Inner products

A. Eremenko

September 7, 2024

An inner product, also known as dot product or scalar product is an operation on vectors of a vector space which from any two vectors $\mathbf{x}$ and $\mathbf{y}$ produces a number which we denote by $(\mathbf{x}, \mathbf{y})$. **In this lecture, all vector spaces are defined over real numbers**. The operation is described by the following three properties:

$$(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x}), \tag{1}$$

$$(c_1\mathbf{x}_1 + c_2\mathbf{x}_2, \mathbf{y}) = c_1(\mathbf{x}_1, \mathbf{y}) + c_2(\mathbf{x}_2, \mathbf{y}), \tag{2}$$

that is the operation is linear with respect to the first argument. From (1) and (2) follows that it is also linear with respect to the second argument:

$$(\mathbf{x}, c_1\mathbf{y}_1 + c_2\mathbf{y}_2) = c_1(\mathbf{x}, \mathbf{y}_1) + c_2(\mathbf{x}, y_2).$$

The final property is

$$(\mathbf{x}, \mathbf{x}) \geq 0, \quad \text{and} \quad (\mathbf{x}, \mathbf{x}) = 0 \quad \text{only when} \quad \mathbf{x} = \mathbf{0}. \tag{3}$$

Examples of dot products. There can be many different inner products on a given vector space. Our primary example is the *standard dot product* on $\mathbf{R}^n$:

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \ldots + x_ny_n.$$

Check that is has the three required properties (1)–(3). For another example, let us fix some *positive* numbers $a_1, \ldots, a_n$ and define the dot product by

$$(\mathbf{x}, \mathbf{y}) = a_1x_1y_2 + \ldots + a_nx_ny_n. \tag{4}$$

Again all three properties (1)–(3) hold.

Now we consider examples on spaces of functions. We recall that a function defined on an interval $[a, b]$ is called *piecewise-continuous* if there are points $a = t_0 < t_1 < t_2 < t_n = b$ such that the function is continuous on each interval (t_k, t_{k+1}) and the one-sided limits at each t_k exist (from both sides of t_k). It is easy to see that the set of all such functions is a vector space. We will call it $PC(a, b)$. A dot product on this space can be defined as

$$(f, g) = \int_a^b f(x)g(x)dx.$$

Check that properties (1)-(3) are satisfied. A generalization similar to (4) can be obtained if we fix a strictly positive continuous function w, and define

$$(f, g) = \int_a^b f(x)g(x)w(x)dx.$$

From now we assume that some dot product is fixed on a vector space V.

The *length* of a vector $\mathbf{x}$ is defined as the positive square root

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}.$$

Let us compute:

$$\|\mathbf{x} + \mathbf{y}\|^2 = (\mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y}) = (\mathbf{x}, \mathbf{x}) + (\mathbf{y}, \mathbf{y}) + 2(\mathbf{x}, \mathbf{y}), \tag{5}$$

and, switching the sign of $\mathbf{y}$:

$$\|x - \mathbf{y}\|^2 = (\mathbf{x}, \mathbf{x}) + (\mathbf{y}, \mathbf{y}) - 2(\mathbf{x}, \mathbf{y}). \tag{6}$$

Adding (5) and (6) we obtain the *polarization identity*

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2. \tag{7}$$

In words: the sum of the squares of diagonals of a parallelogram is equal to the sum of the squares of the sides. Then we have

Cauchy–Schwarz inequality. $|(\mathbf{x}, \mathbf{y})| \leq \|\mathbf{x}\|\|\mathbf{y}\|$. *The equality can hold only if* $\mathbf{x}$ *and* $\mathbf{y}$ *are collinear.*

Proof. Let t be a number, and consider

$$(\mathbf{x} + t\mathbf{y}, \mathbf{x} + ty) = (\mathbf{x}, \mathbf{x}) + 2t(\mathbf{x}, \mathbf{y}) + t^2(\mathbf{y}, \mathbf{y}),$$

where we used (1), (3). By (3), this must be non-negative for all real t. But this is a quadratic polynomial in t, ands since $(\mathbf{y},\mathbf{y})$ is non-negative, this quadratic polynomial is non-negative if and only if the discriminant is non-positive:

$$(\mathbf{x},\mathbf{y}) - (\mathbf{x},\mathbf{x})(\mathbf{y},\mathbf{y}) \leq 0.$$

This gives our inequality. Suppose now that we have equality, so our discriminant equals to zero. Then the quadratic polynomial has a root t_0, so we have $(\mathbf{x}+t_0\mathbf{y},\mathbf{x}+t_0\mathbf{y}) = 0$, and by property (3), second part, this is only possible if $\mathbf{x}+t_0\mathbf{y} = 0$ that is $\mathbf{x}$ is collinear to $\mathbf{y}$.

