Inner product spaces

Laurence Barker, Department of Mathematics, Bilkent University

version: 13 May 2024

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

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

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

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.

Note that, given a subspace U of a real vector space V, then any inner product on V restricts to an inner product on U. Thus, any subspace of a real inner product space becomes, itself, a real inner product space.

♣ ♡ ♢ ♠

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: *Given a real inner product space V and $x, y \in V$, then:*

(1) (Cauchy–Schwarz Inequality.) *We have*

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

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

(2) (Triangle Inequality.) *We have*

$$\|x + y\| \leq \|x\| + \|y\|$$

with equality if and only if $x = 0$ or $y = 0$ or $x = \lambda y$ for some positive real number y.

Proof: For part (1), 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 of part (1) follows.

For part (2), an application of part (1) yields

$$\|x + y\|^2 = \|x\|^2 + 2\langle x \,|\, y\rangle + \|y\|^2 \leq (\|x\| + \|y\|)^2 \,.$$

We have equality if and only if $\langle x \,|\, y\rangle = \|x\|.\|y\|$, equivalently, $\langle x \,|\, y\rangle \geq 0$ and $\langle x \,|\, y\rangle = \|x\|.\|y\|$. The rider of part (2) now follows from the rider of part (1). $\square$

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, alternative version.) *Given a real inner product space and $u, v, w \in V$, then*

$$\|u - w\| \leq \|u - v\| + \|v - w\| \,.$$

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

Conversely, the latter version of the Triangle Inequality can be quickly deduced from the former version by putting $u = x$ and $v = 0$ and $w = -y$. So the two versions of the Triangle Inequality are two different ways of expressing one and the same assertion.

For $u, v \in V$, we define

$$d(u, v) = \|u - v\| \,.$$

We call $d(u, v)$ the **distance** between u and v. The Triangle Inequality says that, for all $u, v, w \in V$, we have

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

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.

♣ ♡ ♢ ♠

For any positive integer n, we shall equip the real vector space $\mathbb{R}^n$ with an inner product called the dot product. Thus, $\mathbb{R}^n$ will become a real inner product space. This is an important example. Indeed, as we shall later explain, every finite-dimensional real inner product space is, in a sense which we shall make precise, a copy of $\mathbb{R}^n$ equipped with the dot product.

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 $x = \sum_s x_s e_s$.

We define the **dot product** on $\mathbb{R}^n$ to be the 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

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

When we speak of the inner product space $\mathbb{R}^n$, it is to be understood that, unless otherwise stated, the inner product under consideration is the dot product. Note that the norm of x is

$$\|x\| = \sqrt{\sum\nolimits_s x_s^2} \ .$$

The distance between x and y is

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

Often, as an alternative notation, the dot product is expressed using a dot, $x.y$ written instead of $\langle x \,|\, y \rangle$. That alternative notation can potentially be ambiguous because, in many different contexts, a dot is sometime used just to separate expressions that are being multiplied together, as in $2.3 = 6$. Still, if no multiplication operation on $\mathbb{R}^n$ is under consideration then, for $x, y \in \mathbb{R}^n$, the notation $x.y$ is unambiguous.

It is worth restating, in the special case $V = \mathbb{R}^n$, the two main results in the previous section.

Theorem 1.3: *Let x_1, ..., x_n, y_1, ..., y_n be real numbers. Then:*

(1) (Cauchy–Schwarz Inequality.) *We have* $\left(\sum_r x_r y_r\right)^2 \leq \left(\sum_s x_s^2\right)\left(\sum_t y_t^2\right)$.

(2) (Triangle Inequality.) *We have* $\sqrt{\sum_r (x_r + y_r)^2} \leq \sqrt{\sum_s x_s^2} + \sqrt{\sum_t y_t^2}$.

