Chapter 7
Normed and Inner Product Spaces

7.1 Normed Vector Spaces

Recall that the length of a vector $\mathbf{x} = (x_1, x_2, x_3)$ in R^3 is

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

Norm is a generalization of the concept of length to vectors of abstract spaces.

7.1.1 Vector Norms

Let $\mathbf{X}$ be a vector space over F, where F is either R or C. A function which associates with every vector $\mathbf{x} \in \mathbf{X}$ a real value denoted $\|\mathbf{x}\|$ is called a **norm** on $\mathbf{X}$ if it satisfies the following.

N1. $\|\mathbf{x}\| > 0$ for any $\mathbf{x} \neq \mathbf{0}$.

N2. $\|c\mathbf{x}\| = |c|\,\|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbf{X}$ and $c \in \mathsf{F}$.

N3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{X}$.

A vector space with a norm defined on it is called a **normed vector space**.

Note that property N2 implies that $\|\mathbf{0}\| = 0$. Property N3 is known as the **triangle inequality**.

If $\mathbf{x}$ is a nonzero vector in $\mathbf{X}$, then $\frac{1}{\|\mathbf{x}\|}\mathbf{x}$ has unity norm, and is called a **unit vector**.

Example 7.1

A simple norm on R^n is

$$\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_n|$$

which is called the **uniform norm**. Obviously, it satisfies properties N1 and N2, and N3 follows from the property of the absolute value that $|a + b| \leq |a| + |b|$ for $a, b \in \mathsf{R}$.

In fact, for any real number $p \geq 1$

$$\|(x_1, \ldots, x_n)\|_p = \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p} \tag{7.1}$$

is a norm on R^n, which reduces to the uniform norm for $p = 1$. Again, properties N1 and N2 are satisfied trivially. Proof of the triangle inequality for $p > 1$ is left to the reader (see Exercises 7.3 and 7.4).

In particular,

$$\| \mathbf{x} \|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2}$$

is a norm, called the **Euclidean norm**. The Euclidean norm is a straightforward generalization of length in $\mathbf{R}^2$ or $\mathbf{R}^3$.

Letting $p \to \infty$, we observe that

$$\| \mathbf{x} \|_\infty = \max \{ |x_1|, \ldots, |x_n| \}$$

is also a norm on $\mathbf{R}^n$, called the **infinity norm**.

As an illustration, if $\mathbf{x} = (4, -12, 3)$ then

$$\begin{aligned} \| \mathbf{x} \|_1 &= 4 + 12 + 3 &&= 19 \\ \| \mathbf{x} \|_2 &= \sqrt{16 + 144 + 9} &&= 13 \\ \| \mathbf{x} \|_\infty &= \max \{4, 12, 3\} &&= 12 \end{aligned}$$

Corresponding norms on $\mathbf{C}^n$, $\mathbf{R}^{n\times 1}$ and $\mathbf{C}^{n\times 1}$ are defined similarly.

Example 7.2

Recall from Example 3.4 that a real n-tuple $(x_1, x_2, \ldots, x_n)$ can be viewed as a function $f : \mathbf{n} \to \mathbf{R}$ such that

$$f[k] = x_k \,, \quad k \in \mathbf{n} \tag{7.2}$$

Hence for any $p \geq 1$

$$\| f \| = (\sum_{k=1}^{n} |f(k)|^p)^{1/p}$$

defines a norm on the function space $\mathcal{F}(\mathbf{n}, \mathbf{R})$.

Now consider the function space $\mathcal{C}_0(\mathcal{I}, \mathbf{R})$ of real-valued continuous functions defined on a closed interval $\mathcal{I} = [\, a, b\,]$. Replacing the summation in (7.2) with an integral, we observe that for any $p \geq 1$

$$\| f \|_p = (\int_a^b |f(t)|^p \, dt \,)^{1/p} \tag{7.3}$$

is a norm on $\mathcal{C}_0(\mathcal{I}, \mathbf{R})$ (see Exercise 7.5).[1] In particular,

$$\begin{aligned} \| f \|_1 &= \int_a^b |f(t)| \, dt \\ \| f \|_2 &= (\int_a^b |f(t)|^2 \, dt)^{1/2} \\ \| f \|_\infty &= \max_{a \leq t \leq b} \{ |f(t)| \} \end{aligned}$$

are norms on $\mathcal{C}_0(\mathcal{I}, \mathbf{R})$, which are also called the uniform, Euclidean and infinity norms, respectively.

[1] The reason for restricting our attention to $\mathcal{C}_0(\mathcal{I}, \mathbf{R})$ rather than $\mathcal{F}(\mathcal{I}, \mathbf{R})$ is to guarantee the convergence of the integral in (7.3).

As an illustration, if

$$f(t) = \sin t\,, \quad -\pi \le t \le \pi$$

then

$$\begin{aligned}
\|\, f\,\|_1 &= \int_{-\pi}^{\pi} |\sin t|\, dt = 2\int_0^{\pi} \sin t\, dt = 4 \\
\|\, f\,\|_2 &= (\int_{-\pi}^{\pi} \sin^2 t\, dt)^{1/2} = (\,\frac{1}{2}\int_{-\pi}^{\pi} (1-\cos 2t)\, dt\,)^{1/2} = \sqrt{\pi} \\
\|\, f\,\|_\infty &= \max_{-\pi\le t\le\pi} \{\, |\sin t|\,\} = 1
\end{aligned}$$

Example 7.3

Consider the vector space $\mathcal{C}_0(\mathcal{I}, \mathbf{R}^{n\times 1})$, where $\mathcal{I} = [\,a, b\,]$ is a closed interval. Recall from Example 3.5 that a vector-valued function $\mathbf{f} \in \mathcal{C}_0(\mathcal{I}, \mathbf{R}^{n\times 1})$ can be viewed as a stack of scalar functions $f_1, \ldots, f_n$ such that $\mathbf{f}(t) = \mathrm{col}\,[\, f_1(t), \ldots, f_n(t)\,]$ for every $t \in \mathcal{I}$.

Let

$$\|\,\mathbf{f}\,\| = \max_{t\in\mathcal{I}} \{\sum_{i=1}^{n} |\, f_i(t)\,|\,\} \tag{7.4}$$

Then $\|\cdot\|$ trivially satisfies properties N1 and N2 of a norm. Since

$$\begin{aligned}
\|\,\mathbf{f}+\mathbf{g}\,\| &= \max_{t\in\mathcal{I}} \{\sum_{i=1}^{n} |\, f_i(t)+g_i(t)\,|\,\} \\
&\le \max_{t\in\mathcal{I}} \{\sum_{i=1}^{n} |\, f_i(t)\,| + \sum_{i=1}^{n} |\, g_i(t)\,|\,\} \\
&\le \max_{t\in\mathcal{I}} \{\sum_{i=1}^{n} |\, f_i(t)\,|\,\} + \max_{t\in\mathcal{I}} \{\sum_{i=1}^{n} |\, g_i(t)\,|\,\} \quad = \quad \|\,\mathbf{f}\,\| + \|\,\mathbf{g}\,\|
\end{aligned}$$

it also satisfies property N3. Hence it is a norm on $\mathcal{C}_0(\mathcal{I}, \mathbf{R}^{n\times 1})$.

The summation in (7.4) is the uniform norm of the vector $\mathbf{f}(t) \in \mathbf{R}^{n\times 1}$. Let us denote its value by $\nu^{\mathbf{f}}(t)$ to indicate its dependence on $\mathbf{f}$ and t:

$$\nu^{\mathbf{f}}(t) = \|\,\mathbf{f}(t)\,\|_1 \quad \text{for all } \ t \in \mathcal{I} \tag{7.5}$$

(7.5) defines a scalar continuous function $\nu^{\mathbf{f}} \in \mathcal{C}_0(\mathcal{I}, \mathbf{R})$. The maximum value of $\nu^{\mathbf{f}}$ on $\mathcal{I}$ is its infinity norm. Thus (7.4) can be rewritten as

$$\|\,\mathbf{f}\,\| = \|\,\nu^{\mathbf{f}}\,\|_\infty \tag{7.6}$$

The norms in (7.5) and (7.6) bear no special significance in defining $\|\,\mathbf{f}\,\|$, and they can be replaced with arbitrary norms. By letting

$$\nu_p^{\mathbf{f}}(t) = \|\,\mathbf{f}(t)\,\|_p \quad \text{for all } \ t \in \mathcal{I} \tag{7.7}$$

and

$$\|\,\mathbf{f}\,\|_{p,q} = \|\,\nu_p^{\mathbf{f}}\,\|_q \tag{7.8}$$

for arbitrary $p, q \ge 1$, we can define many different norms on $\mathcal{C}_0(\mathcal{I}, \mathbf{R}^{n\times 1})$. For details, the reader is referred to Exercises 7.7 and 7.8.

7.1.2 Matrix Norms

Since $\mathbf{C}^{m\times n}$ is a vector space, we may attempt to define a norm for matrices. For example, it is rather easy to show that

$$\| A \|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} \tag{7.9}$$

is a matrix norm, called the **Frobenius norm**. For $m = 1$ or $n = 1$, it reduces to the Euclidean norm. In fact, Frobenius norm of an $m \times n$ matrix A is the Euclidean norm of an $mn \times 1$ column vector formed by stacking the columns of A. In addition to the properties of a norm, the Frobenius norm also satisfies the consistency condition

$$\| AB \|_F \leq \| A \|_F \| B \|_F$$

which is useful and often desired in matrix operations.

For any $p \geq 1$, let

$$\| A \|_p = \max_{\mathbf{x} \neq \mathbf{0}} \left\{ \frac{\| A\mathbf{x} \|_p}{\| \mathbf{x} \|_p} \right\} = \max_{\| \mathbf{x} \|_p = 1} \{ \| A\mathbf{x} \|_p \} \tag{7.10}$$

Then we have the following properties of $\| \cdot \|_p$.

a) Since $\| A\mathbf{x} \|_p \geq 0$ and $\| \mathbf{x} \|_p > 0$ for $\mathbf{x} \neq \mathbf{0}$, $\| A \|_p \geq 0$. If $A \neq O$, then there exists $\mathbf{x} \neq \mathbf{0}_{n\times 1}$ such that $A\mathbf{x} \neq \mathbf{0}_{m\times 1}$, so that

$$0 < \| A\mathbf{x} \|_p \leq \| A \|_p \| \mathbf{x} \|_p$$

that is $\| A \|_p > 0$.

b) For any scalar c

$$\| cA \|_p = \max_{\| \mathbf{x} \|_p = 1} \{ | c | \, \| A\mathbf{x} \|_p \} = | c | \max_{\| \mathbf{x} \|_p = 1} \{ \| A\mathbf{x} \|_p \} = | c | \, \| A \|_p$$

c) Since $\| (A + B)\mathbf{x} \|_p \leq \| A\mathbf{x} \|_p + \| B\mathbf{x} \|_p$, we have

$$\begin{aligned} \| A + B \|_p &= \max_{\| \mathbf{x} \|_p = 1} \{ \| A\mathbf{x} + B\mathbf{x} \|_p \} \\ &\leq \max_{\| \mathbf{x} \|_p = 1} \{ \| A\mathbf{x} \|_p + \| B\mathbf{x} \|_p \} \\ &\leq \max_{\| \mathbf{x} \|_p = 1} \{ \| A\mathbf{x} \|_p \} + \max_{\| \mathbf{x} \|_p = 1} \{ \| B\mathbf{x} \|_p \} \\ &= \| A \|_p + \| B \|_p \end{aligned}$$

Hence $\| \cdot \|_p$ is a norm on $\mathbf{C}^{m\times n}$, called the **matrix norm subordinate to the p-vector norm**.[2]

[2] In this definition, a matrix is interpreted as a mapping between two vector spaces rather than a vector: For $\mathbf{x} \neq \mathbf{0}$, the ratio $\| A\mathbf{x} \|_p / \| \mathbf{x} \|_p$ is the factor by which the strength $\| \mathbf{x} \|_p$ of the vector $\mathbf{x}$ (as measured by its p-norm) changes while undergoing the transformation represented by the matrix A. Hence $\| A \|_p$ represents the maximum possible change in the strength of a vector $\mathbf{x}$ when it is transformed into $A\mathbf{x}$. This interpretation of the norm of a matrix can be generalized to linear transformations between arbitrary normed vector spaces (see Exercise 7.13).

