Inner product spaces

Laurence Barker, Department of Mathematics, Bilkent University

version: 16 April 2024

In current form, these notes are for supplementary material for two courses, MATH 220 Linear Algebra and MATH 224 Linear Algebra 2.

Actually, this particular draft, 16 April, is a first draft and doubtless has many slips. I shall later be making changes, some of which I have in mind already.

1: Real inner product spaces

For vector spaces over arbitrary fields, there is no general notion of distance or angle. In this section, we shall be confining attention to real vector spaces, that is, vector spaces over the field of real numbers $\mathbb{R}$. Sometimes, not always, when dealing with real vector spaces, consideration of distances or angles may be useful or necessary. In that case, the real vector spaces involved have to be viewed as something more than just real vector spaces. They have to be equipped with some or other kind of further structure. In this section, we shall be considering real vector spaces equipped with a further structure called an inner product.

Let V be a real vector space, we mean, a vector space over the field of real numbers $\mathbb{R}$. We define a **bilinear form** on V to be a function $\mathbb{R} \leftarrow V \times V$, written $\langle x \,|\, y \rangle \hookleftarrow (x, y)$, such that the following two conditions hold:

Left linearity: For all $y \in V$, the function $\mathbb{R} \leftarrow V$ given by $\langle x \,|\, y \rangle \hookleftarrow x$ is a linear map.

Right linearity: For all $x \in V$, the function $\mathbb{R} \leftarrow V$ given by $\langle x \,|\, y \rangle \hookleftarrow y$ is a linear map.

Explicitly, the conditions are that, for all $x, x', y, y' \in V$ and $\lambda, \lambda \in \mathbb{R}$, we have

$$\langle \lambda x + \lambda' x' \,|\, y \rangle = \lambda \langle x \,|\, y \rangle + \lambda' \langle x' \,|\, y \rangle \ , \qquad\qquad \langle x \,|\, \lambda y + \lambda' y' \rangle = \lambda \langle x \,|\, y \rangle + \lambda' \langle x \,|\, y' \rangle \ .$$

We shall be considering two further conditions:

• The bilinear form is said to be **symmetric** provided $\langle x \,|\, y \rangle = \langle y \,|\, x \rangle$ for all $x, y \in V$.

• The bilinear form is said to be **positive-definite** provided $\langle x \,|\, x \rangle > 0$ for all $x \in V$ with $x \neq 0$.

We define an **inner product** on V to be a positive-definite symmetric bilinear form on V. We define a **real inner product space** to be a real vector space equipped with an inner product.

Let V be a real inner product space. For $x \in V$, we define the **norm** of x to be the non-negative real number

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

Recall that, given $\lambda \in \mathbb{R}$, the **modulus** of λ is defined to be the non-negative real number $|\lambda|$ such that $|\lambda| = \lambda$ when $\lambda \geq 0$ whereas $|\lambda| = -\lambda$ when $\lambda \leq 0$. Thus, for all $\lambda \in \mathbb{R}$, we have $|\lambda| = \sqrt{\lambda^2}$.

Theorem 1.1: (Cauchy–Schwarz Inequality.) *Given a real inner product space V and $x, y \in V$, then*

$$|\langle x \,|\, y \rangle| \leq \|x\|.\|y\|$$

with equality if and only if $x = 0$ or $y = 0$ or $\mathbb{R}x = \mathbb{R}y$.

Proof: We may assume that $y \neq 0$. Given $t \in \mathbb{R}$, then

$$at^2 + bt + c = \|x + ty\|^2 \geq 0$$

where $a = \|y\|^2$ and $b = 2|\langle x \,|\, y \langle|$ and $c = \|x\|^2$. Note that $a \neq 0$. The graph of the function $at^2 + bt + c \hookleftarrow t$ is a parabola ranging in the upper half place, possibly touching the horizontal axis, but not below that axis. So the discriminant of the quadratic equation $at^2 + bt + c = 0$ must be non-positive, $b^2 - 4ac \leq 0$, in other words, $(b/2)^2 \leq ac$. Taking a square root, we obtain the asserted inequality.