For example, with the standard dot product on $\mathbf{R}^n$ the Cauchy–Schwarz inequality reads:

$$|x_1y_1 + x_2y_2 + \ldots + x_ny_n| \leq \sqrt{(x_1^2 + \ldots + x_n^2)(y_1^2 + \ldots + y_n^2)}.$$

and on the space of functions

$$\left|\int_a^b f(x)g(x)dx\right| \leq \left(\int_a^b f^2(x)dx\right)^{1/2}\left(\int_a^b g^2(x)dx\right)^{1/2}.$$

Cauchy–Schwarz inequality together with (5) give

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|y\|,$$

and this is called the *triangle ineuality*: the length of side of a triangle is at most the sum of the lengths of two other sides. By recalling the case of equality in the Sauchy–Schwarz inequality we concluse that the straight line gives the shortest distance between two points.

Using Cauchy–Schwarz inequality, we can also define the angle between two vectors:

$$\angle(\mathbf{x},\mathbf{y}) = \arccos\frac{(\mathbf{x},\mathbf{y})}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

So the angle is always between 0 and π. This is a different meaning of the word "angle" from the meaning used in Lecture 8! With this notion of angle and (6) we also obtain the familiar cosine theorem:

$$\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|y\|\cos\alpha$$

which shows that our geometry coincides with that studied in high school, except that now we have all those notions in arbitrary dimension.

Vectors are called *orthogonal* if the angle between them is 90°, that is if $(\mathbf{x}, \mathbf{y}) = 0$. Sometimes this property is recorded as $\mathbf{x} \perp \mathbf{y}$.

Suppose that some non-zero vectors $\mathbf{v}_1, \ldots$ are mutually orthogonal, that is

$$(\mathbf{v}_i, \mathbf{v}_j) = 0 \quad \text{for all} \quad i \neq j.$$

Such a set of vectors is called an *orthogonal system.* It does not have to be finite.

Remember, the zero vector can not be included into an orthogonal system, by definition! Then we have

Theorem. *Any orthogonal system is linearly independent.*

Proof. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are mutually orthogonal, and

$$c_1\mathbf{v}_1 + \ldots, c_n\mathbf{v}_n = \mathbf{0}.$$

For any j, we dot-multiply this on $\mathbf{v}_j$. Then all products vanish, except one, and we obtain $c_j(\mathbf{v}_j, \mathbf{v}_j) = 0$, but $(\mathbf{v}_j, \mathbf{v}_j) \neq 0$ since $\mathbf{v}_j \neq 0$ and by property (3). Therefore $c_j = 0$, and this holds for every j.

So an orthogonal system which spans a finite-dimensional space is automatically a basis of this space. Such a system is called an *orthogonal basis.*

Let $\mathbf{x}$ be a vector in the span of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ so we have

$$\mathbf{x} = c_1\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n. \tag{8}$$

In general, to find the coefficients c_j we need to solve the system of linear equations

$$A\mathbf{c} = \mathbf{x},$$

where $A = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $c = (c_1, \ldots, c_n)^T$. However orthogonality allows us to write the answer immediately: dot-multiply equation (8) on $\mathbf{v}_j$. We obtain

$$c_j = \frac{(\mathbf{x}, \mathbf{v}_j)}{\|\mathbf{v}_j\|^2}. \tag{9}$$

So in particular it is useful to have an orthogonal basis in the space, since expansion of vectors in this basis is easy. We will later return to the question how to find an orthogonal basis.

When an orthogonal system has the additional property that $\|\mathbf{v}_j\| = 1$ for all vectors of this system, formula (9) simplifies:

$$c_j = (\mathbf{x}, \mathbf{v}_j). \tag{10}$$

Such systems of vectors are called *orthonormal systems.* In particular we have orthonormal bases. Equations (9) and (10) are called *Fourier formulas.*

Examples. 1. The standard basis $e_1, e_2, \ldots, e_n$ is an orthonormal basis in $\mathbf{R}^n$ with the standard dot product.

2. In the space $C[0,\pi]$ of continuous functions on $[0,\pi]$ with standard dot product, consider the vectors $\phi_n(t) = \sin nt,\ n = 1, 2, \ldots$. This is an orthogonal system. Indeed,

$$\begin{aligned}(\phi_m, \phi_n) &= \int_0^\pi \sin mt \sin nt \, dt \\ &= \frac{1}{2}\int_0^\pi \cos(m-n)t - \cos(m+n)t \, dt \\ &= \begin{cases} 0 & \text{if} \quad m \neq n, \\ \pi/2 & \text{if} \quad n = n. \end{cases}\end{aligned}$$

So the system $(\phi_n)_{n=1}^\infty$ is orthogonal, and the length of each vector is $\sqrt{\pi/2}$. Thus

$$\sqrt{\frac{2}{\pi}} \sin nt, \quad n = 1, 2, 3, \ldots$$

is an orthonormal system.

Suppose that a function f belongs to the span of this system, that is

$$f = \sum_{n=1}^N c_n \phi_n.$$

Then Fourier formulas become

$$c_n = \frac{2}{\pi}\int_0^\pi f(t) \sin nt \, dt.$$

Let us (dot)-multiply the expansion (8) on itself. Using orthogonality, we see that all products of the type $(\mathbf{v}_i, \mathbf{v}_j)$ vanish. What remains is

$$\|\mathbf{x}\|^2 = c_1^2\|\mathbf{v}_1\|^2 + \ldots + c_n^2\|\mathbf{v}_n\|^2.$$

This is a generalization of Pythagorean theorem to arbitrary dimension.

Orthogonal subspaces. Two subspaces X and Y of a vector space V with a dot product are called *orthogonal* if every $\mathbf{x} \in X$ is orthogonal to every $\mathbf{y} \in Y$. Zero-vector is the only common vector of such subspaces. Indeed, if a vector $\mathbf{x} \in X \cap Y$ then $\mathbf{x}$ is orthogonal to itself, and this is only possible when $\mathbf{x} = \mathbf{0}$, by property (3).

So this use of the word "orthogonal" in Linear Algebra is different from the use of the word "perpendicular" in everyday life: a wall in a room is not orthogonal to the floor; they have a common line!

Let V be a space with an inner product, and U a subspace. The *orthogonal complement* $U^\perp$ of U is defined as the set of all vectors in V which are orthogonal to every vector in U:

$$U^\perp := \{\mathbf{x} \in V : (\mathbf{x}, \mathbf{y}) = 0 \text{ for all } \mathbf{y} \in U\}.$$

For example, in $\mathbf{R}^3$ the line through e_3 is orthogonal to the plane spanned by e_1 and e_2.

The main example of orthogonal complement in $\mathbf{R}^n$ with the standard dot product is the following:

For every matrix A, $N(A^T) = C(A)^\perp$, and $N(A) = (R(A)^T)^\perp$.

This follows immediately from the definitions: $\mathbf{x}^T A = 0$ means exactly that every column of A is orthogonal to the row $\mathbf{x}^T$. $N(A^T)$ by definition consists of all rows $\mathbf{x}^T$ which are orthogonal to every column of A.

It is important that the converse is also true:

Every subspace U of $\mathbf{R}^n$ is the column space of some matrix.

Indeed, choose a basis in U, use as columns of a matrix.

So rank and nullity theorem gives a relation between dimensions of a subspace U of V and its orthogonal complement

$$\dim U + \dim U^\perp = \dim V. \tag{11}$$

This relation allows one to prove another important formula:

$$(U^\perp)^\perp = U.$$

Proof. It is an immediate consequence of the definition that $U \subset (U^\perp)^\perp$. But the dimension formula (11) implies that U and $(U^\perp)^\perp$ have the same dimensions. So by Theorem 4 in Lecture 5 these spaces are equal.

Since U and $U^{\perp}$ have only the zero vector in common, we can take any basis $\mathbf{u}_1, \ldots, \mathbf{u}_k$ in U and a basis $\mathbf{u}'_1, \ldots, \mathbf{u}'_\ell$ in $U^{\perp}$. Then together these bases form a linearly independent system: indeed, if

$$\sum_{j=1}^{k} c_j \mathbf{u}_j + \sum_{j=1}^{\ell} c'_j \mathbf{u}'_j = \mathbf{0},$$

then each of the two sums must be equal to zero since the first sum belongs to U and the second one to $U^{\perp}$. So all c_j and c'_j must be zero since both $(\mathbf{u}_j)$ and $(\mathbf{u}'_j)$ are bases in their spaces. Now dimension formula (11) implies that $(\mathbf{u}_j)$ and $(\mathbf{u}'_j)$ together must form a basis in the whole space.

So we obtain: *If $U \subset V$ is a subspace, then any vector $\mathbf{x} \in V$ can be written as*

$$\mathbf{x} = \mathbf{u} + \mathbf{w}, \quad \textit{where} \quad \mathbf{u} \in U \quad \textit{and} \quad \mathbf{w} \in U^{\perp}$$

in a unique way.

Now let $\mathbf{u}_1, \ldots \mathbf{u}_k$ be any orthogonal system in a space V, and let U be its span. Then we can write any vector $\mathbf{x} \in V$ as

$$\mathbf{x} = \mathbf{u} + \mathbf{w} = \sum_{j=1}^{k} c_j \mathbf{u}_j + \mathbf{w}, \quad \text{where} \quad \mathbf{w} \in U^{\perp}.$$

Let us square this:

$$(\mathbf{x}, \mathbf{x}) = \sum_{j=1}^{k} c_j^2 \|\mathbf{u}_j\|^2 + \|\mathbf{w}\|^2.$$

All mixed products $(\mathbf{x}_i, \mathbf{x}_j)$ and $(\mathbf{x}_j, \mathbf{w})$ vanish because of orthogonality. So we have

$$\sum_{j=1}^{k} c_j^2 \mathbf{u}_j{}^2 \leq \|\mathbf{x}\|^2,$$

and this holds for any orthogonal system and any vector, and c_j are given by formulas (9). This is called the *Bessel Inequality.* When the system $\mathbf{u}_1, \ldots, \mathbf{u}_k$ spans the space U, so that $U^{\perp} = \{\mathbf{0}\}$, Bessel's inequality becomes an equality which is called the *Parseval equality* (the generalization of Pythagorean theorem mentioned above).