All matrix p-norms satisfy the consistency condition. This follows from the fact that

$$\| AB\mathbf{x} \|_p \le \| A \|_p \| B\mathbf{x} \|_p \le \| A \|_p \| B \|_p \| \mathbf{x} \|_p$$

so that

$$\| AB \|_p = \max_{\mathbf{x}\neq\mathbf{0}} \{ \frac{\| AB\mathbf{x} \|_p}{\| \mathbf{x} \|_p} \} \le \| A \|_p \| B \|_p$$

It is left to the reader (see Exercise 7.11) to show that the matrix norm subordinate to the uniform vector norm is the maximum column sum

$$\| A \|_1 = \max_{1\le j\le n} \{ \sum_{i=1}^{m} |a_{ij}| \} = \max_{1\le j\le n} \{ \| \mathbf{a}_j \|_1 \}$$

where $\mathbf{a}_j$ denotes the jth column of A, and the matrix norm subordinate to the infinity vector norm is the maximum row sum

$$\| A \|_\infty = \max_{1\le i\le m} \{ \sum_{j=1}^{n} |a_{ij}| \} = \max_{1\le i\le m} \{ \| \boldsymbol{\alpha}_i \|_1 \}$$

where $\boldsymbol{\alpha}_i$ denotes the ith row of A.[3] Note that

$$\| A^h \|_\infty = \| A \|_1 , \quad \| A^h \|_1 = \| A \|_\infty$$

We shall consider the matrix norm subordinate to the Euclidean vector norm in Chapter 8.

MATLAB provides a built-in function to compute the uniform, Euclidean, infinity and Frobenius norms of vectors and matrices. The function `norm(X,p)` returns the p-norm of X, where $p = 1, 2, \infty$ for p-norms or 'fro' for Frobenius norm.

Example 7.4

Let

$$A = \begin{bmatrix} 2 & 3 & 5 \\ 3 & 4 & 1 \end{bmatrix}$$

Then

$$\begin{aligned} \| A \|_F &= \sqrt{4+9+25+9+16+1} &&= 8 \\ \| A \|_1 &= \max \{ (2+3), (3+4), (5+1) \} &&= 7 \\ \| A \|_\infty &= \max \{ (2+3+5), (3+4+1) \} &&= 10 \end{aligned}$$

[3]These definitions are valid for $m \times n$ matrices with $m \ge 2$. For a row matrix $\boldsymbol{\alpha} = [\, a_1, \ldots , a_n \,]$, the definition would result in inconsistent identities $\| \boldsymbol{\alpha} \|_\infty = \| \boldsymbol{\alpha} \|_1$ and $\| \boldsymbol{\alpha} \|_1 = \| \boldsymbol{\alpha} \|_\infty$ where $\| \cdot \|$ denotes the vector norm on the right and the subordinate matrix norm on the left.

7.2 Inner Product Spaces

Let $\mathbf{X}$ be a vector space over F, where F is either R or C. A function which associates with every pair of vectors $\mathbf{x}, \mathbf{y} \in \mathbf{X}$ a complex scalar denoted $\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle$ is called an **inner product** on $\mathbf{X}$ if it satisfies the following.

I1. $\langle\, \mathbf{x} \,|\, \mathbf{x} \,\rangle \geq 0$ for all $\mathbf{x} \in \mathbf{X}$, and $\langle\, \mathbf{x} \,|\, \mathbf{x} \,\rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

I2. $\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle^* = \langle\, \mathbf{y} \,|\, \mathbf{x} \,\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{X}$.

I3. $\langle\, \mathbf{x} \,|\, a\mathbf{y} + b\mathbf{z} \,\rangle = a\,\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle + b\,\langle\, \mathbf{x} \,|\, \mathbf{z} \,\rangle$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{X}$ and $a, b \in \mathsf{F}$

A vector space with an inner product defined on it is called an **inner product space**. If $\mathbf{X}$ is a real vector space, then property I2 reduces to $\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle = \langle\, \mathbf{y} \,|\, \mathbf{x} \,\rangle$.

The following properties of inner product are immediate consequences of the definition.

a) $\langle\, \mathbf{x} \,|\, \mathbf{0} \,\rangle = \langle\, \mathbf{0} \,|\, \mathbf{x} \,\rangle$ for all $\mathbf{x} \in \mathbf{X}$

b) $\langle\, a\mathbf{x} + b\mathbf{y} \,|\, \mathbf{z} \,\rangle = a^*\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle + b^*\langle\, \mathbf{x} \,|\, \mathbf{z} \,\rangle$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{X}$, and $a, b \in \mathsf{F}$

Example 7.5

The standard inner product on $\mathsf{C}^{n\times 1}$ is defined as

$$\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle = \mathbf{x}^h\mathbf{y} = \sum_{i=1}^{n} x_i^* y_i$$

Clearly,

$$\langle\, \mathbf{x} \,|\, \mathbf{x} \,\rangle = \sum_{i=1}^{n} |x_i|^2 \geq 0$$

and $\langle\, \mathbf{x} \,|\, \mathbf{x} \,\rangle = 0$ if and only if $x_i = 0$ for all i, or equivalently, $\mathbf{x} = \mathbf{0}$. Also

$$\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle^* = (\mathbf{x}^h\mathbf{y})^h = \mathbf{y}^h\mathbf{x} = \langle\, \mathbf{y} \,|\, \mathbf{x} \,\rangle$$

and

$$\langle\, \mathbf{x} \,|\, a\mathbf{y} + b\mathbf{z} \,\rangle = \mathbf{x}^h(a\mathbf{y} + b\mathbf{z}) = a\,\mathbf{x}^h\mathbf{y} + b\,\mathbf{x}^h\mathbf{z} = a\,\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle + b\,\langle\, \mathbf{x} \,|\, \mathbf{z} \,\rangle$$

Similarly, the standard inner product for $\mathsf{R}^{n\times 1}$ is

$$\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle = \mathbf{x}^t\mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

Another common example is the standard inner product

$$\langle\, f \,|\, g \,\rangle = \int_a^b f^*(t)g(t)\,dt \tag{7.11}$$

defined on $\mathcal{C}_0([\,a, b\,], \mathsf{C})$. All three properties of the inner product are immediate consequences of the properties of the definite integral.

The following theorem states an important property of inner product.

Theorem 7.1 (The Schwarz Inequality) *In an inner product space* $\mathbf{X}$

$$|\langle \mathbf{x} \,|\, \mathbf{y} \rangle| \leq \sqrt{\langle \mathbf{x} \,|\, \mathbf{x} \rangle}\,\sqrt{\langle \mathbf{y} \,|\, \mathbf{y} \rangle}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbf{X}$*, where equality holds if and only if* $\mathbf{x}$ *and* $\mathbf{y}$ *are linearly dependent.*

Proof If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent, then $\mathbf{y} \neq \mathbf{0}$ and $\mathbf{x} - c\mathbf{y} \neq \mathbf{0}$ for any $c \in \mathsf{F}$. Then

$$0 < \langle \mathbf{x} - c\mathbf{y} \,|\, \mathbf{x} - c\mathbf{y} \rangle = \langle \mathbf{x} \,|\, \mathbf{x} \rangle - c\langle \mathbf{x} \,|\, \mathbf{y} \rangle - c^*\langle \mathbf{y} \,|\, \mathbf{x} \rangle + |c|^2\langle \mathbf{y} \,|\, \mathbf{y} \rangle$$

With $c = \langle \mathbf{y} \,|\, \mathbf{x} \rangle / \langle \mathbf{y} \,|\, \mathbf{y} \rangle$, we get

$$0 < \langle \mathbf{x} \,|\, \mathbf{x} \rangle - \frac{|\langle \mathbf{x} \,|\, \mathbf{y} \rangle|^2}{\langle \mathbf{y} \,|\, \mathbf{y} \rangle}$$

from which the Schwarz inequality follows.

If $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent, then either $\mathbf{y} = \mathbf{0}$, in which case we have

$$|\langle \mathbf{x} \,|\, \mathbf{y} \rangle| = 0 = \sqrt{\langle \mathbf{x} \,|\, \mathbf{x} \rangle}\,\sqrt{\langle \mathbf{y} \,|\, \mathbf{y} \rangle}$$

or $\mathbf{x} = c\mathbf{y}$ for some $c \in \mathsf{F}$ so that

$$|\langle \mathbf{x} \,|\, \mathbf{y} \rangle| = |c|\,|\langle \mathbf{y} \,|\, \mathbf{y} \rangle| = |c|\,\sqrt{\langle \mathbf{y} \,|\, \mathbf{y} \rangle}\,\sqrt{\langle \mathbf{y} \,|\, \mathbf{y} \rangle} = \sqrt{\langle \mathbf{x} \,|\, \mathbf{x} \rangle}\,\sqrt{\langle \mathbf{y} \,|\, \mathbf{y} \rangle}$$

An important consequence of the Schwarz inequality is that if $\mathbf{X}$ is an inner product space then

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x} \,|\, \mathbf{x} \rangle}$$

is a norm on $\mathbf{X}$. Properties N1 and N2 follow immediately from the definition of inner product, and property N3 follows by taking the square root of both sides of the inequality

$$\begin{aligned}
\|\mathbf{x} + \mathbf{y}\|^2 &= \langle \mathbf{x} + \mathbf{y} \,|\, \mathbf{x} + \mathbf{y} \rangle = \langle \mathbf{x} \,|\, \mathbf{x} \rangle + \langle \mathbf{x} \,|\, \mathbf{y} \rangle + \langle \mathbf{y} \,|\, \mathbf{x} \rangle + \langle \mathbf{y} \,|\, \mathbf{y} \rangle \\
&= \|\mathbf{x}\|^2 + 2\,\mathrm{Re}\,\{\langle \mathbf{x} \,|\, \mathbf{y} \rangle\} + \|\mathbf{y}\|^2 \\
&\leq \|\mathbf{x}\|^2 + 2\,|\langle \mathbf{x} \,|\, \mathbf{y} \rangle| + \|\mathbf{y}\|^2 \\
&\leq \|\mathbf{x}\|^2 + 2\,\|\mathbf{x}\|\,\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\
&= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2
\end{aligned}$$

Example 7.6

The norm defined by the standard inner product in $\mathsf{R}^{n\times 1}$ or $\mathsf{C}^{n\times 1}$ is the Euclidean norm

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^h\mathbf{x}} = \sqrt{|x_1|^2 + \cdots + |x_n|^2}$$

Similarly, the standard inner product in $\mathcal{C}_0([\,a, b\,], \mathsf{C})$ defines the Euclidean norm

$$\|f\|_2 = \left(\int_a^b |f(t)|^2\,dt\right)^{1/2}$$

Example 7.7

Let for $A, B \in \mathbf{C}^{m\times n}$

$$\langle\, A \,|\, B \,\rangle = \operatorname{tr}\,(A^h B) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^* b_{ij}$$

Then

$$\langle\, A \,|\, A \,\rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^* a_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} |\, a_{ij} \,|^2 \geq 0$$

and $\langle\, A \,|\, A \,\rangle = 0$ if and only if $A = O$. Thus property I1 of inner product is satisfied. Properties I2 and I3 are obvious from the definition. Hence $\langle\, A \,|\, B \,\rangle = \operatorname{tr}\,(A^h B)$ is an inner product on $\mathbf{C}^{m\times n}$.