♣ ♡ ♢ ♠

We write $\mathbb{R}^{\mathbb{N}}$ to denote the set of infinite sequences $(x_0, x_1, ...)$ with each $x_m \in \mathbb{R}$. We shall be needing a little bit of real analysis. For $(x_0, ...) \in \mathbb{R}^{\mathbb{N}}$, the formal written expression "$\sum_{n \in \mathbb{N}} x_n$" is called the **summation** of the sequence $(x_0, ...)$. Usually, when discussing summations, the quote marks are omitted. The summation is said to **converge** provided there exists $s \in \mathbb{R}$ such that the difference between s and $\sum_{n \in [0,m]} x_n$ is arbitrarily small for sufficiently large m. In that case, s is unique, we call s the **sum** of $(x_0, ...)$ and we write $s = \sum_{n \in \mathbb{N}} x_n$. Thus, the sum $\sum_{n \in \mathbb{N}} x_n$, when it exists, is a real number. It is not to be confused with the summation $\sum_{n \in \mathbb{N}} x_n$, which always exists and is just something that can be written down.

In the language of limits, the summation $\sum_{n \in \mathbb{N}} x_n$ converges when $\sum_{m \in [0,n]} x_m$ has a limt as $n \to \infty$, in which case the sum is the limit

$$\sum_{n \in \mathbb{N}} x_n = \lim_{n \to \infty} \sum_{m \in [0,n]} x_m \ .$$

When there exists $b \in \mathbb{R}$ such that, for all $m \in \mathbb{N}$, we have $\sum_{n \in [0,m]} |x_n| < b$, the summation $\sum_{n \in \mathbb{N}} x_n$ is said to **converge absolutely**. A fundamental theorem in introductory real analysis asserts that any absolutely convergent summation is convergent.

We make $\mathbb{R}^{\mathbb{N}}$ become a real vector space with addition and scalar multiplication operations such that

$$(x_0, ...) + (y_0, ...) = (x_0 + y_0, ...) \qquad \lambda(x_0, ...) = (\lambda x_0, ...)$$

for $(x_0, ...), (y_0, ...) \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. We define

$$\ell(\mathbb{R}) = \left\{ (x_0, ...) \in \mathbb{R}^{\mathbb{N}} : \sum_{n \in \mathbb{N}} x_n^2 < \infty \right\} .$$

A little work will be needed before we can conclude that $\ell^2(\mathbb{R})$ is a subspace of $\mathbb{R}^{\mathbb{N}}$.

Theorem 1.4: *Let* $(x_0, ...), (y_0, ...) \in \ell^2(\mathbb{R})$*. Then:*

(1) (Cauchy–Schwarz Inequality.) *The summation* $\sum_{r \in \mathbb{N}} x_r y_r$ *converges and*

$$\left(\sum_{r \in \mathbb{N}} x_r y_r \right)^2 \leq \left(\sum_{s \in \mathbb{N}} x_s^2 \right) \left(\sum_{t \in \mathbb{N}} y_t^2 \right) .$$

(2) (Triangle Inequality.) *The summation* $\sum_{r \in \mathbb{N}} (x_r + y_r)^2$ *converges and*

$$\sqrt{\sum_{r \in \mathbb{N}} (x_r + y_r)^2} \leq \sqrt{\sum_{s \in \mathbb{N}} x_s^2} + \sqrt{\sum_{t \in \mathbb{N}} y_t^2} .$$

Proof: Applying Theorem 1.3 to the sequences $(|x_0|, ...)$ and $(|y_0|, ...)$, we deduce that the summation $\sum_{r \in \mathbb{N}} x_r y_r$ converges. Then applying Theorem 1.3 to the sequences $(x_0, ...)$ and $(y_0, ...)$, we deduce part (1). A similar argument yields part (2). $\square$

Plainly, the subset $\ell^2(\mathbb{R}) \subseteq \mathbb{R}^n$ is closed under scalar multiplication, we mean to say, given $\lambda \in \mathbb{R}$ and $x \in \ell^2(\mathbb{R})$, then $\lambda x \in \ell^2(\mathbb{R})$. Part (2) of the latest theorem tells us that $\ell^2(\mathbb{R})$ is closed under addition, we mean, given $x, y \in \ell^2(\mathbb{R})$, then $x + y \in \ell^2(\mathbb{R})$. Therefore, $\ell^2(\mathbb{R})$ is a subspace of $\mathbb{R}^{\mathbb{N}}$.

We make $\ell^2(\mathbb{R})$ become a real inner product space with inner product given by

$$\langle (x_0, ...), (y_0, ...) \rangle = \sum_{n \in \mathbb{N}} x_n y_n .$$

2: Orthonormal bases for real inner product spaces

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. We shall be discussing, for real inner product spaces, a special kind of basis with the virtue that the coefficients of the linear combination can be calculated in a different way.