The argument also shows that $x + ty = 0$ for some t if and only if $(b/2)^2 = ac$. The rider follows. $\square$

For nonzero x and y in V, we define the **angle** between x and y to be the real number θ in the interval $0 \leq \theta < \pi$ such that

$$\cos(\theta) = \langle x \,|\, y \rangle / \|x\|.\|y\| \ .$$

The Cauchy–Schwarz Inequality assures us that θ is well-defined.

Theorem 1.2: (Triangle Inequality, first version.) *Given a real inner product space and $u, v \in V$, then*

$$\|u + v\| \leq \|u\| + \|v\| \ .$$

Proof: Applying the previous theorem, $\|u+v\|^2 = \|u\|^2 + 2\langle u \,|\, v \rangle + \|v\|^2 \leq (\|u\| + \|v\|)^2$. □

Theorem 1.3: (Triangle Inequality, second version.) *Given a real inner product space and $x, y, z \in V$, then*

$$\|x - z\| \leq \|x - y\| + \|y - z\| \ .$$

Proof: Putting $u = x - y$ and $v = y - z$, then $u + v = x - z$. Thus, we have reduced to the previous version of the theorem. □

Conversely, Theorem 1.2 can be quickly deduced from Theorem 1.3 by putting $x = u$ and $y = 0$ and $z = -v$. So the two theorems are really just two different ways of expressing one and the same theorem.

For $x, y \in V$, we define

$$d(x, y) = \|x - y\| \ .$$

We call $d(x, y)$ the **distance** between x and y. The Triangle Inequality says that

$$d(x, z) \leq d(x, y) + d(y, z) \ .$$

Put roughly, without digressing into any precise definition of a triangle, the theorem says that, for any triangle with vertices in a real inner product space, the length of any one of the edges is at most the sum of the lengths of the other two edges.

(Which of the above two versions of the Triangle Inequalty should we memorize for the exam? If one is asking that question, then one is doing it wrong. The trick is to understand, not to memorize. Everything about the Triangle Inequality can be recovered from just understanding what it is saying about the three edges of a triangle. That way, one sees that Theorems 1.2 and 1.3 have the same content. In a similar way, everything about the Cauchy–Schwartz Inequality can be recovered from just understanding how it tells us that the angle between two non-zero vectors is well-defined.)

♣ ♡ ♢ ♠

We now discuss the most important example of a finite-dimensional inner product space.

Let n be a positive integer and consider the real vector space $\mathbb{R}^n$. Recall, the **standard basis** for $\mathbb{R}^n$ is defined to be the basis $\{e_1, ..., e_n\}$ such that, given $x \in \mathbb{R}^n$ and writing $x = (x_1, ..., x_n)$ with each $x_s \in \mathbb{R}$, then $x = \sum_s x_s e_s$. The following notation is useful. For $s, t \in [1, n]$, the **Kronecker delta** symbol $\delta_{s,t}$ is defined to be the real number

$$\delta_{s,t} = \begin{cases} 1 & \text{if } s = t, \\ 0 & \text{if } s \neq t. \end{cases}$$

Thus, $e_s = (\delta_{s,1}, ..., \delta_{s,n})$. For example, the standard basis of $\mathbb{R}^3$ is $\{e_1, e_2, e_3\}$ where $e_1 = (1,0,0)$ and $e_2 = (0,1,0)$ and $e_3 = (0,0,1)$.