The norm

$$\|\, A \,\| = \sqrt{\operatorname{tr}\,(A^h A)} = (\sum_{i=1}^{m} \sum_{j=1}^{n} |\, a_{ij} \,|^2\,)^{1/2}$$

defined by this inner product is nothing but the Frobenius norm defined previously.

7.3 Orthogonality

Let $\mathbf{X}$ be an inner product space. Two vectors $\mathbf{x}, \mathbf{y} \in \mathbf{X}$ are said to be **orthogonal**, denoted $\mathbf{x} \perp \mathbf{y}$, if $\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle = 0$. A vector $\mathbf{x}$ is said to be orthogonal to a set of vectors $\mathbf{R}$, denoted $\mathbf{x} \perp \mathbf{R}$, if it is orthogonal to every $\mathbf{r} \in \mathbf{R}$. Two sets $\mathbf{R}$ and $\mathbf{S}$ are said to be orthogonal, denoted $\mathbf{R} \perp \mathbf{S}$, if $\mathbf{r} \perp \mathbf{s}$ for all $\mathbf{r} \in \mathbf{R}$ and $\mathbf{s} \in \mathbf{S}$. A finite set $\mathbf{R} = \{\, \mathbf{r}_1, \cdots, \mathbf{r}_k \,\}$ is said to be orthogonal if $\mathbf{r}_i \neq \mathbf{0}$ for all i and $\mathbf{r}_i \perp \mathbf{r}_j$ for all $i \neq j$. If, in addition, $\|\, \mathbf{r}_i \,\| = \langle\, \mathbf{r}_i \,|\, \mathbf{r}_i \,\rangle^{1/2} = 1$ for all i, then $\mathbf{R}$ is said to be an **orthonormal** set.

We have the following properties concerning orthogonality.

a) $\mathbf{0}$ is orthogonal to every vector in $\mathbf{X}$, and it is the only vector that is orthogonal to every vector in $\mathbf{X}$.

b) If $\mathbf{x} \perp \mathbf{y}$ then $\|\, \mathbf{x} + \mathbf{y} \,\|^2 = \|\, \mathbf{x} \,\|^2 + \|\, \mathbf{y} \,\|^2$.

c) If $\mathbf{x} \perp \mathbf{R}$ then $\mathbf{x} \perp \operatorname{span}(\mathbf{R})$.

That $\mathbf{0}$ is orthogonal to every vector in $\mathbf{X}$ is obvious. If $\mathbf{x}$ is a vector that is orthogonal to every vector in $\mathbf{X}$, then $\|\, \mathbf{x} \,\|^2 = \langle\, \mathbf{x} \,|\, \mathbf{x} \,\rangle = 0$, and therefore, $\mathbf{x} = \mathbf{0}$. This proves (a). If $\mathbf{x} \perp \mathbf{y}$, then

$$\begin{aligned}
\|\, \mathbf{x} + \mathbf{y} \,\|^2 &= \langle\, \mathbf{x} + \mathbf{y} \,|\, \mathbf{x} + \mathbf{y} \,\rangle \\
&= \langle\, \mathbf{x} \,|\, \mathbf{x} \,\rangle + \langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle + \langle\, \mathbf{y} \,|\, \mathbf{x} \,\rangle + \langle\, \mathbf{y} \,|\, \mathbf{y} \,\rangle \\
&= \langle\, \mathbf{x} \,|\, \mathbf{x} \,\rangle + \langle\, \mathbf{y} \,|\, \mathbf{y} \,\rangle \;=\; \|\, \mathbf{x} \,\|^2 + \|\, \mathbf{y} \,\|^2
\end{aligned}$$

proving (b). Note that property (b) is a generalization of the Pythagorean theorem. Finally, if $\mathbf{x} \perp \mathbf{R}$ and $\mathbf{y} = c_1\mathbf{r}_1 + \cdots + c_k\mathbf{r}_k$ for some $\mathbf{r}_1, \ldots, \mathbf{r}_k \in \mathbf{R}$, then

$$\langle\, \mathbf{x} \,|\, \mathbf{y} \,\rangle = c_1 \,\langle\, \mathbf{x} \,|\, \mathbf{r}_1 \,\rangle + \cdots + c_k \,\langle\, \mathbf{x} \,|\, \mathbf{r}_k \,\rangle = 0$$

as $\mathbf{x} \perp \mathbf{r}_i$ for all i, and therefore, $\mathbf{x} \perp \mathbf{y}$.

Example 7.8

In $\mathbf{R}^{3\times 1}$ the vector $\mathbf{r}_3 = \mathrm{col}\,[\,1,1,1\,]$ is orthogonal to each of the vectors

$$\mathbf{r}_1 = \mathrm{col}\,[\,1,-1,0\,] \quad \text{and} \quad \mathbf{r}_2 = \mathrm{col}\,[\,0,-1,1\,]$$

with respect to the standard inner product, as $\mathbf{r}_3^t\mathbf{r}_1 = \mathbf{r}_3^t\mathbf{r}_2 = 0$. Therefore, $\mathbf{r}_3 \perp \mathrm{span}\,(\mathbf{r}_1, \mathbf{r}_2)$. Indeed, for any

$$\mathbf{x} = c_1\mathbf{r}_1 + c_2\mathbf{r}_2 = \begin{bmatrix} c_1 \\ -c_1 - c_2 \\ c_2 \end{bmatrix}$$

$\mathbf{r}_3^t\mathbf{x} = 0$.

Clearly, orthogonality of two vectors depends on the particular inner product chosen (see Exercise 7.17). In the rest of this chapter we shall use the standard inner product for $\mathbf{C}^{n\times 1}$ ($\mathbf{R}^{n\times 1}$) or $\mathcal{C}_0([\,a,b\,],\mathbf{C})$, and unless we indicate otherwise, we shall use the notation $\|\cdot\|$ to denote the Euclidean norm defined by the standard inner product.

Let $\mathbf{R} = (\,\mathbf{r}_1,\cdots,\mathbf{r}_k\,)$ be a finite ordered set. The matrix $G = [\,\langle\,\mathbf{r}_i\,|\,\mathbf{r}_j\,\rangle\,]_{k\times k}$ is called the **Gram matrix** of the vectors $\mathbf{r}_1,\cdots,\mathbf{r}_k$. We claim that

d) $\mathbf{R}$ is linearly independent if and only if G is nonsingular

e) if $\mathbf{R}$ is orthogonal then $\mathbf{R}$ is linearly independent.

If $\mathbf{R}$ is linearly dependent then there exists scalars c_i, not all zero, such that

$$\sum_{j=1}^{k} c_j\mathbf{r}_j = \mathbf{0}$$

Taking inner product of both sides with $\mathbf{r}_i$, we obtain

$$\sum_{j=1}^{k}\langle\,\mathbf{r}_i\,|\,\mathbf{r}_j\,\rangle c_j = 0\,, \quad i = 1,\ldots,k$$

or equivalently, $G\mathbf{c} = \mathbf{0}$, where $\mathbf{c} = \mathrm{col}\,[\,c_1,\ldots,c_k\,] \neq \mathbf{0}$. This implies that G is singular. Conversely, if G is singular then $G\mathbf{c} = \mathbf{0}$ for some $\mathbf{c} \neq \mathbf{0}$. Then

$$0 = \mathbf{c}^hG\mathbf{c} = \sum_{i=1}^{k}\sum_{j=1}^{k} c_i^*\langle\,\mathbf{r}_i\,|\,\mathbf{r}_j\,\rangle c_j = \langle\,\sum_{i=1}^{k} c_i\mathbf{r}_i\,|\,\sum_{j=1}^{k} c_j\mathbf{r}_j\,\rangle = \|\,\sum_{i=1}^{k} c_i\mathbf{r}_i\,\|^2$$

and therefore

$$\sum_{i=1}^{k} c_i\mathbf{r}_i = \mathbf{0}$$

implying that $\mathbf{R}$ is linearly dependent. This proves (d). (e) follows from (d) on noting that if $\mathbf{R}$ is orthogonal then $G = \mathrm{diag}\,[\,\|\,\mathbf{r}_1\,\|^2,\ldots,\|\,\mathbf{r}_k\,\|^2\,]$.

Let $\dim(\mathbf{X}) = n$ and let $\mathbf{R}$ be an orthogonal (orthonormal) set containing n vectors. Then since $\mathbf{R}$ is linearly independent, by Corollary 3.1 it is a basis for $\mathbf{X}$, called an **orthogonal (orthonormal) basis**.

Example 7.9

Let $\mathbf{r}_1, \ldots, \mathbf{r}_k \in \mathsf{C}^{n \times 1}$ be arranged as the column of an $n \times k$ matrix:

$$R = [\, \mathbf{r}_1 \cdots \mathbf{r}_k \,]$$

Since $\langle\, \mathbf{r}_i \,|\, \mathbf{r}_j \,\rangle = \mathbf{r}_i^h \mathbf{r}_j$ it follows that the Gram matrix is

$$G = R^h R$$

Consider the vectors in Example 7.8. Constructing

$$R = \begin{bmatrix} 1 & 0 & 1 \\ -1 & -1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad G = R^t R = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

we observe that G is nonsingular. Hence, $(\, \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \,)$ is linearly independent, and therefore, is a basis for $\mathsf{R}^{3 \times 1}$.

7.4 The Projection Theorem

Let $\mathbf{X}$ be an inner product space with $\dim(\mathbf{X}) = n$ and let $\mathbf{U} \subset \mathbf{X}$ be a subspace with $\dim(\mathbf{U}) = k$. The set

$$\mathbf{U}^\perp = \{\, \mathbf{x} \,|\, \mathbf{x} \perp \mathbf{U} \,\}$$

is called the **orthogonal complement** of $\mathbf{U}$.