♣ ♡ ♢ ♠

To avoid confusion over some fundamental notions that involve some subtleties in the case of infinite-dimensional vector spaces, let us make a careful review of some background. Even for real or complex vector spaces, infinite sums of vectors do not always make sense, indeed, they can only make sense under suitable topological constraints. Such matters lie within the realm of functional analysis, and they lie outside our scope.

Recall that, for any subset $\mathcal{S}$ of a vector space X over any field F, we define a **linear combination** of elements of $\mathcal{S}$ to be a sum having the form

$$\sum_{s \in S} \lambda_s s$$

where each $\lambda_s \in F$ and there are only finitely many s such that $\lambda_s \neq 0$. Thus, although $\mathcal{S}$ may be infinite, the sum makes sense because it can be viewed as the finite sum of the nonzero terms. When the condition

$$\sum_s \lambda_s s = 0$$

implies that each $\lambda_s = 0$, we call $\mathcal{S}$ **linearly independent**. The set of linear combinations of elements of $\mathcal{S}$, denoted $\mathrm{span}(\mathcal{S})$, is a subspace of X. When $\mathrm{span}(\mathcal{S}) = X$, we say call $\mathcal{S}$ a **spanning set** for X. When $\mathcal{S}$ is a linearly independent spanning set for X, we call $\mathcal{S}$ a **basis** for X.

In functional analysis and its applications, a basis for X, as defined above, is sometimes called an **algebraic basis**, in distinction from another concept, called a **topological basis**, for which infinite sums may be considered.

♣ ♡ ♢ ♠

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. We shall be discussing a special kind of basis where the coefficients of the linear combination can be calculated in a different way.

A subset $\mathcal{U}$ of a real inner product space V is said to be **orthogonal** provided $\langle u \,|\, u' \rangle = 0$ for all $u, u' \in S$ with $u \neq u'$. Obviously, any orthogonal subset of V is linearly independent. When $\mathcal{U}$ is an orthogonal subset of V and also a basis for V, we call $\mathcal{U}$ an **orthogonal basis** for V.

Remark 2.1: *Let V be a real inner product space with orthogonal basis $\mathcal{U}$. Given $x \in V$, then*

$$x = \sum_{u \in \mathcal{U}} \frac{\langle u \,|\, x \rangle}{\|u\|^2} \, u \, .$$

Proof: Since $\mathcal{U}$ spans V, we can write $x = \sum_u x_u u$ with each $x_u \in \mathbb{R}$, only finitely many of the x_u being nonzero. The required conclusion follows upon evaluating $\langle u \,|\, x \rangle$. □

An element $v \in V$ is called a **unit vector** in V provided $\|v\| = 1$. Note that, for any nonzero vector v in V, the vector $v/\|v\|$ is a unit vector. In casual language, the replacement of v with $v/\|v\|$ is called **normalization**.

A subset $\mathcal{W}$ of V is called **orthonormal** provided $\mathcal{W}$ is orthogonal and every element of $\mathcal{W}$ is a unit vector. When $\mathcal{W}$ is an orthonormal subset of V and also a basis for V, we call $\mathcal{W}$ an **orthonormal basis** for V.

Note that any orthogonal basis $\mathcal{V}$ for V gives rise to an orthonormal basis $\mathcal{W}$ for V given by

$$\mathcal{W} = \{v/\|v\| : v \in \mathcal{V}\} \ .$$

Thus, $\mathcal{W}$ is obtained from $\mathcal{V}$ by normalizing the elements. The next remark is just a special case of the previous one.