On $\mathbb{R}^n$, we define an inner product $\langle - | - \rangle$ such that, given $x, y \in \mathbb{R}^n$ and writing $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$, then

$$\langle x \,|\, y \rangle = \sum_{s \in [1,n]} x_s y_s \ .$$

We call this inner product the **dot product**. Often, as an alternative notation, we write $x.y = \langle x \,|\, y \rangle$. Note that

$$\|x\| = \sqrt{x.x} = \sqrt{\sum\nolimits_s x_s^2} \ , \qquad\qquad d(x,y) = \sqrt{\sum\nolimits_s (x_s - y_s)^2} \ .$$

When e speak of the inner product space $\mathbb{R}^n$, it is to be understood that the inner product under consideration is the dot product.

Of course, a dot is often used just to separate expressions that are being multiplied together, as in $2.3 = 6$. Fortunately, there will be no ambiguity in the notation $x.y$ for $x, y \in \mathbb{R}^n$, since we shall only be considering a multiplication operation on $\mathbb{R}^n$ when $n = 1$, in which case the two interpretations of $x.y$ coincide.

It is worth restating, in the special case $V = \mathbb{R}^n$, the two theorems in the previous section. The proofs are by substituting $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ into the statements of those theorems.

Theorem 2.1: (Cauchy–Schwarz Inequality, classical version.) *Given real numbers x_1, ..., x_n, y_1, ..., y_n, then*

$$\left(\sum\nolimits_r x_r y_r\right)^2 \leq \left(\sum\nolimits_s x_s^2\right) \left(\sum\nolimits_t y_t^2\right) \ .$$

Proof: Substitute $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ into Theorem 1.1. □

Theorem 2.2: (Triangle Inequality, classical version.) *Given real numbers u_1, ..., u_n, v_1, ..., v_n, then*

$$\sqrt{\sum\nolimits_r (u_r + v_r)^2} \leq \sqrt{\sum\nolimits_s u_s^2} + \sqrt{\sum\nolimits_t v_t^2} \ .$$

Proof: Substitute $u = (u_1, ..., u_n)$ and $v = (v_1, ..., v_n)$ into Theorem 1.2. □

♣ ♡ ♢ ♠

Often, a system of linear equations has to be laboriously solved in order to express a given vector as a linear combination of the elements of a given basis. The next remark gives a quick trick for finding the coefficients. The trick is applicable for suitable basis of a real inner product space.

Let V be a finite-dimensional real inner product space. Vectors $v_1, ..., v_n \in V$ are said to be **mutually orthogonal** provided $\langle v_s \,|\, v_t \rangle = 0$ whenever $s \neq t$. A basis $\{v_1, ..., v_n\}$ for V is called an **orthogonal basis** for V provided the elements v_1, ..., v_n are mutually orthogonal.

Remark 2.3: *Let $\{u_1, ...u_n\}$ be an orthogonal basis for an inner product space V. Given $x \in V$, then*

$$x = \sum_{s \in [1,n]} \frac{\langle u_s \,|\, x \rangle}{\|u_s\|^2} \, u_s \ .$$

Proof: Since $\{u_1, ..., u_n\}$ spans V, we can write $x = \sum_s x_s u_s$ with each $x_s \in \mathbb{R}$. The required conclusion follows upon evaluating $\langle u_s \,|\, x \rangle$. $\square$

A basis $\{w_1, ..., w_n\}$ for V is said to be **orthonormal** provided

$$\langle w_s \,|\, w_t \rangle = \delta_{s,t}$$

for all $s, t \in [1, n]$. Plainly, the next remark is a special case of the previous one.

Remark 2.4: *Let $\{w_1, ...w_n\}$ be an orthonormal basis for an inner product space V. Given $x \in V$, then*

$$x = \sum_{s \in [1,n]} \langle w_s \,|\, x \rangle w_s \ .$$

Corollary 2.5: *Let $\{w_1, ...w_n\}$ be an orthonormal basis for an inner product space V. Let $x, y \in V$. For each $s \in [1, n]$, write $x_s = \langle w_s \,|\, x \rangle$ and $y_s = \langle w_s \,|\, y \rangle$. Then:*

$$\langle x \,|\, y \rangle = \sum_s x_s y_s \ .$$

In particular $\|x\| = \sqrt{\sum_s x_s^2}$. Also, $d(x, y) = \sqrt{\sum_s (x_s - y_s)^2}$.