Let $\mathbf{R} = (\, \mathbf{r}_1, \ldots, \mathbf{r}_k \,)$ be an ordered basis for $\mathbf{U}$. If $\mathbf{x} \perp \mathbf{U}$ then obviously $\mathbf{x} \perp \mathbf{r}_i$ for all i. Conversely, if $\mathbf{x} \perp \mathbf{r}_i$ for all i, then $\mathbf{x} \perp \operatorname{span}(\mathbf{R}) = \mathbf{U}$. Thus $\mathbf{U}^\perp$ can also be characterized as

$$\mathbf{U}^\perp = \{\, \mathbf{x} \,|\, \mathbf{x} \perp \mathbf{r}_i \,,\ i = 1, \ldots, k \,\}$$

Using this characterization, it can be shown that $\mathbf{U}^\perp$ is also a subspace of $\mathbf{X}$ (see Exercise 7.19).

If $\mathbf{x} \in \mathbf{U} \cap \mathbf{U}^\perp$ then $\mathbf{x} \perp \mathbf{x}$, and therefore, $\mathbf{x} = \mathbf{0}$. This shows that $\mathbf{U}$ and $\mathbf{U}^\perp$ are linearly independent.

Let $\mathbf{x}$ be an arbitrary vector in $\mathbf{X}$, and consider the $k \times k$ linear system of equation

$$\begin{bmatrix} \langle\, \mathbf{r}_1 \,|\, \mathbf{r}_1 \,\rangle & \cdots & \langle\, \mathbf{r}_1 \,|\, \mathbf{r}_k \,\rangle \\ \vdots & & \vdots \\ \langle\, \mathbf{r}_k \,|\, \mathbf{r}_1 \,\rangle & \cdots & \langle\, \mathbf{r}_k \,|\, \mathbf{r}_k \,\rangle \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} \langle\, \mathbf{r}_1 \,|\, \mathbf{x} \,\rangle \\ \vdots \\ \langle\, \mathbf{r}_k \,|\, \mathbf{x} \,\rangle \end{bmatrix} \tag{7.12}$$

Since the coefficient matrix is the Gram matrix of $\mathbf{r}_1, \ldots, \mathbf{r}_k$, it is nonsingular and hence (7.12) has a unique solution $\mathbf{c} = \boldsymbol{\alpha} = \operatorname{col}[\, \alpha_1, \ldots, \alpha_k \,]$. Let

$$\mathbf{x}_u = \sum_{j=1}^{k} \alpha_j \mathbf{r}_j \,, \quad \mathbf{x}_v = \mathbf{x} - \mathbf{x}_u$$

Then $\mathbf{x}_u \in \mathbf{U}$, and since

$$\langle\, \mathbf{r}_i \,|\, \mathbf{x}_v \,\rangle = \langle\, \mathbf{r}_i \,|\, \mathbf{x} \,\rangle - \langle\, \mathbf{r}_i \,|\, \mathbf{x}_u \,\rangle = \langle\, \mathbf{r}_i \,|\, \mathbf{x} \,\rangle - \sum_{j=1}^{k} \alpha_j \langle\, \mathbf{r}_i \,|\, \mathbf{r}_j \,\rangle = 0 \,, \quad i = 1, \ldots, k$$

$\mathbf{x}_v \in \mathbf{U}^\perp$. This shows that $\mathbf{U} + \mathbf{U}^\perp = \mathbf{X}$. Together with linear independence of $\mathbf{U}$ and $\mathbf{U}^\perp$ proved earlier, we reach the following theorem.

Theorem 7.2 (The Projection Theorem) $\mathbf{X} = \mathbf{U} \oplus \mathbf{U}^\perp$.

The unique vector $\mathbf{x}_u$ is called the **orthogonal projection** of $\mathbf{x}$ on $\mathbf{U}$. Note that $\mathbf{x}_v$ is the orthogonal projection of $\mathbf{x}$ on $\mathbf{U}^\perp$.

As a consequence of the projection theorem we have the following result.

Corollary 7.2.1 *Let $\mathbf{x}_u$ be the orthogonal projection of $\mathbf{x}$ on $\mathbf{U}$, and let $\mathbf{x}_v = \mathbf{x} - \mathbf{x}_u$. Then*

$$\min_{\mathbf{u}\in\mathbf{U}} \{ \| \mathbf{x} - \mathbf{u} \| \} = \| \mathbf{x}_v \|$$

and the minimum is achieved at $\mathbf{u} = \mathbf{x}_u$.

Proof Writing $\mathbf{x} - \mathbf{u} = \mathbf{x}_u - \mathbf{u} + \mathbf{x}_v$ and noting that $\mathbf{x}_u - \mathbf{u} \perp \mathbf{x}_v$, we have

$$\| \mathbf{x} - \mathbf{u} \|^2 = \| \mathbf{x}_u - \mathbf{u} \|^2 + \| \mathbf{x}_v \|^2$$

from which the result follows.

An illustration of the projection theorem and its corollary is given in Figure 7.1 for $\mathbf{X} = \mathbf{R}^{2\times 1}$.

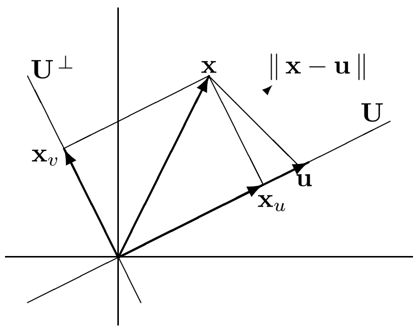

Figure 7.1: Illustration of the projection theorem

Equations (7.12) provide a computational procedure for determining $\mathbf{x}_u$. In particular, if $\mathbf{R}$ is an orthogonal basis for $\mathbf{U}$ then the solution of (7.12) is obtained as $c_j = \alpha_j = \langle \mathbf{r}_j \,|\, \mathbf{x} \rangle / \langle \mathbf{r}_j \,|\, \mathbf{r}_j \rangle$ and we have

$$\mathbf{x}_u = \sum_{j=1}^{k} \frac{\langle \mathbf{r}_j \,|\, \mathbf{x} \rangle}{\langle \mathbf{r}_j \,|\, \mathbf{r}_j \rangle} \mathbf{r}_j$$

If $\mathbf{R}$ is orthonormal then this expression further reduces to

$$\mathbf{x}_u = \sum_{j=1}^{k} \langle \mathbf{r}_j \,|\, \mathbf{x} \rangle \, \mathbf{r}_j \tag{7.13}$$

We can derive a compact formula for computating orthogonal projections in n-spaces. Let $\mathbf{X} = \mathbf{C}^{n\times 1}$ and

$$\mathbf{U} = \operatorname{span}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k) = \operatorname{im}([\,\mathbf{r}_1 \;\; \mathbf{r}_2 \; \cdots \; \mathbf{r}_k\,]) = \operatorname{im}(R)$$

where $\mathbf{r}_1, \ldots, \mathbf{r}_k$ form a basis for $\mathbf{U}$. For a given vector $\mathbf{x}$ let

$$\mathbf{x}_u = c_1\mathbf{r}_1 + \cdots + c_k\mathbf{r}_k = R\mathbf{c}\,, \quad \mathbf{c} = \operatorname{col}[\,c_1, \ldots, c_k\,]$$

Since the Gram matrix of $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k$ is $G = R^h R$, (7.12) becomes

$$R^h R\mathbf{c} = R^h\mathbf{x}$$

from which we obtain

$$\mathbf{c} = (R^h R)^{-1} R^h\mathbf{x} \quad \text{and} \quad \mathbf{x}_u = R(R^h R)^{-1} R^h\mathbf{x}$$

Example 7.10

In $\mathbf{R}^{3\times 1}$ the orthogonal projection of $\mathbf{x} = \operatorname{col}[\,x_1, x_2, x_3\,]$

a) on $\mathbf{E}_1 = \operatorname{span}(\mathbf{e}_1)$ is $\mathbf{x}_1 = x_1\mathbf{e}_1 = \operatorname{col}[\,x_1, 0, 0\,]$

b) on $\mathbf{E}_2 = \operatorname{span}(\mathbf{e}_2)$ is $\mathbf{x}_2 = x_2\mathbf{e}_2 = \operatorname{col}[\,0, x_2, 0\,]$

c) on $\mathbf{E}_{12} = \operatorname{span}(\mathbf{e}_1, \mathbf{e}_2)$ is $\mathbf{x}_{12} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 = \operatorname{col}[\,x_1, x_2, 0\,]$

The reader can interpret $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_{12}$ as the components of the vector $\mathbf{x}$ on the x_1 axis, on the x_2 axis, and on the x_1x_2 plane, respectively.

Now let $\mathbf{u} = \operatorname{col}[\,1, 1, 1\,]$ and $\mathbf{U} = \operatorname{span}(\mathbf{u})$. Then the orthogonal projection of $\mathbf{x}$ on $\mathbf{U}$ is

$$\mathbf{x}_u = \mathbf{u}(\mathbf{u}^t\mathbf{u})^{-1}\mathbf{u}^t\mathbf{x} = \frac{\mathbf{u}^t\mathbf{x}}{\mathbf{u}^t\mathbf{u}}\,\mathbf{u} = \frac{x_1 + x_2 + x_3}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

The reader can easily verify that $\mathbf{x}_u \perp \mathbf{x} - \mathbf{x}_u$.

Orthogonal projections are not restricted to finite dimensional vector spaces as we illustrate by the following example.

* **Example 7.11**

In $\mathcal{C}_0([\,0, 1\,], \mathbf{R})$, let

$$f_1(t) = 1\,, \quad f_2(t) = 2\sqrt{3}(t - 1/2)$$

Since

$$\langle\, f_1 \,|\, f_1 \,\rangle = \int_0^1 dt = 1\,, \quad \langle\, f_2 \,|\, f_2 \,\rangle = \int_0^1 12(t - 1/2)^2\, dt = 1$$

and

$$\langle\, f_1 \,|\, f_2 \,\rangle = \langle\, f_2 \,|\, f_1 \,\rangle = \int_0^1 2\sqrt{3}(t - 1/2)\, dt = 0$$

(f_1, f_2) is an orthonormal set.