Remark 2.2: *Let V be a real inner product space with orthonormal basis $\mathcal{W}$. Given $x \in V$, then*

$$x = \sum_{w \in \mathcal{W}} \langle w \,|\, x \rangle w \ .$$

♣ ♡ ♢ ♠

Now let V be a finite-dimensional real inner product space. Let $n = \dim(V)$. Specializing a definition above, an orthogonal basis for V is a basis having the form $\{u_1, ..., u_n\}$ where $\langle u_s \,|\, u_t \rangle = 0$ for all $s, t \in [1, n]$ with $s \neq t$. Recall, the **Kronecker delta symbol** $\delta_{s,t}$ is defined to be the real number

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

An orthonormal basis for V is a basis having the form $\{w_1, ..., w_n\}$ where

$$\langle u_s \,|\, u_t \rangle = \delta_{s,t} \ .$$

As an example, the standard basis $\{e_1, ..., e_n\}$ of $\mathbb{R}^n$ satisfies

$$\langle e_s \,|\, e_t \rangle = \delta_{s,t}$$

for all $s, t \in [0, n]$. In other words, the standard basis of $\mathbb{R}^n$ is an orthonormal basis for $\mathbb{R}^n$.

For ease of reference, we write down the following specialization of Remarks 2.3 and 2.4.

Remark 2.3: *Let V be a finite-dimensional real inner product space and let $x \in V$. Then:*

(1) *For any orthogonal basis $\{u_1, ..., u_n\}$ of V, we have* $x = \displaystyle\sum_{s \in [1,n]} \frac{\langle u_s \,|\, x \rangle}{\|u_s\|^2} u_s$.

(2) *For any orthonormal basis $\{w_1, ..., w_n\}$ of V, we have* $x = \displaystyle\sum_{s \in [1,n]} \langle w_s \,|\, x \rangle w_s$.

The next result shows that, when using coordinates with respect to an orthonormal basis, the inner product behaves very much like the dot product.

Corollary 2.4: *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,* $d(x, y) = \sqrt{\sum_s (x_s - y_s)^2}$.

♣ ♡ ♢ ♠

We now describe a process, called the **Gram–Schmidt Process**, for replacing any given basis $\mathcal{U}$ of a finite-dimensional real inner product space with an orthogonal basis $\mathcal{V}$ and thence, if desired, an orthonormal basis $\mathcal{W}$.

Let V be a finite-dimensional real inner product space. 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 $2 \le r \le n$. We define a set $\mathcal{W} = \{w_1, ..., w_n\}$ where each w_r is the normalization of v_r. That is to say, $w_r = v_r / \|v_r\|$.

Proposition 2.5: *With the notation above, $\mathcal{V}$ is an orthogonal basis for V. Also, $\mathcal{W}$ is an orthonormal basis for V.*

Proof: For $2 \le s \le n$, 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 $1 \le t < s$. An inductive argument now yields the conclusion that $\mathcal{V}$ is an orthogonal basis for V. It follows immediately that $\mathcal{W}$ is an orthonormal basis for V. □

For any $x \in V$ with $x \ne 0$, we define the **normalization** of x to be the vector $x/\|x\|$. Note that the normalization $x' = x/\|x\|$ has norm $\|x'\| = 1$. We define $\mathcal{W} = \{w_1, ..., w_n\}$ where each w_r is the normalization of v_r. Thus,

$$w_r = v_r / \|v_r\| \;.$$

The previous remark immediately implies the next one.

Proposition 2.6: *With the notation above, $\mathcal{W}$ is an orthonormal basis for V.*

The latest proposition immediately yields the next theorem.

Theorem 2.7: *Any finite-dimensional real inner product space has an orthonormal basis.*

In the previous section, we promised to explain how any finite-dimensional real inner product space can be viewed as a copy of $\mathbb{R}^n$. We now fulfill that promise.

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

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

When there exists an isometry $U \leftarrow V$, we say that U is isometric to V.

Consider inner product spaces U, V, W. The identity operator on U is an isometry. So U is isometric to U. Inverses of isometries are isometries. So if U is isometric to V, then V

is isometric to U. Composites of isometries are isometries. So if U is isometric to V and if V is isometric to W, then U is isometric to W. Thus, isometry of real inner product spaces is a formal equivalence relation. Intuitively, when U is isometric to V, we understand that, as real inner product spaces, U and V have the same structure, in other words, U and V are copies of each other. The next result says that every finite-dimensional real inner product space is, in the above sense, a copy of $\mathbb{R}^n$ for some n.