The corollary reveals that, when working with coordinates with respect to an orthonormal basis, the inner product space V behaves very much like the inner product space $\mathbb{R}^n$. Let us introduce a definition to capture that observation. Given real inner product spaces U and V, we define an **isometry** $\alpha : U \leftarrow V$ to be an isomorphism $U \leftarrow V$ such that

$$\langle \alpha(v) \,|\, \alpha(v') \rangle = \langle v \,|\, v' \rangle \ .$$

The corollary now says that, given an orthonormal basis $\{w_1, ..., w_n\}$ for a real inner product space V, then there is an isometry $\mathbb{R}^n \leftarrow V$ determined by the condition that $e_s \leftarrow\!\shortmid w_s$. However, before we can conclude that this applies to every finite-dimensional real inner product space, we must prove that every finite-dimensional real inner product space has an orthonormal basis. We shall do that in the next subsection.

♣ ♡ ♢ ♠

Again, let V be a finite-dimensional inner product space. We now describe a process, called the **Gram–Schmidt Process**, for replacing any given basis of an real inner product space with a orthonormal basis.

Let $\mathcal{U} = \{u_1, ..., u_n\}$ be any basis for V. We define a set $\mathcal{V} = \{v_1, ..., v_n\}$ recursively by the condition that $v_1 = u_1$ and

$$v_r = u_r - \frac{\langle v_1 \,|\, u_r \rangle}{\|v_1\|^2} v_1 - ... - \frac{\langle v_{r-1} \,|\, u_r \rangle}{\|v_{r-1}\|^2} v_{r-1}$$

for $2 \leq r \leq n$.

Remark 2.6: *With the notation above, $\mathcal{V}$ is an ortogonal basis for V.*

Proof: For $2 \leq s \leq n$, if the vectors $v_1, \ ... \ , v_{s-1}\}$ are mutually orthogonal, then a direct calculation shows that $\langle v_t \,|\, v_s \rangle = 0$ for all $1 \leq t < s$. An inductive argument now yields the required conclusion. $\square$

We define $\mathcal{W} = \{w_1, ..., w_n\}$ where each $w_r = v_r / \|v_r\|$. The previous remark immediately implies the next one.

Remark 2.7: *With the notation above, $\mathcal{W}$ is an orthonormal basis for V.*

The above process for constructing the orthogonal basis $\mathcal{V}$ or the orthonormal basis $\mathcal{W}$ is called the **Gram–Schmidt process**. Combining Corollary 2.5 and Remark 2.7, we obtain the following result.

Corollary 2.8: *Let V be a finite-dimensional inner product space. Then V has an orthonormal basis. Moreover, letting $\{w_1, ..., w_n\}$ be an orthonormal basis for V, then there is an isometry $\mathbb{R}^n \leftarrow V$ given by $(\langle w_1 \,|\, x\rangle, ... \langle w_n \,|\, x\rangle) \leftarrow\!\shortmid x$.*

Procedural exercises on Section 1

1.A: Let $u_1 = (2,1,1)$ and $u_2 = (1,3,1)$ and $u_3 = (1,1,3)$. Let $\{w_1, w_2, w_3\}$ be the orthonormal basis for $\mathbb{R}^3$ obtained from $\{u_1, u_2, u_3\}$ by the Gram–Schmidt process. Evaluate w_1 and w_2 and w_3.

