Scalar Products and Orthogonality

Daoping Zhang

The Chinese University of Hong Kong

dpzhang@math.cuhk.edu.hk

April 27, 2020

Outline

1. Scalar Products
2. Orthogonal Bases
3. Bilinear Maps and Matrices

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 $\langle v, v\rangle \geq 0$, 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.

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, SP2 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.

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

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

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

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

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 having 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

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

As with the case of n-space, we define the projection of v along w to be the vector cw.

Theorem (Schwarz Inequality)

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

$$|\langle v, w\rangle| \leq \|v\|\|w\|.$$

Theorem

For all $v, w \in V$ *then*

$$\|v + w\| \leq \|v\| + \|w\|.$$

Let $v_1, ..., v_n$ 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 $v_1, ..., v_n$. Thus subtracting linear combinations as above orthogonalizes v with respect to $v_1, ..., v_n$. 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 $v_1, ... v_n$.

Theorem

Let $v_1, ..., v_n$ 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 $a_1, ..., a_n$ be numbers. Then

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

Theorem

Bessel Inequality If $v_1, ..., v_n$ are mutually perpendicular unit vectors, and if c_i is the Fourier coefficient of v with respect to v_i, then

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

Proof.

We have

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

From this our inequality follows. □

Orthogonal Bases

Let V be a vector space with a positive definite scalar product throughout this section. A basis $\{v_1, ..., v_n\}$ 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.

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

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

Let w be anther element of V, and write

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

Then

$$\langle v, w \rangle = \sum_{i=1}^{n} x_i y_i$$

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.

Consider $\mathbb{R}^2$. Let

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

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

$$(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}) \quad \text{and} \quad (\tfrac{1}{\sqrt{2}}, \tfrac{-1}{\sqrt{2}}).$$

Theorem

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

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

is a basis of V. Of course, it is not an orthogonal basis. Let W_{m+1} be the space generated by $w_1, ..., w_m, v_{m+1}$. 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 $w_1, ..., w_m$. Thus we let

Proof.

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

Let

$$w_{m+1} = v_{m+1} - c_1 w_1 - \cdots - c_m w_m.$$

Then w_{m+1} is perpendicular to $w_1, \ldots, w_m$. Furthermore, $w_{m+1} \neq O$, and v_{m+1} lies in the space generated by $w_1, \ldots, w_{m+1}$. Hence $\{w_1, \ldots, w_{m+1}\}$ is an orthogonal basis of W_{m+1}. We can now proceed by induction, showing that the space W_{m+s} generated by

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

has orthogonal basis

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

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

Corollary

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.

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

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

Then $\{v_1', ..., v_n'\}$ is an orthogonal basis.
Given an orthogonal basis, we can always obtain an orthonomal basis by dividing each vector by its norm.

Find an orthonormal basis for the vector space generated by the vectors $(1, 1, 0, 1), (1, -2, 0, 0)$ and $(1, 0, -1, 2)$.
Find an orthogonal basis for the space of solutions of the linear equation

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

Theorem

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

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

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

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

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

$$\langle X, AY \rangle = X^t AY = (A^t X)^t Y = \langle A^t X, Y \rangle.$$

Thus we obtain the formula

$$\langle X, AY \rangle = \langle A^t X, Y \rangle.$$

The transpose of the matrix A corresponds to transposing A to A^t 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.

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

$$\begin{aligned} g(u, v_1 + v_2) &= g(u, v_1) + g(u, v_2), \\ g(u, cv) &= cg(u, v), \end{aligned}$$

and similarly for the second condition on the other side.

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

$$g_A : \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathbb{R}$$

by letting

$$g_A(X, Y) = X^t AY,$$

which written out looks like this:

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

This is a bilinear map.

Theorem

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

$$g(X, Y) = X^t AY.$$

Uniqueness

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

$$X^t AY = X^t BY,$$

then $A = B$.

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.