Corollary 2.8: *Let V be a finite-dimensional real inner product space. Let $n = \dim(V)$. Then V is isometric to $\mathbb{R}^n$.*

Proof: By Theorem 2.7, V has an orthonormal basis $\mathcal{W} = \{w_1, ..., w_n\}$. Let $\{e_1, ..., e_n\}$ be the standard basis for $\mathbb{R}^n$. By Corollary 2.4, there is an isometry $U \leftarrow \mathbb{R}^n$ such that $w_s \hookleftarrow e_s$. $\square$

♣ ♡ ♢ ♠

The Gram–Schmidt process adapts in a straightforward way to the case of a real inner product space with an infinite enumerated basis $\mathcal{U} = \{u_1, u_2, ...\}$. Defining $\mathcal{V} = \{v_1, v_2, ...\}$ and $\mathcal{W}\{w_1, w_2, ...\}$ by the same formulas as before, the proof of Proposition 2.5 shows that $\mathcal{V}$ is an orthogonal set and $\mathcal{W}$ is an orthogonal basis. Noting that $\text{span}\{u_1, ...u_n\} = \text{span}\{v_1, ..., v_n\}$ for each n, we see that $\mathcal{V}$ is a basis for V, hence $\mathcal{W}$ is a basis. In conclusion, $\mathcal{V}$ is an orthogonal basis for V and $\mathcal{U}$ is an orthonormal basis for V.

Let us give another example of a real inner product space. Consider an infinite sequence a_0, a_1, ... of real numbers such that $a_n \neq 0$ for only finitely many natural numbers n. Let $f : \mathbb{R} \leftarrow \mathbb{R}$ be the function such that

$$f(t) = a_0 + a_1 t + a_2 t^2 + ... = \sum_{m \in \mathbb{N}} a_m t^m \ .$$

We call f the **real polynomial function** with **coefficients** a_0, a_1, ... Note that f determines the coefficients a_m, indeed,

$$a_m = \frac{1}{m!} f^{(m)}(0)$$

where $f^{(m)}$ denotes the m-th derivative of f. If $a_n \neq 0$ and $a_m = 0$ for all $m \geq n$, then we say that f has **degree** n. In that case,

$$f(t) = a_0 + a_1 t + ... + a_n t^n \ .$$

It is to be understood that if $f(t) = 0$, then f has degree -1.

Let $P(\mathbb{R})$ denote the set of real polynomial functions. We make $P(\mathbb{R})$ become a real vector space such that, given $f, g \in P(\mathbb{R})$ then $(f + g)(t) = f(t) + g(t)$ and, given $a \in \mathbb{R}$, then $(af)(t) = a(f(t)$. Thus, we can write $af(t)$ unambiguously. We make $P(\mathbb{R})$ become a real inner product space with inner product given by

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

Let I_0, I_1, ... be the real polynomial functions such that

$$I_n(t) = t^n \ .$$

Since a real polynomial function determines its coefficients, the set $\{I_0, I_1, ...\}$ is a basis for $P(\mathbb{R})$. But

$$\langle I_m \mid I_n \rangle = \int_{-1}^{1} I_m(t) I_n(t)\,\mathrm{d}t = \int_{-1}^{1} t^{m+n}\,\mathrm{d}t$$

$$= \left.\frac{t^{m+n+1}}{m+n+1}\right|_{t=-1}^{t=1} = \begin{cases} 2/(m+n+1) & \text{if } m+n \text{ is even,} \\ 0 & \text{if } m+n \text{ is odd.} \end{cases}$$

Evidently, $\{I_0, I_1, ...\}$ is not an orthogonal basis for $P(\mathbb{R})$.

For any nonzero real numbers a_0, a_1, ..., the set $\{a_0 I_0, a_1 I_1, ...\}$ is a basis for $P(\mathbb{R})$. in particular, defining

$$K_n(t) = \frac{1}{2^n} \binom{2n}{n} t^n$$

then $\{k_0, K_1, ...\}$ is a basis for $P(\mathbb{R})$. The **Legendre polynomial functions** L_1, L_1, ... are defined to be the real polynomial functions such that $\{L_0, L_1, ...\}$ is the orthogonal basis for $P(\mathbb{R})$ obtained from $\{K_0, K_1, ...\}$ by the Gram–Schmidt process without any normalization. We shall be investigating the Legendre polynomial functions in some of the exercises.

