = I_n$. This implies that $P^{-1} = P^\top$. Moreover, it is not difficult to verify that $P^{-1} = P^\top$ if and only if the columns of P satisfy $\mathbf{u}_i^\top \mathbf{u}_j = 0$ for $i \neq j$ and $\mathbf{u}_j^\top \mathbf{u}_j = 1$.