Rob 501 Handouts

Linear Algebra and Geometry

J.W. Grizzle

Linear Algebra by Gabriel Nagy

You can download the full textbook at the link given on the next page. We have permission to use it. Professor Nagy warns us that there are some typos.

In the following, I have extracted 3 important chapters.

Sources:

- Linear Algebra
  `http://math.msu.edu/~gnagy/teaching/la.pdf`

Chapter 4. Vector spaces

4.1. Spaces and subspaces

A vector space is any set where the linear combination operation is defined on its elements. In a vector space, also called a linear space, the elements are not important. The actual elements that constitute the vector space are left unspecified, only the relation among them is determined. An example of a vector space is the set $\mathbb{F}^n$ of n-vectors with the operation of linear combination studied in Chapter 1. Another example is the set $\mathbb{F}^{m,n}$ of all $m \times n$ matrices with the operation of linear combination studied in Chapter 2. We now define a vector space and comment its main properties. A subspace is introduced later on as a smaller vector space inside the original one. We end this Section with the concept of span of a set of vectors, which is a way to construct a subspace from any subset in a vector space.

Definition 4.1.1. *A set V is a **vector space** over the scalar field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ iff there are two operations defined on V, called vector addition and scalar multiplication with the following properties: For all $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w} \in V$ the vector addition satisfies*

(A1) $\boldsymbol{u} + \boldsymbol{v} \in V$, *(closure of V);*
(A2) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$, *(commutativity);*
(A3) $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$, *(associativity);*
(A4) $\exists\ \boldsymbol{0} \in V\ :\ \boldsymbol{0} + \boldsymbol{u} = \boldsymbol{u}\quad \forall\ \boldsymbol{u} \in V$, *(existence of a neutral element);*
(A5) $\forall\ \boldsymbol{u} \in V\quad \exists\ (-\boldsymbol{u}) \in V\ :\ \boldsymbol{u} + (-\boldsymbol{u}) = \boldsymbol{0}$, *(existence of an opposite element);*

furthermore, for all a, $b \in \mathbb{F}$ the scalar multiplication satisfies

(M1) $a\boldsymbol{u} \in V$, *(closure of V);*
(M2) $1\,\boldsymbol{u} = \boldsymbol{u}$, *(neutral element);*
(M3) $a(b\boldsymbol{u}) = (ab)\boldsymbol{u}$, *(associativity);*
(M4) $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$, *(distributivity);*
(M5) $(a + b)\boldsymbol{u} = a\boldsymbol{u} + b\boldsymbol{u}$, *(distributivity).*

The definition of a vector space does not specify the elements of the set V, it only determines the properties of the vector addition and scalar multiplication operations. We use the convention that elements in a vector space, called vectors, are represented in boldface. Nevertheless, we allow several exceptions, the first two of them are given in Examples 4.1.1 and 4.1.2. We now present several examples of vector spaces.

Example 4.1.1: A vector space is the set $\mathbb{F}^n$ of n-vectors $\mathsf{u} = [u_i]$ with components $u_i \in \mathbb{F}$ and the operations of column vector addition and scalar multiplication given by

$$[u_i] + [v_i] = [u_i + v_i], \qquad a[u_i] = [au_i].$$

This space of column vectors was introduced in Chapter 1. Elements in these vector spaces are not represented in boldface, instead we keep the previous sanserif font, $\mathsf{u} \in \mathbb{F}^n$. The reason for this notation will be clear in Sect. 4.4. ◁

Example 4.1.2: A vector space is the set $\mathbb{F}^{m,n}$ of $m \times n$ matrices $\mathsf{A} = [A_{ij}]$ with matrix coefficients $A_{ij} \in \mathbb{F}$ and the operations of addition and scalar multiplication given by

$$[A_{ij}] + [B_{ij}] = [A_{ij} + B_{ij}], \qquad a[A_{ij}] = [aA_{ij}],$$

These operations were introduced in Chapter 2. As in the previous example, elements in these vector spaces are not represented in boldface, instead we keep the previous capital sanserif font, $\mathsf{A} \in \mathbb{F}^{m,n}$. The reason for this notation will be clear in Sect. 4.4. ◁

Example 4.1.3: Let $\mathbb{P}_n(U)$ be the set of all polynomials having degree $n \geqslant 0$ and domain $U \subset \mathbb{F}$, that is,

$$\mathbb{P}_n(U) = \{\boldsymbol{p}(x) = a_0 + a_1 x + \cdots + a_n x^n, \text{ with } a_0, \cdots, a_n \in \mathbb{F} \text{ and } x \in U \subset \mathbb{F}\}.$$

The set $\mathbb{P}_n(U)$ together with the addition of polynomials $(\boldsymbol{p}+\boldsymbol{q})(x) = \boldsymbol{p}(x) + \boldsymbol{q}(x)$ and the scalar multiplication $(a\boldsymbol{p})(x) = a\,\boldsymbol{p}(x)$ is a vector space. In the case $U = \mathbb{R}$ we use the notation $\mathbb{P}_n = \mathbb{P}_n(\mathbb{R})$. ◁

Example 4.1.4: Let $C^k\big([a,b],\mathbb{F}\big)$ be the set of scalar valued functions with domain $[a,b] \subset \mathbb{R}$ with the k-th derivative being a continuous function, that is,

$$C^k\big([a,b],\mathbb{F}\big) = \{\boldsymbol{f} : [a,b] \to \mathbb{F} \quad \text{such that} \quad \boldsymbol{f}^{(k)} \quad \text{is continuous}\}.$$

The set $C^k\big([a,b],\mathbb{F}\big)$ together with the addition of functions $(\boldsymbol{f}+\boldsymbol{g})(x) = \boldsymbol{f}(x) + \boldsymbol{g}(x)$ and the scalar multiplication $(a\boldsymbol{f})(x) = a\,\boldsymbol{f}(x)$ is a vector space. The particular case $C^k(\mathbb{R},\mathbb{R})$ is denoted simply as C^k. The set of real valued continuous function is then C^0. ◁

Example 4.1.5: Let ℓ be the set of absolute convergent series, that is,

$$\ell = \Big\{\boldsymbol{a} = \sum a_n \;:\; a_n \in \mathbb{F} \text{ and } \sum_{n=0}^{\infty} |a_n| \text{ exists}\Big\}.$$

The set ℓ with the addition of series $\boldsymbol{a} + \boldsymbol{b} = \sum (a_n + b_n)$ and the scalar multiplication $c\boldsymbol{a} = \sum c\, a_n$ is a vector space. ◁

The properties (A1)-(M5) given in the definition of vector space are not redundant. As an example, these properties do not include the condition that the neutral element $\boldsymbol{0}$ is unique, since it follows from the definition.

Theorem 4.1.2. *The element $\boldsymbol{0}$ in a vector space is unique.*

Proof Theorem 4.1.2: Suppose that there exist two neutral elements $\boldsymbol{0}_1$ and $\boldsymbol{0}_2$ in the vector space V, that is,

$$\boldsymbol{0}_1 + \boldsymbol{u} = \boldsymbol{u} \quad \text{and} \quad \boldsymbol{0}_2 + \boldsymbol{u} = \boldsymbol{u} \quad \text{for all} \quad \boldsymbol{u} \in V$$

Taking $\boldsymbol{u} = \boldsymbol{0}_2$ in the first equation above, and $\boldsymbol{u} = \boldsymbol{0}_1$ in the second equation above we obtain that

$$\boldsymbol{0}_1 + \boldsymbol{0}_2 = \boldsymbol{0}_2, \qquad \boldsymbol{0}_2 + \boldsymbol{0}_1 = \boldsymbol{0}_1.$$

These equations above simply that the two neutral elements must be the same, since

$$\boldsymbol{0}_2 = \boldsymbol{0}_1 + \boldsymbol{0}_2 = \boldsymbol{0}_2 + \boldsymbol{0}_1 = \boldsymbol{0}_1;$$

where in the second equation we used that the addition operation is commutative. This establishes the Theorem. □

Theorem 4.1.3. *It holds that $0\,\boldsymbol{u} = \boldsymbol{0}$ for all element $\boldsymbol{u}$ in a vector space V.*

Proof Theorem 4.1.3: For every $\boldsymbol{u} \in V$ holds

$$\boldsymbol{u} = 1\,\boldsymbol{u} = (1+0)\boldsymbol{u} = 1\,\boldsymbol{u} + 0\,\boldsymbol{u} = \boldsymbol{u} + 0\,\boldsymbol{u} = 0\,\boldsymbol{u} + \boldsymbol{u} \quad \Rightarrow \quad \boldsymbol{u} = 0\,\boldsymbol{u} + \boldsymbol{u}.$$

This last equation says that $0\boldsymbol{u}$ is a neutral element, $\boldsymbol{0}$. Theorem 4.1.2 says that the neutral element is unique, so we conclude that, for all $\boldsymbol{u} \in V$ holds that

$$0\,\boldsymbol{u} = \boldsymbol{0}.$$

This establishes the Theorem. □

Also notice that the property (A5) in the definition of vector space says that the opposite element exists, but it does not say whether it is unique. The opposite element is actually unique.

Theorem 4.1.4. *The opposite element* $-\boldsymbol{u}$ *in a vector space is unique.*

Proof Theorem 4.1.4: Suppose there are two opposite elements $-\boldsymbol{u}_1$ and $-\boldsymbol{u}_2$ to the element $\boldsymbol{u} \in V$, that is,

$$\boldsymbol{u} + (-\boldsymbol{u}_1) = \boldsymbol{0}, \qquad \boldsymbol{u} + (-\boldsymbol{u}_2) = \boldsymbol{0}.$$

Therefore,

$$\begin{aligned}(-\boldsymbol{u}_1) &= \boldsymbol{0} + (-\boldsymbol{u}_1)\\ &= u + (-\boldsymbol{u}_2) + (-\boldsymbol{u}_1)\\ &= (-\boldsymbol{u}_2) + \boldsymbol{u} + (-\boldsymbol{u}_1)\\ &= (-\boldsymbol{u}_2) + \boldsymbol{0}\\ &= \boldsymbol{0} + (-\boldsymbol{u}_2)\\ &= (-\boldsymbol{u}_2) \quad \Rightarrow \quad (-\boldsymbol{u}_1) = (-\boldsymbol{u}_2).\end{aligned}$$

This establishes the Theorem. □

Finally, notice that the element $(-\boldsymbol{u})$ opposite to $\boldsymbol{u}$ is actually the element $(-1)\boldsymbol{u}$.

Theorem 4.1.5. *It holds that* $(-1)\,\boldsymbol{u} = (-\boldsymbol{u})$.

Proof Theorem 4.1.5:

$$\boldsymbol{0} = 0\,\boldsymbol{u} = (1-1)\,\boldsymbol{u} = 1\,\boldsymbol{u} + (-1)\,\boldsymbol{u} = \boldsymbol{u} + (-1)\,\boldsymbol{u}.$$

Hence $(-1)\,\boldsymbol{u}$ is an opposite element of $\boldsymbol{u}$. Since Theorem 4.1.4 says that the opposite element is unique, we conclude that $(-1)\,\boldsymbol{u} = (-\boldsymbol{u})$. This establishes the Theorem. □

4.1.1. **Subspaces.** We now introduce the notion of a subspace of a vector space, which is essentially a smaller vector space inside the original vector space. Subspaces are important in physics, since physical processes frequently take place not inside the whole space but in a particular subspace. For instance, planetary motion does not take plane in the whole space but it is confined to a plane.

Definition 4.1.6. *The subset* $W \subset V$ *of a vector space* V *over* $\mathbb{F}$ *is called a* ***subspace*** *of* V *iff for all* $\boldsymbol{u}$, $\boldsymbol{v} \in W$ *and all* a, $b \in \mathbb{F}$ *holds that* $a\boldsymbol{u} + b\boldsymbol{v} \in W$.

A subspace is a particular type of set in a vector space. Is a set where all possible linear combinations of two elements in the set results in another element in the same set. In other words, elements outside the set cannot be reached by linear combinations of elements within the set. For this reason a subspace is called a *closed set under linear combinations*. The following statement is usually helpful

Theorem 4.1.7. *If* $W \subset V$ *is a subspace of a vector space* V*, then* $\boldsymbol{0} \in W$.

This statement says that $\boldsymbol{0} \notin W$ implies that W *is not a subspace*. However, if actually $\boldsymbol{0} \in W$, this fact alone does not prove that W is a subspace. One must show that W is closed under linear combinations.
Proof of Theorem 4.1.7: Since W is closed under linear combinations, given any element $\boldsymbol{u} \in W$, the trivial linear combination $0\,\boldsymbol{u} = \boldsymbol{0}$ must belong to W, hence $\boldsymbol{0} \in W$. This establishes the Theorem. □

EXAMPLE 4.1.6: Show that the set $W \subset \mathbb{R}^3$ given by $W = \{\mathsf{u} = [u_i] \in \mathbb{R}^3 \ : \ u_3 = 0\}$ is a subspace of $\mathbb{R}^3$:

Solution: Given two arbitrary elements u, $\mathsf{v} \in W$ we must show that $a\mathsf{u} + b\mathsf{v} \in W$ for all a, $b \in \mathbb{R}$. Since u, $\mathsf{v} \in W$ we know that

$$\mathsf{u} = \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix}, \quad \mathsf{v} = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix}.$$

Therefore

$$a\mathsf{u} + b\mathsf{v} = \begin{bmatrix} au_1 + bv_1 \\ au_2 + bv_2 \\ 0 \end{bmatrix} \in W,$$

since the third component vanishes, which makes the linear combination an element in W. Hence, W is a subspace of $\mathbb{R}^3$. In Fig. 29 we see the plane $u_3 = 0$. It is a subspace, since not only $\mathsf{0} \in W$, but any linear combination of vectors on the plane stays on the plane. ◁

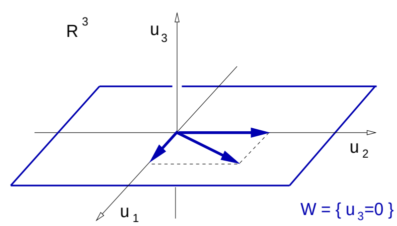

Figure 29. The horizontal plane $u_3 = 0$ is a subspace of $\mathbb{R}^3$.

Example 4.1.7: Show that the set $W = \big\{\mathsf{u} = [u_i] \in \mathbb{R}^2 \ : \ u_2 = 1\big\}$ is not a subspace of $\mathbb{R}^2$.

Solution: The set W is not a subspace, since $\mathsf{0} \notin W$. This is enough to show that W is not a subspace. Another proof is that the addition of two vectors in the set is a vector outside the set, as can be seen by the following calculation,

$$\mathsf{u} = \begin{bmatrix} u_1 \\ 1 \end{bmatrix} \in W, \quad \mathsf{v} = \begin{bmatrix} v_1 \\ 1 \end{bmatrix} \in W \quad \Rightarrow \quad \mathsf{u} + \mathsf{v} = \begin{bmatrix} u_1 + v_1 \\ 2 \end{bmatrix} \notin W.$$

The second component in the addition above is 2, not 1, hence this addition does not belong to W. An example of this calculation is given in Fig. 30. ◁

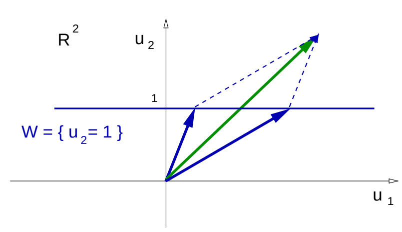

Figure 30. The horizontal line $u_2 = 1$ is not a subspace of $\mathbb{R}^2$.

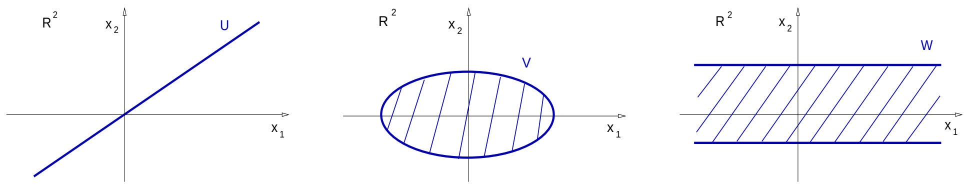

FIGURE 31. Three subsets, U, V, and W, of $\mathbb{R}^2$. Only the set U is a subspace.

EXAMPLE 4.1.8: Determine which one of the sets given in Fig. 31 is a subspace of $\mathbb{R}^2$.

SOLUTION: The set U is a vector space, since any linear combination of vectors parallel to the line is again a vector parallel to the line. The sets V and W are not subspaces, since given a vector $\boldsymbol{u}$ in these spaces, a the vector $a\boldsymbol{u}$ does not belong to these sets for a number $a \in \mathbb{R}$ big enough. This argument is sketched in Fig. 32.

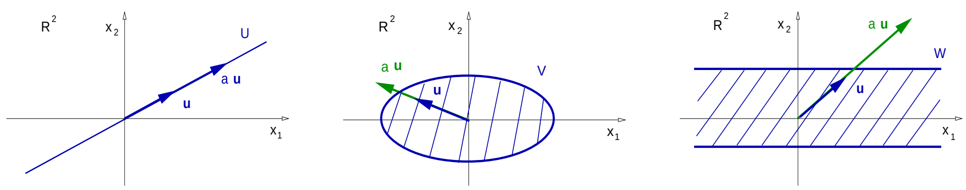

FIGURE 32. Three subsets, U, V, and W, of $\mathbb{R}^2$. Only the set U is a subspace.

$\lhd$

4.1.2. **The span of finite sets.** If a set is not a subspace there is a way to increase it into a subspace. Define a new set including all possible linear combinations of elements in the old set.

Definition 4.1.8. *The* ***span*** *of a finite set* $S = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$ *in a vector space* V *over* $\mathbb{F}$*, denoted as* $\mathrm{Span}(S)$*, is the set given by*

$$\mathrm{Span}(S) = \{\boldsymbol{u} \in V \ : \ \boldsymbol{u} = c_1\boldsymbol{u}_1 + \cdots + c_n\boldsymbol{u}_n, \quad c_1, \cdots, c_n \in \mathbb{F}\}.$$

The following result remarks that the span of a set is a subspace.

Theorem 4.1.9. *Given a finite set* S *in a vector space* V*,* $\mathrm{Span}(S)$ *is a subspace of* V*.*

Proof of Theorem 4.1.9: Since $\mathrm{Span}(S)$ contains all possible linear combinations of the elements in S, then $\mathrm{Span}(S)$ is closed under linear combinations. This establishes the Theorem. $\square$

EXAMPLE 4.1.9: The subspace $\mathrm{Span}(\{\boldsymbol{v}\})$, that is, the set of all possible linear combinations of the vector $\boldsymbol{v}$, is formed by all vectors of the form $c\boldsymbol{v}$, and these vectors belong to a line containing $\boldsymbol{v}$. The subspace $\mathrm{Span}(\{\boldsymbol{v}, \boldsymbol{w}\})$, that is, the set of all linear combinations of two vectors $\boldsymbol{v}$, $\boldsymbol{w}$, is the plane containing both vectors $\boldsymbol{v}$ and $\boldsymbol{w}$. See Fig. 33 for the case of the vector spaces $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. $\lhd$

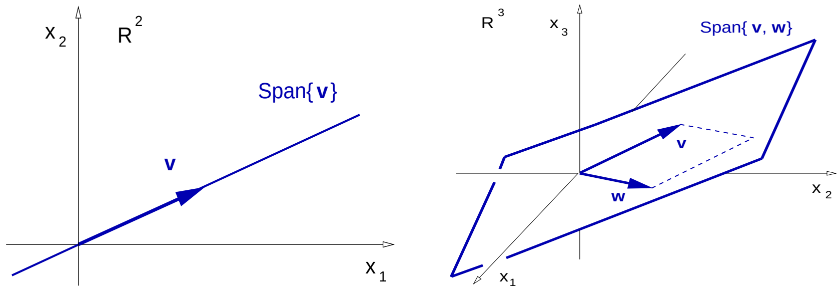

FIGURE 33. Examples of the span of a set of a single vector, and the span of a linearly independent set of two vectors.

4.1.3. **Algebra of subspaces.** We now show that the intersection of two subspaces is again a subspace. However, the union of two subspaces is not, in general, a subspace. The smaller subspace containing the union of two subspaces is precisely the span of the union. We then define the addition of two subspaces as the span of the union of two subspaces.

Theorem 4.1.10. *If W_1 and W_2 are subspaces of a vector space V, then $W_1 \cap W_2 \subset V$ is also a subspace of V.*

Proof of Theorem 4.1.10: Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be any two elements in $W_1 \cap W_2$. This means that $\boldsymbol{u}$, $\boldsymbol{v} \in W_1$, which is a subspace, so any linear combination $(a\boldsymbol{u} + b\boldsymbol{v}) \in W_1$. Since $\boldsymbol{u}$, $\boldsymbol{v}$ belong to $W_1 \cap W_2$ they also belong to W_2, which is a subspace, so any linear combination $(a\boldsymbol{u} + b\boldsymbol{v}) \in W_2$. Therefore, any linear combination $(a\boldsymbol{u} + b\boldsymbol{v}) \in W_1 \cap W_2$. This establishes the Theorem. □

EXAMPLE 4.1.10: The sketch in Fig. 34 shows the intersection of two subspaces in $\mathbb{R}^3$, a plane and a line. In this case the intersection is the former line, so the intersection is a subspace.

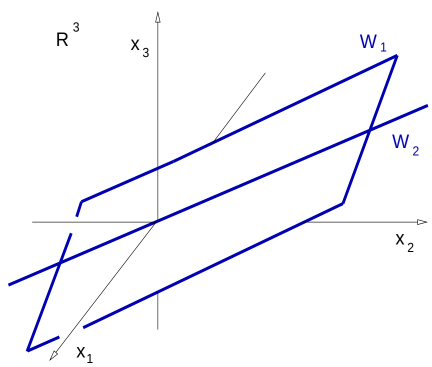

FIGURE 34. Intersection of two subspaces, W_1 and W_2 in $\mathbb{R}^3$. Since the line W_2 is included into the plane W_1, we have that $W_1 \cap W_2 = W_2$.

◁

While the intersection of two subspaces is always a subspace, their union is, in general, not a subspace, unless one subspace is contained into the other.

EXAMPLE 4.1.11: Consider the vector space $V = \mathbb{R}^2$, and the subspaces W_1 and W_2 given by the lines sketched in Fig. 35. Their union is the set formed by these two lines. This set is not a subspace, since the addition of the vectors $\boldsymbol{u}_1 \in W_1$ with $\boldsymbol{u}_2 \in W_2$ does not belong to $W_1 \cup W_2$, as is it shown in Fig. 35.

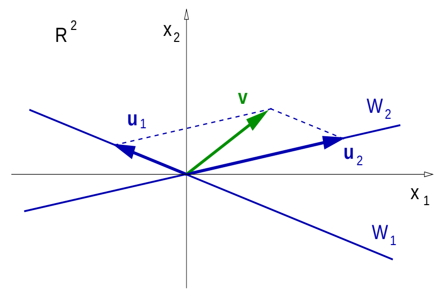

FIGURE 35. The union of the subspaces W_1 and W_2 is the set formed by these two lines. This is not a subspace, since the addition of $\boldsymbol{u}_1 \in W_1$ and $\boldsymbol{u}_2 \in W_2$ is the vector $\boldsymbol{v}$ which does not belong to $W_1 \cup W_2$.

$\lhd$

Although the union of two subspaces is not always a subspace, it is possible to enlarge the union into a subspace. The idea is to incorporate all possible additions of vectors in the two original subspaces, and the result is called the addition of the two subspaces. Here is the precise definition.

Definition 4.1.11. *The* ***addition of the subspaces*** W_1, W_2 *in a vector space* V, *denoted as* $W_1 + W_2$, *is the set given by*

$$W_1 + W_2 = \big\{\boldsymbol{u} \in V \ : \ \boldsymbol{u} = \boldsymbol{w}_1 + \boldsymbol{w}_2 \text{ with } \boldsymbol{w}_1 \in W_1, \ \boldsymbol{w}_2 \in W_2\big\}.$$

The following result remarks that the addition of subspaces is again a subspace.

Theorem 4.1.12. *If* W_1 *and* W_2 *are subspaces of a vector space* V, *then the addition* $W_1 + W_2$ *is also a subspace of* V.

Proof of Theorem 4.1.12: Suppose that $\boldsymbol{x} \in W_1 + W_2$ and $\boldsymbol{y} \in W_1 + W_2$. We must show that any linear combination $a\boldsymbol{x} + b\boldsymbol{y}$ also belongs to $W_1 + W_2$. This is the case, by the following argument. Since $\boldsymbol{x} \in W_1 + W_2$, there exist $\boldsymbol{x}_1 \in W_1$ and $\boldsymbol{x}_2 \in W_2$ such that $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$. Analogously, since $\boldsymbol{y} \in W_1 + W_2$, there exist $\boldsymbol{y}_1 \in W_1$ and $\boldsymbol{y}_2 \in W_2$ such that $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2$. Now any linear combination of $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfies

$$\begin{aligned} a\boldsymbol{x} + b\boldsymbol{y} &= a(\boldsymbol{x}_1 + \boldsymbol{x}_2) + b(\boldsymbol{y}_1 + \boldsymbol{y}_2) \\ &= (a\boldsymbol{x}_1 + b\boldsymbol{y}_1) + (a\boldsymbol{x}_2 + b\boldsymbol{y}_2) \end{aligned}$$

Since W_1 and W_2 are subspaces, $(a\boldsymbol{x}_1 + b\boldsymbol{y}_2) \in W_1$, and $(a\boldsymbol{x}_2 + b\boldsymbol{y}_2) \in W_2$. Therefore, the equation above says that $(a\boldsymbol{x} + b\boldsymbol{y}) \in W_1 + W_2$. This establishes the Theorem. $\square$

EXAMPLE 4.1.12: The sketch in Fig. 36 shows the union and the addition of two subspaces in $\mathbb{R}^3$, each subspace given by a line through the origin. While the union is not a subspace, their addition is the plane containing both lines, which is a subspace. Given any non-zero vector $\boldsymbol{w}_1 \in W_1$ and any other non-zero vector $\boldsymbol{w}_2 \in W_2$, one can verify that the sum of two subspaces is the span of $\big\{\boldsymbol{w}_1, \boldsymbol{w}_2\big\}$, that is,

$$W_1 + W_2 = \text{Span}\big(\{\boldsymbol{w}_1\} \cup \{\boldsymbol{w}_2\}\big).$$

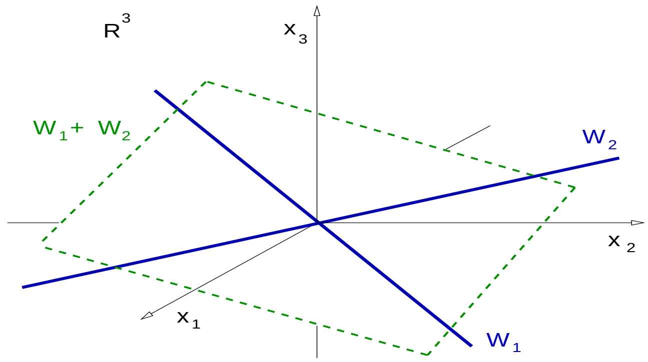

Figure 36. Union and addition of the subspaces W_1 and W_2 in $\mathbb{R}^3$. The union is not a subspace, while the addition is a subspace of $\mathbb{R}^3$.

4.1.4. **Internal direct sums.** This is a particular case of the addition of subspaces. It is called internal direct sum in order to differentiate it from another type of direct sum found in the literature. The latter, also called external direct sum, is a sum of different vector spaces, and it is a way to construct new vector spaces from old ones. We do not discuss this type of direct sums here. From now on, direct sum in these notes means the internal direct sum of subspaces inside a vector space.

Definition 4.1.13. *Given a vector space V, we say that V is the **internal direct sum** of the subspaces W_1, $W_2 \subset V$, denoted as $V = W_1 \oplus W_2$, iff every vector $\boldsymbol{v} \in V$ can be written in a unique way, except for order, as a sum of vectors from W_1 and W_2.*

A crucial part in the definition above is the uniqueness of the decomposition of every vector $\boldsymbol{v} \in V$ as a sum of a vector in W_1 plus a vector in W_2. By uniqueness we mean the following: For every $\boldsymbol{v} \in V$ exist $\boldsymbol{w}_1 \in \mathbb{W}_1$ and $\boldsymbol{w}_2 \in W_2$ such that $\boldsymbol{v} = \boldsymbol{w}_1 + \boldsymbol{w}_2$, and if $\boldsymbol{v} = \tilde{\boldsymbol{w}}_1 + \tilde{\boldsymbol{w}}_2$ with $\tilde{\boldsymbol{w}}_1 \in W_1$ and $\tilde{\boldsymbol{w}}_2 \in W_2$, then $\boldsymbol{w}_1 = \tilde{\boldsymbol{w}}_1$ and $\boldsymbol{w}_2 = \tilde{\boldsymbol{w}}_2$. In the case that $V = W_1 \oplus W_2$ we say that W_1 and W_2 are *direct summands* of V, and we also say that W_1 is the *direct complement* of W_2 in V. There is an useful characterization of internal direct sums.

Theorem 4.1.14. *A vector space V is the direct sum of subspaces W_1 and W_2 iff holds both $V = W_1 + W_2$ and $W_1 \cap W_2 = \{\boldsymbol{0}\}$.*

Proof of Theorem 4.1.14:

($\Rightarrow$) If $V = W_1 \oplus W_2$, then it implies that $V = W_1 + W_2$. Suppose that $\boldsymbol{v} \in W_1 \cap W_2$, then on the one hand, there exists $\boldsymbol{w}_1 \in W_1$ such that $\boldsymbol{v} = \boldsymbol{w}_1 + \boldsymbol{0}$; on the other hand, there is $\boldsymbol{w}_2 \in W_2$ such that $\boldsymbol{v} = \boldsymbol{0} + \boldsymbol{w}_2$. Therefore, $\boldsymbol{w}_1 = \boldsymbol{0}$ and $\boldsymbol{w}_2 = \boldsymbol{0}$, so $W_1 \cap W_2 = \{\boldsymbol{0}\}$.

($\Leftarrow$) Since $V = W_1 + W_2$, for every $\boldsymbol{v} \in V$ there exist $\boldsymbol{w}_1 \in W_1$ and $\boldsymbol{w}_2 \in W_2$ such that $\boldsymbol{v} = \boldsymbol{w}_1 + \boldsymbol{w}_2$. Suppose there exists other vectors $\tilde{\boldsymbol{w}}_1 \in W_1$ and $\tilde{\boldsymbol{w}}_2 \in W_2$ such that $\boldsymbol{v} = \tilde{\boldsymbol{w}}_1 + \tilde{\boldsymbol{w}}_2$. Then,

$$\boldsymbol{0} = (\boldsymbol{w}_1 - \tilde{\boldsymbol{w}}_1) + (\boldsymbol{w}_2 - \tilde{\boldsymbol{w}}_2) \quad \Leftrightarrow \quad (\boldsymbol{w}_1 - \tilde{\boldsymbol{w}}_1) = -(\boldsymbol{w}_2 - \tilde{\boldsymbol{w}}_2),$$

Therefore $(\boldsymbol{w}_1 - \tilde{\boldsymbol{w}}_1) \in W_2$ and so $(\boldsymbol{w}_1 - \tilde{\boldsymbol{w}}_1) \in W_1 \cap W_2$. Since $W_1 \cap W_2 = \{\boldsymbol{0}\}$, we then conclude that $\boldsymbol{w}_1 = \tilde{\boldsymbol{w}}_1$, which also says $\boldsymbol{w}_2 = \tilde{\boldsymbol{w}}_2$. Then $V = W_1 \oplus W_2$. This establishes the Theorem. $\square$

EXAMPLE 4.1.13: Denote by Sym and SkewSym the sets of all symmetric and all skew-symmetric $n \times n$ matrices. Show that $\mathbb{F}^{n,n} = \text{Sym} \oplus \text{SkewSym}$.

SOLUTION: Given any matrix $\mathsf{A} \in \mathbb{F}^{n,n}$, holds

$$\mathsf{A} = \mathsf{A} + \frac{1}{2}\big(\mathsf{A}^T - \mathsf{A}^T\big) = \frac{1}{2}\big(\mathsf{A} + \mathsf{A}^T\big) + \frac{1}{2}\big(\mathsf{A} - \mathsf{A}^T\big).$$

We then can decompose matrix as $\mathsf{A} = \mathsf{B} + \mathsf{C}$, where matrix $\mathsf{B} = \big(\mathsf{A} + \mathsf{A}^T\big)/2 \in \text{Sym}$ while matrix $\mathsf{C} = \big(\mathsf{A} - \mathsf{A}^T\big)/2 \in \text{SkewSym}$. That is, we can write any square matrix as a symmetric matrix plus a skew-symmetric matrix, hence $\mathbb{F}^{n,n} \subset \text{Sym} + \text{SkewSym}$. The other inclusion is obvious, that is, $\text{Sym} + \text{SkewSym} \subset \mathbb{F}^{n,n}$, because each term in the sum is a subset of $\mathbb{F}^{n,n}$. So, we conclude that

$$\mathbb{F}^{n,n} = \text{Sym} + \text{SkewSym}.$$

Now we must show that $\text{Sym} \cap \text{SkewSym} = \{\boldsymbol{0}\}$. This is the case, as the following argument shows. If matrix $\mathsf{D} \in \text{Sym} \cap \text{SkewSym}$, then matrix D is symmetric, $\mathsf{D} = \mathsf{D}^T$, but matrix D is also skew-symmetric, $\mathsf{D} = -\mathsf{D}^T$. This implies that $\mathsf{D} = -\mathsf{D}$, that is, $\mathsf{D} = \mathsf{0}$, proving our assertion that $\text{Sym} \cap \text{SkewSym} = \{\boldsymbol{0}\}$. We then conclude that

$$\mathbb{F}^{n,n} = \text{Sym} \oplus \text{SkewSym}.$$

$\triangleleft$

4.1.5. **Exercises.**

4.1.1.- Determine which of the following subsets of $\mathbb{R}^n$, with $n \geqslant 1$, are in fact subspaces. Justify your answers.

(a) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_i \geqslant 0 \quad i = 1, \cdots, n\}$;
(b) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_1 = 0\}$;
(c) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_1 x_2 = 0 \quad n \geqslant 2\}$;
(d) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_1 + \cdots + x_n = 0\}$;
(e) $\{\mathsf{x} \in \mathbb{R}^n \ : \ x_1 + \cdots + x_n = 1\}$;
(f) $\{\mathsf{x} \in \mathbb{R}^n \ : \ \mathsf{A}\mathsf{x} = \mathsf{b},\ \mathsf{A} \neq \mathsf{0},\ \mathsf{b} \neq \mathsf{0}\}$.

4.1.2.- Determine which of the following subsets of $\mathbb{F}^{n,n}$, with $n \geqslant 1$, are in fact subspaces. Justify your answers.

(a) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A} = \mathsf{A}^T\}$;
(b) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A}$ invertible$\}$;
(c) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A}$ not invertible$\}$;
(d) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A}$ upper-triangular$\}$;
(e) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathsf{A}^2 = \mathsf{A}\}$;
(f) $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ \mathrm{tr}(\mathsf{A}) = 0\}$.
(g) Given a matrix $\mathsf{X} \in \mathbb{F}^{n,n}$, define $\{\mathsf{A} \in \mathbb{F}^{n,n} \ : \ [\mathsf{A}, \mathsf{X}] = \mathsf{0}\}$.

4.1.3.- Find $W_1 + W_2 \subset \mathbb{R}^3$, where W_1 is a plane passing through the origin in $\mathbb{R}^3$ and W_2 is a line passing through the origin in $\mathbb{R}^3$ not contained in W_1.

4.1.4.- Sketch a picture of the subspaces spanned by the following vectors:

(a) $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}, \begin{bmatrix} -2 \\ -4 \\ -6 \end{bmatrix} \right\}$;

(b) $\left\{ \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \right\}$;

(c) $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

4.1.5.- Given two finite subsets S_1, S_2 in a vector space V, show that

$$\mathrm{Span}(S_1 \cup S_2) = \mathrm{Span}(S_1) + \mathrm{Span}(S_2).$$

4.1.6.- Let $W_1 \subset \mathbb{R}^3$ be the subspace

$$W_1 = \mathrm{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\} \right).$$

Find a subspace $W_2 \subset \mathbb{R}^3$ such that $\mathbb{R}^3 = W_1 \oplus W_2$.

4.2. Linear dependence

4.2.1. **Linearly dependent sets.** In this Section we present the notion of a linearly dependent set of vectors. If one of the vectors in the set is a linear combination of the other vectors in the set, then the set is called linearly dependent. If this is not the case, the set is called linearly independent. This notion plays a crucial role in Sect. 4.3 to define a basis of a vector space. Bases are very useful in part because every vector in the vector space can be decomposed in a unique way as a linear combination of the basis elements. Bases also provide a precise way to measure the size of a vector space.

Definition 4.2.1. *A finite set of vectors $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ in a vector space is called* ***linearly dependent*** *iff there exists a set of scalars $\{c_1, \cdots, c_k\}$, not all of them zero, such that,*

$$c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = \boldsymbol{0}. \tag{4.1}$$

On the other hand, the set $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ is called ***linearly independent*** *iff Eq. (4.1) implies that every scalar vanishes, $c_1 = \cdots = c_k = 0$.*

The wording in this definition is carefully chosen to cover the case of the empty set. The result is that the empty set is linearly independent. It might seems strange, but this result fits well with the rest of the theory. On the other hand, the set $\{\boldsymbol{0}\}$ is linearly dependent, since $c_1\boldsymbol{0} = \boldsymbol{0}$ for any $c_1 \neq 0$. Moreover, any set containing the zero vector is also linearly dependent.

Linear dependence or independence are properties of a set of vectors. There is no meaning to a vector to be linearly dependent, or independent. And there is no meaning of a set of linearly dependent vectors, as well as a set of linearly independent vectors. What is meaningful is to talk of a linearly dependent or independent set of vectors.

Example 4.2.1: Show that the set $S \subset \mathbb{R}^2$ below is linearly dependent,

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}.$$

Solution: It is clear that

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \Rightarrow \quad 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since $c_1 = 2$, $c_2 = 3$, and $c_3 = -1$ are non-zero, the set S is linearly dependent. ◁

It will be convenient to have the concept of a linearly dependent or independent set containing infinitely many vectors.

Definition 4.2.2. *An infinite set of vectors $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots\}$ in a vector space V is called* ***linearly independent*** *iff every finite subset of S is linearly independent. Otherwise, the infinite set S is called linearly dependent.*

Example 4.2.2: Consider the vector space $V = C^\infty([-\ell, \ell], \mathbb{R})$, that is, the space of infinitely differentiable real-valued functions defined on the domain $[-\ell, \ell] \subset \mathbb{R}$ with the usual operation of linear combination of functions. This vector space contains linearly independent sets with infinitely many vectors. One example is the infinite sets S_1 below, which is linearly independent,

$$S_1 = \{1, x, x^2, \cdots, x^n, \cdots\}.$$

Another example is the infinite set S_2, which is also linearly independent,

$$S_2 = \left\{1, \cos\left(\frac{n\pi x}{\ell}\right), \sin\left(\frac{n\pi x}{\ell}\right)\right\}_{n=1}^{\infty}.$$

◁

4.2.2. **Main properties.** As we have seen in the Example 4.2.1 above, in a linearly dependent set there is always at least one vector that is a linear combination of the other vectors in the set. This is simple to see from the Definition 4.2.1. Since not all the coefficients c_i are zero in a linearly dependent set, let us suppose that $c_j \neq 0$; then from the Eq. (4.1) we obtain

$$\boldsymbol{v}_j = -\frac{1}{c_j}\big[c_1\boldsymbol{v}_1 + \cdots + c_{j-1}\boldsymbol{v}_{j-1} + c_{j+1}\boldsymbol{v}_{j+1} + \cdots + c_k\boldsymbol{v}_k\big],$$

that is, $\boldsymbol{v}_j$ is a linear combination of the other vectors in the set.

Theorem 4.2.3. *The set* $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ *is linearly dependent with the vector* $\boldsymbol{v}_k$ *being a linear combination of the the remaining* $k-1$ *vectors iff*

$$\operatorname{Span}\big(\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}\big) = \operatorname{Span}\big(\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1}\}\big).$$

This Theorem captures the idea behind the notion of a linearly dependent set: A finite set is linearly dependent iff there exists a smaller set with the same span. In this sense the vector $\boldsymbol{v}_k$ in the Proposition above is redundant with respect to linear combinations.

Proof of Theorem 4.2.3: Let $S_k = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ and $S_{k-1} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1}\}$.

On the one hand, if $\boldsymbol{v}_k$ is a linear combination of the other vectors in S, then for every $\boldsymbol{x} \in \operatorname{Span}(S_k)$ can be expressed as an element in $\operatorname{Span}(S_{k-1})$ simply by replacing $\boldsymbol{v}_k$ in terms of the vectors in $\tilde{S}$. This shows that $\operatorname{Span}(S_k) \subset \operatorname{Span}(S_{k-1})$. The other inclusion is trivial, so $\operatorname{Span}(S_k) = \operatorname{Span}(S_{k-1})$.

On the other hand, if $\operatorname{Span}(S_k) = \operatorname{Span}(S_{k-1})$, this means that $\boldsymbol{v}_k$ is a linear combination of the elements in S_{k-1}. Therefore, the set S_k is linearly dependent. This establishes the Theorem. □

Example 4.2.3: Consider the set $S \subset \mathbb{R}^3$ given by $S = \left\{ \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix}, \begin{bmatrix} -2 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} -4 \\ 1 \\ -3 \end{bmatrix} \right\}$. Find a set $\tilde{S} \subset S$ having the smallest number of vectors such that $\operatorname{Span}(\tilde{S}) = \operatorname{Span}(S)$.

Solution: We have to find all the redundant vectors in S with respect to linear combinations. In other words, with have to find a linearly independent subset of $\tilde{S} \subset S$ such that $\operatorname{Span}(\tilde{S}) = \operatorname{Span}(S)$. The calculation we must do is to find the non-zero coefficients c_i in the solution of the equation

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = c_1 \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix} + c_2 \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix} + c_3 \begin{bmatrix} -2 \\ -3 \\ 2 \end{bmatrix} + c_4 \begin{bmatrix} -4 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}.$$

Hence, we must find the reduced echelon form of the coefficient matrix above, that is,

$$\mathsf{A} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 1 & 2 \\ 0 & -2 & -5 & -3 \\ 0 & 2 & 5 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 1 & 2 \\ 0 & 2 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 1 & \frac{5}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathsf{E_A}.$$

This means that the solution for the coefficients is

$$c_1 = -6c_3 - 5c_4, \qquad c_2 = -\frac{5}{2}c_3 - \frac{3}{2}c_4, \qquad c_3, c_4 \text{ free variables.}$$

Choosing:

$$c_4 = 0,\ c_3 = 2 \quad \Rightarrow \quad c_1 = -12,\ c_2 = -5 \quad \Rightarrow \quad -12 \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix} - 5 \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix} + 2 \begin{bmatrix} -2 \\ -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$c_4 = 2,\ c_3 = 0 \quad \Rightarrow \quad c_1 = -10,\ c_2 = -3 \quad \Rightarrow \quad -10\begin{bmatrix}-2\\2\\-3\end{bmatrix} - 3\begin{bmatrix}4\\-6\\8\end{bmatrix} + 2\begin{bmatrix}-4\\1\\-3\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

We can interpret this result thinking that the third and fourth vectors in matrix A are linear combination of the first two vectors. Therefore, a linearly independent subset of S having its same span is given by

$$\tilde{S} = \left\{ \begin{bmatrix}-2\\2\\-3\end{bmatrix}, \begin{bmatrix}4\\-6\\8\end{bmatrix} \right\}.$$

Notice that all the information to find $\tilde{S}$ is in matrix $\mathsf{E_A}$, the reduced echelon form of matrix A,

$$\mathsf{A} = \begin{bmatrix}-2 & 4 & -2 & -4\\ 2 & -6 & -3 & 1\\ -3 & 8 & 2 & -3\end{bmatrix} \rightarrow \begin{bmatrix}1 & 0 & 6 & 5\\ 0 & 1 & \frac{5}{2} & \frac{3}{2}\\ 0 & 0 & 0 & 0\end{bmatrix} = \mathsf{E_A}.$$

The columns with pivots in $\mathsf{E_A}$ determine the column vectors in A that form a linearly independent set. The non-pivot columns in $\mathsf{E_A}$ determine the column vectors in A that are linear combination of the vectors in the linearly independent set. The factors of these linear combinations are precisely the component of the non-pivot vectors in $\mathsf{E_A}$. For example, the last column vector in $\mathsf{E_A}$ has components 5 and 3/2, and these are precisely the coefficients in the linear combination:

$$\begin{bmatrix}-4\\1\\-3\end{bmatrix} = 5\begin{bmatrix}-2\\2\\-3\end{bmatrix} + \frac{3}{2}\begin{bmatrix}4\\-6\\8\end{bmatrix}.$$

$\triangleleft$

In Example 4.2.3 we answered a question about the linear independence of a set $S = \{\mathsf{v}_1, \cdots, \mathsf{v}_n\} \subset \mathbb{F}^n$ by studying the properties of a matrix having these vectors a column vectors, that is, $\mathsf{A} = [\mathsf{v}_1, \cdots, \mathsf{v}_n]$. It turns out that this is a good idea and the following result summarizes few useful relations.

Theorem 4.2.4. *Given* $\mathsf{A} = [\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}] \in \mathbb{F}^{m,n}$*, the following statements are equivalent:*

(a) The column vectors set $\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\} \subset \mathbb{F}^m$ *is linearly independent;*

(b) $N(\mathsf{A}) = \{\mathbf{0}\}$;

(c) $\operatorname{rank}(\mathsf{A}) = n$

In the case $\mathsf{A} \in \mathbb{F}^{n,n}$*, the set* $\{\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\} \subset \mathbb{F}^n$ *is linearly independent iff* A *is invertible.*

Proof of Theorem 4.2.4: Let us denote by $S = \{\mathsf{v}_1, \cdots, \mathsf{v}_n\} \subset \mathbb{F}^m$ a set of vectors in a vector space, and introduce the matrix $\mathsf{A} = [\mathsf{v}_1, \cdots, \mathsf{v}_n]$. The set S is linearly independent iff only solution $\mathsf{c} \in \mathbb{R}^n$ to the equation $\mathsf{Ac} = \mathsf{0}$ is the trivial solution $\mathsf{c} = \mathsf{0}$. This is equivalent to say that $N(\mathsf{A}) = \{\mathsf{0}\}$. This is equivalent to say that $\mathsf{E_A}$ has n pivot columns, which is equivalent to say that $\operatorname{rank}(\mathsf{A}) = n$. The furthermore part is straightforward, since an $n \times n$ matrix A is invertible iff $\operatorname{rank}(\mathsf{A}) = n$. This establishes the Theorem. $\square$

Further reading. See Section 4.3 in Meyer's book [3].

4.2.3. Exercises.

4.2.1.- Determine which of the following sets is linearly independent. For those who are linearly dependent, express one vector as a linear combination of the other vectors in the set.

(a) $\left\{ \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}2\\1\\0\end{bmatrix}, \begin{bmatrix}1\\5\\9\end{bmatrix} \right\}$;

(b) $\left\{ \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}0\\4\\5\end{bmatrix}, \begin{bmatrix}0\\0\\6\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix} \right\}$;

(c) $\left\{ \begin{bmatrix}3\\2\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}2\\1\\0\end{bmatrix} \right\}$.

4.2.2.- Let $\mathsf{A} = \begin{bmatrix}2&1&1&0\\4&2&1&2\\6&3&2&3\end{bmatrix}$.

(a) Find a linearly independent set containing the largest possible number of columns of A.

(b) Find how many linearly independent sets can be constructed using any number of column vectors of A.

4.2.3.- Show that any set containing the zero vector must be linearly dependent.

4.2.4.- Given a vector space V, prove the following: If the set

$$\{\boldsymbol{v},\ \boldsymbol{w}\} \subset V$$

is linearly independent, then so is

$$\big\{(\boldsymbol{v}+\boldsymbol{w}),\ (\boldsymbol{v}-\boldsymbol{w})\big\}.$$

4.2.5.- Determine whether the set

$$\left\{ \begin{bmatrix}1&2\\2&1\end{bmatrix}, \begin{bmatrix}2&1\\1&1\end{bmatrix}, \begin{bmatrix}4&-1\\-1&1\end{bmatrix} \right\} \subset \mathbb{R}^{2,2}$$

is linearly independent of dependent.

4.2.6.- Show that the following set in $\mathbb{P}_2$ is linearly dependent,

$$\{1,\ x,\ x^2,\ 1+x+x^2\}.$$

4.2.7.- Determine whether $S \subset \mathbb{P}_2$ is a linearly independent set, where

$$S = \big\{1+x+x^2, 2x-3x^2, 2+x\big\}.$$

4.3. Bases and dimension

In this Section we introduce a notion that quantifies the size of a vector space. Before doing that, however, we need to separate two main cases, the vector spaces we call finite dimensional from those called infinite dimensional. In the first case, finite dimensional vector spaces, we introduce the notion of a basis. This is a particular type of set in the vector space that is small enough to be a linearly independent set and big enough to span the whole vector space. A basis of a finite dimensional vector space is not unique. However, every basis contains the same number of vectors. This number, called dimension, quantifies the size of the finite dimensional vector space. In the second case above, infinite dimensional vector spaces, we do not introduce here a concept of basis. More structure is needed in the vector space to be able to determine whether or not an infinite sum of vectors converges. We will not discuss these issues here.

4.3.1. **Basis of a vector space.**

A particular type of finite sets in a vector space, small enough to be linearly independent and big enough to span the whole vector space, is called a basis of that vector space. Vector spaces having a finite set with these properties are essentially small, and they are called finite dimensional. When there is no finite set that spans the whole vector space, we call that space infinite dimensional. We now highlight these ideas in a more precise way.

Definition 4.3.1. *A finite set $\mathcal{S} \subset V$ is called a* ***finite basis*** *of a vector space V iff $\mathcal{S}$ is linearly independent and* $\mathrm{Span}(\mathcal{S}) = V$.

The existence of a finite basis is the property that defines the size of the vector space.

Definition 4.3.2. *A vector space V is* ***finite dimensional*** *iff V has a finite basis or V is one of the following two extreme cases: $V = \emptyset$ or $V = \{\mathbf{0}\}$. Otherwise, the vector space V is called* ***infinite dimensional****.*

In these notes we will often call a finite basis just simply as a basis, without remarking that they contain a finite number of elements. We only study this type of basis, and we do not introduce the concept of an infinite basis. Why don't we define the notion of an infinite basis, since we have already defined the notion of an infinite linearly independent set? Because we do not have any way to define what is the span of an infinite set of vectors. In a vector space, without any further structure, there is no way to know whether an infinite sum converges or not. The notion of convergence needs further structure in a vector space, for example it needs a notion of distance between vectors. So, only in certain vector spaces with a notion of distance it is possible to introduce an infinite basis. We will discuss this subject in a later Chapter.

Example 4.3.1: We now present several examples.

(a) Let $V = \mathbb{R}^2$, then the set $\mathcal{S}_2 = \left\{\mathsf{e}_1 = \begin{bmatrix}1\\0\end{bmatrix}, \mathsf{e}_2 = \begin{bmatrix}0\\1\end{bmatrix}\right\}$ is a basis of $\mathbb{F}^2$. Notice that $\mathsf{e}_i = \mathsf{I}_{:i}$, that is, e_i is the i-th column of the identity matrix I_2. This basis $\mathcal{S}_2$ is called the standard basis of $\mathbb{R}^2$.

(b) A vector space can have infinitely many bases. For example, a second basis for $\mathbb{R}^2$ is the set $\mathcal{U} = \left\{\mathsf{u}_1 = \begin{bmatrix}1\\1\end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix}-1\\1\end{bmatrix}\right\}$. It is not difficult to verify that this set is a basis of $\mathbb{R}^2$, since u is linearly independent, and $\mathrm{Span}(\mathcal{U}) = \mathbb{R}^2$.

(c) Let $V = \mathbb{F}^n$, then the set $\mathcal{S}_n = \left\{\mathsf{e}_1 = \mathsf{I}_{:1}, \cdots, \mathsf{e}_n = \mathsf{I}_{:n}\right\}$ is a basis of $\mathbb{R}^n$, where $\mathsf{I}_{:i}$ is the i-th column of the identity matrix I_n. This set $\mathcal{S}_n$ is called the ***standard basis*** of $\mathbb{F}^n$.

(d) A basis for the vector space $\mathbb{F}^{2,2}$ of all 2×2 matrices is the set $\mathcal{S}_{2,2}$ given by

$$\mathcal{S}_{2,2} = \left\{ \mathsf{E}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathsf{E}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\};$$

This set is linearly independent and $\mathrm{Span}(\mathcal{S}_{2,2}) = \mathbb{F}^{2,2}$, since any element $\mathsf{A} \in \mathbb{F}^{2,2}$ can be decomposed as follows,

$$\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = A_{11} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + A_{12} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + A_{21} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + A_{22} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

(e) A basis for the vector space $\mathbb{F}^{m,n}$ of all $m \times n$ matrices is the following:

$$\mathcal{S}_{m,n} = \{\mathsf{E}_{11}, \mathsf{E}_{12}, \cdots, \mathsf{E}_{mn}\},$$

where each $m \times n$ matrix E_{ij} is a matrix with all coefficients zero except the coefficient (i,j) which is equal to one (see previous example). The set $\mathcal{S}_{m,n}$ is linearly independent, and $\mathrm{Span}(\mathcal{S}_{m,n}) = \mathbb{F}^{m,n}$, since

$$A = [A_{ij}] = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} \mathsf{E}_{ij}.$$

(f) Let $V = \mathbb{P}_n$, the set of all polynomials with domain $\mathbb{R}$ and degree less or equal n. Any element $\boldsymbol{p} \in \mathbb{P}_n$ can be expressed as follows,

$$\boldsymbol{p}(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

that is equivalent to say that the set

$$\mathcal{S} = \{\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2,\ \cdots,\ \boldsymbol{p}_n = x^n\}$$

satisfies $\mathbb{P}_n = \mathrm{Span}(\mathcal{S})$. The set $\mathcal{S}$ is also linearly independent, since

$$\boldsymbol{q}(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n = 0 \quad \Rightarrow \quad c_0 = \cdots = c_n = 0.$$

The proof of the latter statement is simple: Compute the n-th derivative of $\boldsymbol{q}$ above, and obtain the equation $n!\, c_n = 0$, so $c_n = 0$. Add this information into the $(n-1)$-th derivative of $\boldsymbol{q}$ and we conclude that $c_{n-1} = 0$. Continue in this way, and you will prove that all the coefficient c's vanish. Therefore, $\mathcal{S}$ is a basis of $\mathbb{P}_n$, ands it is also called the standard basis of $\mathbb{P}_n$.

◁

EXAMPLE 4.3.2: Show that the set $\mathcal{U} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$.

SOLUTION: We must show that $\mathcal{U}$ is a linearly independent set and $\mathrm{Span}(\mathcal{U}) = \mathbb{R}^3$. Both properties follow from the fact that matrix U below, whose columns are the elements in $\mathcal{U}$, is invertible,

$$\mathsf{U} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{U}^{-1} = \frac{1}{2} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 2 & 0 \\ 1 & 1 & -1 \end{bmatrix}.$$

Let us show that $\mathcal{U}$ is a basis of $\mathbb{R}^3$: Since matrix U is invertible, this implies that its reduced echelon form $\mathsf{E}_{\mathsf{U}} = \mathsf{I}_3$, so its column vectors form a linearly independent set. The existence of U^{-1} implies that the system of equations $\mathsf{U}\mathsf{x} = \mathsf{y}$ has a solution $\mathsf{x} = \mathsf{U}^{-1}\mathsf{y}$ for every $\mathsf{y} \in \mathbb{R}^3$, that is, $\mathsf{y} \in \mathrm{Col}(\mathsf{U}) = \mathrm{Span}(\mathcal{U})$ for all $\mathsf{y} \in \mathbb{R}^3$. This means that $\mathrm{Span}(\mathcal{U}) = \mathbb{R}^3$. Hence, the set $\mathcal{U}$ is a basis of $\mathbb{R}^3$. ◁

The following definitions will be useful to establish important properties of a basis.

Definition 4.3.3. *Let V be a vector space and $S_n \subset V$ be a subset with n elements. The set S_n is a* ***maximal linearly independent*** *set iff S_n is linearly independent and every other set $\tilde{S}_m$ with $m > n$ elements is linearly dependent. The set S_n is a* ***minimal spanning*** *set iff $\operatorname{Span}(S_n) = V$ and every other set $\tilde{S}_m$ with $m < n$ elements satisfies $\operatorname{Span}(\tilde{S}_m) \subsetneq V$.*

A maximal linearly independent set S is the biggest set in a vector space that is linearly independent. A set cannot be linearly independent if it is too big, since the bigger the set the more probable that one element in the set is a linear combination of the other elements in the set. A minimal spanning set is the smallest set in a vector space that spans the whole space. A spanning set, that is, a set whose span is the whole space, cannot be too small, since the smaller the set the more probable that an element in the vector space is outside the span of the set. The following result provides a useful characterization of a basis: *A basis is a set in the vector space that is both maximal linearly independent and minimal spanning.* In this sense, a basis is a set with the right size, small enough to be linearly independent and big enough to span the whole vector space.

Theorem 4.3.4. *Let V be a vector space. The following statements are equivalent:*

(a) $\mathcal{U}$ is a basis of V;
(b) $\mathcal{U}$ is a minimal spanning set in V.
(c) $\mathcal{U}$ is a maximal linearly independent set in V;

EXAMPLE 4.3.3: We showed in Example 4.3.2 above that the set $\mathcal{U} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$. Since this basis has three elements, Theorem 4.3.4 says that any other spanning set in $\mathbb{R}^3$ cannot have less than three vectors, and any other linearly independent set in $\mathbb{R}^3$ cannot have more that three vectors. For example, any subset of $\mathcal{U}$ containing two elements cannot span $\mathbb{R}^3$; the linear combination of two vectors in $\mathcal{U}$ span a plane in $\mathbb{R}^3$. Another example, any set of four vectors in $\mathbb{R}^3$ must be linearly dependent. $\triangleleft$

Proof of Theorem 4.3.4: We first show part (a)-(b).

($\Rightarrow$) Assume that $\mathcal{U}$ is a basis of V. If the set $\mathcal{U}$ is not a minimal spanning set of V, that means there exists $\tilde{\mathcal{U}} = \{\tilde{\boldsymbol{u}}_1, \cdots, \tilde{\boldsymbol{u}}_{n-1}\}$ such that $\operatorname{Span}(\tilde{\mathcal{U}}) = V$. So, every vector in $\mathcal{U}$ can be expressed as a linear combination of vectors in $\tilde{\mathcal{U}}$. Hence, there exists a set of coefficients C_{ij} such that

$$\boldsymbol{u}_j = \sum_{i=1}^{n-1} C_{ij} \tilde{\boldsymbol{u}}_i, \qquad j = 1, \cdots, n.$$

The reason to order the coefficients C_{ij} is this form is that they form a matrix $\mathsf{C} = [C_{ij}]$ which is $(n-1) \times n$. This matrix C defines a function $\mathsf{C} : \mathbb{R}^n \to \mathbb{R}^{n-1}$, and since $\operatorname{rank}(\mathsf{C}) \leqslant (n-1) < n$, this matrix satisfies that $N(\mathsf{C})$ is nontrivial as a subset of $\mathbb{R}^n$. So there exists a nonzero column vector in $\mathbb{R}^n$ with components $\mathsf{z} = [z_j] \in \mathbb{R}^n$, not all components zero, such that $\mathsf{z} \in N(\mathsf{C})$, that is,

$$\sum_{j=1}^{n} C_{ij} z_j = 0, \qquad i = 1, \cdots, (n-1).$$

What we have found is that the linear combination

$$z_1 \boldsymbol{u}_1 + \cdots + z_n \boldsymbol{u}_n = \sum_{j=1}^{n} z_j \boldsymbol{u}_j = \sum_{j=1}^{n} z_j \Big(\sum_{i=1}^{n-1} C_{ij} \tilde{\boldsymbol{u}}_i \Big) = \sum_{i=1}^{n-1} \Big(\sum_{j=1}^{n} C_{ij} z_j \Big) \tilde{\boldsymbol{u}}_i = \boldsymbol{0},$$

with at least one of the coefficients z_j non-zero. This means that the set $\mathcal{U}$ is not linearly independent. But this contradicts that $\mathcal{U}$ is a basis. Therefore, the set $\mathcal{U}$ is a minimal spanning set of V.

($\Leftarrow$) Assume that $\mathcal{U}$ is a minimal spanning set of V. If $\mathcal{U}$ is not a basis, that means $\mathcal{U}$ is not a linearly independent set. At least one element in $\mathcal{U}$ must be a linear combination of the others. Let us arrange the order of the basis vectors such that the vector $\boldsymbol{u}_n$ is a linear combination of the other vectors in $\mathcal{U}$. Then, the set $\tilde{\mathcal{U}} = \{\boldsymbol{u}, \cdots, \boldsymbol{u}_{n-1}\}$ must still span V, that is $\operatorname{Span}(\tilde{\mathcal{U}}) = V$. But this contradicts the assumption that $\mathcal{U}$ is a minimal spanning set of V.

We now show part (a)-(c).

($\Leftarrow$) Assume that $\mathcal{U}$ is a maximal linearly independent set in V. If $\mathcal{U}$ is not a basis, that means $\operatorname{Span}(\mathcal{U}) \subsetneqq V$, so there exists $\boldsymbol{u}_{n+1} \in V$ such that $\boldsymbol{u}_{n+1} \notin \operatorname{Span}(\mathcal{U})$. Hence, the set $\tilde{\mathcal{U}} = \{\boldsymbol{u}, \cdots, \boldsymbol{u}_{n+1}\}$ is a linearly independent set. However, this contradicts the assumption that $\mathcal{U}$ is a maximal linearly independent set. We conclude that $\mathcal{U}$ is a basis of V.

($\Rightarrow$) Assume that $\mathcal{U}$ is a basis of V. If the set $\mathcal{U}$ is not a maximal linearly independent set in V, then there exists a maximal linearly independent set $\tilde{\mathcal{U}} = \{\tilde{\boldsymbol{u}}_1, \cdots, \tilde{\boldsymbol{u}}_k\}$, with $k > n$. By the argument given just above, $\tilde{\mathcal{U}}$ is a basis of V. By part (b) the set $\tilde{\mathcal{U}}$ must be a minimal spanning set of V. However, this is not true, since $\mathcal{U}$ is smaller and spans V. Therefore, $\mathcal{U}$ must be a maximal linearly independent set in V.

This establishes the Theorem. □

4.3.2. **Dimension of a vector space.** The characterization of a basis given in Theorem 4.3.4 above implies that the number of elements in a basis is always the same as in any other basis.

Theorem 4.3.5. *The number of elements in any basis of a finite dimensional vector space is the same as in any other basis.*

Proof of Theorem (4.3.5: Let $\mathcal{V}_n$ and $\mathcal{V}_m$ be two bases of a vector space V with n and m elements, respectively. If $m > n$, the property that $\mathcal{V}_m$ is a minimal spanning set implies that $\operatorname{Span}(\mathcal{V}_n) \subsetneqq \operatorname{Span}(\mathcal{V}_m) = V$. The former inclusion contradicts that $\mathcal{V}_n$ is a basis. Therefore, $n = m$. (A similar proof can be constructed with the maximal linearly independence property of a basis.) This establishes the Theorem. □

The number of elements in a basis of a finite dimensional vector space is a characteristic of the vector space, so we give that characteristic a name.

Definition 4.3.6. *The **dimension** of a finite dimensional vector space V with a finite basis, denoted as $\dim V$, is the number of elements in any basis of V. The extreme cases of $V = \emptyset$ and $V = \{\mathbf{0}\}$ are defined as zero dimensional.*

From the definition above we see that $\dim\{\boldsymbol{0}\} = 0$ and $\dim \emptyset = 0$.

Example 4.3.4: We now present several examples.

(a) The set $\mathcal{S}_n = \{\mathsf{e}_1 = \mathsf{I}_{:1}, \cdots, \mathsf{e}_n = \mathsf{I}_{:n}\}$ is a basis for $\mathbb{F}^n$, so $\dim \mathbb{F}^n = n$.

(b) A basis for the vector space $\mathbb{F}^{2,2}$ of all 2×2 matrices is the set $\mathcal{S}_{2,2}$ is given by

$$\mathcal{S}_{2,2} = \left\{\mathsf{E}_{11} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \mathsf{E}_{12} = \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \mathsf{E}_{21} = \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \mathsf{E}_{22} = \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right\},$$

so we conclude that $\dim \mathbb{F}^{2,2} = 4$.

(c) A basis for the vector space $\mathbb{F}^{m,n}$ of all $m \times n$ matrices is the following:

$$\mathcal{S}_{m,n} = \{\mathsf{E}_{11}, \mathsf{E}_{12}, \cdots, \mathsf{E}_{mn}\},$$

where we recall that each $m \times n$ matrix E_{ij} is a matrix with all coefficients zero except the coefficient (i, j) which is equal to one. Since the basis $\mathcal{S}_{m,n}$ contains mn elements, we conclude that $\dim \mathbb{F}^{m,n} = mn$.

(d) A basis for the vector space $\mathbb{P}_n$ of all polynomial with degree less or equal n is the set $\mathcal{S}$ given by $\mathcal{S} = \{\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2, \cdots, \boldsymbol{p}_n = x^n\}$. This set has $n+1$ elements, so $\dim \mathbb{P}_n = n + 1$.

$\lhd$

REMARK: Any subspace $W \subset V$ of a vector space V is itself a vector space, so the definition of basis also holds for W. Since $W \subset V$, we conclude that $\dim W \leqslant \dim V$.

EXAMPLE 4.3.5: Consider the case $V = \mathbb{R}^3$. It is simple to see in Fig. 37 that $\dim U = 1$ and $\dim W = 2$, where the subspaces U and W are spanned by one vector and by two non-collinear vectors, respectively.

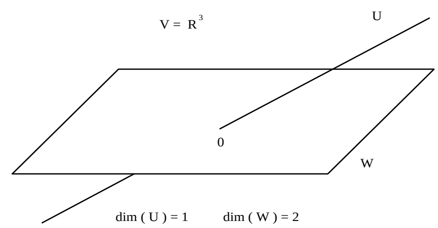

FIGURE 37. Sketch of two subspaces U and W, in the vector space $\mathbb{R}^3$, of dimension one and two, respectively.

$\lhd$

EXAMPLE 4.3.6: Find a basis for $N(\mathsf{A})$ and $R(\mathsf{A})$, where matrix $\mathsf{A} \in \mathbb{R}^{3,4}$ is given by

$$\mathsf{A} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix}. \tag{4.2}$$

SOLUTION: Since $\mathsf{A} \in \mathbb{R}^{3,4}$, then $\mathsf{A} : \mathbb{R}^4 \to \mathbb{R}^3$, which implies that $N(\mathsf{A}) \subset \mathbb{R}^4$ while $R(\mathsf{A}) \subset \mathbb{R}^3$. A basis for $N(\mathsf{A})$ is found as follows: Find all solution of $\mathsf{A}\mathsf{x} = \mathsf{0}$ and express these solutions as the span of a linearly independent set of vectors. We first find E_{A},

$$\mathsf{A} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 1 & \frac{5}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathsf{E}_{\mathsf{A}} \quad \Rightarrow \quad \begin{cases} x_1 = -6x_3 - 5x_4, \\ x_2 = -\dfrac{5}{2}\,x_3 - \dfrac{3}{2}\,x_4, \\ x_3,\ x_4 \quad \text{free variables.} \end{cases}$$

Therefore, every element in $N(\mathsf{A})$ can be expressed as follows,

$$\mathsf{x} = \begin{bmatrix} -6x_3 - 5x_4 \\ -\frac{5}{2}\,x_3 - \frac{3}{2}\,x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix} x_3 + \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix} x_4, \quad \Rightarrow \quad N(\mathsf{A}) = \operatorname{Span}\left(\left\{ \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix} \right\}\right).$$

Since the vectors in the span above form a linearly independent set, we conclude that a basis for $N(\mathsf{A})$ is the set $\mathcal{N}$ given by

$$\mathcal{N} = \left\{ \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix} \right\}.$$

We now find a basis for $R(\mathsf{A})$. We know that $R(\mathsf{A}) = \mathrm{Col}(\mathsf{A})$, that is, the span of the column vectors of matrix A. We only need to find a linearly independent subset of column vectors of A. This information is given in $\mathsf{E_A}$, since the pivot columns in $\mathsf{E_A}$ indicate the columns in A which form a linearly independent set. In our case, the pivot columns in $\mathsf{E_A}$ are the first and second columns, so we conclude that a basis for $R(\mathsf{A})$ is the set $\mathcal{R}$ given by

$$\mathcal{R} = \left\{ \begin{bmatrix} -2 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ -6 \\ 8 \end{bmatrix} \right\}.$$

$\triangleleft$

4.3.3. **Extension of a set to a basis.** We know that a basis of a vector space is not unique, and the following result says that actually any linearly independent set can be extended into a basis of a vector space.

Theorem 4.3.7. *If $S_k = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k\}$ is a linearly independent set in a vector spave V with basis $\mathcal{V} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$, where $k < n$, then, there always exists a basis of V given by an extension of the set S_k of the form $\mathcal{S} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k, \boldsymbol{v}_{i_1}, \cdots, \boldsymbol{v}_{i_{n-k}}\}$.*

The statement above says that a linearly independent set S_k can be extended into a basis $\mathcal{S}$ of a a vector space V simply incorporating appropriate vectors from any basis of V. If a basis $\mathcal{V}$ of V has n vectors, and the set S_k has $k < n$ vectors, then one can always select $n - k$ vectors from the basis $\mathcal{V}$ to enlarge the set S_k into a basis of V.

Proof of Theorem 4.3.7: Introduce the set S_{k+n}

$$s_{k+n} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}.$$

We know that $\mathrm{Span}(S_{k+n}) = V$ since $\mathcal{V} \subset S_{k+n}$. We also know that S_{k+n} is linearly dependent, since the maximal linearly independent set contains n elements and S_{k+n} contains $n + k > n$ elements. The idea is to eliminate the $\boldsymbol{v}_i$ such that $S_k \cup \{\boldsymbol{v}_i\}$ is linearly dependent. Since the maximal linearly independent set contains n elements and the S_k is linearly independent, there are k elements in $\mathcal{V}$ that will be eliminated. The resulting set is $\mathcal{S}$, which is a basis of V containing S_k. This establishes the Theorem. $\square$

EXAMPLE 4.3.7: Given the 3×4 matrix A defined in Eq. (4.2) in Example 4.3.6 above, extend the basis of $N(\mathsf{A}) \subset \mathbb{R}^4$ into a basis of $\mathbb{R}^4$.

SOLUTION: We know from Example 4.3.6 that a basis for the $N(\mathsf{A})$ is the set $\mathcal{N}$ given by

$$\mathcal{N} = \left\{ \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Following the idea in the proof of Theorem 4.3.7, we look for a linear independent set of vectors among the columns of the matrix

$$\mathsf{M} = \begin{bmatrix} -6 & -5 & 1 & 0 & 0 & 0 \\ -\frac{5}{2} & -\frac{3}{2} & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

That is, matrix M include the basis vectors of $N(\mathsf{A})$ and the four vectors e_i of the standard basis of $\mathbb{R}^4$. It is important to place the basis vectors of $N(\mathsf{A})$ in the first columns of M. In this way, the Gauss method will select these first vectors as part of the linearly independent set of vectors. Find now the reduced echelon form matrix $\mathsf{E_M}$,

$$\mathsf{E_M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 6 & 5 \\ 0 & 0 & 0 & 1 & 5 & 3 \end{bmatrix}.$$

Therefore, the first four vectors in M are form a linearly independent set, so a basis of $\mathbb{R}^4$ that includes $\mathcal{N}$ is given by

$$\mathcal{V} = \left\{ \begin{bmatrix} -6 \\ -\frac{5}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -\frac{3}{2} \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

◁

4.3.4. **The dimension of subspace addition.** Recall that the sum and the intersection of two subspaces is again a subspace in a given vector space. The following result relates the dimension of a sum of subspaces with the dimension of the individual subspaces and the dimension of their intersection.

Theorem 4.3.8. *If W_1, $W_2 \subset V$ are subspaces of a vector space V, then holds*

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2).$$

Proof of Theorem 4.3.8: We find the dimension of $W_1 + W_2$ finding a basis of this sum. The key idea is to start with a basis of $W_1 \cap W_2$. Let $\mathcal{B}_0 = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_l\}$ be a basis for $W_1 \cap W_2$. Enlarge that basis into basis $\mathcal{B}_1$ for W_1 and $\mathcal{B}_2$ for W_2 as follows,

$$\mathcal{B}_1 = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_l, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}, \quad \mathcal{B}_2 = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_l, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\}.$$

We use the notation $l = \dim(W_1 \cap W_2)$, $l + n = \dim W_1$ and $l + m = \dim W_2$. We now propose as basis for $W_1 + W_2$ the set

$$\mathcal{B} = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_l, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\}.$$

By construction this set satisfies that $\mathrm{Span}(\mathcal{B}) = W_1 + W_2$. We only need to show that $\mathcal{B}$ is linearly independent. Assume that the set $\mathcal{B}$ is linearly dependent. This means that there is non-zero constants a_i, b_j and c_k solutions of the equation

$$\sum_{i=1}^{n} a_i \boldsymbol{x}_i + \sum_{j=1}^{m} b_j \boldsymbol{y}_j + \sum_{k=1}^{l} c_k \boldsymbol{z}_k = \boldsymbol{0}. \tag{4.3}$$

This implies that the vector $\sum_{i=1}^{n} a_i \boldsymbol{x}_i$, which by definition belongs to W_1, also belongs to W_2, since

$$\sum_{i=1}^{n} a_i \boldsymbol{x}_i = -\Big(\sum_{j=1}^{m} b_j \boldsymbol{y}_j + \sum_{k=1}^{l} c_k \boldsymbol{z}_k\Big) \in W_2.$$

Therefore, $\sum_{i=1}^{n} a_i \boldsymbol{x}_i$ belongs to $W_1 \cap W_2$, and so is a linear combination of the elements of $\mathcal{B}_0$, that is, there exists scalars d_k such that

$$\sum_{i=1}^{n} a_i \boldsymbol{x}_i = \sum_{i=1}^{l} d_k \boldsymbol{z}_k.$$

Since $\mathcal{B}_1$ is a basis of W_1, this implies that all the coefficients a_i and d_k vanish. Introduce this information into Eq. (4.3) and we conclude that

$$\sum_{j=1}^{m} b_j \boldsymbol{y}_j + \sum_{k=1}^{l} c_k \boldsymbol{z}_k = \boldsymbol{0}.$$

Analogously, the set $\mathcal{B}_2$ is a basis, so all the coefficients b_j and c_k must vanish. This implies that the set $\mathcal{B}$ is linearly independent, hence a basis of $W_1 + W_2$. Therefore, the dimension of the sum is given by

$$\dim(W_1 + W_2) = n + m + k = (n + k) + (m + k) - k = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2).$$

This establishes the Theorem. □

The following corollary is immediate.

Corollary 4.3.9. *If a vector space can be decomposed as $V = W_1 \oplus W_2$, then*

$$\dim(W_1 \oplus W_2) = \dim W_1 + \dim W_2.$$

The proof is straightforward from Theorem 4.3.8, since the condition of subspaces direct sum, $W_1 \cap W_2 = \{\boldsymbol{0}\}$, says that $\dim\big(W_1 \cap W_2\big) = 0$.

4.3.5. **Exercises.**

4.3.1.- Find a basis for each of the spaces $N(\mathsf{A})$, $R(\mathsf{A})$, $N(\mathsf{A}^T)$, $R(\mathsf{A}^T)$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 1 & 4 \end{bmatrix}.$$

4.3.2.- Find the dimension of the space spanned by

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 8 \\ -4 \\ 8 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 0 \\ 6 \end{bmatrix} \right\}.$$

4.3.3.- Find the dimension of the following spaces:

(a) The space $\mathbb{P}_n$ of polynomials of degree less or equal n.
(b) The space $\mathbb{F}^{m,n}$ of $m \times n$ matrices.
(c) The space of real symmetric $n \times n$ matrices.
(d) The space of real skew-symmetric $n \times n$ matrices.

4.3.4.- Find an example to show that the following statement is false: Given a basis $\{\mathsf{v}_1, \mathsf{v}_2\}$ of $\mathbb{R}^2$, then every subspace $W \subset \mathbb{R}^2$ has a basis containing at least one of the basis vectors v_1, v_2.

4.3.5.- Given the matrix A and vector v,

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 2 & 0 & 5 \\ 2 & 4 & 3 & 1 & 8 \\ 3 & 6 & 1 & 5 & 5 \end{bmatrix}, \quad \mathsf{v} = \begin{bmatrix} -8 \\ 1 \\ 3 \\ 3 \\ 0 \end{bmatrix},$$

verify that $\mathsf{v} \in N(\mathsf{A})$, and then find a basis of $N(\mathsf{A})$ containing v.

4.3.6.- Determine whether or not the set

$$\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right\}$$

is a basis for the subspace

$$\mathrm{Span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 8 \\ 7 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} \right\} \right) \subset \mathbb{R}^3.$$

4.4. Vector components

4.4.1. **Ordered bases.** In the previous Section we introduced a finite basis in a vector space. Although a vector space can have different bases, every basis has the same number of elements, which provides a measure of the vector space size, called the dimension of the vector space. In this Section we study another property of a basis. Every vector in a finite dimensional vector space can be expressed in a unique way as a linear combination of the basis vectors. This property can be clearly stated in an ordered basis, which is a basis with the basis vectors given in a specific order.

Definition 4.4.1. *An* ***ordered basis*** *of an n-dimensional vector space V is a sequence $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ of vectors such that the set $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ is a basis of V.*

Recall that a sequence is an ordered set, that is, a set with elements given in a particular order.

Example 4.4.1: The following four ordered basis of $\mathbb{R}^3$ are all different,

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right), \quad \left(\begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right), \quad \left(\begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}\right), \quad \left(\begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}\right),$$

however, they determine the same basis $\mathcal{S}_3 = \left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$. ◁

4.4.2. **Vector components in a basis.** The following result states that given a vector space with an ordered basis, there exists a correspondence between vectors and certain sequences of scalars.

Theorem 4.4.2. *Let V be an n-dimensional vector space over the scalar field $\mathbb{F}$ with an ordered basis $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$. Then, every vector $\boldsymbol{v} \in V$ determines a unique scalars' sequence $(v_1, \cdots, v_n) \subset \mathbb{F}$ such that*

$$\boldsymbol{v} = v_1 \boldsymbol{u}_1 + \cdots + v_n \boldsymbol{u}_n. \tag{4.4}$$

And every scalars' sequence $(v_1, \cdots, v_n) \subset \mathbb{F}$ determines a unique vector $\boldsymbol{v} \in V$ by Eq. (4.4).

Proof of Theorem 4.4.2: Denote by $\mathcal{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ the ordered basis of V. Since $\mathcal{U}$ is a basis, $\mathrm{Span}(\mathcal{U}) = V$ and $\mathcal{U}$ is linearly independent. The first property implies that for every $\boldsymbol{v} \in V$ there exist scalars $v_1, \cdots, v_n$ such that $\boldsymbol{v}$ is a linear combination of the basis vectors, that is,

$$\boldsymbol{v} = v_1 \boldsymbol{u}_1 + \cdots + v_n \boldsymbol{u}_n.$$

The second property of a basis implies that the linear combination above is unique. Indeed, if there exists another linear combination

$$\boldsymbol{v} = \nu_1 \boldsymbol{u}_1 + \cdots + \nu_n \boldsymbol{u}_n,$$

then $\boldsymbol{0} = \boldsymbol{v} - \boldsymbol{v} = (v_1 - \nu_1)\boldsymbol{u}_1 + \cdots + (v_n - \nu_n)\boldsymbol{u}_n$. Since $\mathcal{U}$ is linearly independent, this implies that each coefficient above vanishes, so $v_1 = \nu_1, \cdots, v_n = \nu_n$.

The converse statement is simple to show, since the scalars are given in a specific order. Every scalars' sequence $(v_1, \cdots, v_n)$ determines a unique linear combination with an ordered basis $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ given by $v_1 \boldsymbol{u}_1 + \cdots + v_n \boldsymbol{u}_n$. This unique linear combination determines a unique vector in the vector space. This establishes the Theorem. □

Theorem 4.4.2 says that there exists a correspondence between vectors in a vector space with an ordered basis and scalars' sequences. This correspondence is called a coordinate map and the scalars are called vector components in the basis. Here is a precise definition.

Definition 4.4.3. *Let V be an n-dimensional vector space over $\mathbb{F}$ with an ordered basis $\mathcal{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$. The* ***coordinate map*** *is the function $[\,]_u : V \to \mathbb{F}^n$, with $[\boldsymbol{v}]_u = \mathsf{v}_u$, and*

$$\mathsf{v}_u = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{v} = v_1 \boldsymbol{u}_1 + \cdots + v_n \boldsymbol{u}_n.$$

The scalars $v_1, \cdots, v_n$ are called the ***vector components*** *of $\boldsymbol{v}$ in the ordered basis $\mathcal{U}$.*

Therefore, we use the notation $[\boldsymbol{v}]_u = \mathsf{v}_u \in \mathbb{F}^n$ for the components of a vector $\boldsymbol{v} \in V$ in an ordered basis $\mathcal{U}$. We remark that the coordinate map is defined only after an ordered basis is fixed in V. Different ordered bases on V determine different coordinate maps between V and $\mathbb{F}^n$. When the situation under study involves only one ordered basis, we suppress the basis subindex. The coordinate map will be denoted by $[\,] : V \to \mathbb{F}^n$ and the vector components by $\mathsf{v} = [\boldsymbol{v}]$. In the particular case that $V = \mathbb{F}^n$ and the basis is the standard basis $\mathcal{S}_n$, then the coordinate map $[\,]_s$ is the identity map, so $\boldsymbol{v} = [\boldsymbol{v}]_s = \mathsf{v}$. In this case we follow the convention established in the first Chapters, that is, we denote vectors in $\mathbb{F}^n$ by v instead of $\boldsymbol{v}$. When the situation under study involves more than one ordered basis we keep the sub-indices in the coordinate map, like $[\,]_u$, and in the vector components, like v_u, to keep track of the basis attached to these expressions.