3: Orthogonal complements for real inner product spaces

Consider a subspace W of a real inner product space V. We shall define the notion of an orthogonal complement of W in V, and we shall show that, when V is finite-dimensional, there exists a unique orthogonal complement of W in V.

♣ ♡ ♢ ♠

Let us quickly review some background that pertains to vector spaces over arbitrary fields. Let X be a vector space over any field F. Let Y and Z be subspaces of X. Recall that the intersection $Y \cap Z$ and the sum

$$Y + Z = \{y + z : y \in Y, z \in Z\}$$

are subspaces of X. Given $y, y' \in Y$ and $z, z' \in Z$ such that $y + z = y' + z'$, then $y - y' = z' - z \in Y \cap Z$. Therefore, the following two conditions are equivalent:

- We have $Y \cap Z = \{0\}$.
- Every element of $Y + Z$ can be expressed uniquely in the form $y + z$ with $y \in Y$ and $z \in Z$. That is to say, given $x \in Y + Z$, then there exist unique y and unique $z \in Z$ such that $x = y + z$.

When those equivalent conditions hold, we write $Y \oplus Z = Y + Z$ and we call $Y \oplus Z$ the **direct sum** of Y and Z. When $X = Y \oplus Z$, we say that Y and Z are **complementary** to each other in V and we call each a **complement** of the other.

It can be shown that, for any subspace Y of any vector space X, there exists a complement of Y in X. The easy proof is by extending a basis of Y to a basis for X. We mention that, for arbitrary X, not necessarily finite-dimensional, a Zorn's Lemma argument is needed, beforehand, to show that any basis for Y can be extended to a basis for X.

Note that a complement of Y in X need not be unique. For example, if X is any 2-dimensional vector space and Y is any 1-dimensional subspace of X, then every 1-dimensional subspace of X distinct from Y is a complement of Y in X.

♣ ♡ ♢ ♠

We return to a consideration of a subspace W of a real inner product space V. We define the **annihilator** of W in V, denoted $W^\perp$, to be the set of elements $z \in V$ such that, for all $w \in W$, we have $\langle w \,|\, z \rangle = 0$. Obviously, $W^\perp$ is a subspace of V. Given $v \in W \cap W^\perp$, then $\langle v \,|\, v \rangle = 0$, hence $v = 0$. Therefore $W \cap W^\perp = \{0\}$ and we can form the direct sum $W \oplus W^\perp$ as a subspace of V.

When $V = W \oplus W^\perp$, we call $W^\perp$ the **orthogonal complement** of W in V. Thus, when the annihilator is a complement, we call it the orthogonal complement.

Lemma 3.1: *Given a subspace W of a finite-dimensional real inner product space V, then any orthonormal basis for W can be extended to an orthonormal basis for V.*

Proof: We must show that, any orthonormal basis $\{e_1, ..., e_p\}$ of W can be extended to an orthonormal basis $\{e_1, ..., e_p, e_{p+1}, ..., e_q\}$ for V. To obtain the elements e_{p+1}, ..., e_q, we extend to a basis for V, then apply the Gram-Schmidt process. □

Theorem 3.2: *Every subspace of a finite-dimensional real inner product space has an orthogonal complement.*

Proof: We are to show that, given a subspace W of a finite-dimensional real inner product space V, then $V = W \oplus W^{\perp}$. In the notation of the proof of the previous lemma, let $W^{\perp}$ is the span of $\{e_{p+1}, ..., e_q\}$. $\square$

In one of the exercises below, we show that the conclusion of the lemma and the conclusion of the theorem can fail for infinite-dimensional real vector spaces.

$\clubsuit$ $\heartsuit$ $\diamondsuit$ $\spadesuit$

For finite-dimensional real inner product spaces, we now give another characterization of the orthogonal complement of a subspace.

Let V be a finite-dimensional real inner product space. Let W be a subspace of V. We define the **orthogonal projection** to W from V to be the linear map $\pi : W \leftarrow V$ such that $\pi(w + z) = w$ for all $w \in W$ and $z \in W^{\perp}$. The next theorem says that, given $v \in V$, then $\pi(v)$ is the unique element of W that is of minimum distance from v.

Theorem 3.3: *Let π be the orthogonal projection to a subspace W of a real inner product space V. Given $v \in V$, then $d(\pi(v), v) \leq d(w, v)$ for all $w \in W$ with $w \neq \pi(v)$.*

Proof: Let $\{e_1, ..., e_m\}$ and $\{f_1, ..., f_n\}$ be orthonormal basis for W and $W^{\perp}$, respectively. Note that $\{e_1, ..., e_m, f_1, ..., f_n\}$ is a basis for V. Given $v \in V$ and $w \in W$, then $v = \sum_i u_i e_i + \sum_j v_j f_j$ and $w = \sum_i w_i e_i$ with each $u_i, v_j, w_i \in \mathbb{R}$. Since $\pi(v) = \sum_i u_i e_i$, we have

$$d(\pi(v), v) = \sum_j v_j^2 \leq \sum_i (u_i - w_i)^2 + \sum_j v_j^2 = d(w, v)$$

with equality if and only if $\pi(v) = w$. $\square$

A generalization of the latest theorem, without the assumption of finite dimension, appears in the exercises.

X: List of ommitted syllabus topics, Spring 2024

The following can be found in the course textbooks.

For MATH 220 and 224:

• Complex inner product spaces.

• Real symmetric and real orthogonal matrices.

For MATH 224:

• Symmetric, orthogonal, Hermitian, unitary operators and matrices, proofs of their various diagonalizability properties.

Procedural exercises

Not all of the background for the folowing exercises is covered in the completed sections above.

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: Directly from the definition of the Legendre polynomial functions L_n in Section 2, give explicit formulas for $L_0(t)$, $L_1(t)$, $L_2(t)$, $L_3(t)$.

1.C: Consider the matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} .$$

Find a real orthogonal matrix P and a real diagonal matrix D such that $A = PDP^{-1}$. (Hint: first find a basis of eigenvectors of A, then apply the Gram–Schmidt process to find an orthonormal basis of eigenvectors of A.)

Theoretical exercises

1.Z: 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.Y: 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.X: 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.)

1.W: Let V be a real inner product space with subspaces X and Y such that $V = X \oplus Y$. Let $\mathcal{U} = \{u_1, u_2, ...\}$ be a finite or countably infinite basis for V such that, for each index s, either $u_s \in X$ or $u_s \in Y$. Let $\mathcal{V}$ be the orthogonal basis obtained from $\mathcal{U}$ by the Gram–Schmidt process. Show that, for each s, if $u_s \in X$ then $v_s \in X$ whereas if $u_s \in Y$ then $v_s \in Y$.

1.V: A function $f : \mathbb{R} \leftarrow \mathbb{R}$ is said to be **even** provided $f(-t) = f(t)$ for all $t \in \mathbb{R}$, **odd** provided $f(-t) = -f(t)$ for all t. Directly from the definition of the Legendre polynomial functions $L_n$in Section 2, show that if n is even then L_n is even whereas if n is odd then L_n is odd.

1.U: In this question, you may assume the equality

$$\frac{1+m+n}{2^{m+n}} \sum_{a \in [0,m], b \in [0,n]} \binom{m}{a}^2 \binom{n}{b}^2 \int_{-1}^{1} (t-1)^{m+n-a-b}(t+1)^{a+b} \, \mathrm{d}t = \delta_{m,n}$$

for all natural numbers m and n.

(a) Directly from the above equality and the definition of the functions L_n in Section 2, show that

$$L_n(t) = \frac{1}{2^n} \sum_{b \in [0,n]} \binom{n}{b}^2 (t-1)^{n-b}(t+1)^b \ .$$

Hint: Consider the coefficient of t^n in the equality $(t+t)^{2n} = (t+1)^n(t+1)^n$.

(b) Evaluate $\|L_n\|^2$.

(c) Let f be a real polynomial function of degree at most n. Show that

$$f = \sum_{m \in [0,n]} b_m L_m$$

for some real numbers b_0, ..., b_n. Express b_m in terms of f and L_m.

Comment: I do not know of any way of proving the baroque equality above using just the basic techniques of calculus. The equality arises, eventually, after first using some more sophisticated techniques to establish the equivalence of various characterizations of the Legendre functions.

1.S: Let U be a subspace of a real inner product space V. Suppose that, for all $v \in V$, there exists $\pi(v) \in U$ such that $d(\pi(v), v)$ is minimal, in other words, $d(\pi(v), v) \leq d(u, v)$ for all $u \in U$.

(a) Show that $V = U \oplus U^{\perp}$.

(b) Show that π is a linear map.

Solutions to procedural exercises on Section 1

1.A: We have $w_1 = u_1/\|u_1\| = \dfrac{1}{\sqrt{6}}(2,1,1)$. 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 $w_2 = u_2/\|u_2\| = \dfrac{1}{\sqrt{5}}(-1,2,0)$. 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 $w_3 = v_3/\|v_3\| = (-2,-1,5)/\|(-2,-1,5)\| = \dfrac{1}{\sqrt{30}}(-2,-1,5)$. 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: We begin with a little observation that will simplify the calculations. Let $f \in P(\mathbb{R})$ and write $f(t) = \sum_m a_m t^m$ with $a_m \in \mathbb{R}$ and a_m nonzero for only finitely many m. We have $f(t) = -f(-t)$ if and only if $a_m = 0$ for all even m. When those equivalent conditions hold, we say that f is **odd**. In that case,

$$\int_{-1}^{1} f(t)\,\mathrm{d}t = 0\ .$$

We have $\dfrac{1}{2^0}\dbinom{0}{0} = 1$ and $\dfrac{1}{2^1}\dbinom{2}{1} = 1$ and $\dfrac{1}{2^2}\dbinom{4}{2} = \dfrac{3}{2}$ and $\dfrac{1}{2^3}\dbinom{6}{3} = \dfrac{5}{2}$. So

$$K_0(t) = 1\ , \qquad K_1(t) = t\ , \qquad K_2(t) = \frac{3}{2}\,t^2\ , \qquad K_3(t) = \frac{5}{2}\,t^3$$

for all $t \in \mathbb{R}$. We are to calculate

$$L_n = K_n - \frac{\langle L_0 \,|\, K_n\rangle}{\|L_0\|^2}\,L_0 - \ldots - \frac{\langle L_{n-1} \,|\, K_n\rangle}{\|L_{n-1}\|^2}\,L_{n-1}\ .$$

We have $L_0 = K_0$. Since the function $L_0(t)K_1(t) \leftarrow\!\shortmid t$ is odd, we have $\langle L_0 \,|\, K_1\rangle = 0$, hence $L_1 = K_1$. Since the function $L_1(t)K_2(t) \leftarrow\!\shortmid t$ is odd, we have $\langle L_1 \,|\, K_2\rangle = 0$ and

$$L_2 = K_2 - \frac{\langle L_0 \,|\, K_2\rangle}{\|L_0\|^2}\,L_0\ .$$

Now $\|L_0\|^2 = \displaystyle\int_{-1}^{1} \mathrm{d}t = 2$ and $\langle L_0 \,|\, K_2\rangle = \displaystyle\int_{-1}^{1} \frac{3}{2}\,t^2\,\mathrm{d}t = \frac{t^3}{2}\bigg|_{-1}^{1} = 1$. So

$$L_2 = K_2 - L_0/2\ .$$

Since the functions $L_0(t)K_3(t) \hookleftarrow t$ and $L_2(t)K_2(t) \hookleftarrow t$ are odd, we have

$$L_3 = K_3 - \frac{\langle L_1 \mid K_3 \rangle}{\|L_1\|^2} L_1 \ .$$

Noting that $\|L_1\|^2 = \int_{-1}^{1} t^2 \, \mathrm{d}t = \left. \frac{t^3}{3} \right|_{-1}^{1} = \frac{2}{3}$ and $\langle L_1 \mid K_3 \rangle = \int_{-1}^{1} \frac{5}{2} t^4 \, \mathrm{d}t = \left. \frac{t^5}{2} \right|_{-1}^{1} = 1$, we obtain

$$L_3 = K_3 - 3L_1/2 \ .$$

In conclusion, $L_0(t) = 1$ and $L_1(t) = t$ and $L_2(t) = (3t^2 - 1)/2$ and $L_3(t) = (5t^2 - 3)/2$.

Solutions to theoretical exercises on Section 1

1.$\mathcal{A}$:

1.$\mathcal{B}$: