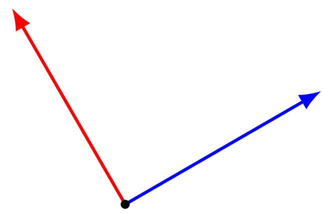

Figure 5.2 Orthonormal vectors in a plane.

Since the vectors $a_i = (b_i, \alpha_i)$ are linearly independent, the equality (5.2) only holds when all the coefficients β_i and γ are all zero. This in turns implies that the vectors $c_1, \ldots, c_{k-1}$ are linearly independent. By the induction hypothesis $k-1 \leq n-1$, so we have established that $k \leq n$.

5.3 Orthonormal vectors

A collection of vectors $a_1, \ldots, a_k$ is *orthogonal* or *mutually orthogonal* if $a_i \perp a_j$ for any i, j with $i \neq j$, $i, j = 1, \ldots, k$. A collection of vectors $a_1, \ldots, a_k$ is *orthonormal* if it is orthogonal and $\|a_i\| = 1$ for $i = 1, \ldots, k$. (A vector of norm one is called *normalized*; dividing a vector by its norm is called *normalizing* it.) Thus, each vector in an orthonormal collection of vectors is normalized, and two different vectors from the collection are orthogonal. These two conditions can be combined into one statement about the inner products of pairs of vectors in the collection: $a_1, \ldots, a_k$ is orthonormal means that

$$a_i^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Orthonormality, like linear dependence and independence, is an attribute of a collection of vectors, and not an attribute of vectors individually. By convention, though, we say "The vectors $a_1, \ldots, a_k$ are orthonormal" to mean "The collection of vectors $a_1, \ldots, a_k$ is orthonormal".

Examples. The standard unit n-vectors $e_1, \ldots, e_n$ are orthonormal. As another example, the 3-vectors

$$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \qquad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \tag{5.3}$$

are orthonormal. Figure 5.2 shows a set of two orthonormal 2-vectors.

Linear independence of orthonormal vectors. Orthonormal vectors are linearly independent. To see this, suppose $a_1, \ldots, a_k$ are orthonormal, and

$$\beta_1 a_1 + \cdots + \beta_k a_k = 0.$$

Taking the inner product of this equality with a_i yields

$$\begin{aligned} 0 &= a_i^T(\beta_1 a_1 + \cdots + \beta_k a_k) \\ &= \beta_1(a_i^T a_1) + \cdots + \beta_k(a_i^T a_k) \\ &= \beta_i, \end{aligned}$$

since $a_i^T a_j = 0$ for $j \neq i$ and $a_i^T a_i = 1$. Thus, the only linear combination of $a_1, \ldots, a_k$ that is zero is the one with all coefficients zero.

Linear combinations of orthonormal vectors. Suppose a vector x is a linear combination of $a_1, \ldots, a_k$, where $a_1, \ldots, a_k$ are orthonormal,

$$x = \beta_1 a_1 + \cdots + \beta_k a_k.$$

Taking the inner product of the left-hand and right-hand sides of this equation with a_i yields

$$a_i^T x = a_i^T(\beta_1 a_1 + \cdots + \beta_k a_k) = \beta_i,$$

using the same argument as above. So if a vector x is a linear combination of orthonormal vectors, we can easily find the coefficients of the linear combination by taking the inner products with the vectors.

For any x that is a linear combination of orthonormal vectors $a_1, \ldots, a_k$, we have the identity

$$x = (a_1^T x)a_1 + \cdots + (a_k^T x)a_k. \tag{5.4}$$

This identity gives us a simple way to check if an n-vector y is a linear combination of the orthonormal vectors $a_1, \ldots, a_k$. If the identity (5.4) holds for y, *i.e.*,

$$y = (a_1^T y)a_1 + \cdots + (a_k^T y)a_k,$$

then (evidently) y is a linear combination of $a_1, \ldots, a_k$; conversely, if y is a linear combination of $a_1, \ldots, a_k$, the identity (5.4) holds for y.

Orthonormal basis. If the n-vectors $a_1, \ldots, a_n$ are orthonormal, they are linearly independent, and therefore also a basis. In this case they are called an *orthonormal basis*. The three examples above (on page 95) are orthonormal bases.

If $a_1, \ldots, a_n$ is an orthonormal basis, then we have, for any n-vector x, the identity

$$x = (a_1^T x)a_1 + \cdots + (a_n^T x)a_n. \tag{5.5}$$

To see this, we note that since $a_1, \ldots, a_n$ are a basis, x can be expressed as a linear combination of them; hence the identity (5.4) above holds. The equation above is sometimes called the *orthonormal expansion formula*; the right-hand side is called the *expansion of x in the basis* $a_1, \ldots, a_n$. It shows that any n-vector can be expressed as a linear combination of the basis elements, with the coefficients given by taking the inner product of x with the elements of the basis.

As an example, we express the 3-vector $x = (1, 2, 3)$ as a linear combination of the orthonormal basis given in (5.3). The inner products of x with these vectors are

$$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}^T x = -3, \qquad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}^T x = \frac{3}{\sqrt{2}}, \qquad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}^T x = \frac{-1}{\sqrt{2}}.$$

It can be verified that the expansion of x in this basis is

$$x = (-3) \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} + \frac{3}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right) + \frac{-1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right).$$

5.4 Gram–Schmidt algorithm

In this section we describe an algorithm that can be used to determine if a list of n-vectors $a_1, \ldots, a_k$ is linearly independent. In later chapters we will see that it has many other uses as well. The algorithm is named after the mathematicians Jørgen Pedersen Gram and Erhard Schmidt, although it was already known before their work.

If the vectors are linearly independent, the Gram–Schmidt algorithm produces an orthonormal collection of vectors $q_1, \ldots, q_k$ with the following properties: For each $i = 1, \ldots, k$, a_i is a linear combination of $q_1, \ldots, q_i$, and q_i is a linear combination of $a_1, \ldots, a_i$. If the vectors $a_1, \ldots, a_{j-1}$ are linearly independent, but $a_1, \ldots, a_j$ are linearly dependent, the algorithm detects this and terminates. In other words, the Gram–Schmidt algorithm finds the first vector a_j that is a linear combination of previous vectors $a_1, \ldots, a_{j-1}$.

Algorithm 5.1 GRAM–SCHMIDT ALGORITHM

given n-vectors $a_1, \ldots, a_k$

for $i = 1, \ldots, k$,

1. *Orthogonalization.* $\tilde{q}_i = a_i - (q_1^T a_i) q_1 - \cdots - (q_{i-1}^T a_i) q_{i-1}$
2. *Test for linear dependence.* if $\tilde{q}_i = 0$, quit.
3. *Normalization.* $q_i = \tilde{q}_i / \|\tilde{q}_i\|$

The orthogonalization step, with $i = 1$, reduces to $\tilde{q}_1 = a_1$. If the algorithm does not quit (in step 2), *i.e.*, $\tilde{q}_1, \ldots, \tilde{q}_k$ are all nonzero, we can conclude that the original collection of vectors is linearly independent; if the algorithm does quit early, say, with $\tilde{q}_j = 0$, we can conclude that the original collection of vectors is linearly dependent (and indeed, that a_j is a linear combination of $a_1, \ldots, a_{j-1}$).

Figure 5.3 illustrates the Gram–Schmidt algorithm for two 2-vectors. The top row shows the original vectors; the middle and bottom rows show the first and second iterations of the loop in the Gram–Schmidt algorithm, with the left-hand side showing the orthogonalization step, and the right-hand side showing the normalization step.