Let $\mathbf{U} = \operatorname{span}(\,f_1, f_2\,)$ and $f(t) = t^2$. The orthogonal projection of f on $\mathbf{U}$ is $f_u = \alpha_1 f_1 + \alpha_2 f_2$, where

$$\alpha_1 = \langle\, f_1 \,|\, f \,\rangle = \int_0^1 t^2\, dt = \frac{1}{3}$$

and

$$\alpha_2 = \langle f_2 \,|\, f \rangle = \int_0^1 2\sqrt{3}(t - 1/2)t^2 \, dt = \frac{1}{2\sqrt{3}}$$

Thus

$$f_u(t) = (1/3) + (t - 1/2) = t - 1/6$$

Referring to Figure 7.1, we observe that the angle between the straight lines that contain the vectors $\mathbf{x}$ and $\mathbf{u}$ is given by

$$\theta(\mathbf{x}, \mathbf{u}) = \cos^{-1} \frac{\| \mathbf{x}_u \|}{\| \mathbf{x} \|} = \cos^{-1} \frac{|\langle \mathbf{u} \,|\, \mathbf{x} \rangle|}{\| \mathbf{x} \| \, \| \mathbf{u} \|}\,, \quad 0 \le \theta \le \pi/2$$

where the second equality follows from

$$\| \mathbf{x}_u \| = \| \frac{\langle \mathbf{u} \,|\, \mathbf{x} \rangle}{\langle \mathbf{u} \,|\, \mathbf{u} \rangle} \| = \frac{|\langle \mathbf{u} \,|\, \mathbf{x} \rangle|}{\| \mathbf{u} \|^2} \| \mathbf{u} \|$$

Note that since $\| \mathbf{x}_u \| \le \| \mathbf{x} \|$, $\theta(\mathbf{x}, \mathbf{u})$ is well-defined. This definition of angle between two vectors in the plane can readily be generalized to vectors of any inner product space. It can be used as a measure of alignment, and therefore, linear independence of two vectors: The larger the angle between two vectors the more linearly independent they are. Thus orthogonal vectors are maximally linearly independent. This explains why the vectors $\mathbf{u}_1 = (2.0, 1.0)$ and $\mathbf{u}_2 = (1.0, 2.0)$ in Example 3.20 can be considered to be more linearly independent than the vectors $\mathbf{v}_1 = (1.1, 1.0)$ and $\mathbf{v}_2 = (1.0, 1.1)$ of the same example: Comparing the angles between vectors of each pair, we see that

$$\theta(\mathbf{u}_1, \mathbf{u}_2) = \cos^{-1} \frac{4.0}{5.0} = 0.6435 \;\; (\approx 36.9^\circ)$$

whereas

$$\theta(\mathbf{v}_1, \mathbf{v}_2) = \cos^{-1} \frac{2.20}{2.21} = 0.0952 \;\; (\approx 5.5^\circ)$$

7.4.1 The Gram-Schmidt Orthogonalization Process

Let $\mathbf{R}_m = (\, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m \,)$ be an ordered linearly independent set in an inner product space $\mathbf{X}$. Define vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ successively as

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{r}_1 \\ \mathbf{u}_i &= \mathbf{r}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{r}_i \,|\, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j \,|\, \mathbf{u}_j \rangle} \mathbf{u}_j\,, \quad i = 2, \ldots, m \end{aligned} \tag{7.14}$$

We claim that

a) $\mathbf{u}_i \ne \mathbf{0}, i = 1, \ldots, m$, so that the process continues to the end without encountering a problem of division by zero

b) the set $\mathbf{U}_m = (\, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \,)$ is orthogonal

c) $\mathrm{span}\,(\mathbf{U}_m) = \mathrm{span}\,(\mathbf{R}_m)$

We prove the claims by induction on m. Since $\mathbf{u}_1 = \mathbf{r}_1 \neq \mathbf{0}$, they are obviously true for $m = 1$. Suppose they are true for $m = k$, and consider the case $m = k + 1$. Since

$$\mathbf{v}_{k+1} = \sum_{j=1}^{k} \frac{\langle\, \mathbf{r}_{k+1} \,|\, \mathbf{u}_j \,\rangle}{\langle\, \mathbf{u}_j \,|\, \mathbf{u}_j \,\rangle}\, \mathbf{u}_j$$

is the orthogonal projection of $\mathbf{r}_{k+1}$ on $\operatorname{span}(\mathbf{U}_k) = \operatorname{span}(\mathbf{R}_k)$ we have

$$\mathbf{u}_{k+1} = \mathbf{r}_{k+1} - \mathbf{v}_{k+1} \neq \mathbf{0}$$

for otherwise, $\mathbf{r}_{k+1} = \mathbf{v}_{k+1} \in \operatorname{span}(\mathbf{R}_k)$ contradicting linear independence of $\mathbf{R}_{k+1}$. Also, since $\mathbf{u}_{k+1} \perp \mathbf{U}_k$ and $\mathbf{U}_k$ is orthogonal by induction hypothesis, $\mathbf{U}_{k+1}$ is also orthogonal. Finally,

$$\begin{aligned} \operatorname{span}(\mathbf{U}_{k+1}) &= \operatorname{span}(\mathbf{U}_k) \oplus \operatorname{span}(\mathbf{u}_{k+1}) \\ &= \operatorname{span}(\mathbf{R}_k) \oplus \operatorname{span}(\mathbf{u}_{k+1}) \\ &= \operatorname{span}(\mathbf{R}_k) \oplus \operatorname{span}(\mathbf{r}_{k+1}) \;=\; \operatorname{span}(\mathbf{R}_{k+1}) \end{aligned}$$

The process described above, which generates an orthogonal set from a linearly independent set, is known as the **Gram-Schmidt orthogonalization process** (GSOP).

The GSOP can also be used to check if a given set is linearly independent: Suppose that the subset $\mathbf{R}_k$ is linearly independent, but $\mathbf{R}_{k+1}$ is not for some $k \leq m$. Then, since $\mathbf{r}_{k+1} \in \operatorname{span}(\mathbf{R}_k) = \operatorname{span}(\mathbf{U}_k)$, the process gives $\mathbf{u}_{k+1} = \mathbf{0}$ at the $(k+1)$st step. Conversely, if the process continues up to the kth step, and gives $\mathbf{u}_{k+1} = \mathbf{0}$, then we conclude that $\mathbf{R}_k$ is linearly independent, but $\mathbf{R}_{k+1}$ is not.

Example 7.12

Let us complete $\{\mathbf{r}\}$ to an orthogonal basis for $\mathsf{R}^{3\times 1}$, where $\mathbf{r} = \operatorname{col}[\,1, 1, 0\,]$.

Referring to Corollary 3.2, all we have to do is to apply the GSOP to the set $(\mathbf{r}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ and obtain an orthogonal set while eliminating the vectors that are linearly dependent on the previous ones. The process continues as follows.

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{r} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\ \mathbf{u}_2 &= \mathbf{e}_1 - \frac{\mathbf{e}_1^t \mathbf{u}_1}{\mathbf{u}_1^t \mathbf{u}_1}\, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} \\ \mathbf{u}_3 &= \mathbf{e}_2 - \frac{\mathbf{e}_2^t \mathbf{u}_1}{\mathbf{u}_1^t \mathbf{u}_1}\, \mathbf{u}_1 - \frac{\mathbf{e}_2^t \mathbf{u}_2}{\mathbf{u}_2^t \mathbf{u}_2}\, \mathbf{u}_2 \\ &= \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{-1/2}{1/2} \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

Since $\mathbf{u}_3 = \mathbf{0}$, we conclude that $\mathbf{e}_2$ is linearly dependent on $\mathbf{u}_1$ and $\mathbf{u}_2$ (equivalently, on $\mathbf{r}$ and $\mathbf{e}_1$), discard $\mathbf{e}_2$, and continue with $\mathbf{e}_3$. Observing that $\mathbf{e}_3 \perp \{\mathbf{u}_1, \mathbf{u}_2\}$, we immediately take

$$\mathbf{u}_3 = \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

7.4.2 The Least-Squares Problem

Let $A \in \mathbf{F}^{m\times n}$ and $\mathbf{y} \in \mathbf{F}^{n\times 1}$. Recall from Section 3.5 that if $\mathbf{y} \notin \operatorname{im}(A)$ then the linear equation

$$A\mathbf{x} = \mathbf{y}$$

has no solution. In such a case, we might be interested in finding an approximate solution $\mathbf{x} = \boldsymbol{\phi}$ such that $A\boldsymbol{\phi}$ is as close to $\mathbf{y}$ as possible with respect to a suitable measure. If we use the Euclidean norm on $\mathbf{F}^{m\times 1}$ as a measure of closeness, the problem can be formulated as

$$\min_{\mathbf{x}\in\mathbf{F}^{n\times 1}} \{ \| \mathbf{y} - A\mathbf{x} \| \} \tag{7.15}$$

Since

$$\{ A\mathbf{x} \,|\, \mathbf{x} \in \mathbf{F}^{n\times 1} \} = \operatorname{im}(A)$$

problem (7.15) is a matter of finding the orthogonal projection of $\mathbf{y}$ on $\operatorname{im}(A)$. Let $\mathbf{y}_A$ be the orthogonal projection of $\mathbf{y}$ on $\operatorname{im}(A)$. Then

$$\min_{\mathbf{x}\in\mathbf{F}^{n\times 1}} \{ \| \mathbf{y} - A\mathbf{x} \| \} = \| \mathbf{y} - \mathbf{y}_A \|$$

and the minimum is achieved at a solution $\mathbf{x} = \boldsymbol{\phi}_{LS}$ of the consistent equation

$$A\mathbf{x} = \mathbf{y}_A \tag{7.16}$$

Such a solution is called a **least-squares solution** of the equation $A\mathbf{x} = \mathbf{y}$, for the reason that it minimizes the sum of the squares of the differences between the elements of $\mathbf{y}$ and those of $A\mathbf{x}$.

Let $r(A) = r$ and let the columns of $R = [\,\mathbf{r}_1 \ \cdots \ \mathbf{r}_r\,]$ be a basis for $\operatorname{im}(A)$. Then the orthogonal projection of $\mathbf{y}$ on $\operatorname{im}(A)$ is

$$\mathbf{y}_A = R\boldsymbol{\alpha} = R(R^hR)^{-1}R^h\mathbf{y}$$

Once $\mathbf{y}_A$ is found, a least-squares solution can be obtained by solving (7.16).

In the special case when $r(A) = n$, we can choose $R = A$. Then

$$\mathbf{y}_A = A(A^hA)^{-1}A^h\mathbf{y}$$

and (7.16) becomes

$$A\mathbf{x} = A(A^hA)^{-1}A^h\mathbf{y}$$

Clearly, the formula

$$\mathbf{x} = \boldsymbol{\phi}_{LS} = (A^hA)^{-1}A^h\mathbf{y} \tag{7.17}$$

gives a least-squares solution.

Example 7.13

A constant quantity x is measured three times, and the values $x_1 = 14.6, x_2 = 15.6, x_3 = 14.8$ are obtained. What is the best estimate of x based on the measurements? In what sense?

The problem can be formulated as a linear equation as

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} x = \begin{bmatrix} 14.6 \\ 15.6 \\ 14.8 \end{bmatrix}$$

which is obviously inconsistent. A least-squares solution can be obtained from (7.17) as

$$x = \phi_{LS} = \frac{1}{3}[1 \quad 1 \quad 1] \begin{bmatrix} 14.6 \\ 15.6 \\ 14.8 \end{bmatrix} = \frac{14.6 + 15.6 + 14.8}{3} = 15.0$$

Observe that the least-squares solution $x = 15.0$ is simply the average of x_1, x_2 and x_3. It is the best estimate in Euclidean norm in the sense that it minimizes the sum-square error

$$e^2 = (x - 14.6)^2 + (x - 15.6)^2 + (x - 14.8)^2$$

If the error were measured with infinity norm, then we would try to minimize the absolute error

$$e = \max\{\, |x - x_1|, |x - x_2|, |x - x_3| \,\}$$

In this case the best estimate would be $x = 15.1$.

* **Example 7.14**

An operation analyst conducts a study to analyze the relationship between production volume and manufacturing expenses in the auto tyre industry. He assumes a linear relation

$$y = ax + b$$

between the number of tyres produced per day (x) and the daily manufacturing cost (y), and collects data (x_i, y_i) from N companies. His problem is to compute a and b such that the assumed model fits best the collected data.

Clearly, he is faced with a least-squares problem involving the linear system

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

the solution of which is given by (7.17). Straightforward computations yield

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{N\sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} N\sum x_i y_i - (\sum x_i)(\sum y_i) \\ (\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i) \end{bmatrix}$$

Letting

$$\mu_x = \frac{1}{N}\sum x_i\,, \qquad \sigma_x^2 = \frac{1}{N}\sum (x_i - \mu_x)^2$$
$$\mu_y = \frac{1}{N}\sum y_i\,, \qquad \sigma_{xy} = \frac{1}{N}\sum (x_i - \mu_x)(y_i - \mu_y)$$

the least-squares solution above can be manipulated into

$$a = \frac{\sigma_{xy}}{\sigma_x^2}\,, \quad b = \mu_y - a\mu_x$$

As a numerical example, suppose that the operation analyst collects the following data from $N = 10$ selected firms, where y is in thousands of dollars:

x :	600	700	825	925	1050	1125	1200	1275	1400	1500
y :	14.8	15.8	16.9	18.0	19.5	19.9	22.4	25.0	26.3	28.7

The solution of the least-squares problem yields a linear model

$$y = 0.01392\,x + 5.975$$

the graph of which is shown in Figure 7.2 together with the data points. The parameters a and b of the linear model are chosen so as to minimize the total sum-square-error

$$e^2 = \sum_{i=1}^{N} d_i^2 = \sum_{i=1}^{N} (ax_i + b - y_i)^2 = \| A\mathbf{x} - \mathbf{y} \|^2$$

Based on this model, the analyst expects the total cost of a firm that produces $x = 1000$ tyres to be

$$y = (0.01392)(1000) + 5.975 = 19.895$$

thousand dollars.

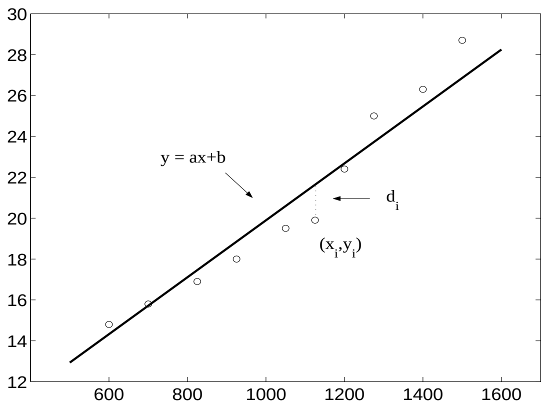

Figure 7.2: Data points and the least-squares linear model

7.4.3 Fourier Series

Let $\mathbf{X}$ be an n-dimensional inner product space, let $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ be an orthonormal basis for $\mathbf{X}$, and let the subspaces $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_n$ be defined as

$$\begin{aligned} \mathbf{V}_1 &= \operatorname{span}(\mathbf{u}_1) \\ \mathbf{V}_2 &= \operatorname{span}(\mathbf{u}_1, \mathbf{u}_2) \\ &\vdots \\ \mathbf{V}_n &= \operatorname{span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) \end{aligned}$$

Consider an arbitrary vector $\mathbf{x} \in \mathbf{X}$. The orthogonal projections of $\mathbf{x}$ on $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_n$ are computed as

$$\begin{aligned} \mathbf{x}_1 &= c_1\mathbf{u}_1 \\ \mathbf{x}_2 &= c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \\ &\vdots \\ \mathbf{x}_n &= c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n \end{aligned} \tag{7.18}$$

where

$$c_p = \langle\, \mathbf{u}_p \,|\, \mathbf{x} \,\rangle\,, \quad p = 1, 2, \ldots, n$$

Note that each $c_p\mathbf{u}_p$ is the orthogonal projection of $\mathbf{x}$ on the one dimensional subspace $\mathbf{U}_p = \operatorname{span}(\mathbf{u}_p)$. To construct the orthogonal projection on $\mathbf{V}_2 = \mathbf{U}_1 \oplus \mathbf{U}_2$, we simply add the orthogonal projections on $\mathbf{U}_1$ and $\mathbf{U}_2$; to construct the orthogonal projection on $\mathbf{V}_3 = \mathbf{U}_1 \oplus \mathbf{U}_2 \oplus \mathbf{U}_3$, we add the orthogonal projections on $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{U}_3$; and so on. This is a consequence of $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ being orthogonal. Otherwise, the orthogonal projection on $\mathbf{U}_1 \oplus \mathbf{U}_2$ would be different from the sum of the orthogonal projections on the individual subspaces (see Exercise 7.34).

From the discussion in the previous section we know that each $\mathbf{x}_q$ is the best approximation to $\mathbf{x}$ in terms of the vectors of $\mathbf{V}_q$, $q = 1, 2, \ldots, n$. Since

$$\mathbf{V}_1 \;\subset\; \mathbf{V}_2 \;\subset\; \cdots \;\subset\; \mathbf{V}_n$$

$\mathbf{x}_2$ is a better approximation to $\mathbf{x}$ than $\mathbf{x}_1$ is, $\mathbf{x}_3$ is better than $\mathbf{x}_2$ is, and so on. That is,

$$\|\, \mathbf{x} - \mathbf{x}_1 \,\| \geq \|\, \mathbf{x} - \mathbf{x}_2 \,\| \geq \cdots \geq \|\, \mathbf{x} - \mathbf{x}_n \,\|$$

In fact, since $\mathbf{V}_n = \mathbf{X}$, we have $\mathbf{x}_n = \mathbf{x}$ so that $\|\, \mathbf{x} - \mathbf{x}_n \,\| = 0$. The situation, which is illustrated in Figure 7.3 for $\mathbf{X} = \mathbf{R}^3$, is easy to understand when $\mathbf{X}$ is finite dimensional. The infinite dimensional case is more interesting.

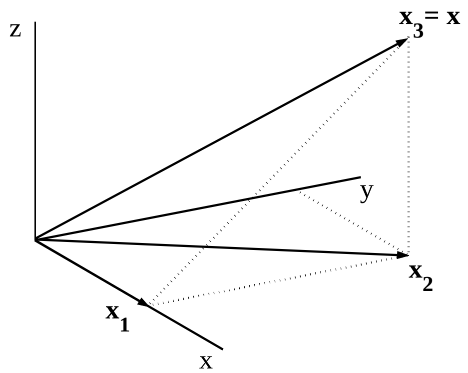

Figure 7.3: Orthogonal projections of a vector

Suppose $\dim(\mathbf{X}) = \infty$, and suppose $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \ldots)$ is an infinite sequence of orthonormal vectors in $\mathbf{X}$.[4] For an arbitrary $\mathbf{x} \in \mathbf{X}$, let us define the subspaces $\mathbf{V}_q$ and the vectors $\mathbf{x}_q$ as in (7.18) and (7.18) with the index q running not just up to n but up to infinity. Then again each $\mathbf{x}_q$ is the orthogonal projection of $\mathbf{x}$ on the q-dimendional subspace $\mathbf{V}_q$, and hence it is the best approximation to $\mathbf{x}$ in terms of the vectors in $\mathbf{V}_q$. That is,

$$\min_{\mathbf{u} \in \mathbf{V}_q} \| \mathbf{x} - \mathbf{u} \| = \| \mathbf{x} - \mathbf{x}_q \|, \quad q = 1, 2, \ldots$$

Since $\mathbf{V}_q \subset \mathbf{V}_{q+1}$, we also have

$$\| \mathbf{x} - \mathbf{x}_{q+1} \| = \min_{\mathbf{u} \in \mathbf{V}_{q+1}} \| \mathbf{x} - \mathbf{u} \| \leq \min_{\mathbf{u} \in \mathbf{V}_q} \| \mathbf{x} - \mathbf{u} \| = \| \mathbf{x} - \mathbf{x}_q \|$$

that is,

$$\| \mathbf{x} - \mathbf{x}_1 \| \geq \| \mathbf{x} - \mathbf{x}_2 \| \geq \cdots \geq \| \mathbf{x} - \mathbf{x}_q \| \geq \cdots$$

The interesting point is that although $\mathbf{x}_q$ approximate $\mathbf{x}$ better and better as q increases, without further information about the set $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \ldots)$, we cannot say that

$$\lim_{q \to \infty} \| \mathbf{x} - \mathbf{x}_q \| = 0 \tag{7.19}$$

However, if (7.19) is true, we formally write

$$\mathbf{x} = \sum_{p=1}^{\infty} c_p \mathbf{u}_p \tag{7.20}$$

The expression on the right-hand-side of (7.20) is known as the **Fourier series** of $\mathbf{x}$ in terms of $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \ldots)$.

Example 7.15

Refer to Example 3.25. Let $\mathbf{X} = \mathcal{F}(\mathbf{D}_N, C)$, and define

$$\langle f \,|\, g \rangle = \frac{1}{N} \sum_{k=0}^{N-1} f^*[k] g[k]$$

which is the standard inner product on $\mathbf{C}^{N \times 1}$ scaled by $1/N$. (Recall that $\mathcal{F}(\mathbf{D}_N, C)$ is essentially the same as $\mathbf{C}^{N \times 1}$.)

Consider the set of functions $\phi_p, p = 0, \ldots, N-1$ defined in Example 3.25. Using the hint in Exercise 3.20, it can easily be shown that $(\,\phi_p\,)$ is an orthonormal set in $\mathcal{F}(\mathbf{D}_N, C)$. Then for a given $f \in \mathcal{F}(\mathbf{D}_N, C)$, (7.20) reduces to the (discrete) Fourier series in (3.7), where

$$c_p = \langle \phi_p \,|\, f \rangle = \frac{1}{N} \sum_{k=0}^{N-1} \phi_p^*[k] f[k]$$

and each $c_p \phi_p$ term is the orthogonal projection of f on $\operatorname{span}(\phi_p)$.

[4] Here we assume that such an infinite orthonormal set exists. Although with our present knowledge we cannot guarantee the existence of such a set, we may try to construct one.

Example 7.16

Let $\mathbf{X}$ be the vector space of piece-wise continuous complex-valued functions defined on a real interval $(0, T)$. Let us define an inner product on $\mathbf{X}$ as

$$\langle f \mid g \rangle = \frac{1}{T} \int_0^T f^*(t) g(t)\, dt \tag{7.21}$$

which is the familiar standard inner product scaled with $1/T$. Consider the following set of functions:

$$\phi_k(t) = e^{ik\frac{2\pi}{T}t}, \quad k = \ldots, -1, 0, 1, \ldots$$

It is left to the reader to prove that (ϕ_k) is an orthonormal set with respect to the inner product in (7.21). Hence the Fourier coefficients of a given function f are computed as

$$\alpha_k = \langle \phi_k \mid f \rangle = \frac{1}{T} \int_0^T f(t) e^{-ik\frac{2\pi}{T}t}\, dt$$

As a specific example consider the piece-wise continuous function

$$f(t) = \begin{cases} 1, & 0 < t < 0.5 \\ 0, & 0.5 < t < 1 \end{cases}$$

defined on the interval $(0, 1)$. Then its Fourier coefficients are computed as

$$\alpha_k = \int_0^{0.5} e^{-i2k\pi t}\, dt = \begin{cases} \frac{1}{2}, & k = 0 \\ \frac{1}{k\pi i}, & k \text{ odd} \\ 0, & k \neq 0, k \text{ even} \end{cases}$$

Hence the orthogonal projection of f on the subspace

$$\mathbf{U}_q = \text{span}\,(\phi_{-q}, \ldots, \phi_{-1}, \phi_0, \phi_1, \ldots, \phi_q)$$

is given as

$$f_q(t) = \frac{1}{2} + \sum_{\substack{k=1 \\ k \text{ odd}}}^{q} \left(\frac{1}{k\pi i} e^{i2k\pi t} - \frac{1}{k\pi i} e^{-i2k\pi t}\right) = \frac{1}{2} + \sum_{\substack{k=1 \\ k \text{ odd}}}^{q} \frac{2}{k\pi} \sin 2k\pi t$$

Plots of f and f_q for $q = 0, 1, 3, 9$ are shown in Figure 7.4.

7.5 Exercises

1. (a) Find the uniform, Euclidean and infinity norms of the following vectors.

$$\mathbf{x} = \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1+i \\ -3i \\ 2 \end{bmatrix}$$

   (b) Repeat part (a) by using MATLAB command `norm`.

2. In $\mathbf{R}^2$ plot the locus of points $\mathbf{x}$ for which $\| \mathbf{x} \|_p = 1$ for $p = 1, 2, \infty$.

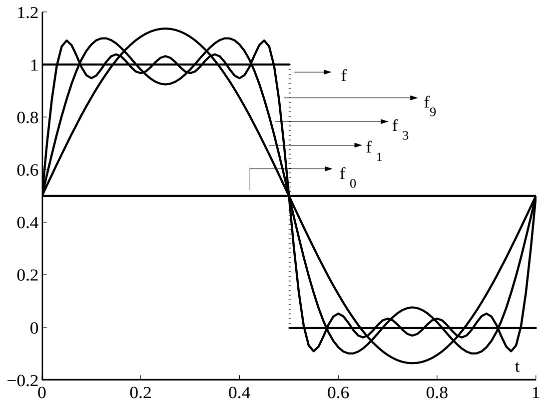

Figure 7.4: Fourier approximations of a function

3. Let $p > 1$ be a real number and $q > 1$ be such that $p^{-1} + q^{-1} = 1$.

   (a) Show that

   $$u^{1/p}v^{1/q} \leq u/p + v/q \quad \text{for all} \;\; u \geq 0\,, \quad v \geq 0$$

   Hint: First show that

   $$(1+x)^p \geq 1 + px \quad \text{for all} \;\; x \geq -1\,, \quad p > 1$$

   and let $1 + x = (u/v)^{1/p}$.

   (b) Prove **Hölder's inequality**

   $$\sum_{i=1}^{n} |x_i y_i| \leq (\sum_{i=1}^{n} |x_i|^p)^{1/p} (\sum_{i=1}^{n} |y_i|^q)^{1/q}$$

   for $\mathbf{x} = \operatorname{col}[x_1, \ldots, x_n]$ and $\mathbf{y} = \operatorname{col}[y_1, \ldots, y_n] \in \mathbf{C}^{n\times 1}$. Hint: Apply the inequality in (a) to

   $$u = u_i = \frac{|x_i|^p}{\sum_{i=1}^{n} |x_i|^p}\,, \quad v = v_i = \frac{|y_i|^q}{\sum_{i=1}^{n} |y_i|^q}$$

   and then take summation on i.

4. Prove Minkowski's inequality

   $$(\sum_{i=1}^{n} |x_i + y_i|^p)^{1/p} \leq (\sum_{i=1}^{n} |x_i|^p)^{1/p} + (\sum_{i=1}^{n} |y_i|^p)^{1/p}$$

   Hint: Take summation of both sides of the inequalities

   $$|x_i + y_i|^p \leq |x_i|\,|x_i + y_i|^{p-1} + |y_i|\,|x_i + y_i|^{p-1}\,, \quad i = 1, \ldots, n$$

   and use Hölder's inequality and the identity $q(p-1) = p$.

5. Show that

   $$\|\, f \,\|_p = (\int_a^b |f(t)|^p \, dt)^{1/p}$$

is a norm on $\mathcal{C}_0([a,b],\mathbf{R})$. Hint: Derive integral counterparts of Hölder's and Minkowski's inequalities.

6. Find the uniform, Euclidean and infinity norms of the following functions.

   (a) $f(t) = t - 1\,, \quad 0 \le t \le 2$

   (b) $g(t) = e^{i\omega t}\,, \quad -\pi/\omega \le t \le \pi/\omega$

   (c) $h(t) = 1/t\,, \quad 1 \le t \le T\,, \quad T \to \infty$

7. Refer to Example 7.3. Since the components of any $\mathbf{f} \in \mathcal{C}_0(\mathcal{I},\mathbf{R}^{n\times 1})$ are continuous, the function $\nu_p^{\mathbf{f}}$ defined in (7.7) is a continuous function for any $p \ge 1$. Hence, $\|\,\mathbf{f}\,\|_{p,q}$ in (7.8) is a well-defined quantity for any $q \ge 1$. Now, $\mathbf{f} \ne \mathbf{0}$ implies $\nu_p^{\mathbf{f}} \ne 0$, which in turn implies $\|\,\mathbf{f}\,\|_{p,q} > 0$. Also, for any $c \in \mathbf{R}$,

$$\nu_p^{c\mathbf{f}}(t) = \|\,(c\mathbf{f})(t)\,\|_p = \|\,c\mathbf{f}(t)\,\|_p = |\,c\,|\,\|\,\mathbf{f}(t)\,\|_p = |\,c\,|\,\nu_p^{\mathbf{f}}(t) \quad \text{for all } \; t \in \mathcal{I}$$

so that

$$\|\,c\mathbf{f}\,\|_{p,q} = \|\,\nu_p^{c\mathbf{f}}\,\|_q = \|\,|\,c\,|\,\nu_p^{\mathbf{f}}\,\|_q = |\,c\,|\,\|\,\nu_p^{\mathbf{f}}\,\|_q = |\,c\,|\,\|\,\mathbf{f}\,\|_{p,q}$$

Thus $\|\,\mathbf{f}\,\|_{p,q}$ satisfies the first two properties of a norm. Show that it also satisfies the triangle inequality, so that it is a norm on $\mathcal{C}_0(\mathcal{I},\mathbf{R}^{n\times 1})$.

8. Refer to Example 7.3.

   (a) Let $\mathbf{f}(t) = \mathrm{col}\,[\,1, t\,]\,, \quad 0 \le t \le 1$. Find $\|\,\mathbf{f}\,\|_{1,2}$ and $\|\,\mathbf{f}\,\|_{2,\infty}$

   (b) Let $\mathbf{h}(t) = \mathrm{col}\,[\,t, t-1\,]\,, \quad 0 \le t \le 1$. Find $\|\,\mathbf{h}\,\|_{1,\infty}$ and $\|\,\mathbf{h}\,\|_{\infty,1}$

9. Two norms $\|\cdot\|$ and $\|\cdot\|'$ on $\mathbf{X}$ are said to be equivalent if there exist $0 < c_1 \le c_2$ such that

$$c_1\,\|\,\mathbf{x}\,\| \le \|\,\mathbf{x}\,\|' \le c_2\,\|\,\mathbf{x}\,\| \quad \text{for all } \; \mathbf{x} \in \mathbf{X}$$

in which case

$$\frac{1}{c_2}\,\|\,\mathbf{x}\,\|' \le \|\,\mathbf{x}\,\| \le \frac{1}{c_1}\,\|\,\mathbf{x}\,\|' \quad \text{for all } \; \mathbf{x} \in \mathbf{X}$$

   (a) Show that all p-norms on $\mathbf{C}^{n\times 1}$ (including $p = \infty$) are equivalent. Hint: First show that all p-norms are equivalent to the ∞-norm.

   (b) Does a corresponding result hold for the p-norms on $\mathcal{C}_0([\,0,1\,],\mathbf{R})$? Hint: Suppose that there exist $0 < c_1 \le c_2$ such that

$$c_1\,\|\,f\,\|_1 \le \|\,f\,\|_\infty \le c_2\,\|\,f\,\|_1 \quad \text{for all } \; f \in \mathcal{C}_0([\,0,1\,],\mathbf{R})$$

   Let

$$f(t) = \begin{cases} 1 - nt, & 0 \le t \le 1/n \\ 0, & 1/n \le t \le 1 \end{cases}$$

   and show that the second equality is violated for sufficiently large n.

10. Show that the following are norms on $\mathbf{F}^{m\times n}$.

   (a) $$\|\,A\,\| = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|$$

   (b) $$\|\,A\,\| = \max_{i,j}\,\{\,|a_{ij}|\,\}$$

11. (a) Show that the matrix norm subordinate to the uniform vector norm is

$$\| A \|_1 = \max_{1 \le j \le n} \{ \sum_{i=1}^{m} |a_{ij}| \} = \max_{1 \le j \le n} \{ \| \mathbf{a}_j \|_1 \}$$

where $\mathbf{a}_j$ denotes the jth column of A Hint: Suppose that the maximum of the right-hand side is achieved for $j = q$. Show that for arbitrary $\mathbf{x} = \mathrm{col}\,[\, x_1, \ldots, x_n \,]$

$$\| A\mathbf{x} \|_1 \le (|x_1| + \cdots + |x_n|) \cdot \| \mathbf{a}_q \|_1 = \| \mathbf{a}_q \|_1 \| \mathbf{x} \|_1$$

with equality holding for $\mathbf{x} = \mathbf{e}_q$.

(b) Show that the matrix norm subordinate to the infinity vector norm is

$$\| A \|_\infty = \max_{1 \le i \le m} \{ \sum_{j=1}^{n} |a_{ij}| \} = \max_{1 \le i \le m} \{ \| \boldsymbol{\alpha}_i \|_1 \}$$

where $\boldsymbol{\alpha}_i$ denotes the ith row of A. Hint: Suppose that the maximum of the right-hand side is achieved for $i = p$. Show that for arbitrary $\mathbf{x} = \mathrm{col}\,[\, x_1, \ldots, x_n \,]$

$$\| A\mathbf{x} \|_\infty \le \max_i \{ \sum_{j=1}^{n} |a_{ij}| \} \cdot \max_j \{ |x_j| \} = \| \boldsymbol{\alpha}_p \|_1 \| \mathbf{x} \|_\infty$$

with equality holding for

$$\mathbf{x} = \mathrm{col}\,[\, \mathrm{sign}(a_{p1}), \ldots, \mathrm{sign}(a_{pn}) \,]$$

12. (a) Find the uniform and infinity norms of the following matrices.

(b) Use MATLAB command `norm` to verify your results.

$$A = \begin{bmatrix} -1 & 1 \\ 0 & 3 \\ 2 & -1 \end{bmatrix}, \quad B = A^t$$

13. Let $\mathcal{A} : \mathbf{R}^{2\times 2} \to \mathbf{R}^{2\times 2}$ be defined as $\mathcal{A}(X) = X^t$. Clearly, $\mathcal{A}$ is a linear transformation. Find

$$\| \mathcal{A} \|_1 = \max_{X \ne O} \frac{\| X^t \|_1}{\| X \|_1}$$

14. Let $\mathbf{X}$ be a normed vector space with a norm $\| \cdot \|$. An infinite sequence of vectors $(\, \mathbf{x}_n \,)$ is said to **converge** (in the norm $\| \cdot \|$) to a limit vector $\mathbf{x}$, denoted as

$$\lim_{n \to \infty} \mathbf{x}_n = \mathbf{x}$$

if the sequence of real numbers $(\, \| \mathbf{x}_n - \mathbf{x} \| \,)$ converges to 0. Equivalently, $(\, \mathbf{x}_n \,)$ converges to $\mathbf{x}$ if for any $\epsilon > 0$ there exists an integer $N > 0$ such that

$$\| \mathbf{x}_n - \mathbf{x} \| < \epsilon \quad \text{for all} \quad n \ge N$$

Check if the sequence $(\, \mathbf{x}_n \,)$, where

$$\mathbf{x}_n = \begin{bmatrix} 1 - \frac{1}{(-2)^n} \\ \frac{1}{3^n} \end{bmatrix}$$

converges in $\mathbf{R}^{2\times 1}$, and find its limit if it does. Does your answer depend on the particular norm you choose for $\mathbf{R}^{2\times 1}$?

15. A sequence $(\mathbf{x}_n)$ in a normed vector space $\mathbf{X}$ is called a **Cauchy sequence** if

$$\lim_{m,n\,\to\,\infty} \|\,\mathbf{x}_n - \mathbf{x}_m\,\| = 0$$

Show that every convergent sequence in $\mathbf{X}$ is a Cauchy sequence.

16. (a) Find all orthogonal pairs of the following vectors in $\mathbf{R}^{3\times 1}$ w.r.t. the standard inner product:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \ \mathbf{x}_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

(b) Repeat (a) for the following vectors in $\mathbf{C}^{3\times 1}$:

$$\mathbf{z}_1 = \begin{bmatrix} 1 \\ i \\ 1+i \end{bmatrix}, \ \mathbf{z}_2 = \begin{bmatrix} 3i \\ 1 \\ 1-i \end{bmatrix}, \ \mathbf{z}_3 = \begin{bmatrix} 0 \\ 1+i \\ -1 \end{bmatrix}, \ \mathbf{z}_4 = \begin{bmatrix} -1 \\ -i \\ 1+i \end{bmatrix}$$

(c) Repeat (a) for the following vectors in $\mathcal{F}([\,0,1\,],\mathbf{R})$:

$$f_1 = 1\,,\ f_2(t) = t+1\,,\ f_3(t) = 2t-1\,,\ f_4(t) = 6t^2 - 6t + 1$$

17. For $\mathbf{x},\mathbf{y} \in \mathbf{R}^{2\times 1}$ let

$$\langle\,\mathbf{x}\,|\,\mathbf{y}\,\rangle_Q = x_1y_1 + x_1y_2 + x_2y_1 + 2x_2y_2 = \mathbf{x}^t Q\mathbf{y}$$

where

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

(a) Show that $\langle\,\cdot\,|\,\cdot\,\rangle_Q$ is an inner product. Hint:

$$\langle\,\mathbf{x}\,|\,\mathbf{x}\,\rangle_Q = (x_1+x_2)^2 + x_2^2$$

(b) Are $\mathbf{e}_1$ and $\mathbf{e}_2$ orthogonal with respect to this inner product? Find a vector that is orthogonal to $\mathbf{e}_1$ and a vector orthogonal to $\mathbf{e}_2$.

(c) The norm defined by this inner product is

$$\|\,\mathbf{x}\,\|_Q = \sqrt{\mathbf{x}^t Q\mathbf{y}} = \sqrt{(x_1+x_2)^2 + x_2^2}$$

Compute $\|\,\mathbf{e}_1\,\|_Q$ and $\|\,\mathbf{e}_2\,\|_Q$.

18. Show that Schwarz inequality for $\mathbf{R}^{n\times 1}$ and $\mathcal{F}(\mathcal{I},\mathbf{R})$ is a special case of Hölder's inequality.

19. Let $\mathbf{U}$ be a subset of an inner product space $\mathbf{X}$. Prove that $\mathbf{U}^{\perp}$ is a subspace of $\mathbf{X}$. Hint: Show that $\mathbf{U}^{\perp}$ is closed under vector addition and scalar multiplication.

20. In $\mathbf{R}^{3\times 1}$, let

$$\mathbf{U} = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)$$

Find bases for $\mathbf{U}^{\perp}$ and $(\mathbf{U}^{\perp})^{\perp}$.

21. In $\mathcal{C}_0([\,-T,T\,],\mathbf{R})$, let

$$\mathbf{U} = \{\, f \mid f(-t) = f(t)\,\}$$

Characterize the orthogonal complement of $\mathbf{U}$ with respect to the inner product in (7.11).

22. Apply GSOP to

$$\mathbf{u}_1 = \text{col}\,[\,1,1,0\,]\,, \quad \mathbf{u}_2 = \text{col}\,[\,0,2,1\,]\,, \quad \mathbf{u}_3 = \text{col}\,[\,4,0,1\,]$$

to generate an orthogonal basis for $\mathbf{R}^{3\times 1}$.

23. The MATLAB command `orth(A)` finds an orthonormal basis for $\text{im}\,(A)$. Thus if

$$A = [\,\mathbf{a}_1 \ \cdots \ \mathbf{a}_n\,] \quad \text{and} \quad B = \ \texttt{orth}(A) = [\,\mathbf{b}_1 \ \cdots \ \mathbf{b}_q\,]$$

then $(\,\mathbf{b}_1,\ldots,\mathbf{b}_q\,)$ is an orthonormal set generated by $(\,\mathbf{a}_1,\ldots,\mathbf{a}_n\,)$.

(a) Use `orth` to generate an orthogonal basis for $\mathbf{R}^{3\times 1}$ from the vectors in Exercise 7.22.

(b) Use `orth` to compute the rank of the matrix

$$A = \begin{bmatrix} 1 & 2 & 1 & -3 \\ -1 & 1 & 2 & -3 \\ 2 & -1 & -3 & 4 \end{bmatrix}$$

24. Let $\mathbf{x}_1 = \text{col}\,[\,1,1,0\,]$ and $\mathbf{x}_2 = \text{col}\,[\,1,1,1\,]$.

(a) Apply GSOP to $(\,\mathbf{x}_1,\mathbf{x}_2\,)$ to generate an orthonormal set $(\,\mathbf{v}_1,\mathbf{v}_2\,)$.

(b) Find orthogonal projections of $\mathbf{x} = \text{col}\,[\,0,1,1\,]$ on $\mathbf{S}_1 = \text{span}\,(\,\mathbf{x}_1\,)$ and on $\mathbf{S}_2 = \text{span}\,(\,\mathbf{x}_1,\mathbf{x}_2\,)$.

(c) Repeat (a) and (b) for $\mathbf{x}_1 = \text{col}\,[\,1,-1,0\,]$, $\mathbf{x}_2 = \text{col}\,[\,0,1,1\,]$ and $\mathbf{x} = \text{col}\,[\,1,1,1\,]$

25. In $\mathbf{R}^3$ find the minimum distance from the origin to the plane

$$2x_1 + 3x_2 - x_3 = -5$$

and also find the point on the plane closest to the origin. Hint: The given plane has a normal $\mathbf{n} = (2,3,-1)$.

26. A mirror lies in the plane defined by

$$-2x_1 + 3x_2 + x_3 = 0$$

which defines a subspace of $\mathbf{R}^3$. Find the reflected image of the vector $\mathbf{x} = \text{col}\,[\,5,2,-3\,]$.

27. Consider $\mathcal{C}_0([\,-1,1\,],\mathbf{R})$ with the inner product given in (7.11).

(a) Apply GSOP to the set $(1,t,t^2)$ to generate an orthonormal set of functions.

(b) Find the orthogonal projection of $g(t) = t^3$ on $\text{span}\,(\,1,t,t^2\,)$.

28. Let $A \in \mathcal{C}^{n\times n}$, and $\mathbf{U} \in \mathcal{C}^{n\times 1}$ be a proper subspace. Show that if $\mathbf{U}$ is A-invariant, then $\mathbf{U}^\perp$ is A^h-invariant.

29. (a) Find a least squares solution to

$$\begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 6 \end{bmatrix}$$

(b) Use MATLAB command `x=pinv(A)*b` to verify your result.

30. Consider the linear equation

$$\begin{bmatrix} 1 & 1 & 2 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$$

(a) Characterize all least squares solutions $\mathbf{x}_{LS}$ of the given equation.

(b) Among all least squares solutions, find the one with minimum norm.

31. (a) In the xy plane plot the locus of all vectors (points) of the form $\mathbf{x} = \mathbf{p} + c\,\mathbf{q}$, $c \in \mathbf{R}$, where

$$\mathbf{p} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

(b) Determine geometrically the value of c such that $\|\, \mathbf{p} + c\,\mathbf{q} \,\|$ is minimum.

(c) Formulate and solve the problem in part (b) as a least-squares problem.

32. (a) Find a, b, c which minimize

$$\int_{-1}^{1} (t^3 - a - bt - ct^2)^2 \, dt$$

(b) Formulate and solve the problem as a least-squares problem in $\mathcal{C}_0([-1,1], \mathbf{R})$.

33. Let $f(t)$ be a continuous function defined on an interval $0 \le t \le 1$. Consider the problem of approximating f by an $(n-1)$st degree polynomial

$$p(t) = \sum_{k=0}^{n-1} p_k t^k$$

whose coefficients $p_k, k = 0, \ldots, n-1$, are to be determined such that the error

$$E = \int_0^1 [p(t) - f(t)]^2 \, dt$$

is minimized.

(a) Obtain a system of n linear equations in the n unknowns $p_k, k = 0, \ldots, n-1$, by setting

$$\frac{\partial E}{\partial p_k} = 0\,, \quad k = 0, \ldots, n-1$$

Show that the coefficient matrix of the resulting linear system $H\mathbf{p} = \mathbf{b}$ is a Hilbert matrix of order n.

(b) Interpret the problem as a least-squares problem.

34. Let

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

(a) Find the orthogonal projections $\mathbf{x}_1$ and $\mathbf{x}_2$ of $\mathbf{x}$ on span$(\mathbf{u}_1)$ and span$(\mathbf{u}_2)$.

(b) Find the orthogonal projection $\mathbf{x}_{12}$ of $\mathbf{x}$ on span$(\mathbf{u}_1, \mathbf{u}_2)$. Is $\mathbf{x}_{12} = \mathbf{x}_1 + \mathbf{x}_2$? Explain.

35. Refer to Example 7.16. Obtain the Fourier series of the function

$$f(t) = t\,, \quad 0 \le t \le 1$$

Use MATLAB to compute and plot the Fourier series truncated at $k = 0, 1, 5, 10$.

36. Let $(\mathbf{x}_i, i = 1, \ldots, k)$ be an orthonormal set in an inner product space $\mathbf{X}$. Show that for any $\mathbf{x} \in \mathbf{X}$

$$\sum_{i=1}^{k} |\, \langle\, \mathbf{x} \,|\, \mathbf{x}_i \,\rangle\, |^2 \le \|\, \mathbf{x} \,\|^2$$

The inequality above is known as the **Bessel's inequality** and is true whether $\mathbf{X}$ is finite or infinite dimensional.