Hence $|(\alpha|\beta)|^2 \leq ||\alpha||^2||\beta||^2$. Now using (c) we find that

$$\begin{aligned}||\alpha + \beta||^2 &= ||\alpha||^2 + (\alpha|\beta) + (\beta|\alpha) + ||\beta||^2 \\ &= ||\alpha||^2 + 2 \operatorname{Re} (\alpha|\beta) + ||\beta||^2 \\ &\leq ||\alpha||^2 + 2\,||\alpha||\,||\beta|| + ||\beta||^2 \\ &= (||\alpha|| + ||\beta||)^2.\end{aligned}$$

Thus, $||\alpha + \beta|| \leq ||\alpha|| + ||\beta||$. ∎

The inequality in (iii) is called the **Cauchy-Schwarz inequality.** It has a wide variety of applications. The proof shows that if (for example) α is non-zero, then $|(\alpha|\beta)| < ||\alpha||\,||\beta||$ unless

$$\beta = \frac{(\beta|\alpha)}{||\alpha||^2}\,\alpha.$$

Thus, equality occurs in (iii) if and only if α and β are linearly dependent.

Example 7. If we apply the Cauchy-Schwarz inequality to the inner products given in Examples 1, 2, 3, and 5, we obtain the following:

(a) $$|\Sigma\, x_k\bar{y}_k| \leq (\Sigma\, |x_k|^2)^{1/2}(\Sigma\, |y_k|^2)^{1/2}$$

(b) $$|x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2| \leq ((x_1 - x_2)^2 + 3x_2^2)^{1/2}((y_1 - y_2)^2 + 3y_2^2)^{1/2}$$

(c) $$|\operatorname{tr}(AB^*)| \leq (\operatorname{tr}(AA^*))^{1/2}(\operatorname{tr}(BB^*))^{1/2}$$

(d) $$\left|\int_0^1 f(x)\overline{g(x)}\,dx\right| \leq \left(\int_0^1 |f(x)|^2\,dx\right)^{1/2}\left(\int_0^1 |g(x)|^2\,dx\right)^{1/2}.$$

Definitions. *Let α and β be vectors in an inner product space V. Then α is* **orthogonal** *to β if $(\alpha|\beta) = 0$; since this implies β is orthogonal to α, we often simply say that α and β are orthogonal. If S is a set of vectors in V, S is called an* **orthogonal set** *provided all pairs of distinct vectors in S are orthogonal. An* **orthonormal set** *is an orthogonal set S with the additional property that $||\alpha|| = 1$ for every α in S.*

The zero vector is orthogonal to every vector in V and is the only vector with this property. It is appropriate to think of an orthonormal set as a set of mutually perpendicular vectors, each having length 1.

Example 8. The standard basis of either R^n or C^n is an orthonormal set with respect to the standard inner product.

Example 9. The vector (x, y) in R^2 is orthogonal to $(-y, x)$ with respect to the standard inner product, for

$$((x, y)|(-y, x)) = -xy + yx = 0.$$

However, if R^2 is equipped with the inner product of Example 2, then (x, y) and $(-y, x)$ are orthogonal if and only if

$$y = \tfrac{1}{2}(-3 \pm \sqrt{13})x.$$

EXAMPLE 10. Let V be $C^{n\times n}$, the space of complex $n \times n$ matrices, and let E^{pq} be the matrix whose only non-zero entry is a 1 in row p and column q. Then the set of all such matrices E^{pq} is orthonormal with respect to the inner product given in Example 3. For

$$(E^{pq}|E^{rs}) = \operatorname{tr}(E^{pq}E^{sr}) = \delta_{qs}\operatorname{tr}(E^{pr}) = \delta_{qs}\delta_{pr}.$$

EXAMPLE 11. Let V be the space of continuous complex-valued (or real-valued) functions on the interval $0 \leq x \leq 1$ with the inner product

$$(f|g) = \int_0^1 f(x)\overline{g(x)}\,dx.$$

Suppose $f_n(x) = \sqrt{2}\ \cos 2\pi nx$ and that $g_n(x) = \sqrt{2}\ \sin 2\pi nx$. Then $\{1, f_1, g_1, f_2, g_2, \ldots\}$ is an infinite orthonormal set. In the complex case, we may also form the linear combinations

$$\frac{1}{\sqrt{2}}(f_n + ig_n), \qquad n = 1, 2, \ldots.$$

In this way we get a new orthonormal set S which consists of all functions of the form

$$h_n(x) = e^{2\pi inx}, \qquad n = \pm 1, \pm 2, \ldots.$$

The set S' obtained from S by adjoining the constant function 1 is also orthonormal. We assume here that the reader is familiar with the calculation of the integrals in question.

The orthonormal sets given in the examples above are all linearly independent. We show now that this is necessarily the case.

Theorem 2. *An orthogonal set of non-zero vectors is linearly independent.*

Proof. Let S be a finite or infinite orthogonal set of non-zero vectors in a given inner product space. Suppose $\alpha_1, \alpha_2, \ldots, \alpha_m$ are distinct vectors in S and that

$$\beta = c_1\alpha_1 + c_2\alpha_2 + \cdots + c_m\alpha_m.$$

Then

$$\begin{aligned}(\beta|\alpha_k) &= (\sum_j c_j\alpha_j|\alpha_k)\\ &= \sum_j c_j(\alpha_j|\alpha_k)\\ &= c_k(\alpha_k|\alpha_k).\end{aligned}$$

Since $(\alpha_k|\alpha_k) \neq 0$, it follows that

$$c_k = \frac{(\beta|\alpha_k)}{||\alpha_k||^2}, \qquad 1 \leq k \leq m.$$

Thus when $\beta = 0$, each $c_k = 0$; so S is an independent set. ∎

Corollary. *If a vector β is a linear combination of an orthogonal sequence of non-zero vectors $\alpha_1, \ldots, \alpha_m$, then β is the particular linear combination*

$$\beta = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{||\alpha_k||^2} \alpha_k. \tag{8-8}$$

This corollary follows from the proof of the theorem. There is another corollary which although obvious, should be mentioned. If $\{\alpha_1, \ldots, \alpha_m\}$ is an orthogonal set of non-zero vectors in a finite-dimensional inner product space V, then $m \leq \dim V$. This says that the number of mutually orthogonal directions in V cannot exceed the algebraically defined dimension of V. The maximum number of mutually orthogonal directions in V is what one would intuitively regard as the geometric dimension of V, and we have just seen that this is not greater than the algebraic dimension. The fact that these two dimensions are equal is a particular corollary of the next result.

Theorem 3. *Let V be an inner product space and let $\beta_1, \ldots, \beta_n$ be any independent vectors in V. Then one may construct orthogonal vectors $\alpha_1, \ldots, \alpha_n$ in V such that for each $k = 1, 2, \ldots, n$ the set*

$$\{\alpha_1, \ldots, \alpha_k\}$$

is a basis for the subspace spanned by $\beta_1, \ldots, \beta_k$.

Proof. The vectors $\alpha_1, \ldots, \alpha_n$ will be obtained by means of a construction known as the **Gram-Schmidt orthogonalization process.** First let $\alpha_1 = \beta_1$. The other vectors are then given inductively as follows: Suppose $\alpha_1, \ldots, \alpha_m$ $(1 \leq m < n)$ have been chosen so that for every k

$$\{\alpha_1, \ldots, \alpha_k\}, \qquad 1 \leq k \leq m$$

is an orthogonal basis for the subspace of V that is spanned by $\beta_1, \ldots, \beta_k$. To construct the next vector α_{m+1}, let

$$\alpha_{m+1} = \beta_{m+1} - \sum_{k=1}^{m} \frac{(\beta_{m+1}|\alpha_k)}{||\alpha_k||^2} \alpha_k. \tag{8-9}$$

Then $\alpha_{m+1} \neq 0$. For otherwise β_{m+1} is a linear combination of $\alpha_1, \ldots, \alpha_m$ and hence a linear combination of $\beta_1, \ldots, \beta_m$. Furthermore, if $1 \leq j \leq m$, then

$$\begin{aligned}(\alpha_{m+1}|\alpha_j) &= (\beta_{m+1}|\alpha_j) - \sum_{k=1}^{m} \frac{(\beta_{m+1}|\alpha_k)}{||\alpha_k||^2} (\alpha_k|\alpha_j) \\ &= (\beta_{m+1}|\alpha_j) - (\beta_{m+1}|\alpha_j) \\ &= 0.\end{aligned}$$

Therefore $\{\alpha_1, \ldots, \alpha_{m+1}\}$ is an orthogonal set consisting of $m + 1$ non-zero vectors in the subspace spanned by $\beta_1, \ldots, \beta_{m+1}$. By Theorem 2, it is a basis for this subspace. Thus the vectors $\alpha_1, \ldots, \alpha_n$ may be constructed one after the other in accordance with (8-9). In particular, when $n = 4$, we have

$$
\begin{aligned}
\alpha_1 &= \beta_1 \\
\alpha_2 &= \beta_2 - \frac{(\beta_2|\alpha_1)}{||\alpha_1||^2}\,\alpha_1 \\
\alpha_3 &= \beta_3 - \frac{(\beta_3|\alpha_1)}{||\alpha_1||^2}\,\alpha_1 - \frac{(\beta_3|\alpha_2)}{||\alpha_2||^2}\,\alpha_2 \\
\alpha_4 &= \beta_4 - \frac{(\beta_4|\alpha_1)}{||\alpha_1||^2}\,\alpha_1 - \frac{(\beta_4|\alpha_2)}{||\alpha_2||^2}\,\alpha_2 - \frac{(\beta_4|\alpha_3)}{||\alpha_3||^2}\,\alpha_3.
\end{aligned}
\tag{8-10}
$$ ∎

Corollary. *Every finite-dimensional inner product space has an orthonormal basis.*

Proof. Let V be a finite-dimensional inner product space and $\{\beta_1, \ldots, \beta_n\}$ a basis for V. Apply the Gram-Schmidt process to construct an orthogonal basis $\{\alpha_1, \ldots, \alpha_n\}$. Then to obtain an orthonormal basis, simply replace each vector α_k by $\alpha_k/||\alpha_k||$. ∎

One of the main advantages which orthonormal bases have over arbitrary bases is that computations involving coordinates are simpler. To indicate in general terms why this is true, suppose that V is a finite-dimensional inner product space. Then, as in the last section, we may use Equation (8-5) to associate a matrix G with every ordered basis $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ of V. Using this matrix

$$G_{jk} = (\alpha_k|\alpha_j),$$

we may compute inner products in terms of coordinates. If $\mathcal{B}$ is an orthonormal basis, then G is the identity matrix, and for any scalars x_j and y_k

$$(\sum_j x_j\alpha_j|\sum_k y_k\alpha_k) = \sum_j x_j\bar{y}_j.$$

Thus in terms of an orthonormal basis, the inner product in V looks like the standard inner product in F^n.

Although it is of limited practical use for computations, it is interesting to note that the Gram-Schmidt process may also be used to test for linear dependence. For suppose $\beta_1, \ldots, \beta_n$ are linearly dependent vectors in an inner product space V. To exclude a trivial case, assume that $\beta_1 \neq 0$. Let m be the largest integer for which $\beta_1, \ldots, \beta_m$ are independent. Then $1 \leq m < n$. Let $\alpha_1, \ldots, \alpha_m$ be the vectors obtained by applying the orthogonalization process to $\beta_1, \ldots, \beta_m$. Then the vector α_{m+1} given by (8-9) is necessarily 0. For α_{m+1} is in the subspace spanned by $\alpha_1, \ldots, \alpha_m$ and orthogonal to each of these vectors; hence it is 0 by (8-8). Conversely, if $\alpha_1, \ldots, \alpha_m$ are different from 0 and $\alpha_{m+1} = 0$, then $\beta_1, \ldots, \beta_{m+1}$ are linearly dependent.

EXAMPLE 12. Consider the vectors

$$\begin{aligned} \beta_1 &= (3, 0, 4) \\ \beta_2 &= (-1, 0, 7) \\ \beta_3 &= (2, 9, 11) \end{aligned}$$

in R^3 equipped with the standard inner product. Applying the Gram-Schmidt process to $\beta_1, \beta_2, \beta_3$, we obtain the following vectors.

$$\begin{aligned} \alpha_1 &= (3, 0, 4) \\ \alpha_2 &= (-1, 0, 7) - \frac{((-1, 0, 7)|(3, 0, 4))}{25}(3, 0, 4) \\ &= (-1, 0, 7) - (3, 0, 4) \\ &= (-4, 0, 3) \\ \alpha_3 &= (2, 9, 11) - \frac{((2, 9, 11)|(3, 0, 4))}{25}(3, 0, 4) \\ &\quad - \frac{((2, 9, 11)|(-4, 0, 3))}{25}(-4, 0, 3) \\ &= (2, 9, 11) - 2(3, 0, 4) - -4, 0, 3) \\ &= (0, 9, 0). \end{aligned}$$

These vectors are evidently non-zero and mutually orthogonal. Hence $\{\alpha_1, \alpha_2, \alpha_3\}$ is an orthogonal basis for R^3. To express an arbitrary vector (x_1, x_2, x_3) in R^3 as a linear combination of $\alpha_1, \alpha_2, \alpha_3$ it is *not* necessary to solve any linear equations. For it suffices to use (8-8). Thus

$$(x_1, x_2, x_3) = \frac{3x_1 + 4x_3}{25}\alpha_1 + \frac{-4x_1 + 3x_3}{25}\alpha_2 + \frac{x_2}{9}\alpha_3$$

as is readily verified. In particular,

$$(1, 2, 3) = \tfrac{3}{5}(3, 0, 4) + \tfrac{1}{5}(-4, 0, 3) + \tfrac{2}{9}(0, 9, 0).$$

To put this point in another way, what we have shown is the following: The basis $\{f_1, f_2, f_3\}$ of $(R^3)^*$ which is dual to the basis $\{\alpha_1, \alpha_2, \alpha_3\}$ is defined explicitly by the equations

$$\begin{aligned} f_1(x_1, x_2, x_3) &= \frac{3x_1 + 4x_3}{25} \\ f_2(x_1, x_2, x_3) &= \frac{-4x_1 + 3x_3}{25} \\ f_3(x_1, x_2, x_3) &= \frac{x_2}{9} \end{aligned}$$

and these equations may be written more generally in the form

$$f_j(x_1, x_2, x_3) = \frac{((x_1, x_2, x_3)|\alpha_j)}{||\alpha_j||^2}.$$

Finally, note that from α_1, α_2, α_3 we get the orthonormal basis

$$\tfrac{1}{5}\,(3, 0, 4), \quad \tfrac{1}{5}\,(-4, 0, 3), \quad (0, 1, 0).$$

EXAMPLE 13. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ where a, b, c, and d are complex numbers. Set $\beta_1 = (a, b)$, $\beta_2 = (c, d)$, and suppose that $\beta_1 \neq 0$. If we apply the orthogonalization process to β_1, β_2, using the standard inner product in C^2, we obtain the following vectors:

$$\begin{aligned} \alpha_1 &= (a, b) \\ \alpha_2 &= (c, d) - \frac{((c, d)|(a, b))}{|a|^2 + |b|^2}\,(a, b) \\ &= (c, d) - \frac{(c\bar{a} + d\bar{b})}{|a|^2 + |b|^2}\,(a, b) \\ &= \left(\frac{cb\bar{b} - d\bar{b}a}{|a|^2 + |b|^2}, \frac{d\bar{a}a - c\bar{a}b}{|a|^2 + |b|^2}\right) \\ &= \frac{\det A}{|a|^2 + |b|^2}\,(-\bar{b}, \bar{a}). \end{aligned}$$

Now the general theory tells us that $\alpha_2 \neq 0$ if and only if β_1, β_2 are linearly independent. On the other hand, the formula for α_2 shows that this is the case if and only if $\det A \neq 0$.

In essence, the Gram-Schmidt process consists of repeated applications of a basic geometric operation called orthogonal projection, and it is best understood from this point of view. The method of orthogonal projection also arises naturally in the solution of an important approximation problem.

Suppose W is a subspace of an inner product space V, and let β be an arbitrary vector in V. The problem is to find a best possible approximation to β by vectors in W. This means we want to find a vector α for which $||\beta - \alpha||$ is as small as possible subject to the restriction that α should belong to W. Let us make our language precise.

A **best approximation** to β by vectors in W is a vector α in W such that

$$||\beta - \alpha|| \leq ||\beta - \gamma||$$

for every vector γ in W.

By looking at this problem in R^2 or in R^3, one sees intuitively that a best approximation to β by vectors in W ought to be a vector α in W such that $\beta - \alpha$ is perpendicular (orthogonal) to W and that there ought to