1.B: Let V be the vector space whose vectors are the polynomial functions $\mathbb{R} \leftarrow \mathbb{R}$ of degree at most 4. That is to say, the vectors in V are the functions $f : \mathbb{R} \leftarrow \mathbb{R}$ that can be expressed in the form

$$f(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

with each $a_s \in \mathbb{R}$. We make V become a real inner product space such that

$$\langle f \,|\, g \rangle = \int_{-1}^{1} f(t) g(t) \, \mathrm{d}t \ .$$

Let $\{u_0, u_1, u_2, u_3\}$ be the basis for V such that

$$u_0(t) = 1 \ , \qquad u_1(t) = t \ , \qquad u_2(t) = t^2 \ , \qquad u_3(t) = t^3 \ .$$

Let $\{w_0, w_1, w_2, w_3\}$ be the orthogonal basis obtained from $\{u_0, u_1, u_3, u_4\}$ without any normalization. Give explicit formulas for $w_0(t)$, $w_1(t)$, $w_2(t)$, $w_3(t)$.

1.C:

Theoretical exercises on Section 1

1.$\mathcal{A}$: Let $\langle - | - \rangle$ be a bilinear form on a real vector space V. We say that $\langle - | - \rangle$ is **non-degenerate** provided the following two conditions hold:

Left non-degeneracy: all $x \in V$ with $x \neq 0$, there exists $y \in V$ satisfying $\langle x \,|\, y \rangle \neq 0$.

Right non-degeneracy: for all $y \in V$ with $y \neq 0$ there exists $x \in V$ satisfying $\langle x \,|\, y \rangle \neq 0$.

(a) Suppose V is finite-dimensional. Show that $\langle - | - \rangle$ is left non-degenerate if and only if $\langle - | - \rangle$ is right non-degenerate.

(b) Give an example of an infinite-dimensional V such that the conclusion of part (a) fails.

1.$\mathcal{B}$: Let $\langle - | - \rangle$ be a non-degenerate bilinear form on a real vector space V. Show that $\langle x \,|\, x \rangle \neq 0$ for some $x \in V$.

1.$\mathcal{C}$: Let V be a finite-dimensional inner product space and $\alpha : V \to V$ a function such that $\alpha(x, y) = d(\alpha(x), \alpha(y))$ for all $x, y \in V$. Show that α is an isometry. (Warning: most of the work is in showing that α is a linear map.)

Solutions to procedural exercises on Section 1

1.A: We have $w_1 = u_1/\|u_1\| = \frac{1}{\sqrt{6}}(2,1,1)$. Let $v_2 = u_2 - (w_1.u_2)w_1$. Then

$$v_2 = (1,3,1) - \left(\frac{1}{\sqrt{6}}(2,1,1).(1,3,1)\right)\frac{1}{\sqrt{6}}(2,1,1) = (1,3,1) - (2,1,1) = (-1,2,0)\ .$$

We have $w_2 = u_2/\|u_2\| = \frac{1}{\sqrt{5}}(-1,2,0)$. Let $v_3 = u_3 - (w_1.u_3)w_1 - (w_2.u_3)w_2$. Then

$$\begin{aligned} v_3 &= (1,1,3) - \left(\frac{1}{\sqrt{6}}(2,1,1).(1,1,3)\right)\frac{1}{\sqrt{6}}(2,1,1) - \left(\frac{1}{\sqrt{5}}(-1,2,0).(1,1,3)\right)\frac{1}{\sqrt{5}}(-1,2,0) \\ &= (1,1,3) - (2,1,1) - \frac{1}{5}(-1,2,0) = \frac{1}{5}\big((-5,0,10) - (-1,2,0)\big) = \frac{2}{5}(-2,-1,5)\ . \end{aligned}$$

We have $w_3 = v_3/\|v_3\| = (-2,-1,5)/\|(-2,-1,5)\| = \frac{1}{\sqrt{30}}(-2,-1,5)$. In conclusion,

$$w_1 = \frac{1}{\sqrt{6}}(2,1,1)\ , \qquad w_2 = \frac{1}{\sqrt{5}}(-1,2,0)\ , \qquad w_3 = \frac{1}{\sqrt{30}}(-2,-1,5)\ .$$

1.B:

Solutions to theoretical exercises on Section 1

1.$\mathcal{A}$:

1.$\mathcal{B}$: