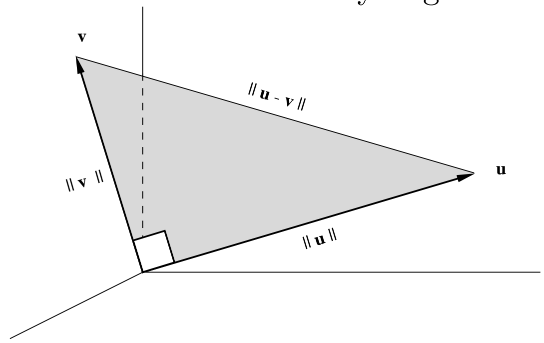
5.4 ORTHOGONAL VECTORS

Two vectors in $\Re^3$ are ***orthogonal*** (perpendicular) if the angle between them is a right angle (90°). But the visual concept of a right angle is not at our disposal in higher dimensions, so we must dig a little deeper. The essence of perpendicularity in $\Re^2$ and $\Re^3$ is embodied in the classical Pythagorean theorem,



which says that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u}-\mathbf{v}\|^2$. But [39] $\|\mathbf{u}\|^2 = \mathbf{u}^T\mathbf{u}$ for all $\mathbf{u} \in \Re^3$, and $\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u}$, so we can rewrite the Pythagorean statement as

$$\begin{aligned} \mathbf{0} &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2 = \mathbf{u}^T\mathbf{u} + \mathbf{v}^T\mathbf{v} - (\mathbf{u}-\mathbf{v})^T(\mathbf{u}-\mathbf{v}) \\ &= \mathbf{u}^T\mathbf{u} + \mathbf{v}^T\mathbf{v} - \left(\mathbf{u}^T\mathbf{u} - \mathbf{u}^T\mathbf{v} - \mathbf{v}^T\mathbf{u} + \mathbf{v}^T\mathbf{v}\right) = 2\mathbf{u}^T\mathbf{v}. \end{aligned}$$

Therefore, $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors in $\Re^3$ if and only if $\mathbf{u}^T\mathbf{v} = 0$. The natural extension of this provides us with a definition in more general spaces.

Orthogonality

In an inner-product space $\mathcal{V}$, two vectors $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ are said to be ***orthogonal*** (to each other) whenever $\langle \mathbf{x}|\mathbf{y} \rangle = 0$, and this is denoted by writing $\mathbf{x} \perp \mathbf{y}$.

- For $\Re^n$ with the standard inner product, $\mathbf{x} \perp \mathbf{y} \iff \mathbf{x}^T\mathbf{y} = 0$.
- For $\mathcal{C}^n$ with the standard inner product, $\mathbf{x} \perp \mathbf{y} \iff \mathbf{x}^*\mathbf{y} = 0$.

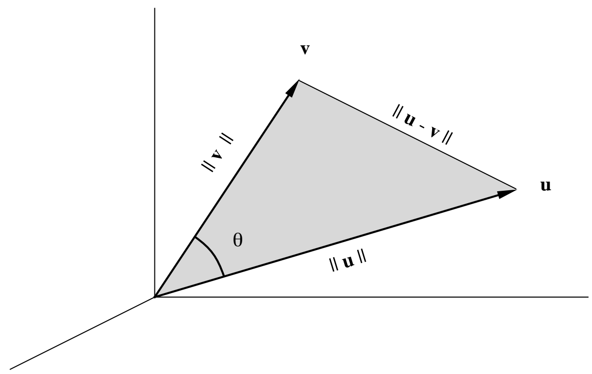
Example 5.4.1

$$\mathbf{x} = \begin{pmatrix} 1 \\ -2 \\ 3 \\ -1 \end{pmatrix} \text{ is orthogonal to } \mathbf{y} = \begin{pmatrix} 4 \\ 1 \\ -2 \\ -4 \end{pmatrix} \text{ because } \mathbf{x}^T\mathbf{y} = 0.$$

[39] Throughout this section, only norms generated by an underlying inner product $\|\star\|^2 = \langle \star|\star \rangle$ are used, so distinguishing subscripts on the norm notation can be omitted.

In spite of the fact that $\mathbf{u}^T\mathbf{v} = 0$, the vectors $\mathbf{u} = \begin{pmatrix} \mathrm{i} \\ 3 \\ 1 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} \mathrm{i} \\ 0 \\ 1 \end{pmatrix}$ are *not* orthogonal because $\mathbf{u}^*\mathbf{v} \neq 0$.

Now that "right angles" in higher dimensions make sense, how can more general angles be defined? Proceed just as before, but use the law of cosines rather than the Pythagorean theorem. Recall that



the ***law of cosines*** in $\Re^2$ or $\Re^3$ says $\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\,\|\mathbf{u}\|\;\|\mathbf{v}\|\cos\theta$. If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, then this reduces to the Pythagorean theorem. But, in general,

$$\cos\theta = \frac{\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2}{2\,\|\mathbf{u}\|\;\|\mathbf{v}\|} = \frac{\mathbf{u}^T\mathbf{u} + \mathbf{v}^T\mathbf{v} - (\mathbf{u}-\mathbf{v})^T(\mathbf{u}-\mathbf{v})}{2\,\|\mathbf{u}\|\;\|\mathbf{v}\|}$$

$$= \frac{2\mathbf{u}^T\mathbf{v}}{2\,\|\mathbf{u}\|\;\|\mathbf{v}\|} = \frac{\mathbf{u}^T\mathbf{v}}{\|\mathbf{u}\|\;\|\mathbf{v}\|}.$$

This easily extends to higher dimensions because if $\mathbf{x}, \mathbf{y}$ are vectors from any real inner-product space, then the general CBS inequality (5.3.4) on p. 287 guarantees that $\langle\mathbf{x}|\mathbf{y}\rangle / \|\mathbf{x}\|\;\|\mathbf{y}\|$ is a number in the interval $[-1,\,1]$, and hence there is a unique value θ in $[0,\,\pi]$ such that $\cos\theta = \langle\mathbf{x}|\mathbf{y}\rangle / \|\mathbf{x}\|\;\|\mathbf{y}\|$.

Angles

In a real inner-product space $\mathcal{V}$, the radian measure of the ***angle*** between nonzero vectors $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ is defined to be the number $\theta \in [0,\,\pi]$ such that

$$\cos\theta = \frac{\langle\mathbf{x}|\mathbf{y}\rangle}{\|\mathbf{x}\|\;\|\mathbf{y}\|}. \qquad (5.4.1)$$

Example 5.4.2

In $\Re^n$, $\cos\theta = \mathbf{x}^T\mathbf{y}/\|\mathbf{x}\|\,\|\mathbf{y}\|$. For example, to determine the angle between $\mathbf{x} = \begin{pmatrix} -4 \\ 2 \\ 1 \\ 2 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \end{pmatrix}$, compute $\cos\theta = 2/(5)(3) = 2/15$, and use the inverse cosine function to conclude that $\theta = 1.437$ radians (rounded).

Example 5.4.3

Linear Correlation. Suppose that an experiment is conducted, and the resulting observations are recorded in two data vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \text{and let } \mathbf{e} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Problem: Determine to what extent the y_i 's are linearly related to the x_i 's. That is, measure how close $\mathbf{y}$ is to being a linear combination $\beta_0\mathbf{e} + \beta_1\mathbf{x}$.

Solution: The cosine as defined in (5.4.1) does the job. To understand how, let $\mu_\mathbf{x}$ and $\sigma_\mathbf{x}$ be the ***mean*** and ***standard deviation*** of the data in $\mathbf{x}$. That is,

$$\mu_\mathbf{x} = \frac{\sum_i x_i}{n} = \frac{\mathbf{e}^T\mathbf{x}}{n} \quad \text{and} \quad \sigma_\mathbf{x} = \sqrt{\frac{\sum_i (x_i - \mu_\mathbf{x})^2}{n}} = \frac{\|\mathbf{x} - \mu_\mathbf{x}\mathbf{e}\|_2}{\sqrt{n}}.$$

The mean is a measure of central tendency, and the standard deviation measures the extent to which the data is spread. Frequently, raw data from different sources is difficult to compare because the units of measure are different—e.g., one researcher may use the metric system while another uses American units. To compensate, data is almost always first "standardized" into unitless quantities. The ***standardization*** of a vector $\mathbf{x}$ for which $\sigma_\mathbf{x} \neq 0$ is defined to be

$$\mathbf{z_x} = \frac{\mathbf{x} - \mu_\mathbf{x}\mathbf{e}}{\sigma_\mathbf{x}}.$$

Entries in $\mathbf{z_x}$ are often referred to as ***standard scores*** or ***z-scores.*** All standardized vectors have the properties that $\|\mathbf{z}\| = \sqrt{n}$, $\mu_\mathbf{z} = 0$, and $\sigma_\mathbf{z} = 1$. Furthermore, it's not difficult to verify that for vectors $\mathbf{x}$ and $\mathbf{y}$ such that $\sigma_\mathbf{x} \neq 0$ and $\sigma_\mathbf{y} \neq 0$, it's the case that

$$\begin{aligned} \mathbf{z_x} = \mathbf{z_y} &\iff \exists \text{ constants } \beta_0, \beta_1 \text{ such that } \mathbf{y} = \beta_0\mathbf{e} + \beta_1\mathbf{x}, \quad \text{where} \quad \beta_1 > 0, \\ \mathbf{z_x} = -\mathbf{z_y} &\iff \exists \text{ constants } \beta_0, \beta_1 \text{ such that } \mathbf{y} = \beta_0\mathbf{e} + \beta_1\mathbf{x}, \quad \text{where} \quad \beta_1 < 0. \end{aligned}$$

- In other words, $\mathbf{y} = \beta_0\mathbf{e} + \beta_1\mathbf{x}$ for some β_0 and β_1 if and only if $\mathbf{z_x} = \pm\mathbf{z_y}$, in which case we say $\mathbf{y}$ is ***perfectly linearly correlated*** with $\mathbf{x}$.

Since $\mathbf{z_x}$ varies continuously with $\mathbf{x}$, the existence of a "near" linear relationship between $\mathbf{x}$ and $\mathbf{y}$ is equivalent to $\mathbf{z_x}$ being "close" to $\pm\mathbf{z_y}$ in some sense. The fact that $\|\mathbf{z_x}\| = \|\pm\mathbf{z_y}\| = \sqrt{n}$ means $\mathbf{z_x}$ and $\pm\mathbf{z_y}$ differ only in orientation, so a natural measure of how close $\mathbf{z_x}$ is to $\pm\mathbf{z_y}$ is $\cos\theta$, where θ is the angle between $\mathbf{z_x}$ and $\mathbf{z_y}$. The number

$$\rho_{\mathbf{xy}} = \cos\theta = \frac{\mathbf{z_x}^T\mathbf{z_y}}{\|\mathbf{z_x}\|\,\|\mathbf{z_y}\|} = \frac{\mathbf{z_x}^T\mathbf{z_y}}{n} = \frac{(\mathbf{x}-\mu_{\mathbf{x}}\mathbf{e})^T(\mathbf{y}-\mu_{\mathbf{y}}\mathbf{e})}{\|\mathbf{x}-\mu_{\mathbf{x}}\mathbf{e}\|\,\|\mathbf{y}-\mu_{\mathbf{y}}\mathbf{e}\|}$$

is called the ***coefficient of linear correlation,*** and the following facts are now immediate.

- $\rho_{\mathbf{xy}} = 0$ if and only if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, in which case we say that $\mathbf{x}$ and $\mathbf{y}$ are ***completely uncorrelated.***
- $|\rho_{\mathbf{xy}}| = 1$ if and only if $\mathbf{y}$ is *perfectly* correlated with $\mathbf{x}$. That is, $|\rho_{\mathbf{xy}}| = 1$ if and only if there exists a linear relationship $\mathbf{y} = \beta_0\mathbf{e} + \beta_1\mathbf{x}$.
  - ▷ When $\beta_1 > 0$, we say that $\mathbf{y}$ is ***positively correlated*** with $\mathbf{x}$.
  - ▷ When $\beta_1 < 0$, we say that $\mathbf{y}$ is ***negatively correlated*** with $\mathbf{x}$.
- $|\rho_{\mathbf{xy}}|$ measures the degree to which $\mathbf{y}$ is linearly related to $\mathbf{x}$. In other words, $|\rho_{\mathbf{xy}}| \approx 1$ if and only if $\mathbf{y} \approx \beta_0\mathbf{e} + \beta_1\mathbf{x}$ for some β_0 and β_1.
  - ▷ Positive correlation is measured by the degree to which $\rho_{\mathbf{xy}} \approx 1$.
  - ▷ Negative correlation is measured by the degree to which $\rho_{\mathbf{xy}} \approx -1$.

If the data in $\mathbf{x}$ and $\mathbf{y}$ are plotted in $\Re^2$ as points (x_i, y_i), then, as depicted in Figure 5.4.1, $\rho_{\mathbf{xy}} \approx 1$ means that the points lie near a straight line with positive slope, while $\rho_{\mathbf{xy}} \approx -1$ means that the points lie near a line with negative slope, and $\rho_{\mathbf{xy}} \approx 0$ means that the points do not lie near a straight line.

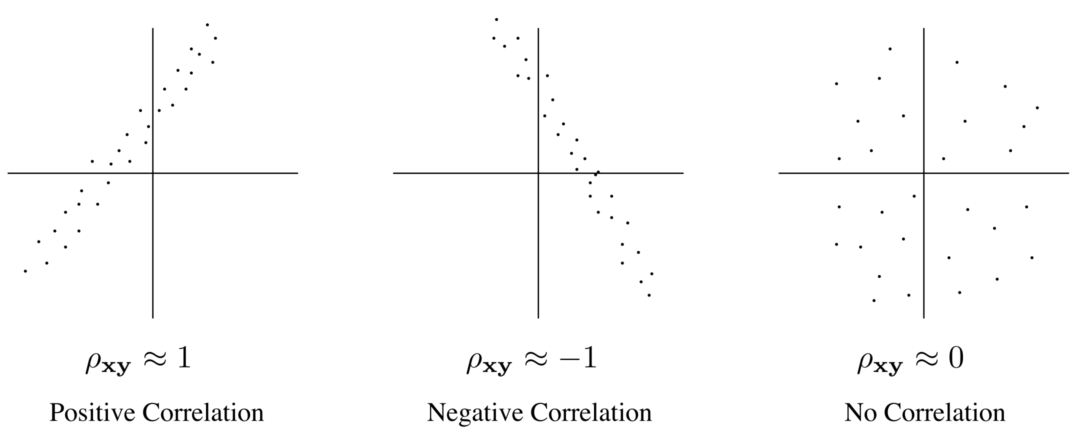


FIGURE 5.4.1

If $|\rho_{\mathbf{xy}}| \approx 1$, then the theory of least squares as presented in §4.6 can be used to determine a "best-fitting" straight line.

