CHAPTER VI

Scalar Products and Orthogonality

VI, §1. Scalar Products

Let V be a vector space. A **scalar product** on V is an association which to any pair of elements (v, w) of V associates a number, denoted by $\langle v, w\rangle$, satisfying the following properties:

SP 1. *We have* $\langle v, w\rangle = \langle w, v\rangle$ *for all* v, w *in* V.

SP 2. *If* u, v, w *are elements of* V, *then*

$$\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle.$$

SP 3. *If* x *is a number, then*

$$\langle xu, v\rangle = x\langle u, v\rangle = \langle u, xv\rangle.$$

We shall also assume that the scalar product satisfies the condition:

SP 4. *For all* v *in* V *we have* $\langle v, v\rangle \geqq 0$, *and* $\langle v, v\rangle > 0$ *if* $v \neq O$.

A scalar product satisfying this condition is called **positive definite**.

For the rest of this section we assume that V *is a vector space with a positive definite scalar product.*

Example 1. Let $V = \mathbf{R}^n$, and define

$$\langle X, Y \rangle = X \cdot Y$$

for elements X, Y of $\mathbf{R}^n$. Then this is a positive definite scalar product.

Example 2. Let V be the space of continuous real-valued functions on the interval $[-\pi, \pi]$. If f, g are in V, we define

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

Simple properties of the integral show that this is a scalar product, which is in fact positive definite.

In calculus, we study the second example, which gives rise to the theory of Fourier series. Here we discuss only general properties of scalar products and applications to euclidean spaces. The notation $\langle\ ,\ \rangle$ is used because in dealing with vector spaces of functions, a dot $f \cdot g$ may be confused with the ordinary product of functions.

As in the case of the dot product, we define elements v, w of V to be **orthogonal**, or **perpendicular**, and write $v \perp w$, if $\langle v, w \rangle = 0$. If S is a subset of V, we denote by $S^{\perp}$ the set of all elements w in V which are perpendicular to all elements of S, i.e. such that $\langle w, v \rangle = 0$ for all v in S. Then using **SP 1**, **SP 2**, and **SP 3**, one verifies at once that $S^{\perp}$ is a subspace of V, called the **orthogonal space** of S. If w is perpendicular to S, we also write $w \perp S$. Let U be the subspace of V generated by the elements of S. If w is perpendicular to S, and if v_1, v_2 are in S, then

$$\langle w, v_1 + v_2 \rangle = \langle w, v_1 \rangle + \langle w, v_2 \rangle = 0.$$

If c is a number, then

$$\langle w, cv_1 \rangle = c\langle w, v_1 \rangle = 0.$$

Hence w is perpendicular to linear combinations of elements of S, and hence w is perpendicular to U.

Example 3. Let (a_{ij}) be an $m \times n$ matrix, and let $A_1, \dots, A_m$ be its row vectors. Let $X = (x_1, \dots, x_n)$ as usual. The system of homogeneous linear equations

$$(**) \qquad \begin{array}{ccccc} a_{11}x_1 & + \cdots + & a_{1n}x_n & = & 0 \\ \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + \cdots + & a_{mn}x_n & = & 0 \end{array}$$

can also be written in abbreviated form using the dot product, as

$$A_1 \cdot X = 0, \quad \dots, \quad A_m \cdot X = 0.$$

The set of solutions X of this homogeneous system is therefore the set of all vectors perpendicular to $A_1, \dots, A_m$. It is therefore the subspace of $\mathbf{R}^n$ which is the orthogonal subspace to the space generated by $A_1, \dots, A_m$. If U is the space of solutions, and if W denotes the space generated by $A_1, \dots, A_m$, we have

$$U = W^{\perp}.$$

We call dim U the **dimension of the space of solutions of the system of linear equations.**

As in Chapter I, we define the **length**, or **norm** of an element $v \in V$ by

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

If c is any number, then we immediately get

$$\|cv\| = |c| \ \|v\|,$$

because

$$\|cv\| = \sqrt{\langle cv, cv \rangle} = \sqrt{c^2 \langle v, v \rangle} = |c| \ \|v\|.$$

Thus we see the same type of arguments as in Chapter I apply here. In fact, any argument given in Chapter I which does not use coordinates applies to our more general situation. We shall see further examples as we go along.

As before, we say that an element $v \in V$ is a **unit vector** if $\|v\| = 1$. If $v \in V$ and $v \neq O$, then $v/\|v\|$ is a unit vector.

The following two identities follow directly from the definition of the length.

The Pythagoras theorem. *If v, w are perpendicular, then*

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

The parallelogram law. *For any v, w we have*

$$\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2.$$

The proofs are trivial. We give the first, and leave the second as an exercise. For the first, we have

$$\begin{aligned}\|v + w\|^2 = \langle v + w, v + w\rangle &= \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle \\ &= \|v\|^2 + \|w\|^2.\end{aligned}$$

Let w be an element of V such that $\|w\| \neq 0$. For any v there exists a unique number c such that $v - cw$ is perpendicular to w. Indeed, for $v - cw$ to be perpendicular to w we must have

$$\langle v - cw,\ w\rangle = 0,$$

whence $\langle v, w\rangle - \langle cw, w\rangle = 0$ and $\langle v, w\rangle = c\langle w, w\rangle$. Thus

$$c = \frac{\langle v, w\rangle}{\langle w, w\rangle}.$$

Conversely, letting c have this value shows that $v - cw$ is perpendicular to w. We call c the **component of v along w**.

In particular, if w is a unit vector, then the component of v along w is simply

$$c = \langle v,\ w\rangle.$$

Example 4. Let $V = \mathbf{R}^n$ with the usual scalar product, i.e. the dot product. If E_i is the i-th unit vector, and $X = (x_1, \ldots, x_n)$ then the component of X along E_i is simply

$$X \cdot E_i = x_i,$$

that is, the i-th component of X.

Example 5. Let V be the space of continuous functions on $[-\pi, \pi]$. Let f be the function given by $f(x) = \sin kx$, where k is some integer > 0. Then

$$\begin{aligned}\|f\| = \sqrt{\langle f, f\rangle} &= \left(\int_{-\pi}^{\pi} \sin^2 kx\, dx\right)^{1/2} \\ &= \sqrt{\pi}.\end{aligned}$$

If g is any continuous function on $[-\pi, \pi]$, then the component of g along f is also called the **Fourier coefficient of g with respect to f**, and is equal to

$$\frac{\langle g, f\rangle}{\langle f, f\rangle} = \frac{1}{\pi}\int_{-\pi}^{\pi} g(x) \sin kx\ dx.$$

As with the case of n-space, we define the **projection of v along w** to be the vector cw, because of our usual picture:

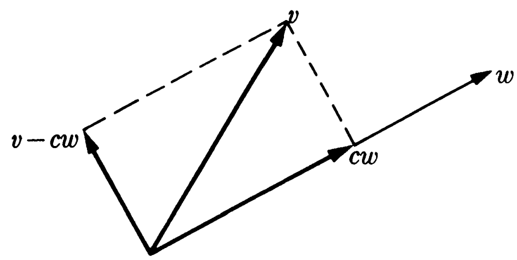

Figure 1

Exactly the same arguments which we gave in Chapter I can now be used to get the **Schwarz inequality**, namely:

Theorem 1.1 *For all v, $w \in V$ we have*

$$|\langle v, w\rangle| \leqq \|v\| \; \|w\|.$$

Proof. If $w = O$, then both sides are equal to 0 and our inequality is obvious. Next, assume that $w \neq O$. Let c be the component of v along w. We write

$$v = v - cw + cw.$$

Then $v - cw$ is perpendicular to cw, so by Pythagoras,

$$\begin{aligned}\|v\|^2 &= \|v - cw\|^2 + \|cw\|^2 \\ &= \|v - cw\|^2 + |c|^2 \; \|w\|^2.\end{aligned}$$

Therefore $|c|^2 \; \|w\|^2 \leqq \|v\|^2$ and taking square roots yields

$$|c| \; \|w\| \leqq \|v\|.$$

But $c = \langle v, w\rangle/\|w\|^2$. Then one factor $\|w\|$ cancels, and cross multiplying by $\|w\|$ yields

$$|\langle v, w\rangle| \leqq \|v\| \; \|w\|,$$

thereby proving the theorem.

Theorem 1.2 *If v, $w \in V$, then*

$$\|v + w\| \leqq \|v\| + \|w\|.$$

Proof. Exactly the same as that of the analogous theorem in Chapter I, §4.

Let $v_1, \ldots, v_n$ be non-zero elements of V which are mutually perpendicular, that is $\langle v_i, v_j \rangle = 0$ if $i \neq j$. Let c_i be the component of v along v_i. Then

$$v - c_1 v_1 - \cdots - c_n v_n$$

is perpendicular to $v_1, \ldots, v_n$. To see this, all we have to do is to take the product with v_j for any j. All the terms involving $\langle v_i, v_j \rangle$ will give 0 if $i \neq j$, and we shall have two remaining terms

$$\langle v, v_j \rangle - c_j \langle v_j, v_j \rangle$$

which cancel. Thus subtracting linear combinations as above orthogonalizes v with respect to $v_1, \ldots, v_n$. The next theorem shows that $c_1 v_1 + \cdots + c_n v_n$ gives the closest approximation to v as a linear combination of $v_1, \ldots, v_n$.

Theorem 1.3 *Let $v_1, \ldots, v_n$ be vectors which are mutually perpendicular, and such that $\|v_i\| \neq 0$ for all i. Let v be an element of V, and let c_i be the component of v along v_i. Let $a_1, \ldots, a_n$ be numbers. Then*

$$\left\| v - \sum_{k=1}^{n} c_k v_k \right\| \leqq \left\| v - \sum_{k=1}^{n} a_k v_k \right\|.$$

Proof. We know that

$$v - \sum_{k=1}^{n} c_k v_k$$

is perpendicular to each v_i, $i = 1, \ldots, n$. Hence it is perpendicular to any linear combination of $v_1, \ldots, v_n$. Now we have:

$$\begin{aligned} \|v - \sum a_k v_k\|^2 &= \|v - \sum c_k v_k + \sum (c_k - a_k) v_k\|^2 \\ &= \|v - \sum c_k v_k\|^2 + \|\sum (c_k - a_k) v_k\|^2 \end{aligned}$$

by the Pythagoras theorem. This proves that

$$\|v - \sum c_k v_k\|^2 \leqq \|v - \sum a_k v_k\|^2,$$

and thus our theorem is proved.

Example 6. Consider the vector space V of all continuous functions on the interval $[0, 2\pi]$. Let

$$g_k(x) = \cos kx, \qquad \text{for} \quad k = 0, 1, 2, \ldots.$$

We use the scalar product

$$\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\, dx.$$

Then it is easily verified that

$$\|g_0\| = \sqrt{2\pi} \qquad \text{and} \qquad \|g_k\| = \sqrt{\pi} \qquad \text{for} \quad k > 0.$$

The Fourier coefficient of f with respect to g_k is

$$c_k = \frac{1}{\pi}\int_0^{2\pi} f(x) \cos kx,\, dx, \qquad \text{for} \quad k > 0.$$

If we take $v_k = g_k$ for $k = 1,\dots,n$ then Theorem 1.3 tells us that the linear combination

$$c_0 + c_1 \cos x + c_2 \cos 2x + \cdots + c_n \cos nx$$

gives the best approximation to the function f among all possible linear combinations

$$a_0 + a_1 \cos x + \cdots + a_n \cos nx$$

with arbitrary real numbers $a_0, a_1,\dots,a_n$. Such a sum is called a **partial sum** of the Fourier series.

Similarly, we could take linear combinations of the functions $\sin kx$. This leads into the theory of Fourier series. We do not go into this deeper here. We merely wanted to point out the analogy, and the usefulness of the geometric language and formalism in dealing with these objects.

The next theorem is known as the **Bessel inequality**.

Theorem 1.4. *If $v_1,\dots,v_n$ are mutually perpendicular unit vectors, and if c_i is the Fourier coefficient of v with respect to v_i, then*

$$\sum_{i=1}^{n} c_i^2 \leqq \|v\|^2.$$

Proof. We have

$$\begin{aligned} 0 &\leqq \langle v - \sum c_i v_i, v - \sum c_i v_i\rangle \\ &= \langle v, v\rangle - \sum 2c_i\langle v, v_i\rangle + \sum c_i^2 \\ &= \langle v, v\rangle - \sum c_i^2. \end{aligned}$$

From this our inequality follows.

Exercises VI, §1

1. Let V be a vector space with a positive definite scalar product. Let $v_1,\dots,v_r$ be non-zero elements of V which are mutually perpendicular, meaning that $\langle v_i, v_j\rangle = 0$ if $i \neq j$. Show that $v_1,\dots,v_r$ are linearly independent.

The following exercise gives an important example of a scalar product.

2. Let A be a symmetric $n \times n$ matrix. Given two column vectors X, $Y \in \mathbf{R}^n$, define
$$\langle X, Y\rangle = {}^tXAY.$$
   (a) Show that this symbol satisfies the first three properties of a scalar product.
   (b) Give an example of a 1×1 matrix and a non-zero 2×2 matrix such that the fourth property is not satisfied. If this fourth property is satisfied, that is ${}^tXAX > 0$ for all $X \neq O$, then the matrix A is called **positive definite**.
   (c) Give an example of a 2×2 matrix which is symmetric and positive definite.
   (d) Let $a > 0$, and let
$$A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}.$$
   Prove that A is positive definite if and only if $ad - b^2 > 0$. [*Hint*: Let $X = {}^t(x, y)$ and complete the square in the expression ${}^tXAX$.]
   (e) If $a < 0$ show that A is not positive definite.

3. Determine whether the following matrices are positive definite.

(a) $\begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}$ (b) $\begin{pmatrix} -2 & 1 \\ 1 & 5 \end{pmatrix}$

(c) $\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$ (d) $\begin{pmatrix} 4 & 4 \\ 4 & 1 \end{pmatrix}$

(e) $\begin{pmatrix} 4 & 1 \\ 1 & 10 \end{pmatrix}$ (f) $\begin{pmatrix} 4 & -1 \\ -1 & 10 \end{pmatrix}$

The trace of a matrix

4. Let A be an $n \times n$ matrix. Define the **trace** of A to be the sum of the diagonal elements. Thus if $A = (a_{ij})$, then
$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$
For instance, if
$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

then $\operatorname{tr}(A) = 1 + 4 = 5$. If

$$A = \begin{pmatrix} 1 & -1 & 5 \\ 2 & 1 & 3 \\ 1 & -4 & 7 \end{pmatrix},$$

then $\operatorname{tr}(A) = 9$. Compute the trace of the following matrices:

(a) $\begin{pmatrix} 1 & 7 & 3 \\ -1 & 5 & 2 \\ 2 & 3 & -4 \end{pmatrix}$ (b) $\begin{pmatrix} 3 & -2 & 4 \\ 1 & 4 & 1 \\ -7 & -3 & -3 \end{pmatrix}$ (c) $\begin{pmatrix} -2 & 1 & 1 \\ 3 & 4 & 4 \\ -5 & 2 & 6 \end{pmatrix}$.

5. (a) For any square matrix A show that $\operatorname{tr}(A) = \operatorname{tr}({}^tA)$.
   (b) Show that the trace is a linear map.

6. If A is a symmetric square matrix, show that $\operatorname{tr}(AA) \geqq 0$, and $=0$ if and only if $A = O$.

7. Let A, B be the indicated matrices. Show that

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

(a) $A = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 4 & 1 \\ 3 & 0 & 1 \end{pmatrix}$ $B = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix}$

(b) $A = \begin{pmatrix} 1 & 7 & 3 \\ -1 & 5 & 2 \\ 2 & 3 & -4 \end{pmatrix}$ $B = \begin{pmatrix} 3 & -2 & 4 \\ 1 & 4 & 1 \\ -7 & -3 & 2 \end{pmatrix}$.

8. (a) Prove in general that if A, B are square $n \times n$ matrices, then

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

   (b) If C is an $n \times n$ matrix which has an inverse, then $\operatorname{tr}(C^{-1}AC) = \operatorname{tr}(A)$.

9. Let V be the vector space of symmetric $n \times n$ matrices. For $A, B \in V$ define the symbol

$$\langle A, B\rangle = \operatorname{tr}(AB),$$

where tr is the trace (sum of the diagonal elements). Show that the previous properties in particular imply that this defines a positive definite scalar product on V.

Exercises 10 through 13 deal with the scalar product in the context of calculus.

10. Let V be the space of continuous functions on $[0, 2\pi]$, and let the scalar product be given by the integral over this interval as in the text, that is

$$\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\, dx.$$

Let $g_n(x) = \cos nx$ for $n \geqq 0$ and $h_m(x) = \sin mx$ for $m \geqq 1$.
(a) Show that $\|g_0\| = \sqrt{2\pi}$, $\|g_n\| = \|h_n\| = \sqrt{\pi}$ for $n \geqq 1$.
(b) Show that $g_n \perp g_m$ if $m \neq n$ and that $g_n \perp h_m$ for all m, n. *Hint*: Use formulas like

$$\sin A \cos B = \tfrac{1}{2}[\sin(A + B) + \sin(A - B)]$$

$$\cos A \cos B = \tfrac{1}{2}[\cos(A + B) + \cos(A - B)].$$

11. Let $f(x) = x$ on the interval $[0, 2\pi]$. Find $\langle f, g_n \rangle$ and $\langle f, h_n \rangle$ for the functions g_n, h_n of Exercise 10. Find the Fourier coefficients of f with respect to g_n and h_n.

12. Same question as in Exercise 11 if $f(x) = x^2$. (Exercises 10 through 13 give you a review of some elementary integrals from calculus.)

13. (a) Let $f(x) = x$ on the interval $[0, 2\pi]$. Find $\|f\|$.
(b) Let $f(x) = x^2$ on the same interval. Find $\|f\|$.

VI, §2. Orthogonal Bases

Let V be a vector space with a positive definite scalar product throughout this section. A basis $\{v_1, \ldots, v_n\}$ of V is said to be **orthogonal** if its elements are mutually perpendicular, i.e. if $\langle v_i, v_j \rangle = 0$ whenever $i \neq j$. If in addition each element of the basis has norm 1, then the basis is called **orthonormal**.

Example 1. The standard unit vectors

$$E_1, \ldots, E_n \quad \text{in} \quad \mathbf{R}^n$$

form an orthonormal basis of $\mathbf{R}^n$. Indeed, each has norm 1, and they are mutually orthogonal, that is

$$E_i \cdot E_j = 0 \qquad \text{if} \quad i \neq j.$$

Of course there are many other orthonormal bases of $\mathbf{R}^n$.

This example is typical in the following sense.

Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of V. Any vector $v \in V$ can be written in terms of coordinates

$$v = x_1 e_1 + \cdots + x_n e_n \qquad \text{with} \quad x_i \in \mathbf{R}.$$

Let w be another element of V, and write

$$w = y_1 e_1 + \cdots + y_n e_n \qquad \text{with} \quad y_i \in \mathbf{R}.$$

Then

$$\begin{aligned}\langle v, w\rangle &= \langle x_1e_1 + \cdots + x_ne_n, y_1e_1 + \cdots + y_ne_n\rangle \\ &= \sum_{i,j=1}^{n} \langle x_ie_i, y_je_j\rangle \\ &= \sum_{i=1}^{n} x_iy_i\end{aligned}$$

because $\langle e_i, e_j\rangle = 0$ for $i \neq j$. Hence if X is the coordinate n-tuple of v and Y the coordinate n-tuple of w, then

$$\langle v, w\rangle = X \cdot Y$$

so the scalar product is given precisely as the dot product of the coordinates. This is one of the uses of orthonormal bases: to identify the scalar product with the old-fashioned dot product.

Example 2. Consider $\mathbf{R}^2$. Let

$$A = (1, 1) \qquad \text{and} \qquad B = (1, -1).$$

Then $A \cdot B = 0$, so A is orthogonal to B, and A, B are linearly independent. Therefore they form a basis of $\mathbf{R}^2$, and in fact they form an orthogonal basis of $\mathbf{R}^2$. To get an orthonormal basis from them, we divide each by its norm, so an orthonormal basis is given by

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \qquad \text{and} \qquad \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right).$$

In general, suppose we have a subspace W of $\mathbf{R}^n$, and let $A_1, \ldots, A_r$ be any basis of W. We want to get an orthogonal basis of W. We follow a stepwise orthogonalization process. We start with $A_1 = B_1$. Then we take A_2 and subtract its projection on A_1 to get a vector B_2. Then we take A_3 and subtract its projections on B_1 and B_2 to get a vector B_3. Then we take A_4 and subtract its projections on B_1, B_2, B_3 to get a vector B_4. We continue in this way. This will eventually lead to an orthogonal basis of W.

We state this as a theorem and prove it in the context of vector spaces with a scalar product.

Theorem 2.1. *Let V be a finite dimensional vector space, with a positive definite scalar product. Let W be a subspace of V, and let $\{w_1, \ldots, w_m\}$ be an orthogonal basis of W. If $W \neq V$, then there exist elements $w_{m+1}, \ldots, w_n$ of V such that $\{w_1, \ldots, w_n\}$ is an orthogonal basis of V.*

Proof. The method of proof is as important as the theorem, and is called the **Gram–Schmidt orthogonalization process**. We know from Chapter III, §3 that we can find elements $v_{m+1},\dots,v_n$ of V such that

$$\{w_1,\dots,w_m,\ v_{m+1},\dots,v_n\}$$

is a basis of V. Of course, it is not an orthogonal basis. Let W_{m+1} be the space generated by $w_1,\dots,w_m,\ v_{m+1}$. We shall first obtain an orthogonal basis of W_{m+1}. The idea is to take v_{m+1} and subtract from it its projection along $w_1,\dots,w_m$. Thus we let

$$c_1 = \frac{\langle v_{m+1}, w_1\rangle}{\langle w_1, w_1\rangle},\quad \dots,\quad c_m = \frac{\langle v_{m+1}, w_m\rangle}{\langle w_m, w_m\rangle}.$$

Let

$$w_{m+1} = v_{m+1} - c_1w_1 - \cdots - c_mw_m.$$

Then w_{m+1} is perpendicular to $w_1,\dots,w_m$. Furthermore, $w_{m+1} \neq O$ (otherwise v_{m+1} would be linearly dependent on $w_1,\dots,w_m$), and v_{m+1} lies in the space generated by $w_1,\dots,w_{m+1}$ because

$$v_{m+1} = w_{m+1} + c_1w_1 + \cdots + c_mw_m.$$

Hence $\{w_1,\dots,w_{m+1}\}$ is an orthogonal basis of W_{m+1}. We can now proceed by induction, showing that the space W_{m+s} generated by

$$w_1,\dots,w_m,\ v_{m+1},\dots,v_{m+s}$$

has orthogonal basis

$$\{w_1,\dots,w_{m+1},\dots,w_{m+s}\}$$

with $s = 1,\dots,n-m$. This concludes the proof.

Corollary 2.2. *Let V be a finite dimensional vector space with a positive definite scalar product. Assume that $V \neq \{O\}$. Then V has an orthogonal basis.*

Proof. By hypothesis, there exists an element v_1 of V such that $v_1 \neq O$. We let W be the subspace generated by v_1, and apply the theorem to get the desired basis.

We summarize the procedure of Theorem 2.1 once more. Suppose we are given an arbitrary basis $\{v_1,\dots,v_n\}$ of V. We wish to orthogonalize it. We proceed as follows. We let

$$\begin{aligned}
v_1' &= v_1,\\
v_2' &= v_2 - \frac{\langle v_2, v_1'\rangle}{\langle v_1', v_1'\rangle} v_1',\\
v_3' &= v_3 - \frac{\langle v_3, v_2'\rangle}{\langle v_2', v_2'\rangle} v_2' - \frac{\langle v_3, v_1'\rangle}{\langle v_1', v_1'\rangle} v_1',\\
&\vdots \qquad \vdots\\
v_n' &= v_n - \frac{\langle v_n, v_{n-1}'\rangle}{\langle v_{n-1}', v_{n-1}'\rangle} v_{n-1}' - \cdots - \frac{\langle v_n, v_1'\rangle}{\langle v_1', v_1'\rangle} v_1'.
\end{aligned}$$

Then $\{v_1',\dots,v_n'\}$ is an orthogonal basis.

Given an orthogonal basis, we can always obtain an orthonormal basis by dividing each vector by its norm.

Example 3. Find an orthonormal basis for the vector space generated by the vectors $(1, 1, 0, 1)$, $(1, -2, 0, 0)$, and $(1, 0, -1, 2)$.

Let us denote these vectors by A, B, C. Let

$$B' = B - \frac{B\cdot A}{A\cdot A} A.$$

In other words, we subtract from B its projection along A. Then B' is perpendicular to A. We find

$$B' = \tfrac{1}{3}(4, -5, 0, 1).$$

Now we subtract from C its projection along A and B', and thus we let

$$C' = C - \frac{C\cdot A}{A\cdot A} A - \frac{C\cdot B'}{B'\cdot B'} B'.$$

Since A and B' are perpendicular, taking the scalar product of C' with A and B' shows that C' is perpendicular to both A and B'. We find

$$C' = \tfrac{1}{7}(-4, -2, -7, 6).$$

The vectors A, B', C' are non-zero and mutually perpendicular. They lie in the space generated by A, B, C. Hence they constitute an orthogonal basis for that space. If we wish an orthonormal basis, then we divide these vectors by their norm, and thus obtain

$$\frac{A}{\|A\|} = \frac{1}{\sqrt{3}}(1, 1, 0, 1),$$

$$\frac{B'}{\|B'\|} = \frac{1}{\sqrt{42}}(4, -5, 0, 1),$$

$$\frac{C'}{\|C'\|} = \frac{1}{\sqrt{105}}(-4, -2, -7, 6),$$

as an orthonormal basis.

Example 4. Find an orthogonal basis for the space of solutions of the linear equation

$$3x - 2y + z = 0.$$

First we find a basis, not necessarily orthogonal. For instance, we give z an arbitrary value, say $z = 1$. Thus we have to satisfy

$$3x - 2y = -1.$$

By inspection, we let $x = 1$, $y = 2$ or $x = 3$, $y = 5$, that is

$$A = (1, 2, 1) \qquad \text{and} \qquad B = (3, 5, 1).$$

Then it is easily verified that A, B are linearly independent. By Theorem 4.3 of Chapter 4, the space of solutions has dimension 2, so A, B form a basis of that space of solutions. To get an orthogonal basis, we start with A. Then we let

$$\begin{aligned} C &= B - \text{projection of } B \text{ along } A \\ &= B - \frac{B \cdot A}{A \cdot A} A \\ &= \left(\frac{2}{3}, \frac{1}{3}, \frac{-4}{3}\right). \end{aligned}$$

Then $\{A, C\}$ is an orthogonal basis of the space of solutions. It is sometimes convenient to get rid of the denominator. We may use

$$A = (1, 2, 1) \qquad \text{and} \qquad D = (2, 1, -4)$$

equally well for an orthogonal basis of that space. As a check, substitute back in the original equation to see that these vectors give solutions, and also verify that $A \cdot D = 0$, so that they are perpendicular to each other.

Example 5. Find an orthogonal basis for the space of solutions of the homogeneous equations

$$\begin{aligned} 3x - 2y + z + w &= 0, \\ x + y \qquad + 2w &= 0. \end{aligned}$$

Let W be the space of solutions in $\mathbf{R}^4$. Then W is the space orthogonal to the two vectors

$$(3, -2, 1, 1) \qquad \text{and} \qquad (1, 1, 0, 2).$$

These are obviously linearly independent (by any number of arguments, you can prove at once that the matrix

$$\begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix}$$

has rank 2, for instance). Hence

$$\dim W = 4 - 2 = 2.$$

Next we find a basis for the space of solutions. Let us put $w = 1$, and solve

$$\begin{aligned} 3x - 2y + z &= -1, \\ x + y \qquad &= -2, \end{aligned}$$

by ordinary elimination. If we put $y = 0$, then we get a solution with $x = -2$, and

$$z = -1 - 3x + 2y = 5.$$

If we put $y = 1$, then we get a solution with $x = -3$, and

$$z = -1 - 3x + 2y = 10.$$

Thus we get the two solutions

$$A = (-2, 0, 5, 1) \qquad \text{and} \qquad B = (-3, 1, 10, 1).$$

(As a check, substitute back in the original system of equations to see that no computational error has been made.) These two solutions are linearly independent, because for instance the matrix

$$\begin{pmatrix} -2 & 0 \\ -3 & 1 \end{pmatrix}$$

has rank 2. Hence $\{A, B\}$ is a basis for the space of solutions. To find an orthogonal basis, we orthogonalize B, to get

$$B' = B - \frac{B \cdot A}{A \cdot A} A = B - \frac{19}{10} A.$$

We can also clear denominators, and let $C = 10B'$, so

$$\begin{aligned} C &= (-30, 10, 100, 10) - (-38, 0, 95, 19) \\ &= (8, 10, 5, -9) \end{aligned}$$

Then $\{A, C\}$ is an orthogonal basis for the space of solutions. (Again, check by substituting back in the system of equations, and also check perpendicularity by seeing directly that $A \cdot C = 0$.)

One can also find an orthogonal basis without guessing solutions by inspection or elimination at the start, as follows.

Example 6. Find a basis for the space of solutions of the equation

$$3x - 2y + z = 0.$$

The space of solutions is the space orthogonal to the vector $(3, -2, 1)$ and hence has dimension 2. There are of course many bases for this space. To find one, we first extend $(3, -2, 1) = A$ to a basis of $\mathbf{R}^3$. We do this by selecting vectors B, C such that A, B, C are linearly independent. For instance, take

$$B = (0, 1, 0)$$

and

$$C = (0, 0, 1).$$

Then A, B, C are linearly independent. To see this, we proceed as usual. If a, b, c are numbers such that

$$aA + bB + cC = 0,$$

then

$$\begin{aligned} 3a \quad\quad &= 0, \\ -2a + b &= 0, \\ a + c &= 0. \end{aligned}$$

This is easily solved to see that $a = b = c = 0$, so A, B, C are linearly independent. Now we must orthogonalize these vectors.

Let

$$B' = B - \frac{\langle B, A\rangle}{\langle A, A\rangle} A = \left(\frac{3}{7}, \frac{5}{7}, \frac{1}{7}\right),$$

$$\begin{aligned} C' &= C - \frac{\langle C, A\rangle}{\langle A, A\rangle} A - \frac{\langle C, B'\rangle}{\langle B', B'\rangle} B' \\ &= (0, 0, 1) - \tfrac{1}{14}(3, -2, 1) - \tfrac{1}{35}(3, 5, 1). \end{aligned}$$

Then $\{B', C'\}$ is a basis for the space of solutions of the given equation. As you see, this procedure is slightly longer than the one used by guessing first, and involves two orthogonalizations rather than one as in Example 4.

In Theorem 2.1 we obtained an orthogonal basis for V by starting with an orthogonal basis for a subspace. Let us now look at the situation more symmetrically.

Theorem 2.3. *Let V be a vector space of dimension n, with a positive definite scalar product. Let $\{w_1, \dots, w_r, u_1, \dots, u_s\}$ be an orthogonal basis for V. Let W be the subspace generated by $w_1, \dots, w_r$ and let U be the subspace generated by $u_1, \dots, u_s$. Then $U = W^\perp$, or by symmetry, $W = U^\perp$. Hence for any subspace W of V we have the relation*

$$\dim W + \dim W^\perp = \dim V.$$

Proof. We shall prove that $W^\perp \subset U$ and $U \subset W^\perp$, so $W^\perp = U$.

First let $v \in W^\perp$. There exist numbers a_i $(i = 1, \dots, r)$ and b_j $(j = 1, \dots, s)$ such that

$$v = \sum_{i=1}^{r} a_i w_i + \sum_{j=1}^{s} b_j u_j.$$

Since v is perpendicular to all elements of W, we have for any $k = 1, \dots, r$:

$$\begin{aligned} 0 = v \cdot w_k &= \sum a_i w_i \cdot w_k + \sum b_j u_j \cdot w_k \\ &= a_k w_k \cdot w_k \end{aligned}$$

because $w_i \cdot w_k = 0$ if $i \neq k$ and $u_j \cdot w_k = 0$ for all j. Since $w_k \cdot w_k \neq 0$ it follows that $a_k = 0$ for all $k = 1, \ldots, r$ so v is a linear combination of $u_1, \ldots, u_s$ and $v \in U$. Thus $W^\perp \subset U$.

Conversely, let $v \in U$, so v is a linear combination of $u_1, \ldots, u_s$. Since $\{w_1, \ldots, w_r, u_1, \ldots, u_s\}$ is an orthogonal basis of V it follows that each u_j is perpendicular to W so v itself is perpendicular to W, so $U \subset W^\perp$. Therefore we have proved that $U = W^\perp$.

Finally, Theorem 2.1 shows that the previous situation applies to any subspace W of V, and by the definition of dimension,

$$\dim V = r + s = \dim W + \dim W^\perp,$$

thus concluding the proof of the theorem.

Example 7. Consider $\mathbf{R}^3$. Let A, B be two linearly independent vectors in $\mathbf{R}^3$. Then the space of vectors which are perpendicular to both A and B is a 1-dimensional space. If $\{N\}$ is a basis for this space, any other basis for this space is of type $\{tN\}$, where t is a number $\neq 0$.

Again in $\mathbf{R}^3$, let N be a non-zero vector. The space of vectors perpendicular to N is a 2-dimensional space, i.e. a plane, passing through the origin O.

Remark. Theorem 2.3 gives a new proof of the fact that the row rank of a matrix is equal to its column rank. Indeed, let $A = (a_{ij})$ be an $m \times n$ matrix. Let S be the space of solutions of the equation $AX = O$, so $S = \operatorname{Ker} L_A$. By Theorem 3.2 of Chapter IV, we have

$$\dim S + \text{column rank} = n,$$

because the image of L_A is the space generated by the columns of A.

On the other hand, S is the space of vectors in $\mathbf{R}^n$ perpendicular to the rows of A, so if W is the row space then $S = W^\perp$. Therefore by Theorem 2.3 we get

$$\dim S + \text{row rank} = n.$$

This proves that row rank = column rank. In some ways, this is a more satisfying and conceptual proof of the relation than with the row and column operations that we used before.

We conclude this section by pointing out some useful notation. Let X, $Y \in \mathbf{R}^n$, and view X, Y as column vectors. Let $\langle\ ,\ \rangle$ denote the standard scalar product on $\mathbf{R}^n$. Thus by definition

$$\langle X, Y \rangle = {}^t X Y.$$

Similarly, let A be an $n \times n$ matrix. Then

$$\langle X, AY \rangle = {}^tXAY = {}^t({}^tAX)Y = \langle {}^tAX, Y \rangle.$$

Thus we obtain the formula

$$\boxed{\langle X, AY \rangle = \langle {}^tAX, Y \rangle.}$$

The transpose of the matrix A corresponds to transposing A to ${}^tA$ from one side of the scalar product to the other. This notation is frequently used in applications, which is one of the reasons for mentioning it here.

Exercises VI, §2

1. Find orthonormal bases for the subspaces of $\mathbf{R}^3$ generated by the following vectors:
   (a) $(1, 1, -1)$ and $(1, 0, 1)$,
   (b) $(2, 1, 1)$ and $(1, 3, -1)$.

2. Find an orthonormal basis for the subspace of $\mathbf{R}^4$ generated by the vectors $(1, 2, 1, 0)$ and $(1, 2, 3, 1)$.

3. Find an orthonormal basis for the subspace of $\mathbf{R}^4$ generated by $(1, 1, 0, 0)$, $(1, -1, 1, 1)$, and $(-1, 0, 2, 1)$.

4. Find an orthogonal basis for the space of solutions of the following equations.
   (a) $\begin{aligned} 2x + y - z &= 0 \\ y + z &= 0 \end{aligned}$ (b) $x - y + z = 0$
   (c) $\begin{aligned} 4x + 7y - \pi z &= 0 \\ 2x - y + z &= 0 \end{aligned}$ (d) $\begin{aligned} x + y + z &= 0 \\ x - y &= 0 \\ y + z &= 0 \end{aligned}$

In the next exercises, we consider the vector space of continuous functions on the interval $[0, 1]$. We define the scalar product of two such functions f, g by the rule

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

5. Let V be the subspace of functions generated by the two functions $f(t) = t$ and $g(t) = t^2$. Find an orthonormal basis for V.

6. Let V be the subspace generated by the three functions $1, t, t^2$ (where 1 is the constant function). Find an orthonormal basis for V.

7. Let V be a finite dimensional vector space with a positive definite scalar product. Let W be a subspace. Show that

$$V = W + W^\perp \quad \text{and} \quad W \cap W^\perp = \{0\}.$$

In the terminology of the preceding chapter, this means that V is the direct sum of W and its orthogonal complement. [Use Theorem 2.3.]

8. In Exercise 7, show that $(W^{\perp})^{\perp} = W$. Why is this immediate from Theorem 2.3?

9. (a) Let V be the space of symmetric $n \times n$ matrices. For $A, B \in V$ define

$$\langle A, B \rangle = \operatorname{tr}(AB),$$

where tr is the trace (sum of diagonal elements). Show that this satisfies all the properties of a positive definite scalar product. (You might already have done this as an exercise in a previous section.)

(b) Let W be the subspace of matrices A such that $\operatorname{tr}(A) = 0$. What is the dimension of the orthogonal complement of W, relative to the scalar product in part (a)? Give an explicit basis for this orthogonal complement.

10. Let A be a symmetric $n \times n$ matrix. Let $X, Y \in \mathbf{R}^n$ be eigenvectors for A, that is suppose that there exist numbers a, b such that $AX = aX$ and $AY = bY$. Assume that $a \neq b$. Prove that X, Y are perpendicular.

VI, §3. Bilinear Maps and Matrices

Let U, V, W be vector spaces, and let

$$g: U \times V \to W$$

be a map. We say that g is **bilinear** if for each fixed $u \in U$ the map

$$v \mapsto g(u, v)$$

is linear, and for each fixed $v \in V$, the map

$$u \mapsto g(u, v)$$

is linear. The first condition written out reads

$$g(u, v_1 + v_2) = g(u, v_1) + g(u, v_2),$$
$$g(u, cv) = cg(u, v),$$

and similarly for the second condition on the other side.

Example. Let A be an $m \times n$ matrix, $A = (a_{ij})$. We can define a map

$$g_A: \mathbf{R}^m \times \mathbf{R}^n \to \mathbf{R}$$

by letting

$$g_A(X, Y) = {}^tXAY,$$

which written out looks like this:

$$(x_1, \dots, x_m)\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Our vectors X and Y are supposed to be column vectors, so that ${}^tX$ is a row vector, as shown. Then ${}^tXA$ is a row vector, and ${}^tXAY$ is a 1×1 matrix, i.e. a number. Thus g_A maps pairs of vectors into the reals. Such a map g_A satisfies properties similar to those of a scalar product. If we fix X, then the map $Y \mapsto {}^tXAY$ is linear, and if we fix Y, then the map $X \mapsto {}^tXAY$ is also linear. In other words, say fixing X, we have

$$g_A(X, Y + Y') = g_A(X, Y) + g_A(X, Y'),$$
$$g_A(X, cY) = cg_A(X, Y),$$

and similarly on the other side. This is merely a reformulation of properties of multiplication of matrices, namely

$${}^tXA(Y + Y') = {}^tXAY + {}^tXAY',$$
$${}^tXA(cY) = c{}^tXAY.$$

It is convenient to write out the multiplication ${}^tXAY$ as a sum. Note that

$$j\text{-th component of } {}^tXA = \sum_{i=1}^{m} x_i a_{ij},$$

and thus

$${}^tXAY = \sum_{j=1}^{n} \sum_{i=1}^{m} x_i a_{ij} y_j = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i y_j.$$

Example. Let

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

If $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ then

$${}^tXAY = x_1y_1 + 2x_1y_2 + 3x_2y_1 - x_2y_2.$$

Theorem 3.1. *Given a bilinear map $g\colon \mathbf{R}^m \times \mathbf{R}^n \to \mathbf{R}$, there exists a unique matrix A such that $g = g_A$,* i.e. *such that*

$$g(X, Y) = {}^tXAY.$$

Proof. The statement of Theorem 3.1 is similar to the statement representing linear maps by matrices, and its proof is an extension of previous proofs. Remember that we used the standard bases for $\mathbf{R}^n$ to prove these previous results, and we used coordinates. We do the same here. Let $E^1,\dots,E^m$ be the standard unit vectors for $\mathbf{R}^m$, and let $U^1,\dots,U^n$ be the standard unit vectors for $\mathbf{R}^n$. We can then write any $X \in \mathbf{R}^m$ as

$$X = \sum_{i=1}^{m} x_i E^i$$

and any $Y \in \mathbf{R}^n$ as

$$Y = \sum_{j=1}^{n} y_j U^j.$$

Then

$$g(X, Y) = g(x_1E^1 + \cdots + x_mE^m, y_1U^1 + \cdots + y_nU^n).$$

Using the linearity on the left, we find

$$g(X, Y) = \sum_{i=1}^{m} x_i g(E^i, y_1U^1 + \cdots + y_nU^n).$$

Using the linearity on the right, we find

$$g(X, Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j g(E^i, U^j).$$

Let

$$a_{ij} = g(E^i, U^j).$$

Then we see that

$$g(X, Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j,$$

which is precisely the expression we obtained for the product

$$ {}^tXAY, $$

where A is the matrix (a_{ij}). This proves that $g = g_A$ for the choice of a_{ij} given above.

The uniqueness is also easy to see, and may be formulated as follows.

Uniqueness. *If A, B are $m \times n$ matrices such that for all vectors X, Y (of the appropriate dimension) we have*

$$ {}^tXAY = {}^tXBY, $$

then $A = B$.

Proof. Since the above relation holds for *all* vectors X, Y, it holds in particular for the unit vectors. Thus we apply the relation when $X = E^i$ and $Y = U^j$. Then the rule for multiplication of matrices shows that

$$ {}^tE^iAU^j = a_{ij} \qquad \text{and} \qquad {}^tE^iBU^j = b_{ij}. $$

Hence $a_{ij} = b_{ij}$ for all indices i, j. This shows that $A = B$.

Remark. Bilinear maps can be added and multiplied by scalars. The sum of two bilinear maps is again bilinear, and the product by a scalar is again bilinear. Hence bilinear maps form a vector space. Verify the rules

$$ g_{A+B} = g_A + g_B \qquad \text{and} \qquad g_{cA} = cg_A. $$

Then Theorem 3.1 can be expressed by saying that the association

$$ A \mapsto g_A $$

is an isomorphism between the space of $m \times n$ matrices, and the space of bilinear maps from $\mathbf{R}^m \times \mathbf{R}^n$ into $\mathbf{R}$.

Application to calculus. If you have had the calculus of several variables, you have associated with a function f of n variables the matrix of second partial derivatives

$$ \left(\frac{\partial^2 f}{\partial x_i\, \partial x_j}\right). $$

This matrix may be viewed as the matrix associated with a bilinear map, which is called the **Hessian**. Note that this matrix is symmetric since it is proved that for sufficiently smooth functions, the partials commute, that is

$$\frac{\partial^2 f}{\partial x_i \, \partial x_j} = \frac{\partial^2 f}{\partial x_j \, \partial x_i}.$$

Exercises VI, §3

1. Let A be $n \times n$ matrix, and assume that A is symmetric, i.e. $A = {}^tA$. Let $\varphi_A: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ be its associated bilinear map. Show that

$$g_A(X, Y) = g_A(Y, X)$$

for all X, $Y \in \mathbf{R}^n$, and thus that g_A is a scalar product, i.e. satisfies conditions **SP 1**, **SP 2**, and **SP 3**.

2. Conversely, assume that A is an $n \times n$ matrix such that

$$g_A(X, Y) = g_A(Y, X)$$

for all X, Y. Show that A is symmetric.

3. Write out in full in terms of coordinates the expression for ${}^tXAY$ when A is the following matrix, and X, Y are vectors of the corresponding dimension.

(a) $\begin{pmatrix} 2 & -3 \\ 4 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 4 & 1 \\ -2 & 5 \end{pmatrix}$ (c) $\begin{pmatrix} -5 & 2 \\ \pi & 7 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & 2 & -1 \\ -3 & 1 & 4 \\ 2 & 5 & -1 \end{pmatrix}$ (e) $\begin{pmatrix} -4 & 2 & 1 \\ 3 & 1 & 1 \\ 2 & 5 & 7 \end{pmatrix}$ (f) $\begin{pmatrix} -\frac{1}{2} & 2 & -5 \\ 1 & \frac{2}{3} & 4 \\ -1 & 0 & 3 \end{pmatrix}$