Orthonormal Sets

$\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is called an ***orthonormal set*** whenever $\|\mathbf{u}_i\| = 1$ for each i, and $\mathbf{u}_i \perp \mathbf{u}_j$ for all $i \neq j$. In other words,

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

- Every orthonormal set is linearly independent. (5.4.2)
- Every orthonormal set of n vectors from an n-dimensional space $\mathcal{V}$ is an orthonormal basis for $\mathcal{V}$.

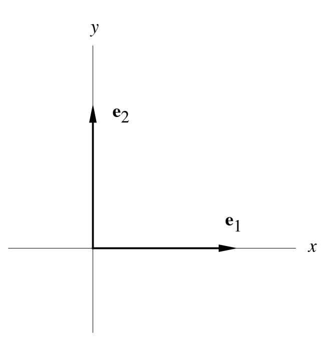
Proof. The second point follows from the first. To prove the first statement, suppose $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is orthonormal. If $\mathbf{0} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_n\mathbf{u}_n$, use the properties of an inner product to write

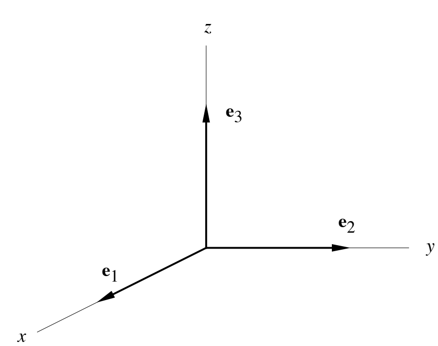
$$\begin{aligned} 0 = \langle \mathbf{u}_i | \mathbf{0} \rangle &= \langle \mathbf{u}_i | \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_n\mathbf{u}_n \rangle \\ &= \alpha_1 \langle \mathbf{u}_i | \mathbf{u}_1 \rangle + \cdots + \alpha_i \langle \mathbf{u}_i | \mathbf{u}_i \rangle + \cdots + \alpha_n \langle \mathbf{u}_i | \mathbf{u}_n \rangle = \alpha_i \|\mathbf{u}_i\|^2 \\ &= \alpha_i \quad \text{for each } i. \quad \blacksquare \end{aligned}$$

Example 5.4.4

The set $\mathcal{B}' = \left\{ \mathbf{u}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \right\}$ is a set of mutually orthogonal vectors because $\mathbf{u}_i^T\mathbf{u}_j = 0$ for $i \neq j$, but $\mathcal{B}'$ is *not* an orthonormal set—each vector does not have unit length. However, it's easy to convert an orthogonal set (not containing a zero vector) into an orthonormal set by simply normalizing each vector. Since $\|\mathbf{u}_1\| = \sqrt{2}$, $\|\mathbf{u}_2\| = \sqrt{3}$, and $\|\mathbf{u}_3\| = \sqrt{6}$, it follows that $\mathcal{B} = \left\{\mathbf{u}_1/\sqrt{2}, \mathbf{u}_2/\sqrt{3}, \mathbf{u}_3/\sqrt{6}\right\}$ is orthonormal.

The most common orthonormal basis is $\mathcal{S} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, the standard basis for $\Re^n$ and $\mathcal{C}^n$, and, as illustrated below for $\Re^2$ and $\Re^3$, these orthonormal vectors are directed along the standard coordinate axes.





Another orthonormal basis $\mathcal{B}$ need not be directed in the same way as $\mathcal{S}$, but that's the only significant difference because it's geometrically evident that $\mathcal{B}$ must amount to some rotation of $\mathcal{S}$. Consequently, we should expect general orthonormal bases to provide essentially the same advantages as the standard basis. For example, an important function of the standard basis $\mathcal{S}$ for $\Re^n$ is to provide coordinate representations by writing

$$\mathbf{x} = [\mathbf{x}]_{\mathcal{S}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{to mean} \quad \mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n.$$

With respect to a general basis $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$, the coordinates of $\mathbf{x}$ are the scalars ξ_i in the representation $\mathbf{x} = \xi_1\mathbf{u}_1 + \xi_2\mathbf{u}_2 + \cdots + \xi_n\mathbf{u}_n$, and, as illustrated in Example 4.7.2, finding the ξ_i 's requires solving an $n \times n$ system, a nuisance we would like to avoid. But if $\mathcal{B}$ is an *orthonormal* basis, then the ξ_i 's are readily available because $\langle \mathbf{u}_i | \mathbf{x} \rangle = \langle \mathbf{u}_i | \xi_1\mathbf{u}_1 + \xi_2\mathbf{u}_2 + \cdots + \xi_n\mathbf{u}_n \rangle = \sum_{j=1}^{n} \xi_j \langle \mathbf{u}_i | \mathbf{u}_j \rangle = \xi_i \|\mathbf{u}_i\|^2 = \xi_i$. This yields the ***Fourier***[40] ***expansion*** of $\mathbf{x}$.

Fourier Expansions

If $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for an inner-product space $\mathcal{V}$, then each $\mathbf{x} \in \mathcal{V}$ can be expressed as

$$\mathbf{x} = \langle \mathbf{u}_1 | \mathbf{x} \rangle \mathbf{u}_1 + \langle \mathbf{u}_2 | \mathbf{x} \rangle \mathbf{u}_2 + \cdots + \langle \mathbf{u}_n | \mathbf{x} \rangle \mathbf{u}_n. \tag{5.4.3}$$

This is called the ***Fourier expansion*** of $\mathbf{x}$. The scalars $\xi_i = \langle \mathbf{u}_i | \mathbf{x} \rangle$ are the coordinates of $\mathbf{x}$ with respect to $\mathcal{B}$, and they are called the ***Fourier coefficients.*** Geometrically, the Fourier expansion resolves $\mathbf{x}$ into n mutually orthogonal vectors $\langle \mathbf{u}_i | \mathbf{x} \rangle \mathbf{u}_i$, each of which represents the orthogonal projection of $\mathbf{x}$ onto the space (line) spanned by $\mathbf{u}_i$. (More is said in Example 5.13.1 on p. 431 and Exercise 5.13.11.)

---

[40] Jean Baptiste Joseph Fourier (1768–1830) was a French mathematician and physicist who, while studying heat flow, developed expansions similar to (5.4.3). Fourier's work dealt with special infinite-dimensional inner-product spaces involving trigonometric functions as discussed in Example 5.4.6. Although they were apparently used earlier by Daniel Bernoulli (1700–1782) to solve problems concerned with vibrating strings, these orthogonal expansions became known as ***Fourier series,*** and they are now a fundamental tool in applied mathematics. Born the son of a tailor, Fourier was orphaned at the age of eight. Although he showed a great aptitude for mathematics at an early age, he was denied his dream of entering the French artillery because of his "low birth." Instead, he trained for the priesthood, but he never took his vows. However, his talents did not go unrecognized, and he later became a favorite of Napoleon. Fourier's work is now considered as marking an epoch in the history of both pure and applied mathematics. The next time you are in Paris, check out Fourier's plaque on the first level of the Eiffel Tower.

Example 5.4.5

Problem: Determine the Fourier expansion of $\mathbf{x} = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$ with respect to the standard inner product and the orthonormal basis given in Example 5.4.4

$$\mathcal{B} = \left\{ \mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{u}_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \right\}.$$

Solution: The Fourier coefficients are

$$\xi_1 = \langle \mathbf{u}_1 | \mathbf{x} \rangle = \frac{-3}{\sqrt{2}}, \quad \xi_2 = \langle \mathbf{u}_2 | \mathbf{x} \rangle = \frac{2}{\sqrt{3}}, \quad \xi_3 = \langle \mathbf{u}_3 | \mathbf{x} \rangle = \frac{1}{\sqrt{6}},$$

so

$$\mathbf{x} = \xi_1 \mathbf{u}_1 + \xi_2 \mathbf{u}_2 + \xi_3 \mathbf{u}_3 = \frac{1}{2} \begin{pmatrix} -3 \\ 3 \\ 0 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} + \frac{1}{6} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}.$$

You may find it instructive to sketch a picture of these vectors in $\Re^3$.

Example 5.4.6

Fourier Series. Let $\mathcal{V}$ be the inner-product space of real-valued functions that are integrable on the interval $(-\pi, \pi)$ and where the inner product and norm are given by

$$\langle f | g \rangle = \int_{-\pi}^{\pi} f(t) g(t) dt \quad \text{and} \quad \|f\| = \left(\int_{-\pi}^{\pi} f^2(t) dt \right)^{1/2}.$$

It's straightforward to verify that the set of trigonometric functions

$$\mathcal{B}' = \{1, \cos t, \cos 2t, \ldots, \sin t, \sin 2t, \sin 3t, \ldots\}$$

is a set of mutually orthogonal vectors, so normalizing each vector produces the orthonormal set

$$\mathcal{B} = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos t}{\sqrt{\pi}}, \frac{\cos 2t}{\sqrt{\pi}}, \ldots, \frac{\sin t}{\sqrt{\pi}}, \frac{\sin 2t}{\sqrt{\pi}}, \frac{\sin 3t}{\sqrt{\pi}}, \ldots \right\}.$$

Given an arbitrary $f \in \mathcal{V}$, we construct its Fourier expansion

$$F(t) = \alpha_0 \frac{1}{\sqrt{2\pi}} + \sum_{k=1}^{\infty} \alpha_k \frac{\cos kt}{\sqrt{\pi}} + \sum_{k=1}^{\infty} \beta_k \frac{\sin kt}{\sqrt{\pi}}, \tag{5.4.4}$$

where the Fourier coefficients are given by

$$\alpha_0 = \left\langle \frac{1}{\sqrt{2\pi}} \middle| f \right\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(t)dt\,,$$
$$\alpha_k = \left\langle \frac{\cos kt}{\sqrt{\pi}} \middle| f \right\rangle = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(t)\cos kt\,dt \quad \text{for } k = 1, 2, 3, \ldots,$$
$$\beta_k = \left\langle \frac{\sin kt}{\sqrt{\pi}} \middle| f \right\rangle = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(t)\sin kt\,dt \quad \text{for } k = 1, 2, 3, \ldots.$$

Substituting these coefficients in (5.4.4) produces the infinite series

$$F(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt)\,, \tag{5.4.5}$$

where

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt \quad \text{and} \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt. \tag{5.4.6}$$

The series $F(t)$ in (5.4.5) is called the ***Fourier series*** expansion for $f(t)$, but, unlike the situation in finite-dimensional spaces, $F(t)$ need not agree with the original function $f(t)$. After all, F is periodic, so there is no hope of agreement when f is not periodic. However, the following statement is true.

- If $f(t)$ is a periodic function with period 2π that is sectionally continuous[41] on the interval $(-\pi, \pi)$, then the Fourier series $F(t)$ converges to $f(t)$ at each $t \in (-\pi, \pi)$, where f is continuous. If f is discontinuous at t_0 but possesses left-hand and right-hand derivatives at t_0, then $F(t_0)$ converges to the average value

$$F(t_0) = \frac{f(t_0^-) + f(t_0^+)}{2},$$

  where $f(t_0^-)$ and $f(t_0^+)$ denote the one-sided limits $f(t_0^-) = \lim_{t\to t_0^-} f(t)$ and $f(t_0^+) = \lim_{t\to t_0^+} f(t)$.

For example, the ***square wave function*** defined by

$$f(t) = \begin{cases} -1 & \text{when } -\pi < t < 0, \\ 1 & \text{when } \quad 0 < t < \pi, \end{cases}$$

[41] A function f is sectionally continuous on (a, b) when f has only a finite number of discontinuities in (a, b) and the one-sided limits exist at each point of discontinuity as well as at the end points a and b.

and illustrated in Figure 5.4.2, satisfies these conditions. The value of f at $t=0$ is irrelevant—it's not even necessary that $f(0)$ be defined.

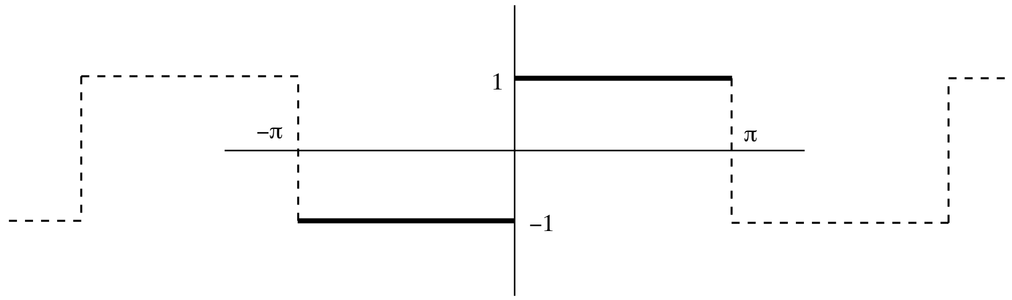


FIGURE 5.4.2

To find the Fourier series expansion for f, compute the coefficients in (5.4.6) as

$$\begin{aligned} a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt = \frac{1}{\pi}\int_{-\pi}^{0} -\cos nt\,dt + \frac{1}{\pi}\int_{0}^{\pi} \cos nt\,dt \\ &= 0, \\ b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt = \frac{1}{\pi}\int_{-\pi}^{0} -\sin nt\,dt + \frac{1}{\pi}\int_{0}^{\pi} \sin nt\,dt \\ &= \frac{2}{n\pi}(1-\cos n\pi) = \begin{cases} 0 & \text{when } n \text{ is even,} \\ 4/n\pi & \text{when } n \text{ is odd,} \end{cases} \end{aligned}$$

so that

$$F(t) = \frac{4}{\pi}\sin t + \frac{4}{3\pi}\sin 3t + \frac{4}{5\pi}\sin 5t + \cdots = \sum_{n=1}^{\infty} \frac{4}{(2n-1)\pi}\sin(2n-1)t.$$

For each $t \in (-\pi, \pi)$, except $t = 0$, it must be the case that $F(t) = f(t)$, and

$$F(0) = \frac{f(0^-) + f(0^+)}{2} = 0.$$

Not only does $F(t)$ agree with $f(t)$ everywhere f is defined, but F also provides a *periodic extension* of f in the sense that the graph of $F(t)$ is the entire square wave depicted in Figure 5.4.2—the values at the points of discontinuity (the jumps) are $F(\pm n\pi) = 0$.

Exercises for section 5.4

5.4.1. Using the standard inner product, determine which of the following pairs are orthogonal vectors in the indicated space.

(a) $\mathbf{x} = \begin{pmatrix} 1 \\ -3 \\ 4 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} -2 \\ 2 \\ 2 \end{pmatrix}$ in $\Re^3$,

(b) $\mathbf{x} = \begin{pmatrix} \mathrm{i} \\ 1+\mathrm{i} \\ 2 \\ 1-\mathrm{i} \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} 0 \\ 1+\mathrm{i} \\ -2 \\ 1-\mathrm{i} \end{pmatrix}$ in $\mathcal{C}^4$,

(c) $\mathbf{x} = \begin{pmatrix} 1 \\ -2 \\ 3 \\ 4 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} 4 \\ 2 \\ -1 \\ 1 \end{pmatrix}$ in $\Re^4$,

(d) $\mathbf{x} = \begin{pmatrix} 1+\mathrm{i} \\ 1 \\ \mathrm{i} \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} 1-\mathrm{i} \\ -3 \\ -\mathrm{i} \end{pmatrix}$ in $\mathcal{C}^3$,

(e) $\mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ in $\Re^n$.

5.4.2. Find two vectors of unit norm that are orthogonal to $\mathbf{u} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$.

5.4.3. Consider the following set of three vectors.

$$\left\{ \mathbf{x}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} -1 \\ -1 \\ 2 \\ 0 \end{pmatrix} \right\}.$$

(a) Using the standard inner product in $\Re^4$, verify that these vectors are mutually orthogonal.

(b) Find a nonzero vector $\mathbf{x}_4$ such that $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$ is a set of mutually orthogonal vectors.

(c) Convert the resulting set into an orthonormal basis for $\Re^4$.

5.4.4. Using the standard inner product, determine the Fourier expansion of $\mathbf{x}$ with respect to $\mathcal{B}$, where

$$\mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} \quad \text{and} \quad \mathcal{B} = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \right\}.$$

5.4.5. With respect to the inner product for matrices given by (5.3.2), verify that the set

$$\mathcal{B} = \left\{ \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},\ \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right\}$$

is an orthonormal basis for $\Re^{2\times 2}$, and then compute the Fourier expansion of $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ with respect to $\mathcal{B}$.

5.4.6. Determine the angle between $\mathbf{x} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

5.4.7. Given an orthonormal basis $\mathcal{B}$ for a space $\mathcal{V}$, explain why the Fourier expansion for $\mathbf{x} \in \mathcal{V}$ is uniquely determined by $\mathcal{B}$.

5.4.8. Explain why the columns of $\mathbf{U}_{n\times n}$ are an orthonormal basis for $\mathcal{C}^n$ if and only if $\mathbf{U}^* = \mathbf{U}^{-1}$. Such matrices are said to be ***unitary***—their properties are studied in a later section.

5.4.9. Matrices with the property $\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$ are said to be ***normal.*** Notice that hermitian matrices as well as real symmetric matrices are included in the class of normal matrices. Prove that if $\mathbf{A}$ is normal, then $R(\mathbf{A}) \perp N(\mathbf{A})$—i.e., every vector in $R(\mathbf{A})$ is orthogonal to every vector in $N(\mathbf{A})$. **Hint:** Recall equations (4.5.5) and (4.5.6).

5.4.10. Using the trace inner product described in Example 5.3.1, determine the angle between the following pairs of matrices.

(a) $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

(b) $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 2 & -2 \\ 2 & 0 \end{pmatrix}$.

5.4.11. Why is the definition for $\cos\theta$ given in (5.4.1) not good for $\mathcal{C}^n$? Explain how to define $\cos\theta$ so that it makes sense in $\mathcal{C}^n$.

5.4.12. If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for an inner-product space $\mathcal{V}$, explain why

$$\langle \mathbf{x}|\mathbf{y}\rangle = \sum_i \langle \mathbf{x}|\mathbf{u}_i\rangle \langle \mathbf{u}_i|\mathbf{y}\rangle$$

holds for every $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.

5.4.13. Consider a real inner-product space, where $\|\star\|^2 = \langle \star|\star \rangle$.
(a) Prove that if $\|\mathbf{x}\| = \|\mathbf{y}\|$, then $(\mathbf{x}+\mathbf{y}) \perp (\mathbf{x}-\mathbf{y})$.
(b) For the standard inner product in $\Re^2$, draw a picture of this. That is, sketch the location of $\mathbf{x}+\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ for two vectors with equal norms.

5.4.14. Pythagorean Theorem. Let $\mathcal{V}$ be a general inner-product space in which $\|\star\|^2 = \langle \star|\star \rangle$.
(a) When $\mathcal{V}$ is a *real* space, prove that $\mathbf{x} \perp \mathbf{y}$ if and only if $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$. (Something would be wrong if this were not true because this is where the definition of orthogonality originated.)
(b) Construct an example to show that one of the implications in part (a) does not hold when $\mathcal{V}$ is a *complex* space.
(c) When $\mathcal{V}$ is a complex space, prove that $\mathbf{x} \perp \mathbf{y}$ if and only if $\|\alpha\mathbf{x}+\beta\mathbf{y}\|^2 = \|\alpha\mathbf{x}\|^2 + \|\beta\mathbf{y}\|^2$ for all scalars α and β.

5.4.15. Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for an inner-product space $\mathcal{V}$, and let $\mathbf{x} = \sum_i \xi_i \mathbf{u}_i$ be the Fourier expansion of $\mathbf{x} \in \mathcal{V}$.
(a) If $\mathcal{V}$ is a real space, and if θ_i is the angle between $\mathbf{u}_i$ and $\mathbf{x}$, explain why

$$\xi_i = \|\mathbf{x}\| \cos\theta_i.$$

Sketch a picture of this in $\Re^2$ or $\Re^3$ to show why the component $\xi_i\mathbf{u}_i$ represents the orthogonal projection of $\mathbf{x}$ onto the line determined by $\mathbf{u}_i$, and thus illustrate the fact that a Fourier expansion is nothing more than simply resolving $\mathbf{x}$ into mutually orthogonal components.
(b) Derive ***Parseval's identity***,[42] which says $\sum_{i=1}^n |\xi_i|^2 = \|\mathbf{x}\|^2$.

5.4.16. Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be an orthonormal set in an n-dimensional inner-product space $\mathcal{V}$. Derive ***Bessel's inequality***,[43] which says that if $\mathbf{x} \in \mathcal{V}$ and $\xi_i = \langle \mathbf{u}_i|\mathbf{x} \rangle$, then

$$\sum_{i=1}^{k} |\xi_i|^2 \le \|\mathbf{x}\|^2.$$

Explain why equality holds if and only if $\mathbf{x} \in span\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$.
Hint: Consider $\|\mathbf{x} - \sum_{i=1}^k \xi_i\mathbf{u}_i\|^2$.

---

[42] This result appeared in the second of the five mathematical publications by Marc-Antoine Parseval des Chênes (1755–1836). Parseval was a royalist who had to flee from France when Napoleon ordered his arrest for publishing poetry against the regime.

[43] This inequality is named in honor of the German astronomer and mathematician Friedrich Wilhelm Bessel (1784–1846), who devoted his life to understanding the motions of the stars. In the process he introduced several useful mathematical ideas.

5.4.17. Construct an example using the standard inner product in $\Re^n$ to show that two vectors $\mathbf{x}$ and $\mathbf{y}$ can have an angle between them that is close to $\pi/2$ without $\mathbf{x}^T\mathbf{y}$ being close to 0. **Hint:** Consider n to be large, and use the vector $\mathbf{e}$ of all 1's for one of the vectors.

5.4.18. It was demonstrated in Example 5.4.3 that $\mathbf{y}$ is linearly correlated with $\mathbf{x}$ in the sense that $\mathbf{y} \approx \beta_0\mathbf{e} + \beta_1\mathbf{x}$ if and only if the standardization vectors $\mathbf{z_x}$ and $\mathbf{z_y}$ are "close" in the sense that they are almost on the same line in $\Re^n$. Explain why simply measuring $\|\mathbf{z_x} - \mathbf{z_y}\|_2$ does not always gauge the degree of linear correlation.

5.4.19. Let θ be the angle between two vectors $\mathbf{x}$ and $\mathbf{y}$ from a real inner-product space.

(a) Prove that $\cos\theta = 1$ if and only if $\mathbf{y} = \alpha\mathbf{x}$ for $\alpha > 0$.

(b) Prove that $\cos\theta = -1$ if and only if $\mathbf{y} = \alpha\mathbf{x}$ for $\alpha < 0$.

Hint: Use the generalization of Exercise 5.1.9.

5.4.20. With respect to the orthonormal set

$$\mathcal{B} = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos t}{\sqrt{\pi}}, \frac{\cos 2t}{\sqrt{\pi}}, \ldots, \frac{\sin t}{\sqrt{\pi}}, \frac{\sin 2t}{\sqrt{\pi}}, \frac{\sin 3t}{\sqrt{\pi}}, \ldots \right\},$$

determine the Fourier series expansion of the ***saw-toothed function*** defined by $f(t) = t$ for $-\pi < t < \pi$. The periodic extension of this function is depicted in Figure 5.4.3.

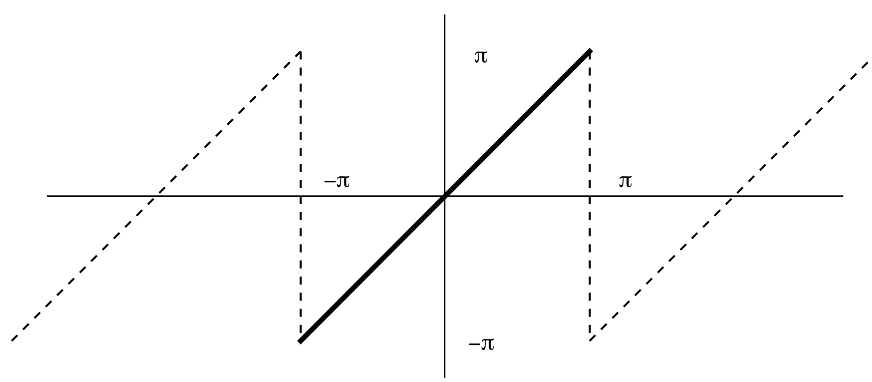


FIGURE 5.4.3

Solutions for exercises in section 5. 4

5.4.1. (a), (b), and (e) are orthogonal pairs.

5.4.2. First find $\mathbf{v} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$ such that $3\alpha_1 - 2\alpha_2 = 0$, and then normalize $\mathbf{v}$. The second must be the negative of $\mathbf{v}$.

5.4.3. (a) Simply verify that $\mathbf{x}_i^T\mathbf{x}_j = 0$ for $i \neq j$.

(b) Let $\mathbf{x}_4^T = (\,\alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \alpha_4\,)$, and notice that $\mathbf{x}_i^T\mathbf{x}_4 = 0$ for $i = 1, 2, 3$ is three homogeneous equations in four unknowns

$$\begin{pmatrix} 1 & -1 & 0 & 2 \\ 1 & 1 & 1 & 0 \\ -1 & -1 & 2 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \beta \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

(c) Simply normalize the set by dividing each vector by its norm.

5.4.4. The Fourier coefficients are

$$\xi_1 = \langle \mathbf{u}_1 | \mathbf{x} \rangle = \frac{1}{\sqrt{2}}, \quad \xi_2 = \langle \mathbf{u}_2 | \mathbf{x} \rangle = \frac{-1}{\sqrt{3}}, \quad \xi_3 = \langle \mathbf{u}_3 | \mathbf{x} \rangle = \frac{-5}{\sqrt{6}},$$

so

$$\mathbf{x} = \xi_1\mathbf{u}_1 + \xi_2\mathbf{u}_2 + \xi_3\mathbf{u}_3 = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{5}{6}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}.$$

5.4.5. If $\mathbf{U}_1$, $\mathbf{U}_2$, $\mathbf{U}_3$, and $\mathbf{U}_4$ denote the elements of $\mathcal{B}$, verify they constitute an orthonormal set by showing that

$$\langle \mathbf{U}_i | \mathbf{U}_j \rangle = trace(\mathbf{U}_i^T\mathbf{U}_j) = 0 \text{ for } i \neq j \quad \text{and} \quad \|\mathbf{U}_i\| = \sqrt{trace(\mathbf{U}_i^T\mathbf{U}_i)} = 1.$$

Consequently, $\mathcal{B}$ is linearly independent—recall (5.4.2)—and therefore $\mathcal{B}$ is a basis because it is a *maximal* independent set—part (b) of Exercise 4.4.4 insures $\dim \Re^{2\times 2} = 4$. The Fourier coefficients $\langle \mathbf{U}_i | \mathbf{A} \rangle = trace(\mathbf{U}_i^T\mathbf{A})$ are

$$\langle \mathbf{U}_1 | \mathbf{A} \rangle = \frac{2}{\sqrt{2}}, \quad \langle \mathbf{U}_2 | \mathbf{A} \rangle = 0, \quad \langle \mathbf{U}_3 | \mathbf{A} \rangle = 1, \quad \langle \mathbf{U}_4 | \mathbf{A} \rangle = 1,$$

so the Fourier expansion of $\mathbf{A}$ is $\mathbf{A} = (2/\sqrt{2})\mathbf{U}_1 + \mathbf{U}_3 + \mathbf{U}_4$.

5.4.6. $\cos\theta = \mathbf{x}^T\mathbf{y}/\,\|\mathbf{x}\|\;\|\mathbf{y}\| = 1/2$, so $\theta = \pi/3$.

5.4.7. This follows because each vector has a unique representation in terms of a basis—see Exercise 4.4.8 or the discussion of coordinates in §4.7.

5.4.8. If the columns of $\mathbf{U} = [\mathbf{u}_1 \,|\, \mathbf{u}_2 \,|\, \cdots \,|\, \mathbf{u}_n]$ are an orthonormal basis for $\mathcal{C}^n$, then

$$[\mathbf{U}^*\mathbf{U}]_{ij} = \mathbf{u}_i^*\mathbf{u}_j = \begin{cases} 1 & \text{when } i = j, \\ 0 & \text{when } i \neq j, \end{cases} \qquad (\ddagger)$$

and, therefore, $\mathbf{U}^*\mathbf{U} = \mathbf{I}$. Conversely, if $\mathbf{U}^*\mathbf{U} = \mathbf{I}$, then (‡) holds, so the columns of $\mathbf{U}$ are orthonormal—they are a basis for $\mathcal{C}^n$ because orthonormal sets are always linearly independent.

5.4.9. Equations (4.5.5) and (4.5.6) guarantee that

$$R(\mathbf{A}) = R(\mathbf{A}\mathbf{A}^*) \quad \text{and} \quad N(\mathbf{A}) = N(\mathbf{A}^*\mathbf{A}),$$

and consequently $\mathbf{r} \in R(\mathbf{A}) = R(\mathbf{A}\mathbf{A}^*) \implies \mathbf{r} = \mathbf{A}\mathbf{A}^*\mathbf{x}$ for some $\mathbf{x}$, and $\mathbf{n} \in N(\mathbf{A}) = N(\mathbf{A}^*\mathbf{A}) \implies \mathbf{A}^*\mathbf{A}\mathbf{n} = \mathbf{0}$. Therefore,

$$\langle \mathbf{r}|\mathbf{n}\rangle = \mathbf{r}^*\mathbf{n} = \mathbf{x}^*\mathbf{A}\mathbf{A}^*\mathbf{n} = \mathbf{x}^*\mathbf{A}^*\mathbf{A}\mathbf{n} = \mathbf{0}.$$

5.4.10. (a) $\pi/4$ (b) $\pi/2$

5.4.11. The number $\mathbf{x}^T\mathbf{y}$ or $\mathbf{x}^*\mathbf{y}$ will in general be complex. In order to guarantee that we end up with a real number, we should take

$$\cos\theta = \frac{|\mathrm{Re}\,(\mathbf{x}^*\mathbf{y})\,|}{\|\mathbf{x}\|\;\|\mathbf{y}\|}.$$

5.4.12. Use the Fourier expansion $\mathbf{y} = \sum_i \langle \mathbf{u}_i|\mathbf{y}\rangle\, \mathbf{u}_i$ together with the various properties of an inner product to write

$$\langle \mathbf{x}|\mathbf{y}\rangle = \left\langle \mathbf{x}\,\middle|\, \sum_i \langle \mathbf{u}_i|\mathbf{y}\rangle\, \mathbf{u}_i \right\rangle = \sum_i \langle \mathbf{x}|\, \langle \mathbf{u}_i|\mathbf{y}\rangle\, \mathbf{u}_i\rangle = \sum_i \langle \mathbf{u}_i|\mathbf{y}\rangle\, \langle \mathbf{x}|\mathbf{u}_i\rangle\,.$$

5.4.13. In a real space, $\langle \mathbf{x}|\mathbf{y}\rangle = \langle \mathbf{y}|\mathbf{x}\rangle$, so the third condition in the definition (5.3.1) of an inner product and Exercise 5.3.2(c) produce

$$\begin{aligned}\langle \mathbf{x}+\mathbf{y}|\mathbf{x}-\mathbf{y}\rangle &= \langle \mathbf{x}+\mathbf{y}|\mathbf{x}\rangle - \langle \mathbf{x}+\mathbf{y}|\mathbf{y}\rangle \\ &= \langle \mathbf{x}|\mathbf{x}\rangle + \langle \mathbf{y}|\mathbf{x}\rangle - \langle \mathbf{x}|\mathbf{y}\rangle - \langle \mathbf{y}|\mathbf{y}\rangle \\ &= \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 = 0.\end{aligned}$$

5.4.14. (a) In a real space, $\langle \mathbf{x}|\mathbf{y}\rangle = \langle \mathbf{y}|\mathbf{x}\rangle$, so the third condition in the definition (5.3.1) of an inner product and Exercise 5.3.2(c) produce

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

and hence $\langle \mathbf{x}|\mathbf{y}\rangle = 0$ if and only if $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$.

(b) In a complex space, $\mathbf{x} \perp \mathbf{y} \implies \|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$, but the converse is not valid—e.g., consider $\mathcal{C}^2$ with the standard inner product, and let $\mathbf{x} = \begin{pmatrix} -\mathrm{i} \\ 1 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix}$.

(c) Again, using the properties of a general inner product, derive the expansion

$$\begin{aligned}\|\alpha\mathbf{x}+\beta\mathbf{y}\|^2 &= \langle\alpha\mathbf{x}+\beta\mathbf{y}|\alpha\mathbf{x}+\beta\mathbf{y}\rangle\\ &= \langle\alpha\mathbf{x}|\alpha\mathbf{x}\rangle+\langle\alpha\mathbf{x}|\beta\mathbf{y}\rangle+\langle\beta\mathbf{y}|\alpha\mathbf{x}\rangle+\langle\beta\mathbf{y}|\beta\mathbf{y}\rangle\\ &= \|\alpha\mathbf{x}\|^2+\overline{\alpha}\beta\langle\mathbf{x}|\mathbf{y}\rangle+\overline{\beta}\alpha\langle\mathbf{y}|\mathbf{x}\rangle+\|\beta\mathbf{y}\|^2.\end{aligned}$$

Clearly, $\mathbf{x}\perp\mathbf{y} \implies \|\alpha\mathbf{x}+\beta\mathbf{y}\|^2=\|\alpha\mathbf{x}\|^2+\|\beta\mathbf{y}\|^2 \ \forall\ \alpha,\beta$. Conversely, if $\|\alpha\mathbf{x}+\beta\mathbf{y}\|^2=\|\alpha\mathbf{x}\|^2+\|\beta\mathbf{y}\|^2 \ \forall\ \alpha,\beta$, then $\overline{\alpha}\beta\langle\mathbf{x}|\mathbf{y}\rangle+\overline{\beta}\alpha\langle\mathbf{y}|\mathbf{x}\rangle=0 \ \forall\ \alpha,\beta$. Letting $\alpha=\langle\mathbf{x}|\mathbf{y}\rangle$ and $\beta=1$ produces the conclusion that $2|\langle\mathbf{x}|\mathbf{y}\rangle|^2=0$, and thus $\langle\mathbf{x}|\mathbf{y}\rangle=0$.

5.4.15. (a) $\cos\theta_i=\langle\mathbf{u}_i|\mathbf{x}\rangle/\|\mathbf{u}_i\|\ \|\mathbf{x}\|=\langle\mathbf{u}_i|\mathbf{x}\rangle/\|\mathbf{x}\|=\xi_i/\|\mathbf{x}\|$

(b) Use the Pythagorean theorem (Exercise 5.4.14) to write

$$\begin{aligned}\|\mathbf{x}\|^2 &= \|\xi_1\mathbf{u}_1+\xi_2\mathbf{u}_2+\cdots+\xi_n\mathbf{u}_n\|^2\\ &= \|\xi_1\mathbf{u}_1\|^2+\|\xi_2\mathbf{u}_2\|^2+\cdots+\|\xi_n\mathbf{u}_n\|^2\\ &= |\xi_1|^2+|\xi_2|^2+\cdots+|\xi_n|^2.\end{aligned}$$

5.4.16. Use the properties of an inner product to write

$$\begin{aligned}\left\|\mathbf{x}-\sum_{i=1}^k\xi_i\mathbf{u}_i\right\|^2 &= \left\langle\mathbf{x}-\sum_{i=1}^k\xi_i\mathbf{u}_i\middle|\mathbf{x}-\sum_{i=1}^k\xi_i\mathbf{u}_i\right\rangle\\ &= \langle\mathbf{x}|\mathbf{x}\rangle-2\sum_i|\xi_i|^2+\left\langle\sum_{i=1}^k\xi_i\mathbf{u}_i\middle|\sum_{i=1}^k\xi_i\mathbf{u}_i\right\rangle\\ &= \|\mathbf{x}\|^2-2\sum_i|\xi_i|^2+\left\|\sum_{i=1}^k\xi_i\mathbf{u}_i\right\|^2,\end{aligned}$$

and then invoke the Pythagorean theorem (Exercise 5.4.14) to conclude

$$\left\|\sum_{i=1}^k\xi_i\mathbf{u}_i\right\|^2=\sum_i\|\xi_i\mathbf{u}_i\|^2=\sum_i|\xi_i|^2.$$

Consequently,

$$0\le\left\|\mathbf{x}-\sum_{i=1}^k\xi_i\mathbf{u}_i\right\|^2=\|\mathbf{x}\|^2-\sum_i|\xi_i|^2 \implies \sum_{i=1}^k|\xi_i|^2\le\|\mathbf{x}\|^2. \qquad (\ddagger)$$

If $\mathbf{x}\in span\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k\}$, then the Fourier expansion of $\mathbf{x}$ with respect to the $\mathbf{u}_i$'s is $\mathbf{x}=\sum_{i=1}^k\xi_i\mathbf{u}_i$, and hence equality holds in $(\ddagger)$. Conversely, if equality holds in $(\ddagger)$, then $\mathbf{x}-\sum_{i=1}^k\xi_i\mathbf{u}_i=\mathbf{0}$.

5.4.17. Choose any unit vector $\mathbf{e}_i$ for $\mathbf{y}$. The angle between $\mathbf{e}$ and $\mathbf{e}_i$ approaches $\pi/2$ as $n \to \infty$, but $\mathbf{e}^T\mathbf{e}_i = 1$ for all n.

5.4.18. If $\mathbf{y}$ is negatively correlated to $\mathbf{x}$, then $\mathbf{z_x} = -\mathbf{z_y}$, but $\|\mathbf{z_x} - \mathbf{z_y}\|_2 = 2\sqrt{n}$ gives no indication of the fact that $\mathbf{z_x}$ and $\mathbf{z_y}$ are on the same line. Continuity therefore dictates that when $\mathbf{y} \approx \beta_0\mathbf{e} + \beta_1\mathbf{x}$ with $\beta_1 < 0$, then $\mathbf{z_x} \approx -\mathbf{z_y}$, but $\|\mathbf{z_x} - \mathbf{z_y}\|_2 \approx 2\sqrt{n}$ gives no hint that $\mathbf{z_x}$ and $\mathbf{z_y}$ are almost on the same line. If we want to use norms to gauge linear correlation, we should use

$$\min\left\{\|\mathbf{z_x} - \mathbf{z_y}\|_2\,,\ \|\mathbf{z_x} + \mathbf{z_y}\|_2\right\}.$$

5.4.19. (a) $\cos\theta = 1 \implies \langle \mathbf{x}|\mathbf{y}\rangle = \|\mathbf{x}\|\ \|\mathbf{y}\| > 0$, and the straightforward extension of Exercise 5.1.9 guarantees that

$$\mathbf{y} = \frac{\langle \mathbf{x}|\mathbf{y}\rangle}{\|\mathbf{x}\|^2}\mathbf{x}, \quad \text{and clearly} \quad \frac{\langle \mathbf{x}|\mathbf{y}\rangle}{\|\mathbf{x}\|^2} > 0.$$

Conversely, if $\mathbf{y} = \alpha\mathbf{x}$ for $\alpha > 0$, then $\langle \mathbf{x}|\mathbf{y}\rangle = \alpha\,\|\mathbf{x}\|^2 \implies \cos\theta = 1.$

(b) $\cos\theta = -1 \implies \langle \mathbf{x}|\mathbf{y}\rangle = -\|\mathbf{x}\|\ \|\mathbf{y}\| < 0$, so the generalized version of Exercise 5.1.9 guarantees that

$$\mathbf{y} = \frac{\langle \mathbf{x}|\mathbf{y}\rangle}{\|\mathbf{x}\|^2}\mathbf{x}, \quad \text{and in this case} \quad \frac{\langle \mathbf{x}|\mathbf{y}\rangle}{\|\mathbf{x}\|^2} < 0.$$

Conversely, if $\mathbf{y} = \alpha\mathbf{x}$ for $\alpha < 0$, then $\langle \mathbf{x}|\mathbf{y}\rangle = \alpha\,\|\mathbf{x}\|^2$, so

$$\cos\theta = \frac{\alpha\,\|\mathbf{x}\|^2}{|\alpha|\ \|\mathbf{x}\|^2} = -1.$$

5.4.20. $F(t) = \sum_n^\infty (-1)^n \frac{2}{n} \sin nt.$

Solutions for exercises in section 5. 5

5.5.1. (a)

$$\mathbf{u}_1 = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{2\sqrt{3}}\begin{pmatrix} 3 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_3 = \frac{1}{\sqrt{6}}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix}$$

(b) First verify this is an orthonormal set by showing $\mathbf{u}_i^T\mathbf{u}_j = \begin{cases} 1 & \text{when } i = j, \\ 0 & \text{when } i \neq j. \end{cases}$

To show that the $\mathbf{x}_i$'s and the $\mathbf{u}_i$'s span the same space, place the $\mathbf{x}_i$'s as rows in a matrix $\mathbf{A}$, and place the $\mathbf{u}_i$'s as rows in a matrix $\mathbf{B}$, and then verify that $\mathbf{E_A} = \mathbf{E_B}$—recall Example 4.2.2.

(c) The result should be the same as in part (a).