EXAMPLE 4.4.2: Let V be the set of points on the plane with a preferred origin. Let $\mathcal{S} = \big(\boldsymbol{e}_1, \boldsymbol{e}_2\big)$ be an ordered basis, pictured in Fig. 38.

(a) Find the components $\mathsf{v}_s = [\boldsymbol{v}]_s \in \mathbb{R}^2$ of the vector $\boldsymbol{v} = \boldsymbol{e}_1 + 3\boldsymbol{e}_2$ in the ordered basis $\mathcal{S}$.
(b) Find the components $\mathsf{v}_u = [\boldsymbol{v}]_u \in \mathbb{R}^2$ of the same vector $\boldsymbol{v}$ given in part (a) but now in the ordered basis $\mathcal{U} = \big(\boldsymbol{u}_1 = \boldsymbol{e}_1 + \boldsymbol{e}_2,\ \boldsymbol{u}_2 = -\boldsymbol{e}_1 + \boldsymbol{e}_2\big)$.

SOLUTION: Part (a) is straightforward to compute, since the definition of component of a vector says that the numbers multiplying the basis vectors in the equation $\boldsymbol{v} = \boldsymbol{e}_1 + 3\boldsymbol{e}_2$ are the components of the vector, that is,

$$\boldsymbol{v} = \boldsymbol{e}_1 + 3\boldsymbol{e}_2 \quad \Leftrightarrow \quad \mathsf{v}_s = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Part (b) is more involved. We are looking for numbers $\tilde{v}_1$ and $\tilde{v}_2$ such that

$$\boldsymbol{v} = \tilde{v}_1 \boldsymbol{u}_1 + \tilde{v}_2 \boldsymbol{u}_2 \quad \Leftrightarrow \quad \mathsf{v}_u = \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix}. \tag{4.5}$$

From the definition of the basis $\mathcal{U}$ we know the components of the basis vectors in $\mathcal{U}$ in terms of the standard basis, that is,

$$\boldsymbol{u}_1 = \boldsymbol{e}_1 + \boldsymbol{e}_2 \quad \Leftrightarrow \quad \mathsf{u}_{1s} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\boldsymbol{u}_2 = -\boldsymbol{e}_1 + \boldsymbol{e}_2 \quad \Leftrightarrow \quad \mathsf{u}_{2s} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

In other words, we can write the ordered basis $\mathcal{U}$ as the column vectors of the matrix $\mathsf{U}_s = \big[\mathcal{U}\big]_s = \big[[\boldsymbol{u}_1]_s, [\boldsymbol{u}_2]_s\big]$ given by

$$\mathsf{U}_s = \big[\mathsf{u}_{1s}, \mathsf{u}_{2s}\big] = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Expressing Eq. (4.5) in the standard basis means

$$\boldsymbol{e}_1 + 3\boldsymbol{e}_2 = \boldsymbol{v} = \tilde{v}_1(\boldsymbol{e}_1 + \boldsymbol{e}_2) + \tilde{v}_2(-\boldsymbol{e}_1 + \boldsymbol{e}_2) \quad \Leftrightarrow \quad \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \tilde{v}_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \tilde{v}_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The last equation on the right is a matrix equation for the unknowns $\mathsf{v}_u = \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix}$,

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \Leftrightarrow \quad \mathsf{U}_s \boldsymbol{v}_u = \mathsf{v}_s.$$

We find the solution using the Gauss method,

$$\left[\begin{array}{cc|c} 1 & -1 & 1 \\ 1 & 1 & 3 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \end{array}\right] \quad \Rightarrow \quad \mathsf{v}_u = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{v} = 2\boldsymbol{u}_1 + \boldsymbol{u}_2.$$

A sketch of what has been computed is in Fig. 38. In this Figure is clear that the vector $\boldsymbol{v}$ is fixed, and we have only expressed this fixed vector in as a linear combination of two different bases. It is clear in this Fig. 38 that one has to stretch the vector $\boldsymbol{u}_1$ by two and add the result to the vector $\boldsymbol{u}_2$ to obtain $\boldsymbol{v}$. ◁

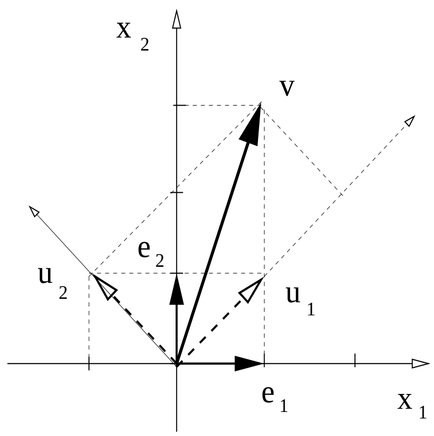

FIGURE 38. The vector $\boldsymbol{v} = \boldsymbol{e}_1 + 3\boldsymbol{e}_2$ expressed in terms of the basis $\mathcal{U} = \{\boldsymbol{u}_1 = \boldsymbol{e}_1 + \boldsymbol{e}_2, \boldsymbol{u}_2 = -\boldsymbol{e}_1 + \boldsymbol{e}_2\}$ is given by $\boldsymbol{v} = 2\boldsymbol{u}_1 + \boldsymbol{u}_2$.

EXAMPLE 4.4.3: Consider the vector space $\mathbb{P}_2$ of all polynomials of degree less or equal two, and let us consider the case of $\mathbb{F} = \mathbb{R}$. An ordered basis is $\mathcal{S} = \big(\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2\big)$. The coordinate map is $[\ ]_s : \mathbb{P}_2 \to \mathbb{R}^3$ defined as follows, $[\boldsymbol{p}]_s = \mathsf{p}_s$, where

$$\mathsf{p}_s = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{p}(x) = a + bx + cx^2.$$

The column vector p_s represents the components of the vector $\boldsymbol{p}$ in the ordered basis $\mathcal{S}$. The equation above defines a correspondence between every element in $\mathbb{P}_2$ and every element in $\mathbb{R}^3$. The coordinate map depend on the choice of the ordered basis. For example, choosing the ordered basis $\tilde{\mathcal{S}} = \big(\boldsymbol{p}_0 = x^2,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = 1\big)$, the corresponding coordinate map is $[\ ]_{\tilde{s}} : \mathbb{P}_2 \to \mathbb{R}^3$ defined by $[\boldsymbol{p}]_{\tilde{s}} = \mathsf{p}_{\tilde{s}}$, where

$$\mathsf{p}_{\tilde{s}} = \begin{bmatrix} c \\ b \\ a \end{bmatrix} \quad \Leftrightarrow \quad p(x) = a + bx + cx^2.$$

The coordinate maps above generalize to the spaces $\mathbb{P}_n$ and $\mathbb{R}^{n+1}$ for all $n \in \mathbb{N}$. Given the ordered basis $\mathcal{S} = \big(\boldsymbol{p}_0 = 1, \boldsymbol{p}_1 = x, \cdots, \boldsymbol{p}_n = x^n\big)$, the corresponding coordinate map $[\ ]_s : \mathbb{P}_n \to \mathbb{R}^{n+1}$ defined by $[\boldsymbol{p}]_s = \mathsf{p}_s$, where

$$\mathsf{p}_s = \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} \quad \Leftrightarrow \quad p(x) = a_0 + \cdots + a_n x^n.$$

$\lhd$

Example 4.4.4: Consider $V = \mathbb{P}_2$ with ordered basis $\mathcal{S} = \big(\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2\big)$.

(a) Find $\mathsf{r}_s = [\boldsymbol{r}]_s$, the components of $\boldsymbol{r}(x) = 3 + 2x + 4x^2$ in the ordered basis $\mathcal{S}$.

(b) Find $\mathsf{r}_q = [\boldsymbol{r}]_q$, the components of the same polynomial $\boldsymbol{r}$ given in part (a) but now in the ordered basis $\mathcal{Q} = \big(\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = 1 + x,\ \boldsymbol{q}_2 = 1 + x + x^2,\big)$.

Solution:

Part (a):This is straightforward to compute, since $\boldsymbol{r}(x) = 3 + 2x + 4x^2$ implies that

$$\boldsymbol{r}(x) = 3\,\boldsymbol{p}_0(x) + 2\,\boldsymbol{p}_1(x) + 4\boldsymbol{p}_2(x) \quad \Leftrightarrow \quad \mathsf{r}_s = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}.$$

Part (b): This is more involved, as in Example 4.4.2. We look for numbers $\tilde{r}_1$, $\tilde{r}_2$, $\tilde{r}_3$ such that

$$\boldsymbol{r}(x) = \tilde{r}_0\,\boldsymbol{q}_0(x) + \tilde{r}_1\,\boldsymbol{q}_1(x) + \tilde{r}_2\,\boldsymbol{q}_2(x) \quad \Leftrightarrow \quad \mathsf{r}_q = \begin{bmatrix} \tilde{r}_0 \\ \tilde{r}_1 \\ \tilde{r}_2 \end{bmatrix}. \tag{4.6}$$

From the definition of the basis $\mathcal{Q}$ we know the components of the basis vectors in $\mathcal{Q}$ in terms of the $\mathcal{S}$ basis, that is,

$$\boldsymbol{q}_0(x) = \boldsymbol{p}_0(x) \quad \Leftrightarrow \quad \mathsf{q}_{0s} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

$$\boldsymbol{q}_1(x) = \boldsymbol{p}_0(x) + \boldsymbol{p}_1(x) \quad \Leftrightarrow \quad \mathsf{q}_{1s} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\boldsymbol{q}_2(x) = \boldsymbol{p}_0(x) + \boldsymbol{p}_1(x) + \boldsymbol{p}_2(x) \quad \Leftrightarrow \quad \mathsf{q}_{2s} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Now we can write the ordered basis $\mathcal{Q}$ in terms of the column vectors of the matrix $\mathsf{Q}_s = \big[\mathcal{Q}\big]_s = \big[\mathsf{q}_{0s}, \mathsf{q}_{1s}, \mathsf{q}_{2s}\big]$, as follows,

$$\mathsf{Q}_s = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Expressing Eq. (4.6) in the standard basis means

$$\begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} = \tilde{r}_0 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \tilde{r}_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \tilde{r}_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The last equation on the right is a matrix equation for the unknowns $\tilde{r}_0$, $\tilde{r}_1$, and $\tilde{r}_2$

$$\begin{bmatrix}1&1&1\\0&1&1\\0&0&1\end{bmatrix}\begin{bmatrix}\tilde{r}_0\\\tilde{r}_1\\\tilde{r}_2\end{bmatrix}=\begin{bmatrix}3\\2\\4\end{bmatrix} \quad\Leftrightarrow\quad \mathsf{Q}_s\mathsf{r}_q=\mathsf{r}_s.$$

We find the solution using the Gauss method,

$$\left[\begin{array}{ccc|c}1&1&1&3\\0&1&1&2\\0&0&1&4\end{array}\right]\to\left[\begin{array}{ccc|c}1&0&0&1\\0&1&0&-2\\0&0&1&4\end{array}\right],$$

hence the solution is

$$\mathsf{r}_q=\begin{bmatrix}1\\-2\\4\end{bmatrix} \quad\Leftrightarrow\quad \boldsymbol{r}(x)=\boldsymbol{q}_0(x)-2\,\boldsymbol{q}_1(x)+4\,\boldsymbol{q}_2(x).$$

We can verify that this is the solution, since

$$\begin{aligned}\boldsymbol{r}(x)&=\boldsymbol{q}_0(x)-2\,\boldsymbol{q}_1(x)+4\,\boldsymbol{q}_2(x)\\&=1-2\,(1+x)+4\,(1+x+x^2)\\&=(1-2+4)+(-2+4)x+4x^2\\&=3+2x+4x^2\\&=3\,\boldsymbol{p}_0(x)+2\,\boldsymbol{p}_1(x)+4\boldsymbol{p}_2(x).\end{aligned}$$

$\lhd$

Example 4.4.5: Given any ordered basis $\mathcal{U}=\big(\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3\big)$ of a 3-dimensional vector space V, find $\mathsf{U}_u=\big[\mathcal{U}\big]_u\subset\mathbb{F}^{3,3}$, that is, find $\mathsf{u}_i=[\boldsymbol{u}_i]_u$ for $i=1,2,3$, the components of the basis vectors $\boldsymbol{u}_i$ in its own basis $\mathcal{U}$.

Solution: The answer is simple: The definition of vector components in a basis says that

$$\boldsymbol{u}_1=\boldsymbol{u}_1+0\,\boldsymbol{u}_2+0\,\boldsymbol{u}_3 \quad\Leftrightarrow\quad \mathsf{u}_{1u}=\begin{bmatrix}1\\0\\0\end{bmatrix}=\mathsf{e}_1,$$

$$\boldsymbol{u}_2=0\,\boldsymbol{u}_1+\boldsymbol{u}_2+0\,\boldsymbol{u}_3 \quad\Leftrightarrow\quad \mathsf{u}_{2u}=\begin{bmatrix}0\\1\\0\end{bmatrix}=\mathsf{e}_2,$$

$$\boldsymbol{u}_3=0\,\boldsymbol{u}_1+0\,\boldsymbol{u}_2+\boldsymbol{u}_3 \quad\Leftrightarrow\quad \mathsf{u}_{3u}=\begin{bmatrix}0\\0\\1\end{bmatrix}=\mathsf{e}_3.$$

In other words, using the coordinate map $\phi_u:V\to\mathbb{F}^3$, we can always write any basis $\mathcal{U}$ as components in its own basis as follows, $\mathsf{U}_u=\big[\mathsf{e}_1,\mathsf{e}_2,\mathsf{e}_3\big]=\mathsf{I}_3$. This example says that there is nothing special about the standard basis $\mathcal{S}=\{\mathsf{e}_1,\cdots,\mathsf{e}_n\}$ of $\mathbb{F}^n$, where $\mathsf{e}_i=\mathsf{I}_{:i}$ is the i-th column of the identity matrix I_n. Given any n-dimensional vector space V over $\mathbb{F}$ with any ordered basis $\mathcal{V}$, the components of the basis vectors expressed on its own basis is always the standard basis of $\mathbb{F}^n$, that is, the result is always $\big[\mathcal{V}\big]_v=\mathsf{I}_n$. $\lhd$

4.4.3. **Exercises.**

4.4.1.- Let $\mathcal{S} = \big(\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3\big)$ be the standard basis of $\mathbb{R}^3$. Find the components of the vector $\mathsf{v} = \mathsf{e}_1 + \mathsf{e}_2 + 2\mathsf{e}_3$ in the ordered basis $\mathcal{U}$

$$\left(\mathsf{u}_{1s} = \begin{bmatrix}1\\0\\1\end{bmatrix}, \mathsf{u}_{2s} = \begin{bmatrix}0\\1\\1\end{bmatrix}, \mathsf{u}_{3s} = \begin{bmatrix}1\\1\\0\end{bmatrix}\right).$$

4.4.2.- Let $\mathcal{S} = \big(\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3\big)$ be the standard basis of $\mathbb{R}^3$. Find the components of the vector

$$\mathsf{v}_s = \begin{bmatrix}8\\7\\4\end{bmatrix}$$

in the ordered basis $\mathcal{U}$ given by

$$\left(\mathsf{u}_{1s} = \begin{bmatrix}1\\1\\1\end{bmatrix}, \mathsf{u}_{2s} = \begin{bmatrix}1\\2\\2\end{bmatrix}, \mathsf{u}_{3s} = \begin{bmatrix}1\\2\\3\end{bmatrix}\right).$$

4.4.3.- Consider the vector space $V = \mathbb{P}_2$ with the ordered basis $\mathcal{S}$ given by

$$\mathcal{S} = \big(\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2\big).$$

(a) Find the components of the polynomial $\boldsymbol{r}(x) = 2 + 3x - x^2$ in the ordered basis $\mathcal{S}$.

(b) Find the components of the same polynomial $\boldsymbol{r}$ given in part (a) but now in the ordered basis $\mathcal{Q}$ given by

$$\big(\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = 1 - x,\ \boldsymbol{q}_2 = x + x^2,\big).$$

4.4.4.- Let $\mathcal{S}$ be the standard ordered basis of $\mathbb{R}^{2,2}$, that is,

$$\mathcal{S} = (\mathsf{E}_{11}, \mathsf{E}_{12}, \mathsf{E}_{21}, \mathsf{E}_{22}) \subset \mathbb{R}^{2,2},$$

with

$$\mathsf{E}_{11} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \quad \mathsf{E}_{12} = \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix},$$

$$\mathsf{E}_{21} = \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \quad \mathsf{E}_{22} = \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}.$$

(a) Show that the ordered set $\mathcal{M}$ below is a basis of $\mathbb{R}^{2,2}$, where

$$\mathcal{M} = (\mathsf{M}_1, \mathsf{M}_2, \mathsf{M}_3, \mathsf{M}_4) \subset \mathbb{R}^{2,2},$$

with

$$\mathsf{M}_1 = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}, \quad \mathsf{M}_2 = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix},$$

$$\mathsf{M}_3 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, \quad \mathsf{M}_4 = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix},$$

where the matrices above are written in the standard basis.

(b) Consider the matrix A written in the standard basis $\mathcal{S}$,

$$\mathsf{A} = \begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}.$$

Find the components of the matrix A in the ordered basis $\mathcal{M}$.

Chapter 6. Inner product spaces

An inner product space is a vector space with an additional structure called inner product. This additional structure is an operation that associates each pair of vectors in the vector space with a scalar. An inner product extends to any vector space the main concepts included in the dot product, which is defined on $\mathbb{R}^n$. These main concepts include the length of a vector, the notioin of perpendicular vectors, and distance between vectors. When these ideas are introduced in function vector spaces, they allow to define the notion of convergence of an infinite sum of vectors. This, in turns, provides a way to evaluate the accuracy of approximate solutions to differential equations.

6.1. Dot product

6.1.1. Dot product in $\mathbb{R}^2$.

We review the definition of the dot product between vectors in $\mathbb{R}^2$, and we describe its main properties, including the Cauchy-Schwarz inequality. We then use the dot product to introduce the notion of length of a vector, distance and angle between vectors, including the special case of perpendicular vectors. We then review that all these notions can be generalized in a straightforward way from $\mathbb{R}^2$ to $\mathbb{F}^n$, $n \geqslant 1$.

Definition 6.1.1. *Given any vectors* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *with components* $\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\mathsf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ *in the standard ordered basis* $\mathcal{S}$. *The* ***dot product*** *on* $\mathbb{R}^2$ *with is the function* $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,

$$\mathsf{x} \cdot \mathsf{y} = x_1 y_1 + x_2 y_2.$$

The ***dot product norm*** *of a vector* $\mathsf{x} \in \mathbb{R}^2$ *is the value of the function* $\| \; \| : \mathbb{R}^2 \to \mathbb{R}$,

$$\|\mathsf{x}\| = \sqrt{\mathsf{x} \cdot \mathsf{x}}.$$

The ***norm distance*** *between* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *is the value of the function* $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,

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

The dot product can be expressed using the transpose of a vector components in the standard basis, as follows,

$$\mathsf{x}^T \mathsf{y} = [x_1, \; x_2] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_1 y_1 + x_2 y_2 = \mathsf{x} \cdot \mathsf{y}.$$

The dot product norm and the norm distance can be expressed in term of vector components in the standard ordered basis $\mathcal{S}$ as follows,

$$\|\mathsf{x}\| = \sqrt{(x_1)^2 + (x_2)^2}, \qquad d(\mathsf{x}, \mathsf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

The geometrical meaning of the norm and distance is clear form this expression in components, as is shown in Fig. 40. The norm of a vector is the Euclidean length from the origin point to the head point of the vector, while the distance between two vectors in the Euclidean distance between the head points of the two vectors.

It is important that we summarize the main properties of the dot product in $\mathbb{R}^2$, since they are the main guide to construct the generalizations of the dot product to other vector spaces.

Theorem 6.1.2. *The dot product on* $\mathbb{R}^2$ *satisfies, for every vector* x, y, $\mathsf{z} \in \mathbb{R}^2$ *and every scalar* a, $b \in \mathbb{R}$, *the following properties:*

(a) $\mathsf{x} \cdot \mathsf{y} = \mathsf{y} \cdot \mathsf{x}$, *(Symmetry);*
(b) $\mathsf{x} \cdot (a\mathsf{y} + b\mathsf{z}) = a\,(\mathsf{x} \cdot \mathsf{y}) + b\,(\mathsf{x} \cdot \mathsf{z})$, *(Linearity on the second argument);*
(c) $\mathsf{x} \cdot \mathsf{x} \geqslant 0$, *and* $\mathsf{x} \cdot \mathsf{x} = 0$ *iff* $\mathsf{x} = 0$, *(Positive definiteness).*

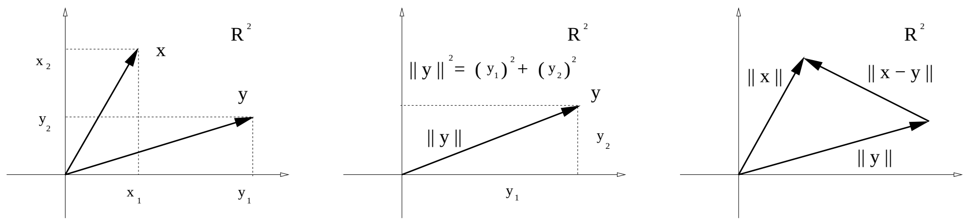

Figure 40. Example of the Euclidean notions of vector length and distance between vectors in $\mathbb{R}^2$.

Proof of Theorem 6.1.2: These properties are simple to obtain from the definition of the dot product.

Part (a): It is simple to see that

$$\mathsf{x} \cdot \mathsf{y} = x_1 y_1 + x_2 y_2 = y_1 x_1 + y_2 x_2 = \mathsf{y} \cdot \mathsf{x}.$$

Part (b): It is also simple to see that

$$\begin{aligned} \mathsf{x} \cdot (a\mathsf{y} + b\mathsf{z}) &= x_1(ay_1 + bz_1) + x_2(ay_2 + bz_2) \\ &= a\,(x_1 y_1 + x_2 y_2) + b\,(x_1 z_1 + x_2 z_2) \\ &= a\,(\mathsf{x} \cdot \mathsf{y}) + b\,(\mathsf{x} \cdot \mathsf{z}). \end{aligned}$$

Part (c): This follows from

$$\mathsf{x} \cdot \mathsf{x} = (x_1)^2 + (x_2)^2 \geqslant 0;$$

furthermore, in the case $\mathsf{x} \cdot \mathsf{x} = 0$ we obtain that

$$(x_1)^2 + (x_2)^2 = 0 \quad \Leftrightarrow \quad x_1 = x_2 = 0.$$

This establishes the Theorem. □

These simple properties are crucial to establish the following result, known as Cauchy-Schwarz inequality for the dot product in $\mathbb{R}^2$. This inequality allows to express the dot product of two vectors in $\mathbb{R}^2$ in terms of the angle between the vectors.

Theorem 6.1.3 (Cauchy-Schwarz). *The properties (a)-(c) in Theorem 6.1.2 imply that for all* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *holds*

$$|\mathsf{x} \cdot \mathsf{y}| \leqslant \|\mathsf{x}\|\,\|\mathsf{y}\|.$$

Proof of Theorem 6.1.3: From the positive definiteness property we know that the following inequality holds for all $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ and for all $a \in \mathbb{R}$,

$$0 \leqslant \|a\mathsf{x} - \mathsf{y}\|^2 = (a\mathsf{x} - \mathsf{y}) \cdot (a\mathsf{x} - \mathsf{y}).$$

The symmetry and the linearity on the second argument imply

$$0 \leqslant (a\mathsf{x} - \mathsf{y}) \cdot (a\mathsf{x} - \mathsf{y}) = a^2\,\|\mathsf{x}\|^2 - 2a\,(\mathsf{x} \cdot \mathsf{y}) + \|\mathsf{y}\|^2. \tag{6.1}$$

Since the inequality above holds for all $a \in \mathbb{R}$, let us choose a particular value of a, the solution of the equation

$$a\,\|\mathsf{x}\|^2 - (\mathsf{x} \cdot \mathsf{y}) = 0 \quad \Rightarrow \quad a = \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\|^2}.$$

Introduce this particular value of a into Eq. (6.1),

$$0 \leqslant -\Big(\frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\|^2}\Big)(\mathsf{x} \cdot \mathsf{y}) + \|\mathsf{y}\|^2 \quad \Rightarrow \quad |\mathsf{x} \cdot \mathsf{y}|^2 \leqslant \|\mathsf{x}\|^2\,\|\mathsf{y}\|^2.$$

This establishes the Theorem. □

The Cauchy-Schwarz inequality implies that we can express the dot product of two vectors in an alternative and more geometrical way, in terms of an angle related with the two vectors. The Cauchy-Schwarz inequality says

$$-1 \leqslant \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\| \, \|\mathsf{y}\|} \leqslant 1,$$

which suggests that the number $(\mathsf{x} \cdot \mathsf{y})/(\|\mathsf{x}\| \, \|\mathsf{y}\|)$ can be expressed as a sine or a cosine of an appropriate angle.

Theorem 6.1.4. *The **angle** between vectors* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ *is the number* $\theta \in [0, \pi]$ *given by*

$$\cos(\theta) = \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\| \, \|\mathsf{y}\|}.$$

Proof of Theorem 6.1.4: It is not difficult to see that given any vectors $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$, the vectors $\mathsf{x}/\|\mathsf{x}\|$ and $\mathsf{y}/\|\mathsf{y}\|$ have unit norm. Indeed,

$$\Big\| \frac{\mathsf{x}}{\|\mathsf{x}\|} \Big\|^2 = \frac{(x_1)^2}{\|\mathsf{x}\|^2} + \frac{(x_2)^2}{\|\mathsf{x}\|^2} = \frac{1}{\|\mathsf{x}\|^2} \big[(x_1)^2 + (x_2)^2\big] = 1.$$

The same holds for the vector $\mathsf{y}/\|\mathsf{y}\|$. The expression

$$\frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\| \, \|\mathsf{y}\|} = \frac{\mathsf{x}}{\|\mathsf{x}\|} \cdot \frac{\mathsf{y}}{\|\mathsf{y}\|},$$

shows that the number $(\mathsf{x} \cdot \mathsf{y})/(\|\mathsf{x}\| \, \|\mathsf{y}\|)$ is the inner product of two vectors in the unit circle, as shown in Fig. 41.

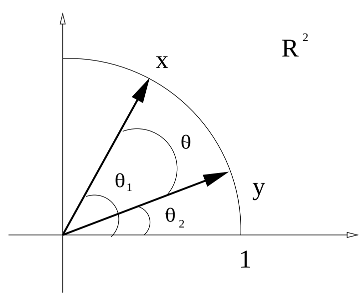

FIGURE 41. The dot product of two vectors $\mathsf{x}, \mathsf{y} \in \mathbb{R}^2$ can be expressed in terms of the angle $\theta = \theta_1 - \theta_2$ between the vectors.

Therefore, we know that

$$\frac{\mathsf{x}}{\|\mathsf{x}\|} = \begin{bmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{bmatrix}, \qquad \frac{\mathsf{y}}{\|\mathsf{y}\|} = \begin{bmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{bmatrix}.$$

Their dot product is given by

$$\frac{\mathsf{x}}{\|\mathsf{x}\|} \cdot \frac{\mathsf{y}}{\|\mathsf{y}\|} = \big[\cos(\theta_1), \, \sin(\theta_1)\big] \begin{bmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{bmatrix} = \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2).$$

Using the formula $\cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2) = \cos(\theta_1 - \theta_2)$, and denoting the angle between the vectors by $\theta = \theta_1 - \theta_2$, we conclude that

$$\frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\| \, \|\mathsf{y}\|} = \cos(\theta).$$

This establishes the Theorem. □

Recall the notion of perpendicular vectors.

Definition 6.1.5. *The vectors* x, $\mathsf{y} \in \mathbb{R}^2$ *are* ***orthogonal****, denoted as* $\mathsf{x} \perp \mathsf{y}$*, iff the angle* $\theta \in [0, \pi]$ *between the vectors is* $\theta = \pi/2$.

The notion of orthogonal vectors in Def. 6.1.5 can be expressed in terms of the dot product, and it is equivalent to Pythagoras Theorem on right triangles.

Theorem 6.1.6. *Let* x, $\mathsf{y} \in \mathbb{R}^2$ *be non-zero vectors, then the following statement holds,*

$$\mathsf{x} \perp \mathsf{y} \quad \Leftrightarrow \quad \mathsf{x} \cdot \mathsf{y} = 0 \quad \Leftrightarrow \quad \|\mathsf{x} - \mathsf{y}\|^2 = \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2.$$

Proof of Theorem 6.1.6: The non-zero vectors x and $\mathsf{y} \in \mathbb{R}^2$ are orthogonal iff $\theta = \pi/2$, which is equivalent to

$$\frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\| \, \|\mathsf{y}\|} = 0 \quad \Leftrightarrow \quad \mathsf{x} \cdot \mathsf{y} = 0.$$

The last part of the Proposition comes from the following calculation,

$$\begin{aligned} \|\mathsf{x} - \mathsf{y}\|^2 &= (x_1 - y_1)^2 + (x_2 - y_2)^2 \\ &= (x_1)^2 + (x_2)^2 + (y_1)^2 + (y_2)^2 - 2(x_1 y_1 + x_2 y_2) \\ &= \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2 - 2\,\mathsf{x} \cdot \mathsf{y}. \end{aligned}$$

Hence, $\mathsf{x} \perp \mathsf{y}$ iff $\mathsf{x} \cdot \mathsf{y} = 0$ iff Pythagoras Theorem holds for the triangle with sides given by x, y and hypotenuse $\mathsf{x} - \mathsf{y}$. This establishes the Theorem. □

EXAMPLE 6.1.1: Find the length of the vectors $\mathsf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathsf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, the angle between them, and then find a non-zero vector z orthogonal to x.

SOLUTION: We first find the length, that is, the norms of x and y,

$$\|\mathsf{x}\|^2 = \mathsf{x} \cdot \mathsf{x} = \mathsf{x}^T \mathsf{x} = \begin{bmatrix} 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 + 4 \quad \Rightarrow \quad \|\mathsf{x}\| = \sqrt{5},$$

$$\|\mathsf{y}\|^2 = \mathsf{y} \cdot \mathsf{y} = \mathsf{y}^T \mathsf{y} = \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 9 + 1 \quad \Rightarrow \quad \|\mathsf{y}\| = \sqrt{10}.$$

We now find the angle between x and y,

$$\cos(\theta) = \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\| \, \|\mathsf{y}\|} = \frac{\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix}}{\sqrt{5}\,\sqrt{10}} = \frac{5}{5\sqrt{2}} = \frac{1}{\sqrt{2}} \quad \Rightarrow \quad \theta = \frac{\pi}{4}.$$

We now find z such that $\mathsf{z} \perp \mathsf{x}$, that is,

$$0 = \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = z_1 + 2z_2 \quad \Rightarrow \quad \left\{ \begin{array}{l} z_1 = -2z_2 \\ z_2 \text{ free variable} \end{array} \right\} \quad \Rightarrow \quad \mathsf{z} = \begin{bmatrix} -2 \\ 1 \end{bmatrix} z_2.$$

◁

6.1.2. **Dot product in $\mathbb{F}^n$.** The notion of dot product reviewed above can be generalized in a straightforward way from $\mathbb{R}^2$ to $\mathbb{F}^n$, $n \geqslant 1$, where $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Definition 6.1.7. *The* ***dot product*** *on the vector space* $\mathbb{F}^n$*, with* $n \geqslant 1$*, is the function* $\cdot : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ *given by*

$$\mathsf{x} \cdot \mathsf{y} = \mathsf{x}^* \mathsf{y},$$

where x, y *denote components in the standard basis of* $\mathbb{F}^n$*. The* ***dot product norm*** *of a vector* $\mathsf{x} \in \mathbb{F}^n$ *is the value of the function* $\| \; \| : \mathbb{F}^n \to \mathbb{R}$*,*

$$\|\mathsf{x}\| = \sqrt{\mathsf{x} \cdot \mathsf{x}}.$$

*The **norm distance** between* x, $\mathsf{y} \in \mathbb{F}^n$ *is the value of the function* $d : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{R}$,

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

The vectors x, $\mathsf{y} \in \mathbb{F}^n$ *are **orthogonal**, denoted as* $\mathsf{x} \perp \mathsf{y}$, *iff holds* $\mathsf{x}\cdot\mathsf{y} = 0$.

Notice that we defined two vectors to be orthogonal by the condition that their dot product vanishes. This is the appropriate generalization to $\mathbb{F}^n$ of the ideas we saw in $\mathbb{R}^2$. The concept of angle is more difficult to study. In the case that $\mathbb{F} = \mathbb{C}$ is not clear what the angle between vectors mean. In the case $\mathbb{F} = \mathbb{R}$ and $n > 3$ we have to define angle by the number $(\mathsf{x}\cdot\mathsf{y})/(\|\mathsf{x}\|\,\|\mathsf{y}\|)$. This will be done after we prove the Cauchy-Schwarz inequality, which then is used to show that the number $(\mathsf{x}\cdot\mathsf{y})/(\|\mathsf{x}\|\,\|\mathsf{y}\|) \in [-1,1]$. The formulas above for the dot product, norm and distance can be expressed in terms of the vector components in the standard basis as follows,

$$\mathsf{x}\cdot\mathsf{y} = \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n,$$
$$\|\mathsf{x}\| = \sqrt{|x_1|^2 + \cdots + |x_n|^2},$$
$$d(\mathsf{x},\mathsf{y}) = \sqrt{|x_1-y_1|^2 + \cdots + |x_n-y_n|^2},$$

where we used the standard notation $\mathsf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $\mathsf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, and $|x_i|^2 = \overline{x}_i x_i$, for $i = 1,\cdots n$. In the particular case that $\mathbb{F} = \mathbb{R}$ all the vector components are real numbers, so $\overline{x}_i = x_i$.

EXAMPLE 6.1.2: Find whether x is orthogonal to y and/or z, where

$$\mathsf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \quad \mathsf{y} = \begin{bmatrix} -5 \\ 4 \\ -3 \\ 2 \end{bmatrix}, \quad \mathsf{z} = \begin{bmatrix} -4 \\ -3 \\ 2 \\ 1 \end{bmatrix}.$$

SOLUTION: We need to compute the dot products $\mathsf{x}\cdot\mathsf{y}$ and $\mathsf{x}\cdot\mathsf{z}$. We obtain

$$\mathsf{x}^T\mathsf{y} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} -5 \\ 4 \\ -3 \\ 2 \end{bmatrix} = -5+8-9+8 \quad\Rightarrow\quad \mathsf{x}\cdot\mathsf{y} = 2 \quad\Rightarrow\quad \mathsf{x} \not\perp \mathsf{y},$$

$$\mathsf{x}^T\mathsf{z} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} -4 \\ -3 \\ 2 \\ 1 \end{bmatrix} = -4-6+6+4 \quad\Rightarrow\quad \mathsf{x}\cdot\mathsf{z} = 0 \quad\Rightarrow\quad \mathsf{x} \perp \mathsf{z}.$$

◁

EXAMPLE 6.1.3: Find $\mathsf{x}\cdot\mathsf{y}$, where $\mathsf{x} = \begin{bmatrix} 2+3i \\ i \\ 1-i \end{bmatrix}$ and $\mathsf{y} = \begin{bmatrix} 2i \\ 1 \\ 1+3i \end{bmatrix}$.

SOLUTION: The first product $\mathsf{x}\cdot\mathsf{y}$ is given by

$$\mathsf{x}^*\mathsf{y} = \begin{bmatrix} 2-3i & -i & 1+i \end{bmatrix} \begin{bmatrix} 2i \\ 1 \\ 1+3i \end{bmatrix} = (2-3i)(2i) - i + (1+i)(1+3i),$$

so $\mathsf{x}\cdot\mathsf{y} = 4i + 6 - i + 1 - 3 + i + 3i$, that is, $\mathsf{x}\cdot\mathsf{y} = 4 + 7i$. ◁

The dot product satisfies the following properties.

Theorem 6.1.8. *The dot product on $\mathbb{F}^n$, with $n \geqslant 1$, satisfies for every vector x, y, $\mathsf{z} \in \mathbb{F}^n$ and every scalar a, $b \in \mathbb{F}$, the following properties:*

(a1) $\mathsf{x} \cdot \mathsf{y} = \mathsf{y} \cdot \mathsf{x}$, *(Symmetry $\mathbb{F} = \mathbb{R}$);*
(a2) $\mathsf{x} \cdot \mathsf{y} = \overline{\mathsf{y} \cdot \mathsf{x}}$, *(Conjugate symmetry, for $\mathbb{F} = \mathbb{C}$);*
(b) $\mathsf{x} \cdot (a\mathsf{y} + b\mathsf{z}) = a\,(\mathsf{x} \cdot \mathsf{y}) + b\,(\mathsf{x} \cdot \mathsf{z})$, *(Linearity on the second argument);*
(c) $\mathsf{x} \cdot \mathsf{x} \geqslant 0$, *and* $\mathsf{x} \cdot \mathsf{x} = 0$ *iff* $\mathsf{x} = 0$, *(Positive definiteness).*

Proof of Theorem 6.1.8: Use the expression of the dot product in terms of the vector components. The property in (a1) can be established as follows,

$$\mathsf{x} \cdot \mathsf{y} = x_1 y_1 + \cdots + x_n y_n = y_1 x_1 + \cdots + y_n x_n = \mathsf{y} \cdot \mathsf{x}.$$

The property in (a2) can be established as follows,

$$\mathsf{x} \cdot \mathsf{y} = \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n = \overline{(\overline{y}_1 x_1 + \cdots + \overline{y}_n x_n)} = \overline{\mathsf{y} \cdot \mathsf{x}}.$$

The property in (b) is shown in a similar way,

$$\mathsf{x} \cdot (a\mathsf{y} + b\mathsf{z}) = \mathsf{x}^*(a\,\mathsf{y} + b\,\mathsf{z}) = a\,\mathsf{x}^*\,\mathsf{y} + b\,\mathsf{x}^*\,\mathsf{z} = a\,(\boldsymbol{x} \cdot \boldsymbol{y}) + b\,(\boldsymbol{x} \cdot \boldsymbol{z}).$$

The property in (c) follows from

$$\mathsf{x} \cdot \mathsf{x} = \mathsf{x}^*\,\mathsf{x} = |x_1|^2 + \cdots + |x_n|^2 \geqslant 0;$$

furthermore, in the case that $\mathsf{x} \cdot \mathsf{x} = 0$ we obtain that

$$|x_1|^2 + \cdots + |x_n|^2 = 0 \quad \Leftrightarrow \quad x_1 = \cdots = x_n = 0 \quad \Leftrightarrow \quad \mathsf{x} = 0.$$

This establishes the Theorem. □

The positive definiteness property (c) above shows that the dot product norm is indeed a real-valued and not a complex-valued function, since $\mathsf{x} \cdot \mathsf{x} \geqslant 0$ implies that $\|\mathsf{x}\| = \sqrt{\mathsf{x} \cdot \mathsf{x}} \in \mathbb{R}$. In the case of $\mathbb{F} = \mathbb{R}$, the symmetry property and the linearity in the second argument property imply that the dot product on $\mathbb{R}^n$ is also linear in the first argument. This is a reason to call the dot product on $\mathbb{R}^n$ a bilinear form. Finally, notice that in the case $\mathbb{F} = \mathbb{C}$, the conjugate symmetry property and the linearity in the second argument imply that *the dot product on $\mathbb{C}^n$ is conjugate linear on the first argument*. The proof is the following:

$$(a\mathsf{y} + b\mathsf{z}) \cdot \mathsf{x} = \overline{\mathsf{x} \cdot (a\mathsf{y} + b\mathsf{z})} = \overline{a\,(\mathsf{x} \cdot \mathsf{y}) + b\,(\mathsf{x} \cdot \mathsf{z})} = \overline{a}\,\overline{(\mathsf{x} \cdot \mathsf{y})} + \overline{b}\,\overline{(\mathsf{x} \cdot \mathsf{z})},$$

that is, for all x, y, $\mathsf{z} \in \mathbb{C}^n$ and all a, $b \in \mathbb{C}$ holds

$$(a\mathsf{y} + b\mathsf{z}) \cdot \mathsf{x} = \overline{a}\,(\mathsf{y} \cdot \mathsf{x}) + \overline{b}\,(\mathsf{z} \cdot \mathsf{x}).$$

Hence we say that the dot product on $\mathbb{C}^n$ is conjugate linear in the first argument.

EXAMPLE 6.1.4: Compute the dot product of $\mathsf{x} = \begin{bmatrix} 2+3i \\ 6i-9 \end{bmatrix}$ with $\mathsf{y} = \begin{bmatrix} 3i \\ 2 \end{bmatrix}$.

SOLUTION: This is a straightforward computation

$$\mathsf{x} \cdot \mathsf{y} = \mathsf{x}^*\mathsf{y} = \begin{bmatrix} 2-3i & -6i-9 \end{bmatrix} \begin{bmatrix} 3i \\ 2 \end{bmatrix} = 6i + 9 - 12i - 18 \quad \Rightarrow \quad \mathsf{x} \cdot \mathsf{y} = -9 - 6i.$$

Notice that $\mathsf{x} = (2+3i)\hat{\mathsf{x}}$, with $\hat{\mathsf{x}} = \begin{bmatrix} 1 \\ 3i \end{bmatrix}$, so we could use the conjugate linearity in the first argument to compute

$$\mathsf{x} \cdot \mathsf{y} = \big((2+3i)\hat{\mathsf{x}}\big) \cdot \mathsf{y} = (2-3i)\,(\hat{\mathsf{x}} \cdot \mathsf{y}) = (2-3i)\begin{bmatrix} 1 & -3i \end{bmatrix} \begin{bmatrix} 3i \\ 2 \end{bmatrix} = (2-3i)(3i-6i),$$

and we obtain the same result, $\mathsf{x} \cdot \mathsf{y} = -9 - 6i$. Finally, notice that $\mathsf{y} \cdot \mathsf{x} = -9 + 6i$. ◁

An important result is that the dot product in $\mathbb{F}^n$ satisfies the Cauchy-Schwarz inequality.

Theorem 6.1.9 (Cauchy-Schwarz). *The properties (a1)-(c) in Theorem 6.1.8 imply that for all* $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ *holds*

$$|\mathsf{x} \cdot \mathsf{y}| \leqslant \|\mathsf{x}\| \, \|\mathsf{y}\|.$$

REMARK: The proof of the Cauchy-Schwarz inequality only uses the three properties of the dot product presented in Theorem 6.1.8. Any other function $f : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ having these three properties also satisfies the Cauchy-Schwarz inequality.

Proof of Theorem 6.1.9: From the positive definiteness property we know that the following inequality holds for all $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ and for all $a \in \mathbb{F}$,

$$0 \leqslant \|a\mathsf{x} - \mathsf{y}\|^2 = (a\mathsf{x} - \mathsf{y}) \cdot (a\mathsf{x} - \mathsf{y}).$$

The symmetry and the linearity on the second argument imply

$$0 \leqslant (a\mathsf{x} - \mathsf{y}) \cdot (a\mathsf{x} - \mathsf{y}) = a\,\overline{a}\,\|\mathsf{x}\|^2 - \overline{a}\,(\mathsf{x} \cdot \mathsf{y}) - a\,(\mathsf{y} \cdot \mathsf{x}) + \|\mathsf{y}\|^2. \tag{6.2}$$

Since the inequality above holds for all $a \in \mathbb{F}$, let us choose a particular value of a, the solution of the equation

$$a\,\overline{a}\,\|\mathsf{x}\|^2 - \overline{a}\,(\mathsf{x} \cdot \mathsf{y}) = 0 \quad \Rightarrow \quad a = \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\|^2}.$$

Introduce this particular value of a into Eq. (6.2),

$$0 \leqslant -\Big(\frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\|^2}\Big)(\overline{\mathsf{x} \cdot \mathsf{y}}) + \|\mathsf{y}\|^2 \quad \Rightarrow \quad |\mathsf{x} \cdot \mathsf{y}|^2 \leqslant \|\mathsf{x}\|^2 \, \|\mathsf{y}\|^2.$$

This establishes the Theorem. □

In the case $\mathbb{F} = \mathbb{R}$, the Cauchy-Schwarz inequality in $\mathbb{R}^n$ implies that the number $(\mathsf{x} \cdot \mathsf{y})/(\|\mathsf{x}\|\,\|\mathsf{y}\|) \in [-1, 1]$, which is a necessary and sufficient condition for the following definition of angle between two vectors in $\mathbb{R}^n$.

Definition 6.1.10. *The* ***angle*** *between vectors* $\mathsf{x}, \mathsf{y} \in \mathbb{R}^n$ *is the number* $\theta \in [0, \pi]$ *given by*

$$\cos(\theta) = \frac{\mathsf{x} \cdot \mathsf{y}}{\|\mathsf{x}\|\,\|\mathsf{y}\|}.$$

The dot product norm function in Definition 6.1.7 satisfies the following properties.

Theorem 6.1.11. *The dot product norm function on* $\mathbb{F}^n$*, with* $n \geqslant 1$*, satisfies for every vector* $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ *and every scalar* $a \in \mathbb{F}$ *the following properties:*

(a) $\|\mathsf{x}\| \geqslant 0$*, and* $\|\mathsf{x}\| = 0$ *iff* $\mathsf{x} = 0$*, (Positive definiteness);*
(b) $\|a\mathsf{x}\| = |a|\,\|\mathsf{x}\|$*, (Scaling);*
(c) $\|\mathsf{x} + \mathsf{y}\| \leqslant \|\mathsf{x}\| + \|\mathsf{y}\|$*, (Triangle inequality).*

Proof of Theorem 6.1.11: Properties (a) and (b) are straightforward to show from the definition of dot product, and their proof is left as an exercise. We show here how to obtain the triangle inequality, property (c). The proof uses the Cauchy-Schwarz inequality presented in Theorem 6.1.9. Given any vectors $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ holds

$$\begin{aligned} \|\mathsf{x} + \mathsf{y}\|^2 &= (\mathsf{x} + \mathsf{y}) \cdot (\mathsf{x} + \mathsf{y}) \\ &= \|\mathsf{x}\|^2 + (\mathsf{x} \cdot \mathsf{y}) + (\mathsf{y} \cdot \mathsf{x}) + \|\mathsf{y}\|^2 \\ &\leqslant \|\mathsf{x}\|^2 + 2\,|\mathsf{x} \cdot \mathsf{y}| + \|\mathsf{y}\|^2 \\ &\leqslant \|\mathsf{x}\|^2 + 2\,\|\mathsf{x}\|\,\|\mathsf{y}\| + \|\mathsf{y}\|^2 = \big(\|\mathsf{x}\| + \|\mathsf{y}\|\big)^2, \end{aligned}$$

We conclude that $\|\mathsf{x} + \mathsf{y}\|^2 \leqslant \big(\|\mathsf{x}\| + \|\mathsf{y}\|\big)^2$. This establishes the Theorem. □

A vector $\mathsf{v} \in \mathbb{F}^n$ is called *normal or unit vector* iff $\|\mathsf{v}\| = 1$. Examples of unit vectors are the standard basis vectors. Unit vectors parallel to a given vector are simple to find.

Theorem 6.1.12. *If* $\mathsf{v} \in \mathbb{F}^n$ *is non-zero, then* $\dfrac{\mathsf{v}}{\|\mathsf{v}\|}$ *is a unit vector parallel to* v*.*

Proof of Theorem 6.1.12: Notice that $\mathsf{u} = \dfrac{\mathsf{v}}{\|\mathsf{v}\|}$ is parallel to v, and it is straightforward to check that u is a unit vector, since

$$\|\mathsf{u}\| = \left\| \frac{\mathsf{v}}{\|\mathsf{v}\|} \right\| = \frac{1}{\|\mathsf{v}\|} \|\mathsf{v}\| = 1.$$

This establishes the Theorem. □

EXAMPLE 6.1.5: Find a unit vector parallel to $\mathsf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

SOLUTION: First compute the norm of x,

$$\|\mathsf{x}\| = \sqrt{1+4+9} = \sqrt{14},$$

therefore $\mathsf{u} = \dfrac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is a unit vector parallel to v. ◁

6.1.3. Exercises.

6.1.1.- Consider the vector space $\mathbb{R}^4$ with standard basis $\mathcal{S}$ and dot product. Find the norm of u and v, their distance and the angle between them, where

$$\mathsf{u} = \begin{bmatrix} 2 \\ 1 \\ -4 \\ -2 \end{bmatrix}, \quad \mathsf{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}.$$

6.1.2.- Use the dot product on $\mathbb{R}^2$ to find two unit vectors orthogonal to

$$\mathsf{x} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

6.1.3.- Use the dot product on $\mathbb{C}^2$ to find a unit vector parallel to

$$\mathsf{x} = \begin{bmatrix} 1+2i \\ 2-i \end{bmatrix}.$$

6.1.4.- Consider the vector space $\mathbb{R}^2$ with the dot product.

(a) Give an example of a linearly independent set $\{\mathsf{x}, \mathsf{y}\}$ with $\mathsf{x} \not\perp \mathsf{y}$.

(b) Give an example of a linearly dependent set $\{\mathsf{x}, \mathsf{y}\}$ with $\mathsf{x} \perp \mathsf{y}$.

6.1.5.- Consider the vector space $\mathbb{F}^n$ with the dot product, and let Re denote the real part of a complex number. Show that for all $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ holds

$$\|\mathsf{x} - \mathsf{y}\|^2 = \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2 - 2\mathsf{Re}(\mathsf{x} \cdot \mathsf{y}).$$

6.1.6.- Use the result in Exercise **6.1.5** above to prove the following generalizations of the Pythagoras Theorem to $\mathbb{F}^n$ with the dot product.

(a) For $\mathsf{x}, \mathsf{y} \in \mathbb{R}^n$ holds

$$\mathsf{x} \perp \mathsf{y} \quad \Leftrightarrow \quad \|\mathsf{x} + \mathsf{y}\|^2 = \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2.$$

(b) For $\mathsf{x}, \mathsf{y} \in \mathbb{C}^n$ holds

$$\mathsf{x} \perp \mathsf{y} \quad \Rightarrow \quad \|\mathsf{x} + \mathsf{y}\|^2 = \|\mathsf{x}\|^2 + \|\mathsf{y}\|^2.$$

6.1.7.- Prove that the parallelogram law holds for the dot product norm in $\mathbb{F}^n$, that is, show that for all $\mathsf{x}, \mathsf{y} \in \mathbb{F}^n$ holds

$$\|\mathsf{x} + \mathsf{y}\|^2 + \|\mathsf{x} - \mathsf{y}\|^2 = 2\|\mathsf{x}\|^2 + 2\|\mathsf{y}\|^2.$$

This law states that the sum of the squares of the lengths of the four sides of a parallelogram formed by x and y equals the sum of the square of the lengths of the two diagonals.

6.2. Inner product

6.2.1. **Inner product.** An inner product on a vector space is a generalization of the dot product on $\mathbb{R}^n$ or $\mathbb{C}^n$ introduced in Sect. 6.1. The inner product is not defined with a particular formula, or requiring a particular basis in the vector space. Instead, the inner product is defined by a list of properties that must satisfy. We did something similar when we introduced the concept of a vector space. In that case we defined a vector space as a set of any kind of elements where linear combinations are possible, instead of defining the set by explicitly giving its elements.

Definition 6.2.1. *Let V be a vector space over the scalar field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A function $\langle\,,\rangle : V \times V \to \mathbb{F}$ is called an **inner product** iff for every $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z} \in V$ and every a, $b \in \mathbb{F}$ the function $\langle\,,\rangle$ satisfies:*

(a1) $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{y}, \boldsymbol{x}\rangle$, *(Symmetry, for $\mathbb{F} = \mathbb{R}$);*
(a2) $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}$, *(Conjugate symmetry, for $\mathbb{F} = \mathbb{C}$);*
(b) $\langle \boldsymbol{x}, (a\boldsymbol{y} + b\boldsymbol{z})\rangle = a\langle \boldsymbol{x}, \boldsymbol{y}\rangle + b\langle \boldsymbol{x}, \boldsymbol{z}\rangle$, *(Linearity on the second argument);*
(c) $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geqslant 0$, *and* $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$ *iff* $\boldsymbol{x} = 0$, *(Positive definiteness).*

*An **inner product space** is a pair $\big(V, \langle\,,\rangle\big)$ of a vector space with an inner product.*

Different inner products can be defined on a given vector space. The dot product is an inner product in $\mathbb{F}^n$. A different inner product can be defined in $\mathbb{F}^n$, as can be seen in the following example.

Example 6.2.1: We show that $\mathbb{R}^n$ can have different inner products.

(a) The dot product on $\mathbb{R}^n$ is an inner product, since the expression

$$\langle \mathsf{x}, \mathsf{y}\rangle = \mathsf{x}_s^T \mathsf{y}_s = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}_s \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_s = x_1 y_1 + \cdots + x_n y_n, \tag{6.3}$$

satisfies all the properties in Definition 6.2.1, with $\mathcal{S}$ the standard ordered basis in $\mathbb{R}^n$.

(b) A different inner product in $\mathbb{R}^n$ can be introduced by a formula similar to the one in Eq. (6.3) by choosing a different ordered basis. If $\mathcal{U}$ is any ordered basis of $\mathbb{R}^n$, then

$$\langle \mathsf{x}, \mathsf{y}\rangle = \mathsf{x}_u^T \mathsf{y}_u.$$

defines an inner product on $\mathbb{R}^n$. The inner product defined using the basis $\mathcal{U}$ is not equal to the inner product defined using the standard basis $\mathcal{S}$. Let $\mathsf{P} = \mathsf{I}_{us}$ be the change of basis matrix, then we know that $\mathsf{x}_u = \mathsf{P}^{-1}\mathsf{x}_s$. The inner product above can be expressed in terms of the $\mathcal{S}$ basis as follows,

$$\langle \mathsf{x}, \mathsf{y}\rangle = \mathsf{x}_s^T \mathsf{M}\, \mathsf{y}_s, \qquad \mathsf{M} = \big(\mathsf{P}^{-1}\big)^T \big(\mathsf{P}^{-1}\big),$$

and in general, $\mathsf{M} \neq I_n$. Therefore, the inner product above is not equal to the dot product. Also, see Example 6.2.2.

◁

Example 6.2.2: Let $\mathcal{S}$ be the standard ordered basis in $\mathbb{R}^2$, and introduce the ordered basis $\mathcal{U}$ as the following rescaling of $\mathcal{S}$,

$$\mathcal{U} = \Big(\boldsymbol{u}_1 = \frac{1}{2}\,\boldsymbol{e}_1, \boldsymbol{u}_2 = \frac{1}{3}\,\boldsymbol{e}_2\Big).$$

Express the inner product $\langle \mathsf{x}, \mathsf{y}\rangle = \mathsf{x}_u^T \mathsf{y}_u$ in terms of x_s and y_s. Is this inner product the same as the dot product?

SOLUTION: The definition of the inner product says that $\langle \mathsf{x}, \mathsf{y} \rangle = \mathsf{x}_u^T \mathsf{y}_u$. Introducing the notation $\mathsf{x}_u = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}_u$ and $\mathsf{y}_u = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{bmatrix}_u$, we obtain the usual expression $\langle \mathsf{x}, \mathsf{y} \rangle = \tilde{x}_1 \tilde{y}_1 + \tilde{x}_2 \tilde{y}_2$. The components x_u and x_s are related by the change of basis formula

$$\mathsf{x}_u = \mathsf{P}^{-1}\mathsf{x}_s, \quad \mathsf{P} = \mathsf{I}_{us} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}_{us} \quad \Rightarrow \quad \mathsf{P}^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}_{su} = \left(\mathsf{P}^{-1}\right)^T.$$

Therefore,

$$\langle \mathsf{x}, \mathsf{y} \rangle = \mathsf{x}_u^T \mathsf{y}_u = \mathsf{x}_s^T \left(\mathsf{P}^{-1}\right)^T \mathsf{P}^{-1} \mathsf{y}_s = \begin{bmatrix} x_1, & x_2 \end{bmatrix}_s \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}_{su} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_s$$

where we used the standard notation $\mathsf{x}_s = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_s$ and $\mathsf{y}_s = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_s$. We conclude that

$$\langle \mathsf{x}, \mathsf{y} \rangle = \mathsf{x}_s^T \left(\mathsf{P}^{-1}\right)^2 \mathsf{y}_s \quad \Leftrightarrow \quad \langle \mathsf{x}, \mathsf{y} \rangle = 4x_1 y_1 + 9x_2 y_2.$$

The inner product $\langle \mathsf{x}, \mathsf{y} \rangle = \mathsf{x}_u^T \mathsf{y}_u$ is different from the dot product $\mathsf{x} \cdot \mathsf{y} = \mathsf{x}_s^T \mathsf{y}_s$. ◁

EXAMPLE 6.2.3: Determine whether the function $\langle \, , \rangle : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ below is an inner product in $\mathbb{R}^3$, where

$$\langle \mathsf{x}, \mathsf{y} \rangle = x_1 y_1 + x_2 y_2 + x_3 y_3 + 3x_1 y_2 + 3x_2 y_1.$$

SOLUTION: The function $\langle \, , \rangle$ seems to be symmetric and linear. It is not so clear whether this function is positive, because of the presence of crossed terms. So, before spending time to prove the symmetry and linearity properties, we first concentrate on the property that might fail, positivity. If positivity fails, we don't need to prove the remaining properties. The crossed terms $3(x_1 y_2 + x_2 y_1)$ in the definition of the inner product suggest that the product is not positive, since the factor 3 makes them too important compared with the other terms. Let us try to find an example, that is a vector $\mathsf{x} \neq \mathsf{0}$ such that $\langle \mathsf{x}, \mathsf{x} \rangle \leqslant 0$. Let us try with

$$\mathsf{x} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \langle \mathsf{x}, \mathsf{x} \rangle = 1^2 + (-1)^2 + 0^2 + 3[(1)(-1) + (-1)(1)] = 2 - 6 = -4.$$

Since $\langle \mathsf{x}, \mathsf{x} \rangle = -4$, the function $\langle \, , \rangle$ is not positive, hence it is not an inner product. ◁

EXAMPLE 6.2.4: Consider the vector space $\mathbb{F}^{m,n}$ of all $m \times n$ matrices. Show that an inner product on that space is the function $\langle \, , \rangle_F : \mathbb{F}^{m,n} \times \mathbb{F}^{m,n} \to \mathbb{F}$

$$\langle \mathsf{A}, \mathsf{B} \rangle_F = \operatorname{tr}(\mathsf{A}^* \mathsf{B}).$$

The inner product is called *Frobenius inner product*.

SOLUTION: We show that the Frobenius function $\langle \, , \rangle_F$ above satisfies the three properties in Def. 6.2.1. We use the component notation $\mathsf{A} = [A_{ij}]$, $\mathsf{B} = [B_{kl}]$, with $i, k = 1, \cdots, m$ and $j, l = 1, \cdots, n$, so

$$(\mathsf{A}^*\mathsf{B})_{jl} = \sum_{i=1}^{m} \left(\overline{A}^T\right)_{ji} B_{il} = \sum_{i=1}^{m} \overline{A}_{ij} B_{il} \quad \Rightarrow \quad \langle \mathsf{A}, \mathsf{B} \rangle_F = \sum_{j=1}^{n} \sum_{i=1}^{n} \overline{A}_{ij} B_{ij}.$$

The first property is satisfied, since

$$\langle \mathsf{A}, \mathsf{A} \rangle_F = \operatorname{tr}(\mathsf{A}^* \mathsf{A}) = \sum_{j=1}^{n} \sum_{i=1}^{m} \left|A_{ij}\right|^2 \geqslant 0,$$

and $\langle \mathsf{A}, \mathsf{A} \rangle_F = 0$ iff $A_{ij} = 0$ for every indices i, j, which is equivalently to $\mathsf{A} = 0$. The second property is satisfied, since

$$\langle \mathsf{A}, \mathsf{B} \rangle_F = \sum_{j=1}^{n} \sum_{i=1}^{n} \overline{A}_{ij} B_{ij} = \overline{\sum_{j=1}^{n} \sum_{i=1}^{n} \overline{B}_{ij} A_{ij}} = \overline{\langle \mathsf{B}, \mathsf{A} \rangle}_F.$$

The same proof can be expressed in index-free notation using the properties of the trace,

$$\operatorname{tr}\big(\mathsf{A}^* \mathsf{B})\big) = \operatorname{tr}\big(\overline{\mathsf{A}}^T \mathsf{B}\big) = \operatorname{tr}\big[(\overline{\mathsf{A}}^T \mathsf{B})^T\big] = \operatorname{tr}\big(\mathsf{B}^T \overline{\mathsf{A}}\big) = \overline{\operatorname{tr}\big(\mathsf{B}^* \mathsf{A}\big)},$$

that is, $\langle \mathsf{A}, \mathsf{B} \rangle_F = \overline{\langle \mathsf{B}, \mathsf{A} \rangle}_F$. The third property comes from the distributive property of the matrix product, that is,

$$\langle \mathsf{A}, (a\mathsf{B} + b\mathsf{C}) \rangle_F = \operatorname{tr}\big(\mathsf{A}^* \, (a\mathsf{B} + b\mathsf{C})\big) = a \operatorname{tr}\big(\mathsf{A}^* \, \mathsf{B}\big) + b \operatorname{tr}\big(\mathsf{A}^* \, \mathsf{C}\big) = a \, \langle \mathsf{A}, \mathsf{B} \rangle_F + b \, \langle \mathsf{A}, \mathsf{C} \rangle_F.$$

This establishes that $\langle \, , \rangle_F$ is an inner product. ◁

EXAMPLE 6.2.5: Compute the Frobenius inner product $\langle \mathsf{A}, \mathsf{B} \rangle_F$ where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \in \mathbb{R}^{2,3}.$$

SOLUTION: Since the matrices have real coefficients, the Frobenius inner product has the form $\langle \mathsf{A}, \mathsf{B} \rangle_F = \operatorname{tr}\big(\mathsf{A}^T \mathsf{B}\big)$. So, we need to compute the diagonal elements in the product

$$\mathsf{A}^T \mathsf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 7 & * & * \\ * & 8 & * \\ * & * & 5 \end{bmatrix} \quad \Rightarrow \quad \langle \mathsf{A}, \mathsf{B} \rangle_F = 7 + 8 + 5 \quad \Rightarrow \quad \langle \mathsf{A}, \mathsf{B} \rangle_F = 20.$$

◁

EXAMPLE 6.2.6: Consider the vector space $\mathbb{P}_n([-1, 1])$ of polynomials with real coefficients having degree less or equal $n \geq 1$ and being defined on the interval $[-1, 1]$. Show that an inner product in this space is the following:

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_{-1}^{1} \boldsymbol{p}(x) \boldsymbol{q}(x) \, dx. \qquad \boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}_n.$$

SOLUTION: We need to verify the three properties in the Definition 6.2.1. The positive definiteness property is satisfied, since

$$\langle \boldsymbol{p}, \boldsymbol{p} \rangle = \int_{-1}^{1} \big[\boldsymbol{p}(x)\big]^2 \, dx \geqslant 0,$$

and in the case $\langle \boldsymbol{p}, \boldsymbol{p} \rangle = 0$ this implies that the integrand must vanish, that is, $[\boldsymbol{p}(x)]^2 = 0$, which is equivalent to $\boldsymbol{p} = 0$. The symmetry property is satisfied, since $\boldsymbol{p}(x)\boldsymbol{q}(x) = \boldsymbol{q}(x)\boldsymbol{p}(x)$, which implies that $\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \langle \boldsymbol{q}, \boldsymbol{p} \rangle$. The linearity property on the second argument is also satisfied, since

$$\begin{aligned} \langle \boldsymbol{p}, (a\boldsymbol{q} + b\boldsymbol{r}) \rangle &= \int_{-1}^{1} \boldsymbol{p}(x) \big[a \, \boldsymbol{q}(x) + b \, \boldsymbol{r}(x)\big] \, dx \\ &= a \int_{-1}^{1} \boldsymbol{p}(x) \boldsymbol{q}(x) \, dx + b \int_{-1}^{1} \boldsymbol{p}(x) \boldsymbol{r}(x) \, dx \\ &= a \, \langle \boldsymbol{p}, \boldsymbol{q} \rangle + b \, \langle \boldsymbol{p}, \boldsymbol{r} \rangle. \end{aligned}$$

This establishes that $\langle \, , \rangle$ is an inner product. ◁

EXAMPLE 6.2.7: Consider the vector space $C^k([a,b],\mathbb{R})$, with $k \geqslant 0$ and $a < b$, of k-times continuously differentiable real-valued functions $\boldsymbol{f} : [a,b] \to \mathbb{R}$. An inner product in this vector space is given by

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_a^b \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx.$$

Any positive function $\mu \in C^0([a,b],\mathbb{R})$ determines an inner product in $C^k([a,b],\mathbb{R})$ as follows

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle_\mu = \int_a^b \mu(x)\,\boldsymbol{f}(x)\boldsymbol{g}(x)\,dx.$$

The function μ is called a weigh function. An inner product in the vector space $C^k([a,b],\mathbb{C})$ of k-times continuously differentiable complex-valued functions $\boldsymbol{f} : [a,b] \subset \mathbb{R} \to \mathbb{C}$ is the following,

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_a^b \overline{\boldsymbol{f}(x)}\boldsymbol{g}(x)\,dx.$$

◁

An inner product satisfies the following inequality.

Theorem 6.2.2 (Cauchy-Schwarz). *If $(V, \langle\,,\rangle)$ is an inner product space over $\mathbb{F}$, then for every $\boldsymbol{x}$, $\boldsymbol{y} \in V$ holds*

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2 \leqslant \langle \boldsymbol{x}, \boldsymbol{x} \rangle \, \langle \boldsymbol{y}, \boldsymbol{y} \rangle.$$

Furthermore, equality holds iff $\boldsymbol{y} = a\boldsymbol{x}$, with $a = \langle \boldsymbol{x}, \boldsymbol{y} \rangle / \langle \boldsymbol{x}, \boldsymbol{x} \rangle$.

Proof of Theorem 6.2.2: From the positive definiteness property we know that for every $\boldsymbol{x}$, $\boldsymbol{y} \in V$ and every scalar $a \in \mathbb{F}$ holds $0 \leqslant \langle (a\boldsymbol{x} - \boldsymbol{y}), (a\boldsymbol{x} - \boldsymbol{y}) \rangle$. The symmetry and the linearity on the second argument imply

$$0 \leqslant \langle (a\boldsymbol{x} - \boldsymbol{y}), (a\boldsymbol{x} - \boldsymbol{y}) \rangle = a\,\overline{a}\,\langle \boldsymbol{x}, \boldsymbol{x} \rangle - \overline{a}\,\langle \boldsymbol{x}, \boldsymbol{y} \rangle - a\,\langle \boldsymbol{y}, \boldsymbol{x} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle. \tag{6.4}$$

Since the inequality above holds for all $a \in \mathbb{F}$, let us choose a particular value of a, the solution of the equation

$$a\,\overline{a}\,\langle \boldsymbol{x}, \boldsymbol{x} \rangle - \overline{a}\,\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0 \quad \Rightarrow \quad a = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}.$$

Introduce this particular value of a into Eq. (6.4),

$$0 \leqslant -\Big(\frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}\Big)\overline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle} + \langle \boldsymbol{y}, \boldsymbol{y} \rangle \quad \Rightarrow \quad |\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2 \leqslant \langle \boldsymbol{x}, \boldsymbol{x} \rangle \, \langle \boldsymbol{y}, \boldsymbol{y} \rangle.$$

Finally, notice that equality holds iff $a\boldsymbol{x} = \boldsymbol{y}$, and in this case, computing the inner product with $\boldsymbol{x}$ we obtain $a\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. This establishes the Theorem. □

6.2.2. **Inner product norm.** The inner product on a vector space determines a particular notion of length, or norm, of a vector, and we call it the inner product norm. After we introduce this norm we show its main properties. In Chapter 8 later on we use these properties to define a more general notion of norm as any function on the vector space satisfying these properties. The inner product norm is just a particular case of this broader notion of length. A normed space is a vector space with any norm.

Definition 6.2.3. *The **inner product norm** determined in an inner product space $(V, \langle\,,\rangle)$ is the function $\|\ \| : V \to \mathbb{R}$ given by*

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

The Cauchy-Schwarz inequality is often expressed using the inner product norm as follows: For every $\boldsymbol{x}$, $\boldsymbol{y} \in V$ holds

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leqslant \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|.$$

A vector $\boldsymbol{x} \in V$ is a *normal or unit vector* iff $\|\boldsymbol{x}\| = 1$.

Theorem 6.2.4. *If $\boldsymbol{v} \neq \boldsymbol{0}$ belongs to $\big(V, \langle\,,\rangle\big)$, then $\dfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$ is a unit vector parallel to $\boldsymbol{v}$.*

The proof is the same of Theorem 6.1.12.

Example 6.2.8: Consider the inner product space $\big(\mathbb{F}^{m,n}, \langle\,,\rangle_F\big)$, where $\mathbb{F}^{m,n}$ is the vector space of all $m \times n$ matrices and $\langle\,,\rangle_F$ is the Frobenius inner product defined in Example 6.2.4. The associated inner product norm is called the *Frobenius norm* and is given by

$$\|\mathsf{A}\|_F = \sqrt{\langle \mathsf{A}, \mathsf{A} \rangle_F} = \sqrt{\operatorname{tr}\big(\mathsf{A}^*\mathsf{A}\big)}.$$

If $\mathsf{A} = [A_{ij}]$, with $i = 1, \cdots, m$ and $j = 1, \cdots, n$, then

$$\|\mathsf{A}\|_F = \Big(\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2\Big)^{1/2}.$$

$\lhd$

Example 6.2.9: Find an explicit expression for the Frobenius norm of any element $\mathsf{A} \in \mathbb{F}^{2,2}$.

Solution: The Frobenius norm of an arbitrary matrix $\mathsf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in \mathbb{F}^{2,2}$ is given by

$$\|\mathsf{A}\|_F^2 = \operatorname{tr}\left(\begin{bmatrix} \overline{A}_{11} & \overline{A}_{21} \\ \overline{A}_{12} & \overline{A}_{22} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \right).$$

Since we are only interested in the diagonal elements of the matrix product in the equation above, we obtain

$$\|\mathsf{A}\|_F^2 = \operatorname{tr} \begin{bmatrix} |A_{11}|^2 + |A_{21}|^2 & * \\ * & |A_{12}|^2 + |A_{22}|^2 \end{bmatrix}$$

which gives the formula

$$\|\mathsf{A}\|_F^2 = |A_{11}|^2 + |A_{12}|^2 + |A_{21}|^2 + |A_{22}|^2.$$

This is the explicit expression of the sum $\|\mathsf{A}\|_F = \Big(\sum_{i=1}^{2} \sum_{j=1}^{2} |A_{ij}|^2\Big)^{1/2}$. $\lhd$

The inner product norm function has the following properties.

Theorem 6.2.5. *The inner product norm introduced in Definition 6.2.3 satisfies that for every $\boldsymbol{x}$, $\boldsymbol{y} \in V$ and every $a \in \mathbb{F}$ holds,*

(a) $\|\boldsymbol{x}\| \geqslant 0$, *and* $\|\boldsymbol{x}\| = 0$ *iff* $\boldsymbol{x} = 0$, *(Positive definiteness);*
(b) $\|a\boldsymbol{x}\| = |a| \, \|\boldsymbol{x}\|$, *(Scaling);*
(c) $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$, *(Triangle inequality).*

Proof of Theorem 6.2.5: Properties (a) and (b) are straightforward to show from the definition of inner product, and their proof is left as an exercise. We show here how to obtain the triangle inequality, property (c). Given any vectors $\boldsymbol{x}$, $\boldsymbol{y} \in V$ holds

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \langle(\boldsymbol{x}+\boldsymbol{y}),(\boldsymbol{x}+\boldsymbol{y})\rangle \\ &= \|\boldsymbol{x}\|^2 + \langle\boldsymbol{x},\boldsymbol{y}\rangle + \langle\boldsymbol{y},\boldsymbol{x}\rangle + \|\boldsymbol{y}\|^2 \\ &\leqslant \|\boldsymbol{x}\|^2 + 2\,|\langle\boldsymbol{x},\boldsymbol{y}\rangle| + \|\boldsymbol{y}\|^2 \\ &\leqslant \|\boldsymbol{x}\|^2 + 2\,\|\boldsymbol{x}\|\,\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 = \big(\|\boldsymbol{x}\| + \|\boldsymbol{y}\|\big)^2,\end{aligned}$$

where the last inequality comes from the Cauchy-Schwarz inequality. We then conclude that $\|\boldsymbol{x}+\boldsymbol{y}\|^2 \leqslant \big(\|\boldsymbol{x}\| + \|\boldsymbol{y}\|\big)^2$. This establishes the Theorem. □

6.2.3. **Norm distance.** The norm on an inner product space determines a particular notion of distance between vectors. After we introduce this norm we show its main properties.

Definition 6.2.6. *The* ***norm distance*** *between two vectors in a vector space V with a norm function $\|\ \|: V \to \mathbb{R}$ is the value of the function $d : V \times V \to \mathbb{R}$ given by*

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

Theorem 6.2.7. *The norm distance in Definition 6.2.6 satisfies for every $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z} \in V$ that*
(a) $d(\boldsymbol{x},\boldsymbol{y}) \geqslant 0$, *and* $d(\boldsymbol{x},\boldsymbol{y}) = 0$ *iff* $\boldsymbol{x} = \boldsymbol{y}$, *(Positive definiteness);*
(b) $d(\boldsymbol{x},\boldsymbol{y}) = d(\boldsymbol{y},\boldsymbol{x})$, *(Symmetry);*
(c) $d(\boldsymbol{x},\boldsymbol{y}) \leqslant d(\boldsymbol{x},\boldsymbol{z}) + d(\boldsymbol{z},\boldsymbol{y})$, *(Triangle inequality).*

Proof of Theorem 6.2.7: Properties (a) and (b) are straightforward from properties (a) and (b), and their proof are left as an exercise. We show how the triangle inequality for the distance comes from the triangle inequality for the norm. Indeed

$$d(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\| = \|(\boldsymbol{x}-\boldsymbol{z}) - (\boldsymbol{y}-\boldsymbol{z})\| \leqslant \|\boldsymbol{x}-\boldsymbol{z}\| + \|\boldsymbol{y}-\boldsymbol{z}\| = d(\boldsymbol{x},\boldsymbol{z}) + d(\boldsymbol{z},\boldsymbol{y}),$$

where we used the symmetry of the distance function on the last term above. This establishes the Theorem. □

The presence of an inner product, and hence a norm and a distance, on a vector space permits to introduce the notion of convergence of an infinite sequence of vectors. We say that the sequence $\{\boldsymbol{x}_n\}_{n=0}^{\infty} \subset V$ *converges* to $\boldsymbol{x} \in V$ iff

$$\lim_{n\to\infty} d(\boldsymbol{x}_n,\boldsymbol{x}) = 0.$$

Some of the most important concepts related to convergence are closeness of a subspace, completeness of the vector space, and the continuity of linear operators and linear transformations. In the case of finite dimensional vector spaces the situation is straightforward. All subspaces are closed, all inner product spaces are complete and all linear operators and linear transformations are continuous. However, in the case of infinite dimensional vector spaces, things are not so simple.

6.2.4. **Exercises.**

6.2.1.- Determine which of the following functions $\langle\,,\rangle : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ defines an inner product on $\mathbb{R}^3$. Justify your answers.

(a) $\langle \mathsf{x}, \mathsf{y}\rangle = x_1y_1 + x_3y_3$;
(b) $\langle \mathsf{x}, \mathsf{y}\rangle = x_1y_1 - x_2y_2 + x_3y_3$;
(c) $\langle \mathsf{x}, \mathsf{y}\rangle = 2x_1y_1 + x_2y_2 + 4x_3y_3$;
(d) $\langle \mathsf{x}, \mathsf{y}\rangle = x_1^2\,y_1^2 + x_2^2\,y_2^2 + x_3^2\,y_3^2$.

We used the standard notation

$$\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathsf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

6.2.2.- Prove that an inner product function $\langle\,,\rangle : V \times V \to \mathbb{F}$ satisfies the following properties:

(a) $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$ for all $\boldsymbol{x} \in V$, then $\boldsymbol{y} = \boldsymbol{0}$.
(b) $\langle a\boldsymbol{x}, \boldsymbol{y}\rangle = \overline{a}\,\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all $\boldsymbol{x}$, $\boldsymbol{y} \in V$.

6.2.3.- Given a matrix $\mathsf{M} \in \mathbb{R}^{2,2}$ introduce the function $\langle\,,\rangle_M : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,

$$\langle \mathsf{y}, \mathsf{x}\rangle_M = \mathsf{y}^T \mathsf{M}\, \mathsf{x}.$$

For each of the matrices M below determine whether $\langle\,,\rangle_M$ defines an inner product or not. Justify your answers.

(a) $\mathsf{M} = \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}$;
(b) $\mathsf{M} = \begin{bmatrix} 4 & -3 \\ 3 & 9 \end{bmatrix}$;
(c) $\mathsf{M} = \begin{bmatrix} 4 & 1 \\ 0 & 9 \end{bmatrix}$.

6.2.4.- Fix any $\mathsf{A} \in \mathbb{R}^{n,n}$ with $N(\mathsf{A}) = \{0\}$ and introduce $\mathsf{M} = \mathsf{A}^T\mathsf{A}$. Prove that $\langle\,,\rangle_M : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, given by

$$\langle \mathsf{y}, \mathsf{x}\rangle = \mathsf{y}^T \mathsf{M}\, \mathsf{x}.$$

is an inner product in $\mathbb{R}^n$.

6.2.5.- Find $k \in \mathbb{R}$ such that the matrices A, $\mathsf{B} \in \mathbb{R}^{2,2}$ are perpendicular in the Frobenius inner product,

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \ \mathsf{B} = \begin{bmatrix} 1 & 2 \\ k & 1 \end{bmatrix}.$$

6.2.6.- Evaluate the Frobenius norm for the matrices

$$\mathsf{A} = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

6.2.7.- Prove that $\|\mathsf{A}\|_F = \|\mathsf{A}^*\|_F$ for all $\mathsf{A} \in \mathbb{F}^{m,n}$.

6.2.8.- Consider the vector space $\mathbb{P}_2([0,1])$ with inner product

$$\langle \boldsymbol{p}, \boldsymbol{q}\rangle = \int_0^1 \boldsymbol{p}(x)\,\boldsymbol{q}(x)\,dx.$$

Find a unit vector parallel to

$$\boldsymbol{p}(x) = 3 - 5x^2.$$

6.3. Orthogonal vectors

6.3.1. **Definition and examples.** In the previous Section we introduced the notion of inner product in a vector space. This structure provides a notion of vector norm and distance between vectors. In this Section we explore another concept provided by an inner product; the notion of perpendicular vectors and the notion of angle between vectors in a real vector space. We start defining perpendicular vectors on any inner product space.

Definition 6.3.1. *Two vectors $\boldsymbol{x}$, $\boldsymbol{y}$ in and inner product space $\big(V, \langle\,,\rangle\big)$ are **orthogonal** or **perpendicular**, denoted as $\boldsymbol{x} \perp \boldsymbol{y}$, iff holds $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$.*

The Pythagoras Theorem holds on any inner product space.

Theorem 6.3.2. *Let $\big(V, \langle\,,\rangle\big)$ be an inner product space over the field $\mathbb{F}$.*
(a) If $\mathbb{F} = \mathbb{R}$, then $\boldsymbol{x} \perp \boldsymbol{y} \quad \Leftrightarrow \quad \|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$;
(b) If $\mathbb{F} = \mathbb{C}$, then $\boldsymbol{x} \perp \boldsymbol{y} \quad \Rightarrow \quad \|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$.

Proof of Theorem 6.3.2: Both statements derive from the following equation:

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{y}\|^2 &= \langle (\boldsymbol{x} - \boldsymbol{y}), (\boldsymbol{x} - \boldsymbol{y})\rangle \\ &= \langle \boldsymbol{x}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle - \langle \boldsymbol{y}, \boldsymbol{x}\rangle \\ &= \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\,\mathsf{Re}\big(\langle \boldsymbol{x}, \boldsymbol{y}\rangle\big). \end{aligned} \tag{6.5}$$

In the case $\mathbb{F} = \mathbb{R}$ holds $\mathsf{Re}\big(\langle \boldsymbol{x}, \boldsymbol{y}\rangle\big) = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$, so Part (a) follows. If $\mathbb{F} = \mathbb{C}$, then $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ implies $\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$, so Part (b) follows. (Notice that the converse statement is not true in the case $\mathbb{F} = \mathbb{C}$, since Eq. (6.5) together with the hypothesis $\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ do not fix $\mathsf{Im}\big(\langle \boldsymbol{x}, \boldsymbol{y}\rangle\big)$.) This establishes the Theorem. □

In the case of real vector space the Cauchy-Schwarz inequality stated in Theorem 6.2.2 allows us to define the angle between vectors.

Definition 6.3.3. *The **angle** between two vectors $\boldsymbol{x}$, $\boldsymbol{y}$ in a real vector inner product space $\big(V, \langle\,,\rangle\big)$ is the number $\theta \in [0, \pi]$ solution of*

$$\cos(\theta) = \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\|\;\|\boldsymbol{y}\|}.$$

Example 6.3.1: Consider the inner product space $\big(\mathbb{R}^{2,2}, \langle\,,\rangle_F\big)$ and show that the following matrices are orthogonal,

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix}, \qquad \mathsf{B} = \begin{bmatrix} -5 & 2 \\ 5 & 1 \end{bmatrix}.$$

Solution: Since we need to compute the Frobenius inner product $\langle \mathsf{A}, \mathsf{B}\rangle_F$, we first compute the matrix

$$\mathsf{A}^T\mathsf{B} = \begin{bmatrix} 1 & -1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -5 & 2 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} -10 & 1 \\ 5 & 10 \end{bmatrix}.$$

Therefore $\langle \mathsf{A}, \mathsf{B}\rangle_F = \operatorname{tr}\big(\mathsf{A}^T\mathsf{B}\big) = 0$, so we conclude that $\mathsf{A} \perp \mathsf{B}$. ◁

Example 6.3.2: Consider the vector space $V = C^\infty\big([-\ell, \ell], \mathbb{R}\big)$ with the inner product

$$\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \int_{-\ell}^{\ell} \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx.$$

Consider the functions $\boldsymbol{u}_n(x) = \cos\left(\frac{n\pi x}{\ell}\right)$ and $\boldsymbol{v}_m(x) = \sin\left(\frac{m\pi x}{\ell}\right)$, where n, m are integers.
(a) Show that $\boldsymbol{u}_n \perp \boldsymbol{v}_m$ for all n, m.
(b) Show that $\boldsymbol{u}_n \perp \boldsymbol{u}_m$ for all $n \neq m$.

(c) Show that $\boldsymbol{v}_n \perp \boldsymbol{v}_m$ for all $n \neq m$.

SOLUTION: Recall the identities

$$\sin(\theta)\cos(\phi) = \frac{1}{2}\big[\sin(\theta-\phi) + \sin(\theta+\phi)\big], \tag{6.6}$$

$$\cos(\theta)\cos(\phi) = \frac{1}{2}\big[\cos(\theta-\phi) + \cos(\theta+\phi)\big], \tag{6.7}$$

$$\sin(\theta)\sin(\phi) = \frac{1}{2}\big[\cos(\theta-\phi) - \cos(\theta+\phi)\big]. \tag{6.8}$$

Part (a): Using identity in Eq. (6.6) is simple to show that

$$\begin{aligned}\langle \boldsymbol{u}_n, \boldsymbol{v}_m\rangle &= \int_{-\ell}^{\ell} \cos\Big(\frac{n\pi x}{\ell}\Big)\,\sin\Big(\frac{m\pi x}{\ell}\Big)\,dx \\ &= \frac{1}{2}\int_{-\ell}^{\ell}\Big[\sin\Big(\frac{(n-m)\pi x}{\ell}\Big) + \sin\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big]. \end{aligned} \tag{6.9}$$

First, assume that both $n-m$ and $n+m$ are non-zero,

$$\langle \boldsymbol{u}_n, \boldsymbol{v}_m\rangle = -\frac{1}{2}\Big[\frac{\ell}{(n-m)\pi}\,\cos\Big(\frac{(n-m)\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell} + \frac{\ell}{(n+m)\pi}\,\cos\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell}\Big]. \tag{6.10}$$

Since $\cos((n\pm m)\pi) = \cos(-(n\pm m)\pi)$, we conclude that both terms above vanish.

Second, in the case that $n-m=0$ the first term in Eq. (6.9) vanishes identically and we need to compute the term with $(n+m)$, which also vanishes by the second term in Eq. (6.10). Analogously, in the case of $(n+m)=0$ the second term in Eq. (6.9) vanishes identically and we need to compute the term with $(n-m)$ which also vanishes by the first term in Eq. (6.10). Therefore, $\langle \boldsymbol{u}_n, \boldsymbol{v}_m\rangle = 0$ for all n, m integers, and so $\boldsymbol{u}_n \perp \boldsymbol{v}_m$ in this case.

Part (b): Using identity in Eq. (6.7) is simple to show that

$$\begin{aligned}\langle \boldsymbol{u}_n, \boldsymbol{u}_m\rangle &= \int_{-\ell}^{\ell} \cos\Big(\frac{n\pi x}{\ell}\Big)\,\cos\Big(\frac{m\pi x}{\ell}\Big)\,dx \\ &= \frac{1}{2}\int_{-\ell}^{\ell}\Big[\cos\Big(\frac{(n-m)\pi x}{\ell}\Big) + \cos\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big]. \end{aligned} \tag{6.11}$$

We know that $n-m$ is non-zero. Now, assume that $n+m$ is non-zero, then

$$\begin{aligned}\langle \boldsymbol{u}_n, \boldsymbol{u}_m\rangle &= \frac{1}{2}\Big[\frac{\ell}{(n-m)\pi}\,\sin\Big(\frac{(n-m)\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell} + \frac{\ell}{(n+m)\pi}\,\sin\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell}\Big] \\ &= \frac{\ell}{(n-m)\pi}\,\sin((n-m)\pi) + \frac{\ell}{(n+m)\pi}\,\sin((n+m)\pi). \end{aligned} \tag{6.12}$$

Since $\sin((n\pm m)\pi) = 0$ for $(n\pm m)$ integer, we conclude that both terms above vanish.

In the case of $(n+m)=0$ the second term in Eq. (6.11) vanishes identically and we need to compute the term with $(n-m)$ which also vanishes by the first term in Eq. (6.12). Therefore, $\langle \boldsymbol{u}_n, \boldsymbol{u}_m\rangle = 0$ for all $n \neq m$ integers, and so $\boldsymbol{u}_n \perp \boldsymbol{u}_m$ in this case.

Part (c): Using identity in Eq. (6.8) is simple to show that

$$\begin{aligned}\langle \boldsymbol{v}_n, \boldsymbol{v}_m\rangle &= \int_{-\ell}^{\ell} \sin\Big(\frac{n\pi x}{\ell}\Big)\,\sin\Big(\frac{m\pi x}{\ell}\Big)\,dx \\ &= \frac{1}{2}\int_{-\ell}^{\ell}\Big[\cos\Big(\frac{(n-m)\pi x}{\ell}\Big) - \cos\Big(\frac{(n+m)\pi x}{\ell}\Big)\Big]. \end{aligned} \tag{6.13}$$

Since the only difference between Eq. (6.13) and (6.11) is the sign of the second term, repeating the argument done in case (b) we conclude that $\langle \boldsymbol{v}_n, \boldsymbol{v}_m \rangle = 0$ for all $n \neq m$ integers, and so $\boldsymbol{v}_n \perp \boldsymbol{v}_m$ in this case. ◁

6.3.2. **Orthonormal basis.** We saw that an important property of a basis is that every vector in a vector space can be decomposed in a unique way in terms of the basis vectors. This decomposition is particularly simple to find in an inner product space when the basis is an orthonormal basis. Before we introduce such basis we define an orthonormal set.

Definition 6.3.4. *The set $U = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p\}$, $p \geqslant 1$, in an inner product space $\big(V, \langle\,,\rangle\big)$ is called an* ***orthonormal set*** *iff for all $i, j = 1, \cdots, p$ holds*

$$\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \begin{cases} 0 & \text{if} \quad i \neq j, \\ 1 & \text{if} \quad i = j. \end{cases}$$

The set U is called an ***orthogonal set*** *iff holds that $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = 0$ if $i \neq j$ and $\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle \neq 0$.*

EXAMPLE 6.3.3: Consider the vector space $V = C^\infty\big([-\ell, \ell], \mathbb{R}\big)$ with the inner product

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int_{-\ell}^{\ell} \boldsymbol{f}(x)\boldsymbol{g}(x)\, dx.$$

Show that the set

$$U = \Big\{ \boldsymbol{u}_0 = \frac{1}{\sqrt{2\ell}},\ \boldsymbol{u}_n(x) = \frac{1}{\sqrt{\ell}} \cos\Big(\frac{n\pi x}{\ell}\Big),\ \boldsymbol{v}_m(x) = \frac{1}{\sqrt{\ell}} \sin\Big(\frac{m\pi x}{\ell}\Big) \Big\}_{n=1}^{\infty}$$

is an orthonormal set.

SOLUTION: We have shown in Example 6.3.2 that U is an orthogonal set. We only need to compute the norm of the vectors $\boldsymbol{u}_0$, $\boldsymbol{u}_n$ and $\boldsymbol{v}_n$, for $n = 1, 2, \cdots$. The norm of the first vector is simple to compute,

$$\langle \boldsymbol{u}_0, \boldsymbol{u}_0 \rangle = \int_{-\ell}^{\ell} \frac{1}{2\ell}\, dx = 1.$$

The norm of the cosine functions is computed as follows,

$$\begin{aligned} \langle \boldsymbol{u}_n, \boldsymbol{u}_n \rangle &= \frac{1}{\ell} \int_{-\ell}^{\ell} \cos^2\Big(\frac{n\pi x}{\ell}\Big)\, dx \\ &= \frac{1}{2\ell} \int_{-\ell}^{\ell} \Big[1 + \cos\Big(\frac{2n\pi x}{\ell}\Big)\Big]\, dx \\ &= 1 + \frac{\ell}{2n\pi} \Big[\sin\Big(\frac{2n\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell}\Big] \quad \Rightarrow \quad \langle \boldsymbol{u}_n, \boldsymbol{u}_n \rangle = 1. \end{aligned}$$

A similar calculation for the sine functions gives the result

$$\begin{aligned} \langle \boldsymbol{v}_n, \boldsymbol{v}_n \rangle &= \frac{1}{\ell} \int_{-\ell}^{\ell} \sin^2\Big(\frac{n\pi x}{\ell}\Big)\, dx \\ &= \frac{1}{2\ell} \int_{-\ell}^{\ell} \Big[1 - \cos\Big(\frac{2n\pi x}{\ell}\Big)\Big]\, dx \\ &= 1 - \frac{\ell}{2n\pi} \Big[\sin\Big(\frac{2n\pi x}{\ell}\Big)\Big|_{-\ell}^{\ell}\Big] \quad \Rightarrow \quad \langle \boldsymbol{v}_n, \boldsymbol{v}_n \rangle = 1. \end{aligned}$$

Therefore, U is an orthonormal set. ◁

A straightforward result is the following:

Theorem 6.3.5. *An orthogonal set in an inner product space is linearly independent.*

Proof of Theorem 6.3.5: Let $U = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p\}$, $p \geqslant 1$, be an orthogonal set. The zero vector is not included since $\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle \neq 0$ for all $i = 1, \cdots, p$. Let $c_1, \cdots, c_p \in \mathbb{F}$ be scalars such that

$$c_1 \boldsymbol{u}_1 + \cdots + c_p \boldsymbol{u}_p = \boldsymbol{0}.$$

Then, for any $\boldsymbol{u}_i \in U$ holds

$$0 = \langle u_i, (c_1 \boldsymbol{u}_1 + \cdots + c_p \boldsymbol{u}_p \rangle = c_1 \langle \boldsymbol{u}_i, \boldsymbol{u}_1 \rangle + \cdots + c_p \langle \boldsymbol{u}_i, \boldsymbol{u}_p \rangle.$$

Since the set U is orthogonal, $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = 0$ for $i \neq j$ and $\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle \neq 0$, so we conclude that

$$c_i \langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle = 0 \quad \Rightarrow \quad c_i = 0, \qquad i = 1, \cdots, p.$$

Therefore, U is a linearly independent set. This establishes the Thorem. □

Definition 6.3.6. *A basis $\mathcal{U}$ of an inner product space is called an **orthonormal basis** (**orthogonal basis**) iff the basis $\mathcal{U}$ is an orthonormal (orthogonal) set.*

Example 6.3.4: Consider the inner product space $(\mathbb{R}^2, \cdot)$. Determine whether the following bases are orthonormal, orthogonal or neither:

$$\mathcal{S} = \left\{ \mathsf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathsf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, \quad \mathcal{U} = \left\{ \mathsf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}, \quad \mathcal{V} = \left\{ \mathsf{v}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \mathsf{v}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}.$$

Solution: The basis $\mathcal{S}$ is orthonormal, since $\mathsf{e}_1 \cdot \mathsf{e}_2 = 0$ and $\mathsf{e}_1 \cdot \mathsf{e}_1 = \mathsf{e}_2 \cdot \mathsf{e}_2 = 1$. The basis $\mathcal{U}$ is orthogonal since $\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 0$, but it is not orthonormal. Finally, the basis $\mathcal{V}$ is neither orthonormal nor orthogonal. ◁

Theorem 6.3.7. *Given the set $U = \{\mathsf{u}_1, \cdots, \mathsf{u}_p\}$, $p \geqslant 1$, in the inner product space $(\mathbb{F}^n, \cdot)$, introduce the matrix $\mathsf{U} = [\mathsf{u}_1, \cdots, \mathsf{u}_p]$. Then the following statements hold:*

(a) U is an orthonormal set iff matrix U satisfies $\mathsf{U}^*\mathsf{U} = \mathsf{I}_p$.
(b) U is an orthonormal basis of $\mathbb{F}^n$ iff matrix U satisfies $\mathsf{U}^{-1} = \mathsf{U}^*$.

Matrices satisfying the property mentioned in part (b) of Theorem 6.3.7 appear frequently in Quantum Mechanics, so they are given a name.

Definition 6.3.8. *A matrix $\mathsf{U} \in \mathbb{F}^{n,n}$ is called **unitary** iff holds $\mathsf{U}^{-1} = \mathsf{U}^*$.*

These matrices are called unitary because they do not change the norm of vectors in $\mathbb{F}^n$ equipped with the dot product. Indeed, given any $\mathsf{x} \in \mathbb{F}^n$ with norm $\|\mathsf{x}\|$, the vector $\mathsf{Ux} \in \mathbb{F}^n$ has the same norm, since

$$\|\mathsf{Ux}\|^2 = \big(\mathsf{Ux}\big)^* \big(\mathsf{Ux}\big) = \mathsf{x}^*\mathsf{U}^*\mathsf{Ux} = \mathsf{x}^*\mathsf{U}^{-1}\mathsf{Ux} = \mathsf{x}^*\mathsf{x} = \|\mathsf{x}\|^2.$$

Proof of Theorem 6.3.7:

Part (a): This is proved by a straightforward computation,

$$\mathsf{U}^*\mathsf{U} = \begin{bmatrix} \mathsf{u}_1^* \\ \vdots \\ \mathsf{u}_p^* \end{bmatrix} \big[\mathsf{u}_1, \cdots, \mathsf{u}_p\big] = \begin{bmatrix} \mathsf{u}_1^*\mathsf{u}_1 & \cdots & \mathsf{u}_1^*\mathsf{u}_p \\ \vdots & & \vdots \\ \mathsf{u}_p^*\mathsf{u}_1 & \cdots & \mathsf{u}_p^*\mathsf{u}_p \end{bmatrix} = \begin{bmatrix} \mathsf{u}_1 \cdot \mathsf{u}_1 & \cdots & \mathsf{u}_1 \cdot \mathsf{u}_p \\ \vdots & & \vdots \\ \mathsf{u}_p \cdot \mathsf{u}_1 & \cdots & \mathsf{u}_p \cdot \mathsf{u}_p \end{bmatrix} = \mathsf{I}_p.$$

Part (b): It follows from part (a): If U is a basis of $\mathbb{F}^n$, then $p = n$; since U is an orthonormal set, part (a) implies $\mathsf{U}^*\mathsf{U} = \mathsf{I}_n$. Since U is an $n \times n$ matrix, it follows that $\mathsf{U}^* = \mathsf{U}^{-1}$. This establishes the Theorem. □

Example 6.3.5: Consider $\mathsf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$, $\mathsf{v}_2 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$, in the inner product space $(\mathbb{R}^3, \cdot)$.

(a) Show that $\mathsf{v}_1 \perp \mathsf{v}_2$;

(b) Find $x \in \mathbb{R}^3$ such that $x \perp v_1$ and $x \perp v_2$.
(c) Rescale the elements of $\{v_1, v_2, x\}$ so that the new set is an orthonormal set.

SOLUTION:
Part (a):

$$\begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = 2 + 0 - 2 = 0 \quad \Rightarrow \quad v_1 \perp v_2.$$

Part (b): We need to find $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ such that

$$v_1 \cdot x = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0, \qquad v_2 \cdot x = \begin{bmatrix} 2 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0,$$

The equations above can be written in matrix notation as

$$A^T x = 0, \quad \text{where} \quad A = \begin{bmatrix} v_1, v_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 2 & -1 \end{bmatrix}.$$

Gauss elimination operation on A^T imply

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1/2 \\ 0 & 1 & 5/2 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = \dfrac{1}{2} x_3, \\ x_2 = -\dfrac{5}{2} x_3, \\ x_3 \text{ free.} \end{cases}$$

There is a solution for any choice of $x_3 \neq 0$, so we choose $x_3 = 2$, that is, $x = \begin{bmatrix} 1 \\ -5 \\ 2 \end{bmatrix}$.

Part (c): The vectors v_1, v_2 and x are mutually orthogonal. Their norms are:

$$\|v_1\| = \sqrt{6}, \qquad \|v_2\| = \sqrt{5}, \qquad \|x\| = \sqrt{30}.$$

Therefore, the orthonormal set is

$$\left\{ u_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \ u_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \ u_3 = \frac{1}{\sqrt{30}} \begin{bmatrix} 1 \\ -5 \\ 2 \end{bmatrix} \right\}.$$

Finally, notice that the inverse of the matrix

$$U = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & -\frac{5}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \end{bmatrix} \quad \text{is} \quad U^{-1} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & 0 & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{30}} & -\frac{5}{\sqrt{30}} & \frac{2}{\sqrt{30}} \end{bmatrix} = U^T.$$

◁

6.3.3. **Vector components.** Given a basis in a finite dimensional vector space, we know that every vector in the vector space can be decomposed in a unique way as a linear combination of the basis vectors. What we do not know is a general formula to compute the vector components in such a basis. However, in the particular case that the vector space admits an inner product and the basis is an orthonormal basis, we have such formula for the vector components. The formula is very simple and is the main result in the following statement.

Theorem 6.3.9. *If $\big(V, \langle\,,\rangle\big)$ is an n-dimensional inner product space with an orthonormal basis $\mathcal{U} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$, then every vector $\boldsymbol{x} \in V$ can be decomposed as*

$$\boldsymbol{x} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\, \boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_n, \boldsymbol{x}\rangle\, \boldsymbol{u}_n. \tag{6.14}$$

Proof of Theorem 6.3.9: Since $\mathcal{U}$ is a basis, we know that for all $\boldsymbol{x} \in V$ there exist scalars $c_1, \cdots, c_n$ such that

$$\boldsymbol{x} = c_1 \boldsymbol{u}_1 + \cdots + c_n \boldsymbol{u}_n.$$

Therefore, the inner product $\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle$ for any $i = 1, \cdots, n$ is given by

$$\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle = c_1 \langle \boldsymbol{u}_i, \boldsymbol{u}_1\rangle + \cdots + c_n \langle \boldsymbol{u}_i, \boldsymbol{u}_n\rangle.$$

Since $\mathcal{U}$ is an orthonormal set, $\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle = c_i$. This establishes the Theorem. □

The result in Theorem 6.3.9 provides a remarkable simple formula for vector components in an orthonormal basis. We will get back to this subject in some depth in Section 7.1. In that Section we will name the coefficients $\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle$ in Eq. (6.14) as *Fourier coefficients* of the vector $\boldsymbol{x}$ in the orthonormal set $\mathcal{U} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$. When the set $\mathcal{U}$ is an orthonormal ordered basis of a vector space V, the coordinate map $[\;]_u : V \to \mathbb{F}^n$ is expressed in terms of the Fourier coefficients as follows,

$$[\boldsymbol{x}]_u = \begin{bmatrix} \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle \\ \vdots \\ \langle \boldsymbol{u}_n, \boldsymbol{x}\rangle \end{bmatrix}.$$

So, the coordinate map has a simple expression when it is defined by an orthonormal basis.

Example 6.3.6: Consider the inner product space $\big(\mathbb{R}^3, \cdot\big)$ with the standard ordered basis $\mathcal{S}$, and find the vector components of $\mathsf{x} = \begin{bmatrix}1\\2\\3\end{bmatrix}$ in the orthonormal ordered basis

$$\mathcal{U} = \left(\mathsf{u}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\1\\2\end{bmatrix}, \mathsf{u}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix}2\\0\\-1\end{bmatrix}, \mathsf{u}_3 = \frac{1}{\sqrt{30}}\begin{bmatrix}1\\-5\\2\end{bmatrix}\right).$$

Solution: The vector components of x in the orthonormal basis $\mathcal{U}$ are given by

$$\mathsf{x}_u = \begin{bmatrix} \langle \mathsf{u}_1, \mathsf{x}\rangle \\ \langle \mathsf{u}_2, \mathsf{x}\rangle \\ \langle \mathsf{u}_3, \mathsf{x}\rangle \end{bmatrix} \quad \Rightarrow \quad \mathsf{u} = \begin{bmatrix} \frac{9}{\sqrt{6}} \\ -\frac{1}{\sqrt{5}} \\ -\frac{3}{\sqrt{30}} \end{bmatrix}.$$

Remark: We have done a change of basis, from the standard basis $\mathcal{S}$ to the $\mathcal{U}$ basis. In fact, we can express the calculation above as follows

$$\mathsf{x}_u = \mathsf{P}^{-1}\mathsf{x}_s, \quad \text{where} \quad \mathsf{P} = \mathsf{I}_{us} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & -\frac{5}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \end{bmatrix} = \mathsf{U}.$$

Since $\mathcal{U}$ is an orthonormal basis, $\mathsf{U}^{-1} = \mathsf{U}^T$, so we conclude that

$$\mathsf{x}_u = \mathsf{U}^T\mathsf{x}_s = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & 0 & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{30}} & -\frac{5}{\sqrt{30}} & \frac{2}{\sqrt{30}} \end{bmatrix} \begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix} \frac{9}{\sqrt{6}} \\ -\frac{1}{\sqrt{5}} \\ -\frac{3}{\sqrt{30}} \end{bmatrix}.$$

◁

6.3.4. **Exercises.**

6.3.1.- Prove that the following form of the Pythagoras Theorem holds on complex vector spaces: Two vectors $\boldsymbol{x}$, $\boldsymbol{y}$ in an inner product space $\big(V, \langle\,,\rangle\big)$ over $\mathbb{C}$ are orthogonal iff for all a, $b \in \mathbb{C}$ holds
$$\|a\boldsymbol{x} + b\boldsymbol{y}\|^2 = \|a\boldsymbol{x}\|^2 + \|b\boldsymbol{y}\|^2.$$

6.3.2.- Consider the vector space $\mathbb{R}^3$ with the dot product. Find all vectors $\mathsf{x} \in \mathbb{R}^3$ which are orthogonal to the vector
$$\mathsf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

6.3.3.- Consider the vector space $\mathbb{R}^3$ with the dot product. Find all vectors $\mathsf{x} \in \mathbb{R}^3$ which are orthogonal to the vectors
$$\mathsf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathsf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

6.3.4.- Let $\mathbb{P}_3([-1,1])$ be the space of polynomials up to degree three defined on the interval $[-1,1] \subset \mathbb{R}$ with the inner product
$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_{-1}^{1} \boldsymbol{p}(x)\boldsymbol{q}(x)\,dx.$$
Show that the set $(\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3)$ is an orthogonal basis of $\mathbb{P}_3$, where
$$\boldsymbol{p}_0(x) = 1,$$
$$\boldsymbol{p}_1(x) = x,$$
$$\boldsymbol{p}_2(x) = \frac{1}{2}(3x^2 - 1),$$
$$\boldsymbol{p}_3(x) = \frac{1}{2}(5x^3 - 3x).$$
(These polynomials are the first four of the Legendre polynomials.)

6.3.5.- Consider the vector space $\mathbb{R}^3$ with the dot product.

(a) Show that the following ordered basis $\mathcal{U}$ is orthonormal,
$$\Big(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}\Big).$$

(b) Use part (a) to find the components in the ordered basis $\mathcal{U}$ of the vector
$$\mathsf{x} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}.$$

6.3.6.- Consider the vector space $\mathbb{R}^{2,2}$ with the Frobenius inner product.

(a) Show that the ordered basis given by $\mathcal{U} = (\mathsf{E}_1, \mathsf{E}_2, \mathsf{E}_3, \mathsf{E}_4)$ is orthonormal, where
$$\mathsf{E}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \mathsf{E}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$
$$\mathsf{E}_3 = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad \mathsf{E}_4 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

(b) Use part (a) to find the components in the ordered basis $\mathcal{U}$ of the matrix
$$\mathsf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

6.3.7.- Consider the inner product space $\big(\mathbb{R}^{2,2}, \langle\,,\rangle_F\big)$, and find the cosine of the angle between the matrices
$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \quad \mathsf{B} = \begin{bmatrix} 2 & -2 \\ 2 & 0 \end{bmatrix}.$$

6.3.8.- Find the third column in matrix U below such that $\mathsf{U}^T = \mathsf{U}^{-1}$, where
$$\mathsf{U} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{14} & U_{13} \\ 1/\sqrt{3} & 2/\sqrt{14} & U_{23} \\ 1/\sqrt{3} & -3/\sqrt{14} & U_{33} \end{bmatrix}.$$

6.4. Orthogonal projections

6.4.1. **Orthogonal projection onto subspaces.** Given any subspace of an inner product space, every vector in the vector space can be decomposed as a sum of two vectors; a vector in the subspace and a vector perpendicular to the subspace. The picture one often has in mind is a plane in $\mathbb{R}^3$ equiped with the dot product, as in Fig. 43. Any vector in $\mathbb{R}^3$ can be decomposed as a vector on the plane plus a vector perpendicular to the plane. In this Section we provide expressions to compute this type of decompositions. We start splitting a vector in orthogonal components with respect to a one dimensional subspace. The study of this simple case describes the main ideas and the main notation used in orthogonal decompositions. Later on we present the decomposition of a vector onto an n-dimensional subspace.

Theorem 6.4.1. *Fix a vector $\boldsymbol{u} \neq \boldsymbol{0}$ in an inner product space $\big(V, \langle\,,\rangle\big)$. Given any vector $\boldsymbol{x} \in V$ decompose it as $\boldsymbol{x} = \boldsymbol{x}_{\shortparallel} + \boldsymbol{x}_{\perp}$ where $\boldsymbol{x}_{\shortparallel} \in \mathrm{Span}(\{\boldsymbol{u}\})$. Then, $\boldsymbol{x}_{\perp} \perp \boldsymbol{u}$ iff holds*

$$\boldsymbol{x}_{\shortparallel} = \frac{\langle \boldsymbol{u}, \boldsymbol{x}\rangle}{\|\boldsymbol{u}\|^2}\,\boldsymbol{u}. \tag{6.15}$$

Furthermore, in the case that $\boldsymbol{u}$ is a unit vector holds $\boldsymbol{x}_{\shortparallel} = \langle \boldsymbol{u}, \boldsymbol{x}\rangle\,\boldsymbol{u}$.

The main idea of this decomposition can be understood in the inner product space $\big(\mathbb{R}^2, \cdot\big)$ and it is sketched in Fig. 42. It is obvious that every vector $\boldsymbol{x}$ can be expressed in many different ways as the sum of two vectors. What is special of the decomposition in Eq. 6.15 is that $\boldsymbol{x}_{\shortparallel}$ has the precise length such that $\boldsymbol{x}_{\perp}$ is orthogonal to $\boldsymbol{x}_{\shortparallel}$ (see Fig. 42).

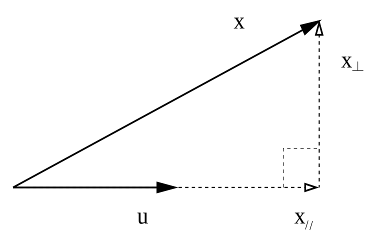

Figure 42. Orthogonal decomposition of the vector $\mathsf{x} \in \mathbb{R}^2$ onto the subspace spanned by vector u.

Proof of Theorem 6.4.1: Since $\boldsymbol{x}_{\shortparallel} \in \mathrm{Span}(\{\boldsymbol{u}\})$, there exists a scalar a such that $\boldsymbol{x}_{\shortparallel} = a\boldsymbol{u}$. Therefore $\boldsymbol{u} \perp \boldsymbol{x}_{\perp}$ iff holds that $\langle \boldsymbol{u}, \boldsymbol{x}_{\perp}\rangle = 0$. A straightforward computation shows,

$$0 = \langle \boldsymbol{u}, \boldsymbol{x}_{\perp}\rangle = \langle \boldsymbol{u}, \boldsymbol{x}\rangle - \langle \boldsymbol{u}, \boldsymbol{x}_{\shortparallel}\rangle = \langle \boldsymbol{u}, \boldsymbol{x}\rangle - a\,\langle \boldsymbol{u}, \boldsymbol{u}\rangle \quad\Leftrightarrow\quad a = \frac{\langle \boldsymbol{u}, \boldsymbol{x}\rangle}{\|\boldsymbol{u}\|^2}.$$

We conclude that the decomposition $\boldsymbol{x} = \boldsymbol{x}_{\shortparallel} + \boldsymbol{x}_{\perp}$ satisfies

$$\boldsymbol{x}_{\perp} \perp \boldsymbol{u} \quad\Leftrightarrow\quad \boldsymbol{x}_{\shortparallel} = \frac{\langle \boldsymbol{u}, \boldsymbol{x}\rangle}{\|\boldsymbol{u}\|^2}\,\boldsymbol{u}.$$

In the case that $\boldsymbol{u}$ is a unit vector holds $\|\boldsymbol{u}\| = 1$, so $\boldsymbol{x}_{\shortparallel}$ is given by $\boldsymbol{x}_{\shortparallel} = \langle \boldsymbol{u}, \boldsymbol{x}\rangle\,\boldsymbol{u}$. This establishes the Theorem. □

Example 6.4.1: Consider the inner product space $\big(\mathbb{R}^3, \cdot\big)$ and decompose the vector x in orthogonal components with respect to the vector u, where

$$\mathsf{x} = \begin{bmatrix} 3\\ 2\\ 1\end{bmatrix}, \qquad \mathsf{u} = \begin{bmatrix} 1\\ 2\\ 3\end{bmatrix}.$$

SOLUTION: We first compute $\mathsf{x}_{\shortparallel} = \dfrac{\mathsf{u}\cdot\mathsf{x}}{\|\mathsf{u}\|^2}\,\mathsf{u}$. Since

$$\mathsf{u}\cdot\mathsf{x} = \begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}3\\2\\1\end{bmatrix} = 3+4+3 = 10, \qquad \|\mathsf{u}\|^2 = \begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} = 1+4+9 = 14,$$

we obtain $\mathsf{x}_{\shortparallel} = \dfrac{5}{7}\begin{bmatrix}1\\2\\3\end{bmatrix}$. We now compute $\mathsf{x}_{\perp}$ as follows,

$$\mathsf{x}_{\perp} = \mathsf{x} - \mathsf{x}_{\shortparallel} = \begin{bmatrix}3\\2\\1\end{bmatrix} - \frac{5}{7}\begin{bmatrix}1\\2\\3\end{bmatrix} = \frac{1}{7}\begin{bmatrix}21\\14\\7\end{bmatrix} - \frac{1}{7}\begin{bmatrix}5\\10\\15\end{bmatrix} = \frac{1}{7}\begin{bmatrix}16\\4\\-8\end{bmatrix} \quad\Rightarrow\quad \mathsf{x}_{\perp} = \frac{4}{7}\begin{bmatrix}4\\1\\-2\end{bmatrix}.$$

Therefore, x can be decomposed as

$$\mathsf{x} = \frac{5}{7}\begin{bmatrix}1\\2\\3\end{bmatrix} + \frac{4}{7}\begin{bmatrix}4\\1\\-2\end{bmatrix}.$$

REMARK: We can verify that this decomposition is orthogonal with respect to u, since

$$\mathsf{u}\cdot\mathsf{x}_{\perp} = \frac{4}{7}\begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}4\\1\\-2\end{bmatrix} = \frac{4}{7}(4+2-6) = 0.$$

◁

We now decompose a vector into orthogonal components with respect to a p-dimensional subspace with $p \geqslant 1$.

Theorem 6.4.2. *Fix an orthogonal set $\mathcal{U} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p\}$, with $p \geqslant 1$, in an inner product space $\big(V, \langle\,,\rangle\big)$. Given any vector $\boldsymbol{x} \in V$, decompose it as $\boldsymbol{x} = \boldsymbol{x}_{\shortparallel} + \boldsymbol{x}_{\perp}$, where $\boldsymbol{x}_{\shortparallel} \in \operatorname{Span}(\mathcal{U})$. Then, $\boldsymbol{x}_{\perp} \perp \boldsymbol{u}_i$, for $i = 1, \cdots, p$ iff holds*

$$\boldsymbol{x}_{\shortparallel} = \frac{\langle \boldsymbol{u}_1, \boldsymbol{x}\rangle}{\|\boldsymbol{u}_1\|^2}\,\boldsymbol{u}_1 + \cdots + \frac{\langle \boldsymbol{u}_p, \boldsymbol{x}\rangle}{\|\boldsymbol{u}_p\|^2}\,\boldsymbol{u}_p. \tag{6.16}$$

Furthermore, in the case that $\mathcal{U}$ is an orthonormal basis holds

$$\boldsymbol{x}_{\shortparallel} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\,\boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_p, \boldsymbol{x}\rangle\,\boldsymbol{u}_p. \tag{6.17}$$

REMARK: The set $\mathcal{U}$ in Theorem 6.4.2 must be an orthogonal set. If the vectors $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p$ are not mutually orthogonal, then the vector $\boldsymbol{x}_{\shortparallel}$ computed in Eq. (6.17) is not the orthogonal projection of vector $\boldsymbol{x}$, that is, $(\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}) \not\perp \boldsymbol{u}_i$ for $i = 1, \cdots, p$. Therefore, before using Eq. (6.17) in a particular application one should verify that the set $\mathcal{U}$ one is working with is in fact an orthogonal set. A particular case of the orthogonal projection of a vector in the inner product space $\big(\mathbb{R}^3, \cdot\big)$ onto a plane is sketched in Fig. 43.

Proof of Theorem 6.4.2: Since $\boldsymbol{x}_{\shortparallel} \in \operatorname{Span}(\mathcal{U})$, there exist scalars a_i, for $i = 1, \cdots, p$ such that $\boldsymbol{x}_{\shortparallel} = a_1\boldsymbol{u}_1 + \cdots + a_p\boldsymbol{u}_p$. The vector $\boldsymbol{x}_{\perp} \perp \boldsymbol{u}_i$ iff holds that $\langle \boldsymbol{u}_i, \boldsymbol{x}_{\perp}\rangle = 0$. A straightforward computation shows that, for $i = 1, \cdots, p$ holds

$$\begin{aligned} 0 = \langle \boldsymbol{u}_i, \boldsymbol{x}_{\perp}\rangle &= \langle \boldsymbol{u}_i, \boldsymbol{x}\rangle - \langle \boldsymbol{u}_i, \boldsymbol{x}_{\shortparallel}\rangle \\ &= \langle \boldsymbol{u}_i, \boldsymbol{x}\rangle - a_1\,\langle \boldsymbol{u}_i, \boldsymbol{u}_1\rangle - \cdots - a_p\,\langle \boldsymbol{u}_i, \boldsymbol{u}_p\rangle \\ &= \langle \boldsymbol{u}_i, \boldsymbol{x}\rangle - a_i\langle \boldsymbol{u}_i, \boldsymbol{u}_i\rangle \quad\Leftrightarrow\quad a_i = \frac{\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle}{\|\boldsymbol{u}_i\|^2}. \end{aligned}$$

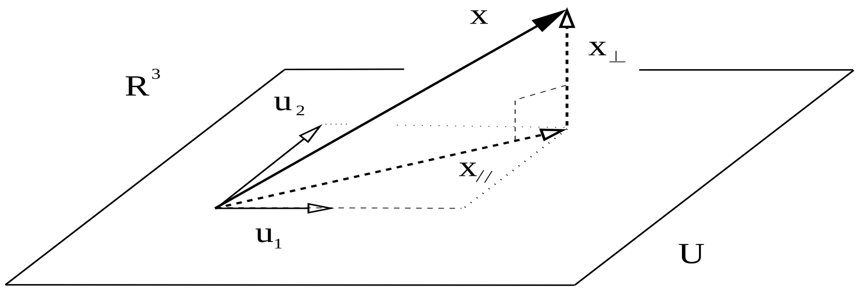

FIGURE 43. Orthogonal decomposition of the vector $\mathsf{x} \in \mathbb{R}^3$ onto the subspace U spanned by the vectors u_1 and u_2.

We conclude that the decomposition $\boldsymbol{x} = \boldsymbol{x}_{\shortparallel} + \boldsymbol{x}_{\perp}$ satisfies $\boldsymbol{x}_{\perp} \perp \boldsymbol{u}_i$ for $i = 1, \cdots, p$ iff holds

$$\boldsymbol{x}_{\shortparallel} = \frac{\langle \boldsymbol{u}_1, \boldsymbol{x}\rangle}{\|\boldsymbol{u}_1\|^2}\,\boldsymbol{u}_1 + \cdots + \frac{\langle \boldsymbol{u}_p, \boldsymbol{x}\rangle}{\|\boldsymbol{u}_p\|^2}\,\boldsymbol{u}_p.$$

In the case that $\mathcal{U}$ is an orthonormal set holds $\|\boldsymbol{u}_i\| = 1$ for $i = 1, \cdots, p$, so $\boldsymbol{x}_{\shortparallel}$ is given by Eq. (6.17). This establishes the Theorem. □

EXAMPLE 6.4.2: Consider the inner product space $(\mathbb{R}^3, \cdot)$ and decompose the vector x in orthogonal components with respect to the subspace U, where

$$\mathsf{x} = \begin{bmatrix}1\\2\\3\end{bmatrix}, \qquad U = \mathrm{Span}\left(\left\{\mathsf{u}_1 = \begin{bmatrix}2\\5\\-1\end{bmatrix}, \mathsf{u}_2 = \begin{bmatrix}-2\\1\\1\end{bmatrix}\right\}\right).$$

SOLUTION: In order to use Eq. (6.16) we need an orthogonal basis of U. So, need need to verify whether u_1 is orthogonal to u_2. This is indeed the case, since

$$\mathsf{u}_1 \cdot \mathsf{u}_2 = \begin{bmatrix}2 & 5 & -1\end{bmatrix}\begin{bmatrix}-2\\1\\1\end{bmatrix} = -4 + 5 - 1 = 0.$$

So now we use u_1 and u_2 to compute $\mathsf{x}_{\shortparallel}$ using Eq. (6.17). We need the quantities

$$\mathsf{u}_1 \cdot \mathsf{x} = \begin{bmatrix}2 & 5 & -1\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} = 9, \qquad \mathsf{u}_1 \cdot \mathsf{u}_1 = \begin{bmatrix}2 & 5 & -1\end{bmatrix}\begin{bmatrix}2\\5\\-1\end{bmatrix} = 30,$$

$$\mathsf{u}_2 \cdot \mathsf{x} = \begin{bmatrix}-2 & 1 & 1\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} = 3, \qquad \mathsf{u}_2 \cdot \mathsf{u}_2 = \begin{bmatrix}-2 & 1 & 1\end{bmatrix}\begin{bmatrix}-2\\1\\1\end{bmatrix} = 6.$$

Now is simple to compute $\mathsf{x}_{\shortparallel}$, since

$$\mathsf{x}_{\shortparallel} = \frac{9}{30}\begin{bmatrix}2\\5\\-1\end{bmatrix} + \frac{3}{6}\begin{bmatrix}-2\\1\\1\end{bmatrix} = \frac{3}{10}\begin{bmatrix}2\\5\\-1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}-2\\1\\1\end{bmatrix} = \frac{1}{10}\begin{bmatrix}6\\15\\-3\end{bmatrix} + \frac{1}{10}\begin{bmatrix}-10\\5\\5\end{bmatrix},$$

therefore,

$$\mathsf{x}_{\shortparallel} = \frac{1}{10}\begin{bmatrix}-4\\20\\2\end{bmatrix} \quad\Rightarrow\quad \mathsf{x}_{\shortparallel} = \frac{1}{5}\begin{bmatrix}-2\\10\\1\end{bmatrix}.$$

The vector $\mathsf{x}_\perp$ is obtained as follows,

$$\mathsf{x}_\perp = \mathsf{x} - \mathsf{x}_{\shortparallel} = \begin{bmatrix}1\\2\\3\end{bmatrix} - \frac{1}{5}\begin{bmatrix}-2\\10\\1\end{bmatrix} = \frac{1}{5}\begin{bmatrix}5\\10\\15\end{bmatrix} - \frac{1}{5}\begin{bmatrix}-2\\10\\1\end{bmatrix} = \frac{1}{5}\begin{bmatrix}7\\0\\14\end{bmatrix} \quad\Rightarrow\quad \mathsf{x}_\perp = \frac{7}{5}\begin{bmatrix}1\\0\\2\end{bmatrix}.$$

We conclude that

$$\mathsf{x} = \frac{1}{5}\begin{bmatrix}-2\\10\\1\end{bmatrix} + \frac{7}{5}\begin{bmatrix}1\\0\\2\end{bmatrix}.$$

REMARK: We can verify that $\mathsf{x}_\perp \perp U$, since

$$\mathsf{u}_1 \cdot \mathsf{x}_\perp = \frac{7}{5}\begin{bmatrix}2 & 5 & -1\end{bmatrix}\begin{bmatrix}1\\0\\2\end{bmatrix} = \frac{7}{5}(2+0-2) = 0,$$

$$\mathsf{u}_2 \cdot \mathsf{x}_\perp = \frac{7}{5}\begin{bmatrix}-2 & 1 & 1\end{bmatrix}\begin{bmatrix}1\\0\\2\end{bmatrix} = \frac{7}{5}(2+0-2) = 0.$$

◁

6.4.2. **Orthogonal complement.** We have seen that any vector in an inner product space can be decomposed as a sum of orthogonal vectors, that is, $\boldsymbol{x} = \boldsymbol{x}_{\shortparallel} + \boldsymbol{x}_\perp$ with $\langle \boldsymbol{x}_{\shortparallel}, \boldsymbol{x}_\perp \rangle = 0$. Inner product spaces can be decomposed in a somehow analogous way as a direct sum of two mutually orthogonal subspaces. In order to understand such decomposition, we need to introduce the notion of the orthogonal complement of a subspace. Then we will show that a finite dimensional inner product space is the direct sum of a subspace and it orthogonal complement. It is precisely this result that motivates the word "complement" in the name orthogonal complement of a subspace.

Definition 6.4.3. *The* ***orthogonal complement*** *of a subspace W in an inner product space $\big(V, \langle\,,\rangle\big)$, denoted as $W^\perp$, is the set $W^\perp = \{\boldsymbol{u} \in V \;:\; \langle \boldsymbol{u}, \boldsymbol{w}\rangle = 0 \quad \forall\ \boldsymbol{w} \in W\}$.*

EXAMPLE 6.4.3: In the inner product space $\big(\mathbb{R}^3, \cdot\big)$, the orthogonal complement to a line is a plane, and the orthogonal complement to a plane is a line, as it is shown in Fig. 44. ◁

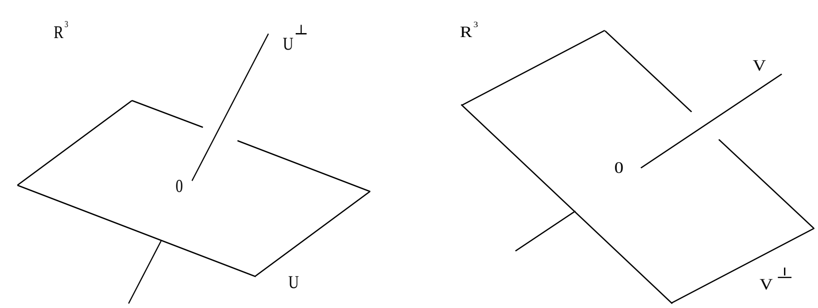

FIGURE 44. The orthogonal complement to the plane U is the line $U^\perp$, and the orthogonal complement to the line V is the plane $V^\perp$.

As the sketch in Fig. 44 suggests, the orthogonal complement of a subspace is a subspace.

Theorem 6.4.4. *The orthogonal complement $W^\perp$ of a subspace W in an inner product space is also a subspace.*

Proof of Theorem 6.4.4: Let $\boldsymbol{u}_1$, $\boldsymbol{u}_2 \in W^\perp$, that is, $\langle \boldsymbol{u}_i, \boldsymbol{w} \rangle = 0$ for all $\boldsymbol{w} \in W$ and $i = 1, 2$. Then, any linear combination $a\boldsymbol{u}_1 + b\boldsymbol{u}_2$ also belongs to $W^\perp$, since

$$\langle (a\boldsymbol{u}_1 + b\boldsymbol{u}_2), \boldsymbol{w} \rangle = \overline{a} \langle \boldsymbol{u}_1, \boldsymbol{w} \rangle + \overline{b} \langle \boldsymbol{u}_2, \boldsymbol{w} \rangle = 0 + 0 \qquad \forall \boldsymbol{w} \in W.$$

This establishes the Theorem. □

Example 6.4.4: Find $W^\perp$ for the subspace $W = \text{Span}\Big(\Big\{\mathsf{w}_1 = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}\Big\}\Big)$ in $(\mathbb{R}^3, \cdot)$.

Solution: We need to find the set of all $\mathsf{x} \in \mathbb{R}^3$ such that $\mathsf{x} \cdot \mathsf{w}_1 = 0$. That is,

$$\begin{bmatrix} -1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \quad \Rightarrow \quad x_1 = 2x_2 + 3x_3.$$

The solution is

$$\mathsf{x} = \begin{bmatrix} 2x_2 + 3x_3 \\ x_2 \\ x_3 \end{bmatrix} \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} x_3,$$

hence we obtain

$$W^\perp = \text{Span}\Big(\Big\{\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}\Big\}\Big).$$

The orthogonal complement of a line is a plane, as sketched in the second picture in Fig. 44.
Remark: We can verify that the result is correct, since

$$\begin{bmatrix} -1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = -2 + 2 + 0 = 0, \qquad \begin{bmatrix} -1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = -3 + 0 + 3 = 0.$$

◁

Example 6.4.5: Find $W^\perp$ for the subspace $W = \text{Span}\Big(\Big\{\mathsf{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathsf{w}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}\Big\}\Big)$ in $(\mathbb{R}^3, \cdot)$.

Solution: We need to find the set of all $\mathsf{x} \in \mathbb{R}^3$ such that $\mathsf{x} \cdot \mathsf{w}_1 = 0$ and $\mathsf{x} \cdot \mathsf{w}_2 = 0$. That is,

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

We an use Gauss elimination to find the solution,

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = x_3, \\ x_2 = -2x_3, \\ x_3 \text{ free.} \end{cases} \quad \Rightarrow \quad \mathsf{x} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} x_3,$$

hence we obtain

$$W^\perp = \text{Span}\Big(\Big\{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\Big\}\Big).$$

So, the orthogonal complement of a pane is a line, as sketched in the first picture in Fig. 44.

REMARK: We can verify that the result is correct, since

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 1 - 4 + 3 = 0, \qquad \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 3 - 4 + 1 = 0.$$

◁

As we mentioned above, the reason for the word "complement" in the name of an orthogonal complement is that the vector space can be split into the sum of two subspaces with zero intersection. We summarize this below.

Theorem 6.4.5. *If W is a subspace in a finite dimensional inner product space V, then*

$$V = W \oplus W^{\perp}.$$

Proof of Theorem 6.4.5: We first show that $V = W + W^{\perp}$. In order to do that, first choose an orthonormal basis $\mathcal{W} = \{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_p\}$ for W, we here we have assumed $\dim W = p \leqslant n = \dim V$. Then, for every vector $\boldsymbol{x} \in V$ holds that it can be decomposed as

$$\boldsymbol{x} = \boldsymbol{x}_{\parallel} + \boldsymbol{x}_{\perp}, \qquad \boldsymbol{x}_{\parallel} = \langle \boldsymbol{w}_1, \boldsymbol{x} \rangle\, \boldsymbol{w}_1 + \cdots + \langle \boldsymbol{w}_p, \boldsymbol{x} \rangle\, \boldsymbol{w}_p, \qquad \boldsymbol{x}_{\perp} = \boldsymbol{x} - \boldsymbol{x}_{\parallel}.$$

Theorem 6.4.2 says that $\boldsymbol{x}_{\perp} \perp \boldsymbol{w}_i$ for $i = 1, \cdots, p$. This implies that $\boldsymbol{x}_{\perp} \in W^{\perp}$ and, since $\boldsymbol{x}$ is an arbitrary vector in V, we have established that $V = W + W^{\perp}$. We now show that $W \cap W^{\perp} = \{\boldsymbol{0}\}$. Indeed, if $\boldsymbol{u} \in W^{\perp}$, then $\langle \boldsymbol{u}, \boldsymbol{w} \rangle = 0$ for all $\boldsymbol{w} \in W$. If $\boldsymbol{u} \in W$, then choosing $\boldsymbol{w} = \boldsymbol{u}$ in the equation above we get $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$, which implies $\boldsymbol{u} = 0$. Therefore, $W \cap W^{\perp} = \{\boldsymbol{0}\}$ and we conclude that $V = W \oplus W^{\perp}$. This establishes the Theorem. □

Since the orthogonal complement $W^{\perp}$ of a subspace W is itself a subspace, we can compute $(W^{\perp})^{\perp}$. The following statement says that the result is the original subspace W.

Theorem 6.4.6. *If W is a subspace in an finite dimensional inner product space, then*

$$\left(W^{\perp}\right)^{\perp} = W.$$

Proof of Theorem 6.4.6: First notice that $W \subset \left(W^{\perp}\right)^{\perp}$. Indeed, given any fixed vector $\boldsymbol{w} \in W$, the definition of $W^{\perp}$ implies that every vector $\boldsymbol{u} \in W^{\perp}$ satisfies $\langle \boldsymbol{w}, \boldsymbol{u} \rangle = 0$. This condition says that $\boldsymbol{w} \in \left(W^{\perp}\right)^{\perp}$, hence we conclude $W \subset \left(W^{\perp}\right)^{\perp}$. Second, Theorem 6.4.2 says that

$$V = W^{\perp} \oplus \left(W^{\perp}\right)^{\perp}.$$

In particular, this decomposition implies that $\dim V = \dim W^{\perp} + \dim\left(W^{\perp}\right)^{\perp}$. Again Theorem 6.4.2 also says that

$$V = W \oplus W^{\perp},$$

which in particular implies that $\dim V = \dim W + \dim W^{\perp}$. These last two results put together imply $\dim W = \dim\left(W^{\perp}\right)^{\perp}$. It is from this last result together with our previous result, $W \subset \left(W^{\perp}\right)^{\perp}$, that we obtain $W = \left(W^{\perp}\right)^{\perp}$. This establishes the Theorem. □

6.4.3. **Exercises.**

6.4.1.- Consider the inner product space $(\mathbb{R}^3, \cdot)$ and use Theorem 6.4.1 to find the orthogonal decomposition of vector x along vector u, where

$$\mathsf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathsf{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

6.4.2.- Consider the subspace W given by

$$\mathrm{Span}\Big(\Big\{\mathsf{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathsf{w}_2 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}\Big\}\Big)$$

in the inner product space $(\mathbb{R}^3, \cdot)$.

(a) Find an orthogonal decomposition of the vector w_2 with respect to the vector w_1. Using this decomposition, find an orthogonal basis for the space W.

(b) Find the decomposition of the vector x below in orthogonal components with respect to the subspace W, where

$$\mathsf{x} = \begin{bmatrix} 4 \\ 3 \\ 0 \end{bmatrix}.$$

6.4.3.- Consider the subspace W given by

$$\mathrm{Span}\Big(\Big\{\mathsf{w}_1 = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}, \mathsf{w}_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix},\Big\}\Big)$$

in the inner product space $(\mathbb{R}^3, \cdot)$. Decompose the vector x below into orthogonal components with respect to W, where

$$\mathsf{x} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

(Notice that $\mathsf{w}_1 \not\perp \mathsf{w}_2$.)

6.4.4.- Given the matrix A below, find a basis for the space $R(\mathsf{A})^{\perp}$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 2 & 0 \end{bmatrix}.$$

6.4.5.- Consider the subspace

$$W = \mathrm{Span}\Big\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \Big\}$$

in the inner product space $(\mathbb{R}^3, \cdot)$.

(a) Find a basis for the space $W^{\perp}$, that is, find a basis for the orthogonal complement of the space W.

(b) Use Theorem 6.4.1 to transform the basis of $W^{\perp}$ found in part (a) into an orthogonal basis.

6.4.6.- Consider the inner product space $(\mathbb{R}^4, \cdot)$, and find a basis for the orthogonal complement of the subspace W given by

$$W = \mathrm{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \\ 6 \end{bmatrix} \Big\}\Big).$$

6.4.7.- Let X and Y be subspaces of a finite dimensional inner product space $(V, \langle\, , \rangle)$. Prove the following:

(a) $X \subset Y \quad \Rightarrow \quad Y^{\perp} \subset X^{\perp}$;

(b) $(X + Y)^{\perp} = X^{\perp} \cap Y^{\perp}$;

(c) Use part (b) to show that

$$(X \cap Y)^{\perp} = X^{\perp} + Y^{\perp}.$$

6.5. Gram-Schmidt method

We now describe the Gram-Schmidt orthogonalization method, which is a method to transform a linearly independent set of vectors into an orthonormal set. The method is based on projecting the i-th vector in the set onto the subspace spanned by the previous $(i-1)$ vectors.

Theorem 6.5.1 (Gram-Schmidt). *Let $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p\}$ be a linearly independent set in an inner product space $\big(V, \langle\,,\rangle\big)$. If the set $Y = \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_p\}$ is defined as follows,*

$$
\begin{aligned}
\boldsymbol{y}_1 &= \boldsymbol{x}_1, \\
\boldsymbol{y}_2 &= \boldsymbol{x}_2 - \frac{\langle \boldsymbol{y}_1, \boldsymbol{x}_2\rangle}{\|\boldsymbol{y}_1\|^2}\,\boldsymbol{y}_1, \\
&\vdots \\
\boldsymbol{y}_p &= \boldsymbol{x}_p - \frac{\langle \boldsymbol{y}_1, \boldsymbol{x}_p\rangle}{\|\boldsymbol{y}_1\|^2}\,\boldsymbol{y}_1 - \cdots - \frac{\langle \boldsymbol{y}_{(p-1)}, \boldsymbol{x}_p\rangle}{\|\boldsymbol{y}_{(p-1)}\|^2}\,\boldsymbol{y}_{(p-1)},
\end{aligned}
$$

then Y is an orthogonal set with $\operatorname{Span}(Y) = \operatorname{Span}(X)$. Furthermore, the set

$$Z = \Big\{ \boldsymbol{z}_1 = \frac{\boldsymbol{y}_1}{\|\boldsymbol{y}_1\|}, \cdots, \boldsymbol{z}_p = \frac{\boldsymbol{y}_p}{\|\boldsymbol{y}_p\|} \Big\}$$

is an orthonormal set with $\operatorname{Span}(Z) = \operatorname{Span}(Y)$.

Remark: Using the notation in Theorem 6.4.2 we can write $\boldsymbol{y}_2 = \boldsymbol{x}_{2\perp}$, where the projection is onto the subspace $\operatorname{Span}(\{\boldsymbol{y}_1\})$. Analogously, $\boldsymbol{y}_i = \boldsymbol{x}_{i\perp}$, for $i = 2, \cdots, p$, where the projection is onto the subspace $\operatorname{Span}(\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{i-1}\})$.

Proof of Theorem 6.5.1: We first show that Y is an orthogonal set. It is simple to see that $\boldsymbol{y}_2 \in \operatorname{Span}(\{\boldsymbol{y}_1\})^\perp$, since

$$\langle \boldsymbol{y}_1, \boldsymbol{y}_2\rangle = \langle \boldsymbol{y}_1, \boldsymbol{x}_2\rangle - \frac{\langle \boldsymbol{y}_1, \boldsymbol{x}_2\rangle}{\|\boldsymbol{y}_1\|^2}\langle \boldsymbol{y}_1, \boldsymbol{y}_1\rangle = 0.$$

Assume that $\boldsymbol{y}_i \in \operatorname{Span}(\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{i-1}\})^\perp$, we then show that $\boldsymbol{y}_{i+1} \in \operatorname{Span}(\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_i\})^\perp$. Indeed, for $j = 1, \cdots, i$ holds

$$\langle \boldsymbol{y}_j, \boldsymbol{y}_{i+1}\rangle = \langle \boldsymbol{y}_j, \boldsymbol{x}_{i+1}\rangle - \frac{\langle \boldsymbol{y}_j, \boldsymbol{x}_{i+1}\rangle}{\|\boldsymbol{y}_j\|^2}\langle \boldsymbol{y}_j, \boldsymbol{y}_j\rangle = 0,$$

where we used that $\boldsymbol{y}_j \in \operatorname{Span}(\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{i-1}\})^\perp$, for all $j = 1, \cdots, i$. Therefore, Y is an orthogonal set (and so, a linearly independent set).

The proof that $\operatorname{Span}(X) = \operatorname{Span}(Y)$ has two steps: On the one hand, the elements in Y are linear combinations of elements in X, hence $\operatorname{Span}(Y) \subset \operatorname{Span}(X)$; on the other hand $\dim \operatorname{Span}(X) = \dim \operatorname{Span}(Y)$, since X and Y are both linearly independent sets with the same number of elements We conclude that $\operatorname{Span}(X) = \operatorname{Span}(Y)$. It is straightforward to see that Z is an orthonormal set, and since every element $\boldsymbol{z}_i \in Z$ is proportional to every $\boldsymbol{y}_i \in Y$, then $\operatorname{Span}(Y) = \operatorname{Span}(Z)$. This establishes the Theorem. □

Example 6.5.1: Use the Gram-Schmidt method to find an orthonormal basis for the inner product space $\big(\mathbb{R}^3, \cdot\big)$ from the ordered basis

$$X = \left(\mathsf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathsf{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \mathsf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right).$$

SOLUTION: We first find an orthogonal basis. The first element is

$$\mathsf{y}_1 = \mathsf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \|\mathsf{y}_1\|^2 = 2.$$

The second element is

$$\mathsf{y}_2 = \mathsf{x}_2 - \frac{\mathsf{y}_1 \cdot \mathsf{x}_2}{\|\mathsf{y}_1\|^2}\, \mathsf{y}_1,$$

where

$$\mathsf{y}_1 \cdot \mathsf{x}_2 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = 3.$$

A simple calculation shows

$$\mathsf{y}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},$$

therefore,

$$\mathsf{y}_2 = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \|\mathsf{y}_2\|^2 = \frac{1}{2}.$$

Finally, the last element is

$$\mathsf{y}_3 = \mathsf{x}_3 - \frac{\mathsf{y}_1 \cdot \mathsf{x}_3}{\|\mathsf{y}_1\|^2}\, \mathsf{y}_1, - \frac{\mathsf{y}_2 \cdot \mathsf{x}_3}{\|\mathsf{y}_2\|^2}\, \mathsf{y}_2,$$

where

$$\mathsf{y}_1 \cdot \mathsf{x}_3 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 2, \qquad \mathsf{y}_2 \cdot \mathsf{x}_3 = \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0.$$

Another simple calculation shows

$$\mathsf{y}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

therefore,

$$\mathsf{y}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \|\mathsf{y}_3\|^2 = 1.$$

The set $Y = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$ is an orthogonal set, while an orthonormal set is given by

$$Z = \Big\{ \mathsf{z}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathsf{z}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathsf{z}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \Big\}.$$

◁

EXAMPLE 6.5.2: Consider the vector space $\mathbb{P}_3([-1,1])$ with the inner product

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_{-1}^{1} \boldsymbol{p}(x)\boldsymbol{q}(x)\, dx.$$

Given the basis $\{\boldsymbol{p}_0 = 1,\ \boldsymbol{p}_1 = x,\ \boldsymbol{p}_2 = x^2,\ \boldsymbol{p}_3 = x^3\}$, use the Gram-Schmidt method starting with the vector $\boldsymbol{p}_0$ to find an orthogonal basis for $\mathbb{P}_3([-1,1])$.

SOLUTION: The first element in the new basis is

$$\boldsymbol{q}_0 = \boldsymbol{p}_0 = 1 \quad \Rightarrow \quad \|\boldsymbol{q}_0\|^2 = \int_{-1}^{1} dx = 2.$$

The second element is

$$\boldsymbol{q}_1 = \boldsymbol{p}_1 - \frac{\langle \boldsymbol{q}_0, \boldsymbol{p}_1 \rangle}{\|\boldsymbol{q}_0\|^2}\, \boldsymbol{q}_0.$$

It is simple to see that

$$\langle \boldsymbol{q}_0, \boldsymbol{p}_1 \rangle = \int_{-1}^{1} x\, dx = \frac{1}{2} x^2 \Big|_{-1}^{1} = 0.$$

So we conclude that

$$\boldsymbol{q}_1 = \boldsymbol{p}_1 = x \quad \Rightarrow \quad \|\boldsymbol{q}_1\|^2 = \int_{-1}^{1} x^2\, dx = \frac{1}{3} x^3 \Big|_{-1}^{1} \quad \Rightarrow \quad \|\boldsymbol{q}_1\|^2 = \frac{2}{3}.$$

The third element in the basis is

$$\boldsymbol{q}_2 = \boldsymbol{p}_2 - \frac{\langle \boldsymbol{q}_0, \boldsymbol{p}_2 \rangle}{\|\boldsymbol{q}_0\|^2}\, \boldsymbol{q}_0 - \frac{\langle \boldsymbol{q}_1, \boldsymbol{p}_2 \rangle}{\|\boldsymbol{q}_1\|^2}\, \boldsymbol{q}_1.$$

It is simple to see that

$$\langle \boldsymbol{q}_0, \boldsymbol{p}_2 \rangle = \int_{-1}^{1} x^2\, dx = \frac{1}{3} x^3 \Big|_{-1}^{1} = \frac{2}{3},$$

$$\langle \boldsymbol{q}_1, \boldsymbol{p}_2 \rangle = \int_{-1}^{1} x^3\, dx = \frac{1}{4} x^4 \Big|_{-1}^{1} = 0.$$

Hence we obtain

$$\boldsymbol{q}_2 = \boldsymbol{p}_2 - \frac{2}{3}\,\frac{1}{2}\,\boldsymbol{q}_0 = x^2 - \frac{1}{3} \quad \Rightarrow \quad \boldsymbol{q}_2 = \frac{1}{3}\,(3x^2 - 1).$$

The norm square of this vector is

$$\begin{aligned} \|\boldsymbol{q}_2\|^2 &= \frac{1}{9} \int_{-1}^{1} (3x^2 - 1)(3x^2 - 1)\, dx \\ &= \frac{1}{9} \int_{-1}^{1} (9x^4 - 6x^2 + 1)\, dx \\ &= \frac{1}{9} \Big(\frac{9}{5} x^5 - 2x^3 + x \Big) \Big|_{-1}^{1} \\ &= \frac{8}{45}. \end{aligned}$$

Finally, the fourth vector of the orthogonal basis is given by

$$\boldsymbol{q}_3 = \boldsymbol{p}_3 - \frac{\langle \boldsymbol{q}_0, \boldsymbol{p}_3 \rangle}{\|\boldsymbol{q}_0\|^2}\, \boldsymbol{q}_0 - \frac{\langle \boldsymbol{q}_1, \boldsymbol{p}_3 \rangle}{\|\boldsymbol{q}_1\|^2}\, \boldsymbol{q}_1 - \frac{\langle \boldsymbol{q}_1, \boldsymbol{p}_3 \rangle}{\|\boldsymbol{q}_2\|^2}\, \boldsymbol{q}_2.$$

It is simple to see that

$$\langle \boldsymbol{q}_0, \boldsymbol{p}_3 \rangle = \int_{-1}^{1} x^3\, dx = \frac{1}{4} x^4 \Big|_{-1}^{1} = 0,$$

$$\langle \boldsymbol{q}_1, \boldsymbol{p}_3 \rangle = \int_{-1}^{1} x^4\, dx = \frac{1}{5} x^5 \Big|_{-1}^{1} = \frac{2}{5},$$

$$\langle \boldsymbol{q}_2, \boldsymbol{p}_3 \rangle = \frac{1}{3} \int_{-1}^{1} (3x^2 - 1)\, x^3\, dx = \frac{1}{3} \Big(\frac{1}{2} x^6 - \frac{1}{4} x^4 \Big) \Big|_{-1}^{1} = 0.$$

Hence we obtain

$$\boldsymbol{q}_3 = \boldsymbol{p}_3 - \frac{2}{5}\,\frac{3}{2}\boldsymbol{q}_1 = x^3 - \frac{3}{5}x \quad \Rightarrow \quad \boldsymbol{q}_3 = \frac{1}{5}(5x^3 - 3x).$$

The orthogonal basis is then given by

$$\Big\{\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = x,\ \boldsymbol{q}_2 = \frac{1}{3}\,(3x^2 - 1),\ \boldsymbol{q}_3 = \frac{1}{5}(5x^3 - 3x)\Big\}.$$

These polynomials are proportional to the first three Legendre polynomials. The Legendre polynomials form an orthogonal set in the space $\mathbb{P}_\infty([-1, 1])$ of polynomials of all degrees. They play an important role in physics, since Legendre polynomials are solution of a particular differential equation that often appears in physics. $\lhd$

6.5.1. **Exercises.**

6.5.1.- Find an orthonormal basis for the subspace of $\mathbb{R}^3$ spanned by the vectors
$$\Big\{ \mathsf{u}_1 = \begin{bmatrix} -2 \\ 2 \\ -1 \end{bmatrix}, \ \mathsf{u}_2 = \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix} \Big\},$$
using the Gram-Schmidt process starting with the vector u_1.

6.5.2.- Let $W \subset \mathbb{R}^3$ be the subspace
$$\mathrm{Span}\Big\{ \mathsf{u}_1 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \ \mathsf{u}_2 = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \Big\}.$$

(a) Find an orthonormal basis for W using the Gram-Schmidt method starting with the vector u_1.

(b) Decompose the vector x below as
$$\mathsf{x} = \mathsf{x}_{\parallel} + \mathsf{x}_{\perp},$$
with $\mathsf{x}_{\parallel} \in W$ and $\mathsf{x}_{\perp} \in W^{\perp}$, where
$$\mathsf{x} = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}.$$

6.5.3.- Knowing that the column vectors in matrix A below form a linearly independent set, use the Gram-Schmidt method to find an orthonormal basis for $R(\mathsf{A})$, where
$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 5 \\ 0 & 2 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

6.5.4.- Use the Gram-Schmidt method to find an orthonormal basis for $R(\mathsf{A})$, where
$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 1 & 0 & 3 \end{bmatrix}.$$

6.5.5.- Consider the vector space $\mathbb{P}_2([0,1])$ with inner product
$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_0^1 \boldsymbol{p}(x)\,\boldsymbol{q}(x)\,dx.$$
Use the Gram-Schmidt method on the ordered basis
$$\big(\boldsymbol{p}_0 = 1, \ \boldsymbol{p}_1 = x, \ \boldsymbol{p}_2 = x^2 \big),$$
starting with vector $\boldsymbol{p}_0$, to obtain an orthogonal basis for $\mathbb{P}_2([0,1])$.

6.6. The adjoint operator

6.6.1. The Riesz Representation Theorem. The Riesz Representation Theorem is a statement concerning linear functionals on an inner product space. Recall that a *linear functional* on a vector space V over a scalar field $\mathbb{F}$ is a linear function $f : V \to \mathbb{F}$, that is, for all $\boldsymbol{x}$, $\boldsymbol{y} \in V$ and all a, $b \in \mathbb{F}$ holds $f(a\boldsymbol{x} + b\boldsymbol{y}) = a\,f(\boldsymbol{x}) + b\,f(\boldsymbol{y}) \in \mathbb{F}$. An example of a linear functional on $\mathbb{R}^3$ is the function

$$\mathbb{R}^3 \ni \mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto f(\mathsf{x}) = 3x_1 + 2x_2 + x_3 \in \mathbb{R}.$$

This function can be expressed in terms of the dot product in $\mathbb{R}^3$ as follows

$$f(\mathsf{x}) = \mathsf{u} \cdot \mathsf{x}, \qquad \mathsf{u} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

The Riesz Representation Theorem says that what we did in this example can be done in the general case. In an inner product space $\big(V, \langle\,,\rangle\big)$ every linear functional f can be expressed in terms of the inner product.

Theorem 6.6.1. *Consider a finite dimensional inner product space* $\big(V, \langle\,,\rangle\big)$ *over the scalar field* $\mathbb{F}$*. For every linear functional* $f : V \to \mathbb{F}$ *there exists a unique vector* $\boldsymbol{u}_f \in V$ *such that holds*

$$f(\boldsymbol{v}) = \langle \boldsymbol{u}_f, \boldsymbol{v} \rangle \qquad \forall \boldsymbol{v} \in V.$$

Proof of Theorem 9.4.1: Introduce the set

$$N = \{ \boldsymbol{v} \in V \;:\; f(\boldsymbol{v}) = 0 \;\} \subset V.$$

This set is the analogous to linear functionals of the null space of linear operators. Since f is a linear function the set N is a subspace of V. (Proof: Given two elements $\boldsymbol{v}_1$, $\boldsymbol{v}_2 \in N$ and two scalars a, $b \in \mathbb{F}$, holds that $f(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) = a\,f(\boldsymbol{v}_1) + b\,f(\boldsymbol{v}_2) = 0 + 0$, so $(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) \in N$.) Introduce the orthogonal complement of N, that is,

$$N^\perp = \{ \boldsymbol{w} \in V \;:\; \langle \boldsymbol{w}, \boldsymbol{v} \rangle = 0 \;\forall \boldsymbol{v} \in V \},$$

which is also a subspace of V. If $N^\perp = \{\boldsymbol{0}\}$, then $N = \big(N^\perp\big)^\perp = \big(\{\boldsymbol{0}\}\big)^\perp = V$. Since the null space of f is the whole vector space, the functional f is identically zero, so only for the choice $\boldsymbol{u}_f = \boldsymbol{0}$ holds $f(\boldsymbol{v}) = \langle \boldsymbol{0}, \boldsymbol{v} \rangle$ for all $\boldsymbol{v} \in V$.

In the case that $N^\perp \neq \{\boldsymbol{0}\}$ we now show that this space cannot be very big, in fact it has dimension one, as the following argument shows. Choose $\tilde{\boldsymbol{u}} \in N^\perp$ such that $f(\tilde{\boldsymbol{u}}) = 1$. Then notice that for every $\boldsymbol{w} \in N^\perp$ the vector $\boldsymbol{w} - f(\boldsymbol{w})\tilde{\boldsymbol{u}}$ is trivially in $N^\perp$ but it is also in N, since

$$f\big(\boldsymbol{w} - f(\boldsymbol{w})\,\tilde{\boldsymbol{u}}\big) = f(\boldsymbol{w}) - f(\boldsymbol{w})\,f(\tilde{\boldsymbol{u}}) = f(\boldsymbol{w}) - f(\boldsymbol{w}) = 0.$$

A vector both in N and $N^\perp$ must vanish, so $\boldsymbol{w} = f(\boldsymbol{w})\,\tilde{\boldsymbol{u}}$. Then every vector in $N^\perp$ is proportional to $\tilde{\boldsymbol{u}}$, so $\dim N^\perp = 1$. This information is used to split any vector $\boldsymbol{v} \in V$ as follows $\boldsymbol{v} = a\,\tilde{\boldsymbol{u}} + \boldsymbol{x}$ where $\boldsymbol{x} \in V$ and $a \in \mathbb{F}$. It is clear that

$$f(\boldsymbol{v}) = f(a\,\tilde{\boldsymbol{u}} + \boldsymbol{x}) = a\,f(\tilde{\boldsymbol{u}}) + f(\boldsymbol{x}) = a\,f(\tilde{\boldsymbol{u}}) = a.$$

However, the function with values $g(\boldsymbol{v}) = \Big\langle \dfrac{\tilde{\boldsymbol{u}}}{\|\tilde{\boldsymbol{u}}\|^2}, \boldsymbol{v} \Big\rangle$ has precisely the same values as f, since for all $\boldsymbol{v} \in V$ holds

$$g(\boldsymbol{v}) = \Big\langle \frac{\tilde{\boldsymbol{u}}}{\|\tilde{\boldsymbol{u}}\|^2}, \boldsymbol{v} \Big\rangle = \Big\langle \frac{\tilde{\boldsymbol{u}}}{\|\tilde{\boldsymbol{u}}\|^2}, (a\,\tilde{\boldsymbol{u}} + \boldsymbol{x}) \Big\rangle = \frac{a}{\|\tilde{\boldsymbol{u}}\|^2}\,\langle \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}} \rangle + \frac{1}{\|\tilde{\boldsymbol{u}}\|^2}\,\langle \tilde{\boldsymbol{u}}, \boldsymbol{x} \rangle = a.$$

Therefore, choosing $\boldsymbol{u}_f = \tilde{\boldsymbol{u}}/\|\tilde{\boldsymbol{u}}\|^2$, holds that

$$f(\boldsymbol{v}) = \langle \boldsymbol{u}_f, \boldsymbol{v}\rangle \qquad \forall \boldsymbol{v} \in V.$$

Since $\dim N^{\perp} = 1$, the choice of $\boldsymbol{u}_f$ is unique. This establishes the Theorem. □

6.6.2. **The adjoint operator.** Given a linear operator defined on an inner product space, a new linear operator can be defined through an equation involving the inner product.

Proposition 6.6.2. *If $\boldsymbol{T} \in L(V)$ is a linear operator on a finite-dimensional inner product space $\big(V, \langle\,,\rangle\big)$, then there exists a unique linear operator $\boldsymbol{T}^* \in L(V)$, called the* **adjoint** *of $\boldsymbol{T}$, such that for all vectors $\boldsymbol{u}$, $\boldsymbol{v} \in V$ holds*

$$\langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{u})\rangle = \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}\rangle.$$

Proof of Proposition 9.4.2: We first establish the following statement: For every vector $\boldsymbol{u} \in V$ there exists a unique vector $\boldsymbol{w} \in V$ such that

$$\langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}\rangle = \langle \boldsymbol{v}, \boldsymbol{w}\rangle \qquad \forall \boldsymbol{v} \in V. \tag{6.18}$$

The proof starts noticing that for a fixed $\boldsymbol{u} \in V$ the scalar-valued function $f_u : V \to \mathbb{F}$ given by $f_u(\boldsymbol{v}) = \langle u, \boldsymbol{T}(\boldsymbol{v})\rangle$ is a linear functional. Therefore, the Riesz Representation Theorem 6.6.1 implies that there exists a unique vector $\boldsymbol{w} \in V$ such that $f_u(\boldsymbol{v}) = \langle \boldsymbol{w}, \boldsymbol{v}\rangle$. This establishes that for every vector $\boldsymbol{u} \in V$ there exists a unique vector $\boldsymbol{w} \in V$ such that Eq. (6.18) holds. Now that this statement is proven we can define a map, that we choose to denote as $\boldsymbol{T}^* : V \to V$, given by $\boldsymbol{u} \mapsto \boldsymbol{T}^*(\boldsymbol{u}) = \boldsymbol{w}$. We now show that this map $\boldsymbol{T}^*$ is linear. Indeed, for all $\boldsymbol{u}_1$, $\boldsymbol{u}_2 \in V$ and all a, $b \in \mathbb{F}$ holds

$$\begin{aligned}
\langle \boldsymbol{v}, \boldsymbol{T}^*(a\boldsymbol{u}_1 + b\boldsymbol{u}_2)\rangle &= \langle \boldsymbol{T}(\boldsymbol{v}), (a\boldsymbol{u}_1 + b\boldsymbol{u}_2)\rangle \qquad \forall \boldsymbol{v} \in V,\\
&= a\,\langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}_1\rangle + b\,\langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}_2)\rangle\\
&= a\langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{u}_1)\rangle + b\langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{u}_2)\rangle\\
&= \big\langle \boldsymbol{v}, \big[a\,\boldsymbol{T}^*(\boldsymbol{u}_1) + b\,\boldsymbol{T}^*(\boldsymbol{u}_2)\big]\big\rangle \qquad \forall \boldsymbol{v} \in V,
\end{aligned}$$

hence $\boldsymbol{T}^*(a\boldsymbol{u}_1 + b\boldsymbol{u}_2) = a\,\boldsymbol{T}^*(\boldsymbol{u}_1) + b\,\boldsymbol{T}^*(\boldsymbol{u}_2)$. This establishes the Proposition. □

The next result relates the adjoint of a linear operator with the concept of the adjoint of a square matrix introduced in Sect. 2.2. Recall that given a basis in the vector space, every linear operator has associated a unique square matrix. Let us use the notation $[\boldsymbol{T}]$ and $[\boldsymbol{T}^*]$ for the matrices on a given basis of the operators $\boldsymbol{T}$ and $\boldsymbol{T}^*$, respectively.

Proposition 6.6.3. *Let $\big(V, \langle\,,\rangle\big)$ be a finite-dimensional vector space, let $\mathcal{V}$ be an orthonormal basis of V, and let $[\boldsymbol{T}]$ be the matrix of the linear operator $\boldsymbol{T} \in L(V)$ in the basis $\mathcal{V}$. Then, the matrix of the adjoint operator $\boldsymbol{T}^*$ in the basis $\mathcal{V}$ is given by $[\boldsymbol{T}^*] = [\boldsymbol{T}]^*$.*

Proposition 6.6.3 says that the matrix of the adjoint operator is the adjoint of the matrix of the operator, however this is true only in the case that the basis used to compute the respective matrices is orthonormal. If the basis is not orthonormal, the relation between the matrices $[\boldsymbol{T}]$ and $[\boldsymbol{T}^*]$ is more involved.

Proof of Proposition 6.6.3: Let $\mathcal{V} = \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ be an orthonormal basis of V, that is,

$$\langle \boldsymbol{e}_i, \boldsymbol{e}_j\rangle = \begin{cases} 0 & \text{if} \quad i \neq j,\\ 1 & \text{if} \quad i = j.\end{cases}$$

The components of two arbitrary vectors $\boldsymbol{u}$, $\boldsymbol{v} \in V$ in the basis $\mathcal{V}$ is denoted as follows

$$\boldsymbol{u} = \sum_i u_i\,\boldsymbol{e}_i, \qquad \boldsymbol{v} = \sum_i v_i\,\boldsymbol{e}_i.$$

The action of the operator $\boldsymbol{T}$ can also be decomposed in the basis $\mathcal{V}$ as follows

$$\boldsymbol{T}(\boldsymbol{e}_j) = \sum_i [\boldsymbol{T}]_{ij}\,\boldsymbol{e}_i, \qquad [\boldsymbol{T}]_{ij} = [\boldsymbol{T}(\boldsymbol{e}_j)]_i.$$

We use the same notation for the adjoint operator, that is,

$$\boldsymbol{T}^*(\boldsymbol{e}_j) = \sum_i [\boldsymbol{T}^*]_{ij}\,\boldsymbol{e}_i, \qquad [\boldsymbol{T}^*]_{ij} = [\boldsymbol{T}^*(\boldsymbol{e}_j)]_i.$$

The adjoint operator is defined such that the equation $\langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{u})\rangle = \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{u}\rangle$ holds for all $\boldsymbol{u}$, $\boldsymbol{v} \in V$. This equation can be expressed in terms of components in the basis $\mathcal{V}$ as follows

$$\sum_{ijk} \big\langle v_i\boldsymbol{e}_i, u_j[\boldsymbol{T}^*(\boldsymbol{e}_j)]_k\boldsymbol{e}_k\big\rangle = \sum_{ijk} \big\langle v_i[\boldsymbol{T}(\boldsymbol{e}_i)]_k\boldsymbol{e}_k, u_j\boldsymbol{e}_j\big\rangle,$$

that is,

$$\sum_{ijk} \overline{v}_i u_j [\boldsymbol{T}^*]_{kj} \langle \boldsymbol{e}_i, \boldsymbol{e}_k\rangle = \sum_{ijk} \overline{v}_i \overline{[\boldsymbol{T}]}_{ki} u_j \langle \boldsymbol{e}_k, \boldsymbol{e}_j\rangle.$$

Since the basis $\mathcal{V}$ is orthonormal we obtain the equation

$$\sum_{ij} \overline{v}_i u_j [\boldsymbol{T}^*]_{ij} = \sum_{ijk} \overline{v}_i \overline{[\boldsymbol{T}]}_{ji} u_j,$$

which holds for all vectors $\boldsymbol{u}$, $\boldsymbol{v} \in V$, so we conclude

$$[\boldsymbol{T}^*]_{ij} = \overline{[\boldsymbol{T}]}_{ji} \quad \Leftrightarrow \quad [\boldsymbol{T}^*] = \overline{[\boldsymbol{T}]}^T \quad \Leftrightarrow \quad [\boldsymbol{T}^*] = [\boldsymbol{T}]^*.$$

This establishes the Proposition. $\square$

Example 6.6.1: Consider the inner product space $\big(\mathbb{C}^3, \cdot\big)$. Find the adjoint of the linear operator $\boldsymbol{T}$ with matrix in the standard basis of $\mathbb{C}^3$ given by

$$[\boldsymbol{T}(\boldsymbol{x})] = \begin{bmatrix} x_1 + 2ix_2 + ix_3 \\ ix_1 - x_3 \\ x_1 - x_2 + ix_3 \end{bmatrix}, \qquad [\boldsymbol{x}] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Solution: The matrix of this operator in the standard basis of $\mathbb{C}^3$ is given by

$$[\boldsymbol{T}] = \begin{bmatrix} 1 & 2i & i \\ i & 0 & -1 \\ 1 & -1 & i \end{bmatrix}.$$

Since the standard basis is an orthonormal basis with respect to the dot product, Proposition 6.6.3 implies that

$$[\boldsymbol{T}^*] = [\boldsymbol{T}]^* = \begin{bmatrix} 1 & 2i & i \\ i & 0 & -1 \\ 1 & -1 & i \end{bmatrix}^* = \begin{bmatrix} 1 & -i & 1 \\ -2i & 0 & -1 \\ -i & -1 & -i \end{bmatrix} \quad \Rightarrow \quad [\boldsymbol{T}^*(\boldsymbol{x})] = \begin{bmatrix} x_1 - ix_2 + x_3 \\ -2ix_1 - x_3 \\ -ix_1 - x_2 - ix_3 \end{bmatrix}.$$

$\triangleleft$

6.6.3. **Normal operators.** Recall now that the commutator of two linear operators $\boldsymbol{T}$, $\boldsymbol{S} \in L(V)$ is the linear operator $[\boldsymbol{T}, \boldsymbol{S}] \in L(V)$ given by

$$[\boldsymbol{T}, \boldsymbol{S}](\boldsymbol{u}) = \boldsymbol{T}(\boldsymbol{S}(\boldsymbol{u})) - \boldsymbol{S}(\boldsymbol{T}(\boldsymbol{u})) \qquad \forall \boldsymbol{u} \in V.$$

Two operators $\boldsymbol{T}$, $\boldsymbol{S} \in L(V)$ are said to commute iff their commutator vanishes, that is, $[\boldsymbol{T}, \boldsymbol{S}] = \boldsymbol{0}$. Examples of operators that commute are two rotations on the plane. Examples of operators that do not commute are two arbitrary rotations in space.

Definition 6.6.4. *A linear operator* $\boldsymbol{T}$ *defined on a finite-dimensional inner product space* $\big(V, \langle\,,\rangle\big)$ *is called a* **normal operator** *iff holds* $[\boldsymbol{T}, \boldsymbol{T}^*] = \boldsymbol{0}$*, that is, the operator commutes with its adjoint.*

An interesting characterization of normal operators is the following: A linear operator $\boldsymbol{T}$ on an inner product space is normal iff $\|\boldsymbol{T}(\boldsymbol{u})\| = \|\boldsymbol{T}^*(\boldsymbol{u})\|$ holds for all $\boldsymbol{u} \in V$. Normal operators are particularly important because for these operators hold the Spectral Theorem, which we study in Chapter 9.

Two particular cases of normal operators are often used in physics. A linear operator $\boldsymbol{T}$ on an inner product space is called a **unitary operator** iff $\boldsymbol{T}^* = \boldsymbol{T}^{-1}$, that is, the adjoint is the inverse operator. Unitary operators are normal operators, since

$$\boldsymbol{T}^* = \boldsymbol{T}^{-1} \quad \Rightarrow \quad \begin{cases} \boldsymbol{T}\boldsymbol{T}^* = \boldsymbol{I}, \\ \boldsymbol{T}^*\boldsymbol{T} = \boldsymbol{I}, \end{cases} \quad \Rightarrow \quad [\boldsymbol{T}, \boldsymbol{T}^*] = \boldsymbol{0}.$$

Unitary operators preserve the length of a vector, since

$$\|\boldsymbol{v}\|^2 = \langle \boldsymbol{v}, \boldsymbol{v}\rangle = \langle \boldsymbol{v}, \boldsymbol{T}^{-1}(\boldsymbol{T}(\boldsymbol{v}))\rangle = \langle \boldsymbol{v}, \boldsymbol{T}^*(\boldsymbol{T}(\boldsymbol{v}))\rangle = \langle \boldsymbol{T}(\boldsymbol{v}), \boldsymbol{T}(\boldsymbol{v})\rangle = \|\boldsymbol{T}(\boldsymbol{v})\|^2.$$

Unitary operators defined on a complex inner product space are particularly important in quantum mechanics. The particular case of unitary operators on a real inner product space are called **orthogonal operators**. So, orthogonal operators do not change the length of a vector. Examples of orthogonal operators are rotations in $\mathbb{R}^3$.

A linear operator $\boldsymbol{T}$ on an inner product space is called an **Hermitian operator** iff $\boldsymbol{T}^* = \boldsymbol{T}$, that is, the adjoint is the original operator. This definition agrees with the definition of Hermitian matrices given in Chapter 2.

EXAMPLE 6.6.2: Consider the inner product space $\big(\mathbb{C}^3, \cdot\big)$ and the linear operator $\boldsymbol{T}$ with matrix in the standard basis of $\mathbb{C}^3$ given by

$$[\boldsymbol{T}(\boldsymbol{x})] = \begin{bmatrix} x_1 - ix_2 + x_3 \\ ix_1 - x_3 \\ x_1 - x_2 + x_3 \end{bmatrix}, \qquad [\boldsymbol{x}] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Show that $\boldsymbol{T}$ is Hermitian.

SOLUTION: We need to compute the adjoint of $\boldsymbol{T}$. The matrix of this operator in the standard basis of $\mathbb{C}^3$ is given by

$$[\boldsymbol{T}] = \begin{bmatrix} 1 & -i & 1 \\ i & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix}.$$

Since the standard basis is an orthonormal basis with respect to the dot product, Proposition 6.6.3 implies that

$$[\boldsymbol{T}^*] = [\boldsymbol{T}]^* = \begin{bmatrix} 1 & -i & 1 \\ i & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix}^* = \begin{bmatrix} 1 & -i & 1 \\ i & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix} = [\boldsymbol{T}].$$

Therefore, $\boldsymbol{T}^* = \boldsymbol{T}$. ◁

6.6.4. Bilinear forms.

Definition 6.6.5. *A* **bilinear form** *on a vector space* V *over* $\mathbb{F}$ *is a function* $a : V \times V \to \mathbb{F}$ *linear on both arguments, that is, for all* $\boldsymbol{u}$*,* $\boldsymbol{v}_1$*,* $\boldsymbol{v}_2 \in V$ *and all* b_1*,* $b_2 \in \mathbb{F}$ *holds*

$$a\big(\boldsymbol{u}, (b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2)\big) = b_1 a(\boldsymbol{u}, \boldsymbol{v}_1) + b_2 a(\boldsymbol{u}, \boldsymbol{v}_2),$$
$$a\big((b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2), \boldsymbol{u}\big) = b_1 a(\boldsymbol{v}_1, \boldsymbol{u}) + b_2 a(\boldsymbol{v}_2, \boldsymbol{u}).$$

The bilinear form $a : V \times V \to \mathbb{F}$ *is called* **symmetric** *iff for all* $\boldsymbol{u}, \boldsymbol{v} \in V$ *holds*

$$a(\boldsymbol{u}, \boldsymbol{v}) = a(\boldsymbol{v}, \boldsymbol{u}).$$

An example of a symmetric bilinear form is any the inner product on a real vector space. Indeed, given a real inner product space $\big(V, \langle\,,\rangle\big)$, the function $\langle\,,\rangle : V \times V \to \mathbb{R}$ is a symmetric bilinear form, since it is symmetric and

$$\langle \boldsymbol{u}, (b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2)\rangle = b_1\langle \boldsymbol{u}, \boldsymbol{v}_1\rangle + b_2\langle \boldsymbol{u}, \boldsymbol{v}_2\rangle.$$

We will shortly see that another example of a bilinear form appears on the weak formulation of the boundary value problem in Eq. (7.30).

On the other hand, an inner product on a complex vector space is not a bilinear form, since it is conjugate linear on the first argument instead of linear. Such functions are called sesquilinear forms. That is, a **sesquilinear form** on a complex vector space V is a function $a : V \times V \to \mathbb{C}$ that is conjugate linear on the first argument and linear on the second argument. Sesquilinear forms are important in the case of studying differential equations involving complex functions.

Definition 6.6.6. *Consider a bilinear form* $a : V \times V \to \mathbb{F}$ *on an inner product space* $\big(V, \langle\,,\rangle\big)$ *over* $\mathbb{F}$. *The bilinear form* a *is called* **positive** *iff there exists a real number* $k > 0$ *such that for all* $\boldsymbol{u} \in V$ *holds*

$$a(\boldsymbol{u}, \boldsymbol{u}) \geqslant k\,\|\boldsymbol{u}\|^2.$$

The bilinear form a *is called* **bounded** *iff there exists a real number* $K > 0$ *such that for all* $\boldsymbol{u}, \boldsymbol{v} \in V$ *holds*

$$a(\boldsymbol{u}, \boldsymbol{v}) \leqslant K\,\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|.$$

An example of a positive bilinear form is any inner product on a real vector space. The Schwarz inequality implies that such inner product is also a bounded bilinear form. In fact, an inner product on a real vector space is a symmetric, positive, bounded bilinear form. We will shortly see that the bilinear form that appears on a weak formulation of the boundary value problem in Eq. (7.30) is symmetric, positive and bounded. We will see that these properties imply the existence and uniqueness of solutions to the weak formulation of the boundary value problem.

6.6.5. **Exercises.**

6.6.1.- .

6.6.2.- .

Chapter 7. Approximation methods

7.1. Best approximation

The first half of this Section is dedicated to show that the Fourier series approximation of a function is a particular case of the orthogonal decomposition of a vector onto a subspace in an inner product space, studied in Section 6.4. Once we realize that it is not difficult to see why such Fourier approximations are useful. We show that the orthogonal projection $\boldsymbol{x}_{\shortparallel}$ of a vector $\boldsymbol{x}$ onto a subspace U is the vector in the subspace U closest to $\boldsymbol{x}$. This is the origin of the name "best approximation" for $\boldsymbol{x}_{\shortparallel}$. What is intuitively clear in $\big(\mathbb{R}^3, \cdot\big)$ is true in every inner product space, hence it is true for the Fourier series approximation of a function. In the second half of this Section we show a deep relation between the Null space of a matrix and the Range space of the adjoint matrix. The former is the orthogonal complement of the latter. A consequence of this relation is a simple proof to the property that a matrix and its adjoint matrix have the same rank. Another consequence is given in the next Section, where we obtain a simple equation, called the normal equation, to find a least squares solution of an inconsistent linear system.

7.1.1. Fourier expansions.

We have seen that orthonormal bases have a practical advantage over arbitrary basis. The components $[\boldsymbol{x}]_u$ of a vector $\boldsymbol{x}$ in an orthonormal basis $\mathcal{U}_n = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$ of the inner product space $\big(V, \langle\,,\rangle\big)$ are given by the simple expression

$$[\boldsymbol{x}]_u = \begin{bmatrix} \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle \\ \vdots \\ \langle \boldsymbol{u}_n, \boldsymbol{x}\rangle \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{x} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\, \boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_n, \boldsymbol{x}\rangle\, \boldsymbol{u}_n.$$

In the case that an orthonormal set $\mathcal{U}_p$ is not a basis of V, that is, $p < \dim V$, one can always introduce the orthogonal projection of a vector $\boldsymbol{x} \in V$ onto the subspace $U_p = \operatorname{Span}(\mathcal{U}_p)$,

$$\boldsymbol{x}_{\shortparallel} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\, \boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_p, \boldsymbol{x}\rangle\, \boldsymbol{u}_p.$$

We have seen that in this case, $\boldsymbol{x}_{\shortparallel} \neq \boldsymbol{x}$. In fact we called the difference vector $\boldsymbol{x}_{\perp} = \boldsymbol{x} - \boldsymbol{x}_{\shortparallel}$, since this vector satisfies that $\boldsymbol{x}_{\perp} \perp \boldsymbol{x}_{\shortparallel}$. We now give the projection vector $\boldsymbol{x}_{\shortparallel}$ a new name.

Definition 7.1.1. *The* ***Fourier expansion*** *of a vector $\boldsymbol{x} \in V$ with respect to an orthonormal set $\mathcal{U}_p = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p\} \subset \big(V, \langle\,,\rangle\big)$ is the unique vector $\boldsymbol{x}_{\shortparallel} \in \operatorname{Span}(\mathcal{U}_p)$ given by*

$$\boldsymbol{x}_{\shortparallel} = \langle \boldsymbol{u}_1, \boldsymbol{x}\rangle\, \boldsymbol{u}_1 + \cdots + \langle \boldsymbol{u}_p, \boldsymbol{x}\rangle\, \boldsymbol{u}_p. \tag{7.1}$$

The scalars $\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle$ are the ***Fourier coefficient*** *of the vector $\boldsymbol{x}$ with respect to the set $\mathcal{U}_p$.*

A reason to the name "Fourier expansion" to the orthogonal projection of a vector onto a subspace is given in the following example.

Example 7.1.1: Given the vector space of continuous functions $V = C\big([-\ell, \ell], \mathbb{R}\big)$ with inner product given by $\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \displaystyle\int_{-\ell}^{\ell} \boldsymbol{f}(x)\boldsymbol{g}(x)\, dx$, find the Fourier expansion of an arbitrary function $\boldsymbol{f} \in V$ with respect to the orthonormal set

$$\mathcal{U}_N = \Big\{ \boldsymbol{u}_0 = \frac{1}{\sqrt{2\ell}},\ \boldsymbol{u}_n = \frac{1}{\sqrt{\ell}} \cos\Big(\frac{n\pi x}{\ell}\Big),\ \boldsymbol{v}_n = \frac{1}{\sqrt{\ell}} \sin\Big(\frac{n\pi x}{\ell}\Big) \Big\}_{n=1}^{N}.$$

Solution: Using Eq. (7.1) on a function $\boldsymbol{f} \in V$ we get

$$\boldsymbol{f}_{\shortparallel} = \langle \boldsymbol{u}_0, \boldsymbol{f}\rangle\, \boldsymbol{u}_0 + \sum_{n=1}^{N} \Big[\langle \boldsymbol{u}_n, \boldsymbol{f}\rangle\, \boldsymbol{u}_n + \langle \boldsymbol{v}_n, \boldsymbol{f}\rangle\, \boldsymbol{v}_n \Big].$$

Introduce the vectors in $\mathcal{U}_N$ explicitly in the expression above,

$$\boldsymbol{f}_{\shortparallel} = \frac{1}{\sqrt{2\ell}}\langle \boldsymbol{u}_0, \boldsymbol{f}\rangle + \sum_{n=1}^{N}\Big[\frac{1}{\sqrt{\ell}}\,\langle \boldsymbol{u}_n, \boldsymbol{f}\rangle\,\cos\Big(\frac{n\pi x}{\ell}\Big) + \frac{1}{\sqrt{\ell}}\,\langle \boldsymbol{v}_n, \boldsymbol{f}\rangle\,\sin\Big(\frac{n\pi x}{\ell}\Big)\Big].$$

Denoting by

$$a_0 = \frac{1}{\sqrt{2\ell}}\,\langle \boldsymbol{u}_0, \boldsymbol{f}\rangle \quad a_n = \frac{1}{\sqrt{\ell}}\,\langle \boldsymbol{u}_n, \boldsymbol{f}\rangle \quad b_n = \frac{1}{\sqrt{\ell}}\,\langle \boldsymbol{u}_n, \boldsymbol{f}\rangle,$$

then we get that

$$\boldsymbol{f}_{\shortparallel}(x) = a_0 + \sum_{n=1}^{N}\Big[a_n\,\cos\Big(\frac{n\pi x}{\ell}\Big) + b_n\,\sin\Big(\frac{n\pi x}{\ell}\Big)\Big],$$

where the coefficients a_0, a_n and b_n, for $n = 1, \cdots, N$ are the usual Fourier coefficients,

$$a_0 = \frac{1}{2\ell}\int_{-\ell}^{\ell}\boldsymbol{f}(x)\,dx, \quad a_n = \frac{1}{\ell}\int_{-\ell}^{\ell}\boldsymbol{f}(x)\cos\Big(\frac{n\pi x}{\ell}\Big)\,dx, \quad b_n = \frac{1}{\ell}\int_{-\ell}^{\ell}\boldsymbol{f}(x)\sin\Big(\frac{n\pi x}{\ell}\Big)\,dx.$$

$\lhd$

The example above is a good reason to name Fourier expansion the orthogonal projection of a vector onto a subspace. We already know that the Fourier expansion $\boldsymbol{x}_{\shortparallel}$ of a vector $\boldsymbol{x}$ with respect to a an orthonormal set $\mathcal{U}_p$ has a particular property, that is, $(\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}) = \boldsymbol{x}_{\perp} \in U_p^{\perp}$. This property means that $\boldsymbol{x}_{\shortparallel}$ is the best approximation of the vector $\boldsymbol{x}$ from within the subspace $\mathrm{Span}(\mathcal{U}_p)$. See Fig. 45 for the case $V = \mathbb{R}^3$, with $\langle\,,\rangle = \cdot$, and $\mathrm{Span}(\mathcal{U}_2)$ two-dimensional. We highlight this property in the following result.

Theorem 7.1.2 (Best approximation). *The Fourier expansion $\boldsymbol{x}_{\shortparallel}$ of a vector $\boldsymbol{x}$ with respect to an orthonormal set $\mathcal{U}_p$ in an inner product space, is the unique vector in the subspace $\mathrm{Span}(\mathcal{U}_p)$ that is closest to $\boldsymbol{x}$, in the sense that*

$$\|\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}\| < \|\boldsymbol{x} - \boldsymbol{y}\| \qquad \forall\ \boldsymbol{y} \in \mathrm{Span}(\mathcal{U}_p) - \{\boldsymbol{x}_{\shortparallel}\}.$$

Proof of Theorem 7.1.2: Recall that $\boldsymbol{x}_{\perp} = \boldsymbol{x} - \boldsymbol{x}_{\shortparallel}$ is orthogonal to $\mathrm{Span}(U)$, that is $(\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}) \perp (\boldsymbol{x}_{\shortparallel} - \boldsymbol{y})$ for all $\boldsymbol{y} \in \mathrm{Span}(U)$. Hence,

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|(\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}) + (\boldsymbol{x}_{\shortparallel} - \boldsymbol{y})\|^2 = \|\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}\|^2 + \|\boldsymbol{x}_{\shortparallel} - \boldsymbol{y}\|^2, \tag{7.2}$$

where the las equality comes from Pythagoras Theorem. Eq. (7.2) says that $\|\boldsymbol{x} - \boldsymbol{y}\|$ is the smallest iff $\boldsymbol{y} = \boldsymbol{x}_{\shortparallel}$ and the smallest value is $\|\boldsymbol{x} - \boldsymbol{x}_{\shortparallel}\|$. This establishes the Theorem. $\square$

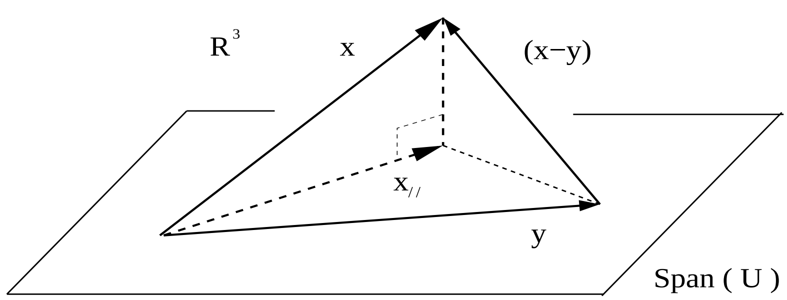

FIGURE 45. The Fourier expansion $\mathsf{x}_{\shortparallel}$ of the vector $\mathsf{x} \in \mathbb{R}^3$ is the best approximation of x from within $\mathrm{Span}(U)$.

Example 7.1.2: Given the vector space of continuous functions $V = C([-1,1],\mathbb{R})$ with the inner product given by $\langle \boldsymbol{f},\boldsymbol{g}\rangle = \int_{-1}^{1} \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx$, find the Fourier expansion of the function $\boldsymbol{f}(x) = x$ with respect to the orthonormal set

$$\mathcal{U}_N = \Big\{ \boldsymbol{u}_0 = \frac{1}{\sqrt{2\ell}},\ \boldsymbol{u}_n = \frac{1}{\sqrt{\ell}} \cos\Big(\frac{n\pi x}{\ell}\Big),\ \boldsymbol{v}_n = \frac{1}{\sqrt{\ell}} \sin\Big(\frac{n\pi x}{\ell}\Big) \Big\}_{n=1}^{N}.$$

Solution: We use the formulas in Example 7.1.1 above to the function $\boldsymbol{f}(x) = x$ on the interval $[-1,1]$, since $\ell = 1$. The coefficient a_0 above is given by

$$a_0 = \frac{1}{2}\int_{-1}^{1} x\,dx = \frac{1}{4}\big(x^2\big|_{-1}^{1}\big) \quad\Rightarrow\quad a_0 = 0.$$

The coefficients a_n, b_n, for $n = 1,\cdots,N$ computed with one integration by parts,

$$\int x\cos(n\pi x)\,dx = \frac{x}{n\pi}\sin(n\pi x) + \frac{1}{n^2\pi^2}\cos(n\pi x),$$

$$\int x\sin(n\pi x)\,dx = -\frac{x}{n\pi}\cos(n\pi x) + \frac{1}{n^2\pi^2}\sin(n\pi x).$$

The coefficients a_n vanish, since

$$a_n = \int_{-1}^{1} x\cos(n\pi x)\,dx = \big[\frac{x}{n\pi}\sin(n\pi x)\big]\Big|_{-1}^{1} + \frac{1}{n^2\pi^2}\cos(n\pi x)\Big|_{-1}^{1} \quad\Rightarrow\quad a_n = 0.$$

The coefficients b_n are given by

$$b_n = \int_{-1}^{1} x\sin(n\pi x)\,dx = -\Big[\frac{x}{n\pi}\cos(n\pi x)\Big]\Big|_{-1}^{1} + \frac{1}{n^2\pi^2}\sin(n\pi x)\Big|_{-1}^{1} \quad\Rightarrow\quad b_n = \frac{2(-1)^{(n+1)}}{n\pi}.$$

Therefore, the Fourier expansion of $\boldsymbol{f}(x) = x$ with respect to $\mathcal{U}_N$ is given by

$$\boldsymbol{f}_{\parallel}(x) = \frac{2}{\pi}\sum_{n=1}^{N} \frac{(-1)^{(n+1)}}{n}\sin(n\pi x).$$

Remark: First, a simpler proof that the coefficients a_0 and a_n vanish is to realized that we are integrating an odd function on the interval $[-1,1]$. The odd function is the product of an odd function times an even function. Second, Theorem 7.1.2 tells us that the function $\boldsymbol{f}_{\parallel}$ above is only combination of the sine and cosine functions in $\mathcal{U}_N$ that approximates best the function $\boldsymbol{f}(x) = x$ on the interval $[-1,1]$. ◁

7.1.2. **Null and range spaces of a matrix.** The null and range spaces associated with a matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ and its adjoint matrix A^* are deeply related.

Theorem 7.1.3. *For every matrix* $\mathsf{A} \in \mathbb{F}^{m,n}$ *holds* $N(\mathsf{A}) = R(\mathsf{A}^*)^{\perp}$ *and* $N(\mathsf{A}^*) = R(\mathsf{A})^{\perp}$.

Since for every subspace W on a finite dimensional inner product space holds $(W^{\perp})^{\perp} = W$, we also have the relations

$$N(\mathsf{A})^{\perp} = R(\mathsf{A}^*), \qquad N(\mathsf{A}^*)^{\perp} = R(\mathsf{A}).$$

In the case of real-valued matrices, the Theorem above says that

$$N(\mathsf{A}) = R(\mathsf{A}^T)^{\perp} \quad\text{and}\quad N(\mathsf{A}^T) = R(\mathsf{A})^{\perp}.$$

Before we state the proof of this Theorem let us review the following notation: Given an $m \times n$ matrix $\mathsf{A} \in \mathbb{F}^{m,n}$, we write it either in terms of column vectors $\mathsf{A}_{:j} \in \mathbb{F}^m$ for $j = 1, \cdots, n$, or in terms of row vectors $\mathsf{A}_{i:} \in \mathbb{F}^n$ for $i = 1, \cdots, m$, as follows,

$$\mathsf{A} = \begin{bmatrix} \mathsf{A}_{:1} & \cdots & \mathsf{A}_{:n} \end{bmatrix}, \qquad \mathsf{A} = \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{m:} \end{bmatrix}.$$

Since the same type of definition holds for the $n \times m$ matrix A^*, that is,

$$\mathsf{A}^* = \begin{bmatrix} (\mathsf{A}^*)_{:1} & \cdots & (\mathsf{A}^*)_{:m} \end{bmatrix}, \qquad \mathsf{A}^* = \begin{bmatrix} (\mathsf{A}^*)_{1:} \\ \vdots \\ (\mathsf{A}^*)_{n:} \end{bmatrix},$$

then we have the relations

$$(\mathsf{A}_{:j})^* = (\mathsf{A}^*)_{j:}, \qquad (\mathsf{A}_{i:})^* = (\mathsf{A}^*)_{:i}.$$

For example, consider the 2×3 matrix

$$\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \Rightarrow \quad \mathsf{A}_{:1} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \ \mathsf{A}_{:2} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \ \mathsf{A}_{:3} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \begin{matrix} \mathsf{A}_{1:} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \\ \mathsf{A}_{2:} = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}. \end{matrix}.$$

The transpose is a 3×2 matrix that can be written as follows

$$\mathsf{A}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \quad \Rightarrow \quad (\mathsf{A}^T)_{:1} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ (\mathsf{A}^T)_{:2} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \quad \begin{matrix} (\mathsf{A}^T)_{1:} = \begin{bmatrix} 1 & 4 \end{bmatrix}, \\ (\mathsf{A}^T)_{2:} = \begin{bmatrix} 2 & 5 \end{bmatrix}, \\ (\mathsf{A}^T)_{3:} = \begin{bmatrix} 3 & 6 \end{bmatrix}. \end{matrix}.$$

So, for example we have the relation

$$(\mathsf{A}_{:3})^T = \begin{bmatrix} 3 & 6 \end{bmatrix} = (\mathsf{A}^T)_{3:}, \qquad (\mathsf{A}_{2:})^T = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = (\mathsf{A}^T)_{:2}.$$

Proof of Theorem 7.1.3: We first show that the $N(\mathsf{A}) = R(\mathsf{A}^*)^\perp$. A vector $\mathsf{x} \in \mathbb{F}^n$ belongs to $N(\mathsf{A})$ iff holds

$$\mathsf{Ax} = 0 \quad \Leftrightarrow \quad \begin{bmatrix} \mathsf{A}_{1:} \\ \vdots \\ \mathsf{A}_{m:} \end{bmatrix} \mathsf{x} = 0 \quad \Leftrightarrow \quad \begin{bmatrix} \big[(\mathsf{A}^*)_{:1}\big]^* \\ \vdots \\ \big[(\mathsf{A}^*)_{:m}\big]^* \end{bmatrix} \mathsf{x} = 0 \quad \Leftrightarrow \quad \begin{cases} (\mathsf{A}^*)_{:1} \cdot \mathsf{x} = 0, \\ \qquad \vdots \\ (\mathsf{A}^*)_{:m} \cdot \mathsf{x} = 0. \end{cases}$$

So, $\mathsf{x} \in N(\mathsf{A})$ iff x is orthogonal to every column vector in A^*, that is, $\mathsf{x} \in R(\mathsf{A}^*)^\perp$.

The equation $N(\mathsf{A}^*) = R(\mathsf{A})^\perp$ comes from $N(\mathsf{B}) = R(\mathsf{B}^*)^\perp$ taking $\mathsf{B} = \mathsf{A}^*$. Nevertheless, we repeat the proof above, just to understand the previous argument. A vector $\mathsf{y} \in \mathbb{F}^m$ belongs to $N(\mathsf{A}^*)$ iff

$$\mathsf{A}^*\mathsf{y} = 0 \quad \Leftrightarrow \quad \begin{bmatrix} (\mathsf{A}^*)_{1:} \\ \vdots \\ (\mathsf{A}^*)_{n:} \end{bmatrix} \mathsf{y} = 0 \quad \Leftrightarrow \quad \begin{bmatrix} (\mathsf{A}_{:1})^* \\ \vdots \\ (\mathsf{A}_{:n})^* \end{bmatrix} \mathsf{y} = 0 \quad \Leftrightarrow \quad \begin{cases} \mathsf{A}_{:1} \cdot \mathsf{y} = 0, \\ \qquad \vdots \\ \mathsf{A}_{:n} \cdot \mathsf{y} = 0. \end{cases}$$

So, $\mathsf{y} \in N(\mathsf{A}^*)$ iff y is orthogonal to every column vector in A, that is, $\mathsf{y} \in R(\mathsf{A})^\perp$. This establishes the Theorem. $\square$

EXAMPLE 7.1.3: Verify Theorem 7.1.3 for the matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$.

SOLUTION: We first find the $N(A)$, that is, all $\mathsf{x} \in \mathbb{R}^3$ solutions of $\mathsf{A}\mathsf{y} = \mathsf{0}$. Gauss operations on matrix A imply

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} x_1 = x_3, \\ x_2 = -2x_3, \\ x_3 \text{ free}, \end{cases} \quad \Rightarrow \quad N(\mathsf{A}) = \operatorname{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \Big\}\Big).$$

It is simple to find $R(\mathsf{A}^T)$, since

$$R(\mathsf{A}^T) = \operatorname{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \Big\}\Big).$$

Theorem 7.1.3 is verified, since

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 1-4+3 = 0, \quad \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 3-4+1 = 0 \quad \Rightarrow \quad N(\mathsf{A}) = R(\mathsf{A}^T)^\perp.$$

Let us verify the same Theorem for A^T. We first find $N(\mathsf{A}^T)$, that is, all $\mathsf{y} \in \mathbb{R}^2$ solutions of $\mathsf{A}^T\mathsf{y} = \mathsf{0}$. Gauss operations on matrix A^T imply

$$\begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \mathsf{y} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad N(\mathsf{A}^T) = \{\mathsf{0}\}.$$

The space $R(\mathsf{A})$ is given by

$$R(\mathsf{A}) = \operatorname{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \Big\}\Big) = \operatorname{Span}\Big(\Big\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix} \Big\}\Big) = \mathbb{R}^2.$$

Since $(\mathbb{R}^2)^\perp = \{\mathsf{0}\}$, Theorem 7.1.3 is verified. ◁

Theorem 7.1.3 provides a simple proof for a result we used in Chapter 2.

Theorem 7.1.4. *For every matrix* $\mathsf{A} \in \mathbb{F}^{m,n}$ *holds that* $\operatorname{rank}(\mathsf{A}) = \operatorname{rank}(\mathsf{A}^*)$.

Proof of Theorem 7.1.4: Recall the Nullity-Rank result in Corollary 5.1.8, which says that for all matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ holds $\dim N(\mathsf{A}) + \dim R(\mathsf{A}) = n$. Equivalently,

$$\dim R(\mathsf{A}) = n - \dim N(\mathsf{A}) = n - \dim R(\mathsf{A}^*)^\perp,$$

since $N(\mathsf{A}) = R(\mathsf{A}^*)^\perp$. From the orthogonal decomposition $\mathbb{F}^n = R(\mathsf{A}^*) \oplus R(\mathsf{A}^*)^\perp$ we know that $\dim R(\mathsf{A}^*) = n - \dim R(\mathsf{A}^*)^\perp$. We then conclude that

$$\dim R(\mathsf{A}) = \dim R(\mathsf{A}^*).$$

This establishes the Theorem. □

7.1.3. Exercises.

7.1.1.- Consider the inner product space $(\mathbb{R}^3, \cdot)$ and the orthonormal set $\mathcal{U}$,

$$\Big\{\mathsf{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix}, \mathsf{u}_2 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}\Big\}.$$

Find the best approximation of x below in the subspace $\mathrm{Span}(\mathcal{U})$, where

$$\mathsf{x} = \begin{bmatrix}1\\0\\-2\end{bmatrix}.$$

7.1.2.- Consider the inner product space $(\mathbb{R}^{2,2}, \langle\,,\rangle_F)$ and the orthonormal set $\mathcal{U} = \{\mathsf{E}_1, \mathsf{E}_2\}$, where

$$\mathsf{E}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}, \ \mathsf{E}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}.$$

Find the best approximation of matrix A below in the subspace $\mathrm{Span}(\mathcal{U})$, where

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

7.1.3.- Consider the inner product space $\mathbb{P}_2([0,1])$, with $\langle \boldsymbol{p}, \boldsymbol{q}\rangle = \int_0^1 \boldsymbol{p}(x)\boldsymbol{q}(x)\,dx$, and the subspace $U = \mathrm{Span}(\mathcal{U})$, where $\mathcal{U} = \{\boldsymbol{q}_0 = 1,\ \boldsymbol{q}_1 = (x - \frac{1}{2})\}$.

(a) Show that $\mathcal{U}$ is an orthogonal set.

(b) Find $\boldsymbol{r}_{\parallel}$, the best approximation with respect to U of the polynomial $\boldsymbol{r}(x) = 2x + 3x^2$.

(c) Verify whether $(\boldsymbol{r} - \boldsymbol{r}_{\parallel}) \in U^{\perp}$ or not.

7.1.4.- Consider the space $C^{\infty}([-\ell, \ell], \mathbb{R})$ with inner product

$$\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \int_{-\ell}^{\ell} \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx,$$

and the orthonormal set $\mathcal{U}$ given by

$$\boldsymbol{u}_0 = \frac{1}{\sqrt{2\ell}}$$
$$\boldsymbol{u}_1 = \frac{1}{\sqrt{\ell}}\cos\Big(\frac{\pi x}{\ell}\Big)$$
$$\boldsymbol{v}_1 = \frac{1}{\sqrt{\ell}}\sin\Big(\frac{\pi x}{\ell}\Big).$$

Find the best approximation of

$$\boldsymbol{f}(x) = \begin{cases} x & 0 \leqslant x \leqslant \ell, \\ -x & -\ell \leqslant x < 0. \end{cases}$$

in the space $\mathrm{Span}(\mathcal{U})$

7.1.5.- For the matrix $\mathsf{A} \in \mathbb{R}^{3,3}$ below, verify that $N(\mathsf{A}) = R(\mathsf{A}^T)^{\perp}$ and that $N(\mathsf{A}^T) = R(\mathsf{A})^{\perp}$, where

$$\mathsf{A} = \begin{bmatrix}2 & 1 & 1\\-1 & -1 & 0\\-2 & -1 & -1\end{bmatrix}.$$

7.2. Least squares

7.2.1. **The normal equation.** We describe the least squares method to find approximate solutions to inconsistent linear systems. The method is often used to find the best parameters that fit experimental data. The parameters are the unknowns of the linear system, and the experimental data determines the matrix of coefficients and the source vector of the system. Such a linear system usually contains more equations than unknowns, and it is inconsistent, since there are no parameters that fit all the data exactly. We start introducing the notion of least squares solution of a possibly inconsistent linear system.

Definition 7.2.1. *Given a matrix* $\mathsf{A} \in \mathbb{F}^{m,n}$ *and a vector* $\mathsf{b} \in \big(\mathbb{F}^m, \cdot\big)$*, the vector* $\hat{\mathsf{x}} \in \mathbb{F}^n$ *is called a* ***least squares solution*** *of the linear system* $\mathsf{Ax} = \mathsf{b}$ *iff holds*

$$\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\| \leqslant \|\mathsf{y} - \mathsf{b}\| \qquad \forall\ \mathsf{y} \in R(\mathsf{A}).$$

The problem we study is to find the least squares solution to an $m \times n$ linear system $\mathsf{Ax} = \mathsf{b}$. In the case that $\mathsf{b} \in R(\mathsf{A})$ the linear system $\mathsf{Ax} = \mathsf{b}$ is consistent and the least squares solution $\hat{\mathsf{x}}$ is the actual solution of the system, hence $\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\| = 0$. In the case that b does not belong to $R(\mathsf{A})$, the linear system $\mathsf{Ax} = \mathsf{b}$ is inconsistent. In such a case the least squares solution $\hat{\mathsf{x}}$ is the vector in $\mathbb{R}^n$ with the property that $\mathsf{A}\hat{\mathsf{x}}$ is a vector in $R(\mathsf{A})$ closest to b in the inner product space $\big(\mathbb{R}^m, \cdot\big)$. A sketch of this situation for a matrix $\mathsf{A} \in \mathbb{R}^{3,2}$ is given in Fig. 46.

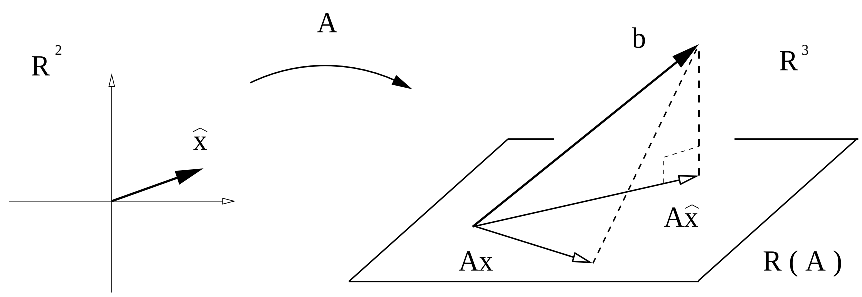

Figure 46. The meaning of the least squares solution $\hat{\mathsf{x}} \in \mathbb{R}^2$ for the 3×2 inconsistent linear system $\mathsf{Ax} = \mathsf{b}$ is that the vector $\mathsf{A}\hat{\mathsf{x}}$ is the closest to b in the inner product space $\big(\mathbb{R}^3, \cdot\big)$.

The solution to the problem of finding a least squares solution to a linear system is summarized in the following result.

Theorem 7.2.2. *Given a matrix* $\mathsf{A} \in \mathbb{F}^{m,n}$ *and a vector* b *in the inner product space* $\big(\mathbb{F}^m, \cdot\big)$*, the vector* $\hat{\mathsf{x}} \in \mathbb{F}^n$ *is a least squares solution of the* $m \times n$ *linear system* $\mathsf{Ax} = \mathsf{b}$ *iff* $\hat{\mathsf{x}}$ *is solution to the* $n \times n$ *linear system, called* ***normal equation****,*

$$\mathsf{A}^*\mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^*\mathsf{b}. \tag{7.3}$$

Furthermore, the least squares solution $\hat{\mathsf{x}}$ *is unique iff the column vectors of matrix* A *form a linearly independent set.*

Remark: In the case that $\mathbb{F} = \mathbb{R}$, the normal equation reduces to $\mathsf{A}^T\mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^T\mathsf{b}$.

Proof of Theorem 7.2.2: We are interested in finding a vector $\hat{\mathsf{x}} \in \mathbb{F}^n$ such that $\mathsf{A}\hat{\mathsf{x}}$ is the best approximation in $R(\mathsf{A})$ of vector $\mathsf{b} \in \mathbb{F}^m$. That is, we want to find $\hat{\mathsf{x}} \in \mathbb{F}^n$ such that

$$\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\| \leqslant \|\mathsf{y} - \mathsf{b}\| \qquad \forall\ \mathsf{y} \in R(\mathsf{A}).$$

Theorem 7.1.2 says that the best approximation of b is when $\mathsf{A}\hat{\mathsf{x}} = \mathsf{b}_{\parallel}$, where $\mathsf{b}_{\parallel}$ is the orthogonal projection of b onto the subspace $R(\mathsf{A})$. This means that

$$(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) \in R(\mathsf{A})^{\perp} = N(\mathsf{A}^*) \quad \Leftrightarrow \quad \mathsf{A}^*(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) = 0.$$

We then conclude that $\hat{\mathsf{x}}$ must be solution of the normal equation

$$\mathsf{A}^*\mathsf{A}\hat{\mathsf{x}} = \mathsf{A}^*\mathsf{b}.$$

The furthermore can be shown as follows. The column vectors of matrix A form a linearly independent set iff $N(\mathsf{A}) = \{\mathsf{0}\}$. Lemma 7.2.3 stated below establishes that, for all matrix A holds that $N(\mathsf{A}) = N(\mathsf{A}^*\mathsf{A})$. This result in our case implies that $N(\mathsf{A}^*\mathsf{A}) = \{\mathsf{0}\}$. Since matrix $\mathsf{A}^*\mathsf{A}$ is a square, $n \times n$, matrix, we conclude that it is invertible. This is equivalent to say that the solution $\hat{\mathsf{x}}$ to the normal equation is unique; moreover, it is given by $\hat{\mathsf{x}} = (\mathsf{A}^*\mathsf{A})^{-1}\mathsf{A}^*\mathsf{b}$. This establishes the Theorem. □

In the proof of Theorem 7.2.2 above we used the following result:

Lemma 7.2.3. *If* $\mathsf{A} \in \mathbb{F}^{m,n}$, *then* $N(\mathsf{A}) = N(\mathsf{A}^*\mathsf{A})$.

Proof of Lemma 7.2.3: We first show that $N(\mathsf{A}) \subset N(\mathsf{A}^*\mathsf{A})$. Indeed,

$$\mathsf{x} \in N(\mathsf{A}) \quad \Rightarrow \quad \mathsf{A}\mathsf{x} = \mathsf{0} \quad \Rightarrow \quad \mathsf{A}^*\mathsf{A}\mathsf{x} = \mathsf{0} \quad \Rightarrow \quad \mathsf{x} \in N(\mathsf{A}^*\mathsf{A}).$$

Now, suppose that there exists $\mathsf{x} \in N(\mathsf{A}^*\mathsf{A})$ such that $\mathsf{x} \notin N(\mathsf{A})$. Therefore, $\mathsf{x} \neq \mathsf{0}$ and $\mathsf{A}\mathsf{x} \neq \mathsf{0}$, which imply that

$$0 \neq \|\mathsf{A}\mathsf{x}\|^2 = \mathsf{x}^*\mathsf{A}^*\mathsf{A}\mathsf{x} \quad \Rightarrow \quad \mathsf{A}^*\mathsf{A}\mathsf{x} \neq \mathsf{0}.$$

However, this last equation contradicts the assumption that $\mathsf{x} \in N(\mathsf{A}^*\mathsf{A})$. Therefore, we conclude that $N(\mathsf{A}) = N(\mathsf{A}^*\mathsf{A})$. This establishes the Lemma. □

EXAMPLE 7.2.1: Show that the 3×2 linear system $\mathsf{A}\mathsf{x} = \mathsf{b}$ is inconsistent; then find a least squares solutions $\hat{\mathsf{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$ to that system, where

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}.$$

SOLUTION: We first show that the linear system above is inconsistent, since Gauss operation on the augmented matrix $[\mathsf{A}|\mathsf{b}]$ imply

$$\left[\begin{array}{cc|c} 1 & 3 & -1 \\ 2 & 2 & 1 \\ 3 & 1 & -1 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 3 & -1 \\ 0 & -4 & 3 \\ 0 & -8 & 2 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 3 & -1 \\ 0 & -4 & 3 \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{array}\right].$$

In order to find the least squares solution to the system above we first construct the normal equation. We need to compute

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 10 \\ 10 & 14 \end{bmatrix}, \qquad \mathsf{A}^T\mathsf{b} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}.$$

Therefore, the normal equation is given by

$$\begin{bmatrix} 14 & 10 \\ 10 & 14 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}.$$

Since the column vectors of A form a linearly independent set, matrix $\mathsf{A}^T\mathsf{A}$ is invertible,

$$(\mathsf{A}^T\mathsf{A})^{-1} = \frac{1}{96} \begin{bmatrix} 14 & -10 \\ -10 & 14 \end{bmatrix} = \frac{1}{48} \begin{bmatrix} 7 & -5 \\ -5 & 7 \end{bmatrix}.$$

The least squares solution is unique and given by

$$\hat{x} = \frac{1}{24}\begin{bmatrix} 7 & -5 \\ -5 & 7 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix} \quad \Rightarrow \quad \hat{x} = \frac{1}{12}\begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

Remark: We now verify that $(A\hat{x} - b) \in R(A)^{\perp}$. Indeed,

$$A\hat{x} - b = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} \frac{1}{12}\begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} = -\frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \quad \Rightarrow \quad A\hat{x} - b = \frac{2}{3}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}.$$

Since

$$\begin{bmatrix} 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 - 4 + 3 = 0, \qquad \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = 3 - 4 + 1 = 0,$$

we have verified that $(A\hat{x} - b) \in R(A)^{\perp}$. ◁

We finish this Subsection with an alternative proof of Theorem 7.2.2 in the particular case that involves real-valued matrices, that is, $\mathbb{F} = \mathbb{R}$. The proof is interesting in its own, since it is based in solving a constrained minimization problem.

Alternative proof of Theorem 7.2.2 for $\mathbb{F} = \mathbb{R}$: The vector $\hat{x} \in \mathbb{R}^n$ is a least squares solution of the system $Ax = b$ iff the function $f : \mathbb{R}^n \to R$ given by $f(x) = \|A\hat{x} - b\|^2$ has a minimum at $x = \hat{x}$. We then find all minima of function f. We first express f as follows,

$$\begin{aligned} f(x) &= (Ax - b) \cdot (Ax - b) \\ &= (Ax) \cdot (Ax) - 2b \cdot (Ax) + b \cdot b \\ &= x^T A^T A x - 2b^T A x + b^T b. \end{aligned}$$

We now need to find all solutions to the equation $\nabla_x f(x) = 0$. Recalling the definition of a gradient vector

$$\nabla_x f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix},$$

it is simple to see that, for any vector $a \in \mathbb{R}^n$, holds,

$$\nabla_x(a^T x) = a, \qquad \nabla_x(x^T a) = a.$$

Therefore, the gradient of f is given by

$$\nabla_x f = 2A^T A x - 2A^T b.$$

We are interested in the stationary points, the $\hat{x}$ solutions of

$$\nabla_x f(\hat{x}) = 0 \quad \Leftrightarrow \quad A^T A\hat{x} = A^T b.$$

We conclude that all stationary points $\hat{x}$ are solutions of the normal equation, Eq. (7.3). These stationary points must be a minimum of f, since f is quadratic on the vector components x_i having the degree two terms all positive coefficients. This establishes the first part of Theorem 7.2.2 in the case that $\mathbb{F} = \mathbb{R}$. □

7.2.2. **Least squares fit.** It is often desirable to construct a mathematical model to describe the results of an experiment. This may involve fitting an algebraic curve to the given experimental data. The least squares method can be used to find the best parameters that fit the data.

EXAMPLE 7.2.2: (**Linear fit**) The simplest situation is the case where the best curve fitting the data is a straight line. More precisely, suppose that the result of an experiment is the following collection of ordered numbers

$$\big\{(t_1, b_1), \cdots, (t_m, b_m)\big\}, \qquad m \geqslant 2,$$

and suppose that a plot on a plane of the result of this experiment is given in Fig. 47. (For example, from measuring the vertical displacement b_i in a spring when a weight t_i is attached to it.) Find the best line $y(t) = \hat{x}_2\, t + \hat{x}_1$ that approximate these points in least squares sense. The latter means to find the numbers $\hat{x}_2$, $\hat{x}_1 \in \mathbb{R}$ such that $\sum_{i=1}^m |\Delta b_i|^2$ is the smallest possible, where

$$\Delta b_i = b_i - y(t_i) \quad \Leftrightarrow \quad \Delta b_i = b_i - (\hat{x}_2\, t_i + \hat{x}_1), \qquad i = 1, \cdots, m.$$

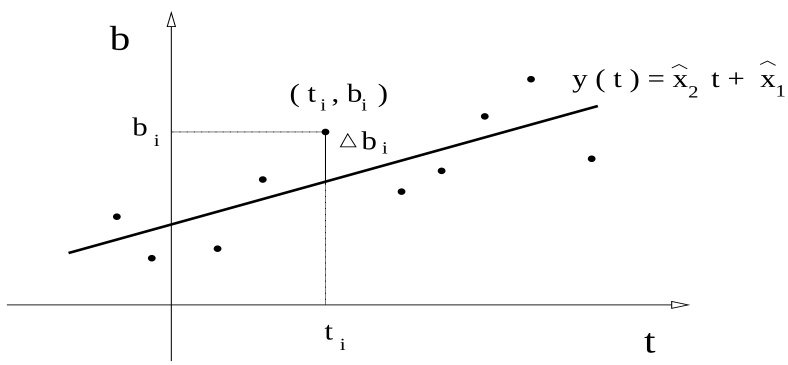

FIGURE 47. Sketch of the best line $y(t) = \hat{x}_2 t + \hat{x}_1$ fitting the set of points (t_i, b_i), for $i = 1, \cdots, 10$.

SOLUTION: Let us rewrite this problem as the least squares solution of an $m \times 2$ linear system, which in general is inconsistent. We are interested to find $\hat{x}_2$, $\hat{x}_1$ solution of the linear system

$$\begin{matrix} y(t_1) = b_1 \\ \vdots \\ y(t_m) = b_m \end{matrix} \quad \Leftrightarrow \quad \begin{matrix} \hat{x}_1 + t_1 \hat{x}_2 = b_1 \\ \vdots \\ \hat{x}_1 + t_m \hat{x}_2 = b_m \end{matrix} \quad \Leftrightarrow \quad \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Introducing the notation

$$\mathsf{A} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \qquad \hat{\mathsf{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix},$$

we are then interested in finding the solution $\hat{\mathsf{x}}$ of the $m \times 2$ linear system $\mathsf{A}\hat{\mathsf{x}} = \mathsf{b}$. Introducing also the vector

$$\Delta\mathsf{b} = \begin{bmatrix} \Delta b_1 \\ \vdots \\ \Delta b_m \end{bmatrix},$$

it is clear that $\mathsf{A}\hat{\mathsf{x}} - \mathsf{b} = \Delta\mathsf{b}$, and so we obtain the important relation

$$\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\|^2 = \|\Delta\mathsf{b}\|^2 = \sum_{i=1}^{m} (\Delta b_i)^2.$$

Therefore, the vector $\hat{\mathsf{x}}$ that minimizes the square of the deviation from the line, $\sum_{i=1}^m (\Delta b_i)^2$, is precisely the same vector $\hat{\mathsf{x}} \in \mathbb{R}^2$ that minimizes the number $\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\|^2$. We studied the latter problem at the begining of this Section. We called it a least squares problem, and the solution $\hat{\mathsf{x}}$ is the solution of the normal equation

$$\mathsf{A}^T \mathsf{A}\, \hat{\mathsf{x}} = \mathsf{A}^T \mathsf{b}.$$

It is simple to see that

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix},$$

$$\mathsf{A}^T\mathsf{b} = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

Therefore, we are interested in finding the solution to the 2×2 linear system

$$\begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

Suppose that at least one of the t_i is different from 1, then matrix $\mathsf{A}^T\mathsf{A}$ is invertible and the inverse is

$$\big(\mathsf{A}^T\mathsf{A}\big)^{-1} = \frac{1}{m \sum t_i^2 - \big(\sum t_i\big)^2} \begin{bmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & m \end{bmatrix}.$$

We conclude that the solution to the normal equation is

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \frac{1}{m \sum t_i^2 - \big(\sum t_i\big)^2} \begin{bmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & m \end{bmatrix} \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

So, the slope $\hat{x}_2$ and vertical intercept $\hat{x}_1$ of the best fitting line are given by

$$\hat{x}_2 = \frac{m \sum t_i b_i - (\sum t_i)(\sum b_i)}{m \sum t_i^2 - \big(\sum t_i\big)^2}, \qquad \hat{x}_1 = \frac{(\sum t_i^2)(\sum b_i) - (\sum t_i)(\sum t_i b_i)}{m \sum t_i^2 - \big(\sum t_i\big)^2}.$$

◁

Example 7.2.3: (**Polynomial fit**) Find the best polynomial of degree $(n-1) \geqslant 0$, say $p(t) = \hat{x}_n\, t^{(n-1)} + \cdots + \hat{x}_1$, that approximates in least squares sense the set of points

$$\big\{(t_1, b_1), \cdots, (t_m, b_m)\big\}, \qquad m \geqslant n.$$

(See Fig. 48 for an example in the case that the fitting curve is a parabola, $n = 3$.) Following Example 7.2.2, the least squares approximation means to find the numbers $\hat{x}_n, \cdots, \hat{x}_1 \in \mathbb{R}$ such that $\sum_{i=1}^m |\Delta b_i|^2$ is the smallest possible, where

$$\Delta b_i = b_i - p(t_i) \quad \Leftrightarrow \quad \Delta b_i = b_i - (\hat{x}_n\, t_i^{(n-1)} + \cdots + \hat{x}_1), \qquad i = 1, \cdots, m.$$

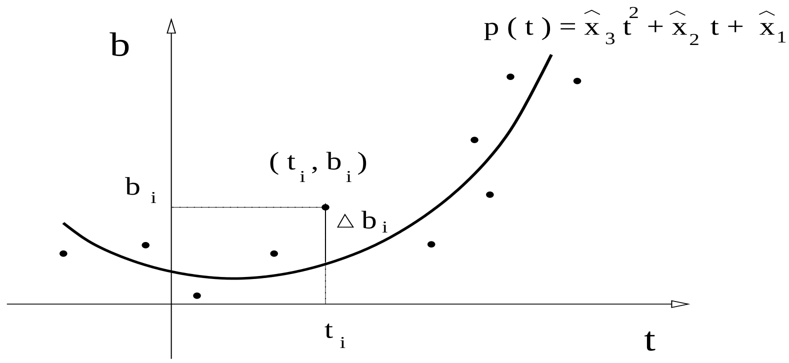

FIGURE 48. Sketch of the best parabola $p(t) = \hat{x}_3 t^2 + \hat{x}_2 t + \hat{x}_1$ fitting the set of points (t_i, b_i), for $i = 1, \cdots, 10$.

SOLUTION: We rewrite this problem as the least squares solution of an $m \times n$ linear system, which in general is inconsistent. We are interested to find $\hat{x}_n, \cdots, \hat{x}_1$ solution of the linear system

$$\begin{matrix} p(t_1) = b_1 \\ \vdots \\ p(t_m) = b_m \end{matrix} \quad \Leftrightarrow \quad \begin{matrix} \hat{x}_1 + \cdots + t_1^{(n-1)} \hat{x}_n = b_1 \\ \vdots \\ \hat{x}_1 + \cdots + t_m^{(n-1)} \hat{x}_n = b_m \end{matrix} \quad \Leftrightarrow \quad \begin{bmatrix} 1 & \cdots & t_1^{(n-1)} \\ \vdots & & \vdots \\ 1 & \cdots & t_m^{(n-1)} \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Introducing the notation

$$\mathsf{A} = \begin{bmatrix} 1 & \cdots & t_1^{(n-1)} \\ \vdots & & \vdots \\ 1 & \cdots & t_m^{(n-1)} \end{bmatrix}, \qquad \hat{\mathsf{x}} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix},$$

we are then interested in finding the solution $\hat{\mathsf{x}}$ of the $m \times n$ linear system $\mathsf{A}\hat{\mathsf{x}} = \mathsf{b}$. Introducing also the vector

$$\Delta\mathsf{b} = \begin{bmatrix} \Delta b_1 \\ \vdots \\ \Delta b_m \end{bmatrix},$$

it is clear that $\mathsf{A}\hat{\mathsf{x}} - \mathsf{b} = \Delta\mathsf{b}$, and so we obtain the important relation

$$\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\|^2 = \|\Delta\mathsf{b}\|^2 = \sum_{i=1}^{m} \big(\Delta b_i\big)^2.$$

Therefore, the vector $\hat{\mathsf{x}}$ that minimizes the square of the deviation from the line, $\sum_{i=1}^{m} (\Delta b_i)^2$, is precisely the same vector $\hat{\mathsf{x}} \in \mathbb{R}^2$ that minimizes the number $\|\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}\|^2$. We studied the latter problem at the begining of this Section. We called it a least squares problem, and the solution $\hat{\mathsf{x}}$ is the solution of the normal equation

$$\mathsf{A}^T \mathsf{A}\, \hat{\mathsf{x}} = \mathsf{A}^T \mathsf{b}. \tag{7.4}$$

It is simple to see that Eq. (7.4) is an $n \times n$ linear system, since

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ t_1^{(n-1)} & \cdots & t_m^{(n-1)} \end{bmatrix} \begin{bmatrix} 1 & \cdots & t_1^{(n-1)} \\ \vdots & & \vdots \\ 1 & \cdots & t_m^{(n-1)} \end{bmatrix},$$

$$\mathsf{A}^T\mathsf{b} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ t_1^{(n-1)} & \cdots & t_m^{(n-1)} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

We do not compute these expressions explicitly here. In the case that the columns of A form a linearly independent set, the solution $\hat{\mathsf{x}}$ to the normal equation is

$$\hat{\mathsf{x}} = \left(\mathsf{A}^T\mathsf{A}\right)^{-1}\mathsf{A}^T\mathsf{b}.$$

The components of $\hat{\mathsf{x}}$ provide the parameters for the best polynomial fitting the data in least squares sense. $\triangleleft$

7.2.3. **Linear correlation.** In statistics a correlation coefficient measures the departure of two random variables from independence. For centered data, that is, for data with zero average, the correlation coefficient can be viewed as the cosine of the angle in an abstract $\mathbb{R}^n$ space between two vectors constructed with the random variables data. We now define and find the correlation coefficient for two variables as given in Example 7.2.2.

Once again, suppose that the result of an experiment is the following collection of ordered numbers

$$\big\{(t_1, b_1), \cdots, (t_m, b_m)\big\}, \qquad m \geqslant 2. \tag{7.5}$$

Introduce the vectors e, t, $\mathsf{b} \in \mathbb{R}^m$ as follows,

$$\mathsf{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad \mathsf{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}, \qquad \mathsf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Before introducing the correlation coefficient, let us use these vectors above to write down the least squares coefficients x found in Example 7.2.2. The matrix of coefficients can be written as $\mathsf{A} = [\mathsf{e}, \mathsf{t}]$, therefore,

$$\mathsf{A}^T\mathsf{A} = \begin{bmatrix} \mathsf{e}^T \\ \mathsf{t}^T \end{bmatrix} \big[\mathsf{e}, \mathsf{t}\big] = \begin{bmatrix} \mathsf{e}\cdot\mathsf{e} & \mathsf{t}\cdot\mathsf{e} \\ \mathsf{e}\cdot\mathsf{t} & \mathsf{t}\cdot\mathsf{t} \end{bmatrix}, \qquad \mathsf{A}^T\mathsf{b} = \begin{bmatrix} \mathsf{e}^T \\ \mathsf{t}^T \end{bmatrix} \mathsf{b} = \begin{bmatrix} \mathsf{e}\cdot\mathsf{b} \\ \mathsf{t}\cdot\mathsf{b} \end{bmatrix}.$$

The least squares solution can be written as follows,

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \frac{1}{(\mathsf{e}\cdot\mathsf{e})(\mathsf{t}\cdot\mathsf{t}) - (\mathsf{t}\cdot\mathsf{e})^2} \begin{bmatrix} \mathsf{t}\cdot\mathsf{t} & -\mathsf{e}\cdot\mathsf{t} \\ -\mathsf{e}\cdot\mathsf{t} & \mathsf{e}\cdot\mathsf{e} \end{bmatrix} \begin{bmatrix} \mathsf{e}\cdot\mathsf{b} \\ \mathsf{t}\cdot\mathsf{b} \end{bmatrix},$$

that is,

$$\hat{\mathsf{x}}_2 = \frac{(\mathsf{e}\cdot\mathsf{e})(\mathsf{t}\cdot\mathsf{b}) - (\mathsf{e}\cdot\mathsf{t})(\mathsf{e}\cdot\mathsf{b})}{(\mathsf{e}\cdot\mathsf{e})(\mathsf{t}\cdot\mathsf{t}) - (\mathsf{t}\cdot\mathsf{e})^2}, \qquad \hat{x}_1 = \frac{(\mathsf{e}\cdot\mathsf{b})(\mathsf{t}\cdot\mathsf{t}) - (\mathsf{e}\cdot\mathsf{t})(\mathsf{t}\cdot\mathsf{b})}{(\mathsf{e}\cdot\mathsf{e})(\mathsf{t}\cdot\mathsf{t}) - (\mathsf{t}\cdot\mathsf{e})^2}.$$

Introduce the average values

$$\bar{t} = \frac{\mathsf{e}\cdot\mathsf{t}}{\mathsf{e}\cdot\mathsf{e}}, \qquad \bar{b} = \frac{\mathsf{e}\cdot\mathsf{b}}{\mathsf{e}\cdot\mathsf{e}}.$$

These are indeed the average values of t_i and b_i, since

$$\mathsf{e}\cdot\mathsf{e} = m, \qquad \mathsf{e}\cdot\mathsf{t} = \sum_{i=1}^m t_i, \qquad \mathsf{e}\cdot\mathsf{b} = \sum_{i=1}^m b_i.$$

Introduce the zero-average vectors $\hat{\mathsf{t}} = (\mathsf{t} - \bar{t}\,\mathsf{e})$ and $\hat{\mathsf{b}} = (\mathsf{b} - \bar{b}\,\mathsf{e})$. The *correlation coefficient* of the data given in (7.5) is given by

$$\operatorname{cor}(\mathsf{t}, \mathsf{b}) = \frac{\hat{\mathsf{t}} \cdot \hat{\mathsf{b}}}{\|\hat{\mathsf{t}}\|\,\|\hat{\mathsf{b}}\|}.$$

Therefore, the correlation coefficient between the data vectors t and b is the angle between the zero-average vectors $\hat{\mathsf{t}}$ and $\hat{\mathsf{b}}$ in $\mathbb{R}^m$.

In order to understand what measures this angle, let us consider the case where all the ordered pairs in (7.5) lies on a line, that is, there exists a solution $\hat{\mathsf{x}}$ of the linear system $\mathsf{A}\hat{\mathsf{x}} = \mathsf{b}$ (a solution, not only a least squares solution). In that case we have

$$\hat{x}_1 \mathsf{e} + \hat{x}_2 \mathsf{t} = \mathsf{b} \quad \Rightarrow \quad \hat{x}_1 + \hat{x}_2 \bar{t} = \bar{b},$$

and this implies that

$$\hat{x}_2\,(\mathsf{t} - \bar{t}\,\mathsf{e}) = (\mathsf{b} - \bar{b}\,\mathsf{e}) \quad \Leftrightarrow \quad \hat{x}_2\,\hat{\mathsf{t}} = \hat{\mathsf{b}} \quad \Leftrightarrow \quad \operatorname{cor}(\mathsf{t}, \mathsf{b}) = 1.$$

That is, in the case that t is linearly related to b we obtain that the zero-average vectors $\hat{\mathsf{t}}$ and $\hat{\mathsf{b}}$ are parallel, so the correlation coefficient is equal one.

7.2.4. **QR-factorization.** The Gram-Schmidt method can be used to factor any $m \times n$ matrix A into a product of an $m \times n$ matrix Q with orthonormal column vectors and an upper triangular $n \times n$ matrix R. We will see that the QR-factorization is useful to solve the normal equation associated to a least squares problem.

Theorem 7.2.4. *If the column vectors of matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ form a linearly independent set, then there exist matrices $\mathsf{Q} \in \mathbb{F}^{m,n}$ and $\mathsf{R} \in \mathbb{F}^{n,n}$ satisfying that $\mathsf{Q}^*\mathsf{Q} = \mathsf{I}_n$, matrix R is upper triangular with positive diagonal elements, and the following equation holds*

$$\mathsf{A} = \mathsf{QR}.$$

Proof of Theorem 7.2.4: Use the Gram-Schmidt method to obtain an orthonormal set $\{\mathsf{q}_1, \cdots, \mathsf{q}_n\}$ from the column vectors of the $m \times n$ matrix $\mathsf{A} = [\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}]$, that is,

$$\begin{aligned}
\mathsf{p}_1 &= \mathsf{A}_{:1} & \mathsf{q}_1 &= \frac{\mathsf{p}_1}{\|\mathsf{p}_1\|},\\
\mathsf{p}_2 &= \mathsf{A}_{:2} - \big(\mathsf{A}_{:2} \cdot \mathsf{q}_1\big)\,\mathsf{q}_1 & \mathsf{q}_2 &= \frac{\mathsf{p}_2}{\|\mathsf{p}_2\|},\\
&\;\vdots & &\;\vdots\\
\mathsf{p}_n &= \mathsf{A}_{:n} - \big(\mathsf{A}_{:n} \cdot \mathsf{q}_1\big)\,\mathsf{q}_1 - \cdots - \big(\mathsf{A}_{:n} \cdot \mathsf{q}_{n-1}\big)\,\mathsf{q}_{n-1} & \mathsf{q}_n &= \frac{\mathsf{p}_n}{\|\mathsf{p}_n\|}.
\end{aligned}$$

Define matrix $\mathsf{Q} = [\mathsf{q}_1, \cdots, \mathsf{q}_n]$, which then satisfies the equation $\mathsf{Q}^*\mathsf{Q} = \mathsf{I}_n$. Notice that the equations above can be expressed as follows,

$$\begin{aligned}
\mathsf{A}_{:1} &= \|\mathsf{p}_1\|\,\mathsf{q}_1,\\
\mathsf{A}_{:2} &= \|\mathsf{p}_2\|\,\mathsf{q}_2 + \big(\mathsf{q}_1 \cdot \mathsf{A}_{:2}\big)\,\mathsf{q}_1\\
&\;\vdots\\
\mathsf{A}_{:n} &= \|\mathsf{p}_n\|\,\mathsf{q}_n + \big(\mathsf{q}_1 \cdot \mathsf{A}_{:n}\big)\,\mathsf{q}_1 + \big(\mathsf{q}_2 \cdot \mathsf{A}_{:n}\big)\,\mathsf{q}_2 + \cdots + \big(\mathsf{q}_{n-1} \cdot \mathsf{A}_{:n}\big)\,\mathsf{q}_{n-1}.
\end{aligned}$$

After some time staring at the equations above, one can rewrite it as a matrix product

$$\big[\mathsf{A}_{:1}, \cdots, \mathsf{A}_{:n}\big] = \big[\mathsf{q}_1, \cdots, \mathsf{q}_n\big] \begin{bmatrix} \|\mathsf{p}_1\| & (\mathsf{q}_1 \cdot \mathsf{A}_{:2}) & \cdots & (\mathsf{q}_1 \cdot \mathsf{A}_{:n}) \\ 0 & \|\mathsf{p}_2\| & \cdots & (\mathsf{q}_2 \cdot \mathsf{A}_{:n}) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & (\mathsf{q}_{n-1} \cdot \mathsf{A}_{:n}) \\ 0 & 0 & \cdots & \|\mathsf{p}_n\| \end{bmatrix} \tag{7.6}$$

Define matrix R by equation above as the matrix satisfying $\mathsf{A} = \mathsf{QR}$. Then, Eq. (7.6) says that matrix R is $n \times n$, upper triangular, with positive diagonal elements. This establishes the Theorem. □

Example 7.2.4: Find the QR-factorization of matrix $\mathsf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

Solution: First use the Gram-Schmidt method to transform the column vectors of matrix A into an orthonormal set. This was done in Example 6.5.1. The result defines the matrix Q as follows

$$\mathsf{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Having matrix A and Q, and knowing that Theorem 7.2.4 is true, then we can compute matrix R by the equation $\mathsf{R} = \mathsf{Q}^T\mathsf{A}$. Since the column vectors of Q form an orthonormal set, we have that $\mathsf{Q}^T = \mathsf{Q}^{-1}$, and in this particular case $\mathsf{Q}^{-1} = \mathsf{Q}$, so matrix R is given by

$$\mathsf{R} = \mathsf{QA} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \Rightarrow \quad \mathsf{R} = \begin{bmatrix} \sqrt{2} & \frac{3}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The QR-factorization of matrix A is then given by

$$\mathsf{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{3}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

◁

The QR-factorization is useful to solve the normal equation in a least squares problem.

Theorem 7.2.5. *Assume that the matrix $\mathsf{A} \in \mathbb{F}^{m,n}$ admits the QR-factorization $\mathsf{A} = \mathsf{QR}$. The vector $\hat{\mathsf{x}} \in \mathbb{F}^n$ is solution of the normal equation $\mathsf{A}^*\mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^*\mathsf{b}$ iff it is solution of*

$$\mathsf{R}\hat{\mathsf{x}} = \mathsf{Q}^*\mathsf{b}.$$

Proof of Theorem 7.2.5: Just introduce the QR-factorization into the normal equation $\mathsf{A}^*\mathsf{A}\,\hat{\mathsf{x}} = \mathsf{A}^*\mathsf{b}$ as follows,

$$\big(\mathsf{R}^*\mathsf{Q}^*\big)\big(\mathsf{QR}\big)\,\hat{\mathsf{x}} = \mathsf{R}^*\mathsf{Q}^*\mathsf{b} \quad \Leftrightarrow \quad \mathsf{R}^*\mathsf{R}\,\hat{\mathsf{x}} = \mathsf{R}^*\mathsf{Q}^*\mathsf{b} \quad \Leftrightarrow \quad \mathsf{R}^*\big(\mathsf{R}\,\hat{\mathsf{x}} - \mathsf{Q}^*\mathsf{b}\big) = 0.$$

Since R is a square, upper triangular matrix with non-zero coefficients, we conclude that R is invertible. Therefore, from the last equation above we conclude that $\hat{\mathsf{x}}$ is solution of the normal equation iff holds

$$\mathsf{R}\hat{\mathsf{x}} = \mathsf{Q}^*\mathsf{b}.$$

This establishes the Theorem. □

7.2.5. **Exercises.**

7.2.1.- Consider the matrix A and the vector b given by

$$\mathsf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the linear system $\mathsf{A}\mathsf{x} = \mathsf{b}$.

(b) Verify that the solution $\hat{\mathsf{x}}$ satisfies

$$(\mathsf{A}\hat{\mathsf{x}} - \mathsf{b}) \in R(\mathsf{A})^{\perp}.$$

7.2.2.- Consider the matrix A and the vector b given by

$$\mathsf{A} = \begin{bmatrix} 2 & 2 \\ 0 & -1 \\ -2 & 0 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

(a) Find the least-squares solution $\hat{\mathsf{x}}$ to the linear system $\mathsf{A}\mathsf{x} = \mathsf{b}$.

(b) Find the orthogonal projection of the source vector b onto the subspace $R(\mathsf{A})$.

7.2.3.- Find all the least-squares solutions $\hat{\mathsf{x}}$ to the linear system $\mathsf{A}\mathsf{x} = \mathsf{b}$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}, \quad \mathsf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

7.2.4.- Find the best line in least-squares sense that fits the measurements, where t_1 is the independent variable and b_i is the dependent variable,

$$t_1 = -2, \quad b_1 = 4,$$
$$t_2 = -1, \quad b_2 = 3,$$
$$t_3 = 0, \quad b_3 = 1,$$
$$t_4 = 2, \quad b_4 = 0.$$

7.2.5.- Find the correlation coefficient corresponding to the measurements given in Exercise **7.2.4** above.

7.2.6.- Use Gram-Schmidt method on the columns of matrix A below to find its QR factorization, where

$$\mathsf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 2 & 1 \end{bmatrix}.$$

7.2.7.- Find the QR factorization of matrix

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

7.3. Finite difference method

A differential equation is an equation where the unknown is a function and both the function itself and its derivatives appear in the equation. The differential equation is called linear iff the unknown function and its derivatives appear linearly in the equation. Solutions of linear differential equations can be approximated by solutions of appropriate $n \times n$ algebraic linear systems in the limit that n approaches infinity. The finite difference method is a way to obtain an $n \times n$ linear system from the original differential equation. *Derivatives are approximated by difference quotients, thus reducing a differential equation to an algebraic linear system.* Since a derivative can be approximated by infinitely many difference quotients, there are infinitely many $n \times n$ linear systems that approximate a differential equation. One tries to choose the linear system whose solution is the best approximation of the solution of the original differential equation. Computers are used to find the vector in $\mathbb{R}^n$ solution of the $n \times n$ linear system. Many approximations of the solution to the differential equation are obtained from this array of n numbers. One way to obtain a function from a vector in $\mathbb{R}^n$ is to find a degree n polynomial that contains all these n points. This is called a polynomial interpolation of the algebraic solution. In this Section we only show how to obtain $n \times n$ algebraic linear systems that approximate a simple differential equation.

7.3.1. **Differential equations.**

A differential equation is an equation where the unknown is a function and both the function and its derivatives appear in the equation. A simple example is the following: Given a continuously differentiable function $f : [0,1] \to \mathbb{R}$, find a function $u : [0,1] \to \mathbb{R}$ solution of the differential equation

$$\frac{du}{dx}(x) = f(x),$$

To find a solution to a differential equation requires to perform appropriate integrations, thus integration constants are introduced in the solution. This suggests that the solution of a differential equation is not unique, and extra conditions must be added to the problem to select only one solution. In the differential equation above the solutions u are given by

$$u(x) = \int_0^x f(t)\,dt + c,$$

with $c \in \mathbb{R}$. An extra condition is needed to obtain a unique solution, for example the condition $u(0) = 1$. Then, the unique solution u is computed as follows

$$1 = u(0) = \int_0^0 f(t)\,dt + c \quad \Rightarrow \quad c = 1 \quad \Rightarrow \quad u(x) = \int_0^x f(t)\,dt + 1.$$

The example above is simple enough that no approximation is needed to obtain the solution.

An *ordinary differential equation* is a differential equation where the unknown function u depends on one variable, as in the example above. A *partial differential equation* is a differential equation where the unknown function depends on more than one variable and the equation contains derivatives of more than one variable. In this Section we use the finite difference method to find a solution to two different problems. The first one involves an ordinary differential equation while the second one involves a partial differential equation, called the heat equation.

The first problem is to find an approximate solution to a boundary value problem for an ordinary differential equation: Given a continuously differentiable function $f : [0,1] \to \mathbb{R}$, find a function $u : [0,1] \to \mathbb{R}$ solution of the boundary value problem

$$\frac{d^2u}{dx^2}(x) + \frac{du}{dx}(x) = f(x), \tag{7.7}$$

$$u(0) = u(1) = 0. \tag{7.8}$$

Finding the function u involves doing two integrations that introduce two integration constants. These constants are determined by the two conditions on $x = 0$ and $x = 1$ above, called boundary conditions, since they are conditions on the boundaries of the interval $[0,1]$. Because of this extra condition on the ordinary differential equation this problem is called a boundary value problem.

The second problem we study in this Section is to find an approximate solution to an initial-boundary value problem for the heat equation: Given the set $D = [0,\pi] \times [0,T]$, the positive constant κ, and infinitely differentiable functions $f : D \to \mathbb{R}$ and $g : [0,\pi] \to \mathbb{R}$, find the function $u : D \to \mathbb{R}$ solution of the problem

$$\frac{\partial u}{\partial t}(x,t) - \kappa \frac{\partial^2 u}{\partial x^2}(x,t) = f(x,t) \qquad (x,t) \in D, \tag{7.9}$$

$$u(0,t) = u(\pi,t) = 0 \qquad t \in [0,T], \tag{7.10}$$

$$u(x,0) = g(x) \qquad x \in [0,\pi]. \tag{7.11}$$

The partial differential equation in Eq. (7.9) is called the one-dimensional heat equation, since in the case that function u is the temperature of a material that depends on time t and one spatial direction x, the equation describes how the material temperature changes in time due to heat propagation. The positive constant κ is called the thermal diffusivity of the material. The condition given in Eq. (7.10) is called a boundary condition, since they are conditions on $x = 0$ and $x = \pi$ that hold for all $t \in [0,T]$, see Fig. 49. The condition given in Eq. (7.11) is called an initial condition, since it is a condition on the initial time $t = 0$ for all $x \in [0,\pi]$, see Fig. 49. Because of these two extra conditions on the partial differential equation this problem is called an initial-boundary value problem for the heat equation.

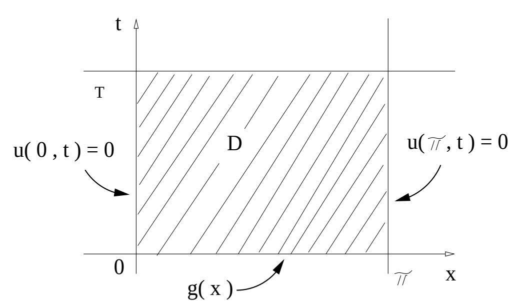

FIGURE 49. The domain $D = [0,\pi] \times [0,T]$ where the initial-boundary value problem for the heat equation is set up. We indicate the boundary data conditions $u(0,t) = 0$ and $u(\pi,t) = 0$, and the initial data function g.

7.3.2. **Difference quotients.** Finite difference methods transform a linear differential equation into an $n \times n$ algebraic linear system by replacing derivatives by difference quotients.

Derivatives can be approximated by difference quotients in many different ways. For example, a derivative of a function u can be expressed in the following equivalent ways,

$$\begin{aligned}\frac{du}{dx}(x) &= \lim_{\Delta x\to 0}\frac{u(x+\Delta x)-u(x)}{\Delta x},\\ &= \lim_{\Delta x\to 0}\frac{u(x)-u(x-\Delta x)}{\Delta x},\\ &= \lim_{\Delta x\to 0}\frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x}.\end{aligned}$$

However, for a fixed nonzero value of Δx, the expressions below are, in general, different.

Definition 7.3.1. *The* ***forward difference quotient*** d_+ *and the* ***backward difference quotient*** d_- *of a continuous function* $u:\mathbb{R}\to\mathbb{R}$ *at* $x\in\mathbb{R}$ *are given by*

$$d_+u(x) = \frac{u(x+\Delta x)-u(x)}{\Delta x}, \qquad d_-u(x) = \frac{u(x)-u(x-\Delta x)}{\Delta x}.$$

The ***centered difference quotient*** d_c *of a continuous function* u *at* $x\in\mathbb{R}$ *is given by*

$$d_cu(x) = \frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x}. \tag{7.12}$$

In the case that the function u has second continuous derivative, the Taylor Expansion Theorem implies that the forward and backward differences differ from the actual derivative by terms order Δx. The proof is not difficult, since

$$\begin{aligned}u(x+\Delta x) = u(x) + \frac{du}{dx}(x)\,\Delta x + O\big((\Delta x)^2\big) \quad &\Rightarrow \quad d_+u(x) = \frac{du}{dx}(x) + O\big(\Delta x\big),\\ u(x-\Delta x) = u(x) - \frac{du}{dx}(x)\,\Delta x + O\big((\Delta x)^2\big) \quad &\Rightarrow \quad d_-u(x) = \frac{du}{dx}(x) + O\big(\Delta x\big),\end{aligned}$$

where $O\big((\Delta x)^n\big)$ denotes a function satisfying $\big[O\big((\Delta x)^n\big)\big]/(\Delta x)^n$ approaches a constant as $\Delta x\to 0$. In the case that the function u has third continuous derivative, the Taylor Expansion Theorem implies that the centered difference quotient differs from the actual derivative by terms of order $(\Delta x)^2$. Again, the proof is not difficult, and it is based on the Taylor expansion of the function u. Compute this expansion in two different ways,

$$u(x+\Delta x) = u(x) + \frac{du}{dx}(x)\,\Delta x + \frac{1}{2}\frac{d^2u}{dx^2}(x)\,(\Delta x)^2 + O\big((\Delta x)^3\big), \tag{7.13}$$

$$u(x-\Delta x) = u(x) - \frac{du}{dx}(x)\,\Delta x + \frac{1}{2}\frac{d^2u}{dx^2}(x)\,(\Delta x)^2 + O\big((\Delta x)^3\big), \tag{7.14}$$

Subtracting the two expressions above we obtain that

$$u(x+\Delta x) - u(x-\Delta x) = 2\frac{du}{dx}(x)\,\Delta x + O\big((\Delta x)^3\big) \quad \Rightarrow \quad d_cu(x) = \frac{du}{dx}(x) + O\big((\Delta x)^2\big),$$

which establishes that centered difference quotients differ from the derivative by order $(\Delta x)^2$. If a function is infinitely continuously differentiable, centered difference quotients are more accurate than forward or backward differences.

Second and higher derivatives of a function can also be approximated by difference quotients. Again, there are infinitely many ways to approximate second derivatives by difference quotients. The freedom to choose difference quotients is higher for second derivatives than for first derivatives. We now present two difference quotients to give an idea of this freedom. On the one hand, one possible approximation is to use the centered difference quotient twice, since a second derivative is the derivative of the derivative function. Using the more precise notation

$$d_{c\Delta x}u(x) = \frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x},$$

it is not difficult to check that $\big(d_{c\Delta x}\big)^2 = d_{c\Delta x}(d_{c\Delta x})$ is given by

$$\big(d_{c\Delta x}\big)^2 u(x) = \frac{1}{(2\Delta x)^2}\big[u(x+2\Delta x) + u(x-2\Delta x) - 2u(x)\big]. \tag{7.15}$$

On the other hand, another approximation for the second derivative of a function can be obtained directly from the Taylor expansion formulas in (7.13)-(7.14). Indeed, add Eqs. (7.13)-(7.14), that is,

$$u(x+\Delta x) + u(x-\Delta x) = 2u(x) + \frac{d^2u}{dx^2}(x)\,(\Delta x)^2 + O\big((\Delta x)^3\big).$$

This equation can be rewritten as

$$\frac{d^2u}{dx^2}(x) = \frac{u(x+\Delta x) + u(x-\Delta x) - 2u(x)}{(\Delta x)^2} + O\big(\Delta x\big).$$

This equation suggests to introduce a second-order centered difference quotient $(d^2)_{c\Delta x}$ as

$$(d^2)_{c\Delta x} u(x) = \frac{1}{(\Delta x)^2}\big[u(x+\Delta x) + u(x-\Delta x) - 2u(x)\big]. \tag{7.16}$$

Using this notation, the equation above is given by

$$\frac{d^2u}{dx^2}(x) = (d^2)_{c\Delta x} u(x) + O\big(\Delta x\big).$$

Therefore, both $\big(d_{c\Delta x}\big)^2$ and $(d^2)_{c\Delta x}$ are approximations of the second derivative of a function. However, they are not the same approximation, since comparing Eqs. (7.15) and (7.16) it is not difficult to see that

$$\big(d_{c(\Delta x)/2}\big)^2 = (d^2)_{c\Delta x}.$$

We conclude that there are many different ways to approximate second derivatives by difference quotients, many more ways than those to approximate first order derivatives. In this Section we use the difference quotient in Eq. (7.16), and we use the simplified notation given by $d_c^2 = (d^2)_{c\Delta x}$.

7.3.3. **Method of finite differences.** We now describe the finite difference method using two examples. In the first example we find an approximate solution for the boundary value problem in Eqs. (7.7)-(7.8). In the second example we find an approximate solution for the initial-boundary value problem in Eqs. (7.9)-(7.11).

EXAMPLE 7.3.1: Consider the boundary value problem for the ordinary differential equation given in Eqs. (7.7)-(7.8).

(a) Divide the interval $[0,1]$ into $n > 1$ equal intervals and use the finite difference method to find a vector $\mathsf{u} = [u_i] \in \mathbb{R}^{n+1}$ that approximates the function $u : [0,1] \to \mathbb{R}$ solution of that boundary value problem. Use centered difference quotients to approximate the first and second derivatives of the unknown function u.
(b) Find the explicit form of the linear system in the case $n = 6$.
(c) Find the degree n polynomial p_n that interpolates the approximate solution vector $\mathsf{u} = [u_i] \in \mathbb{R}^{n+1}$, which includes the boundary points.

SOLUTION:

Part (a): Fix a positive integer $n \in \mathbb{N}$, define the grid step size $h = 1/n$, and introduce the uniform grid

$$\{x_i\}, \qquad x_i = ih, \qquad i = 0, 1, \cdots, n, \qquad \text{on } [0,1].$$

(See Fig. 50.) Introduce the numbers $f_i = f(x_i)$. Finally, denote $u_i = u(x_i)$. The numbers x_i and f_i are known from the problem, and so are the $u_0 = u(0) = 0$ and $u_n = u(1) = 0$, while the u_i for $i = 1, \cdots, n-1$ are the unknowns. We now use the original differential equation to construct an $(n-1) \times (n-1)$ linear system for these unknowns u_i.

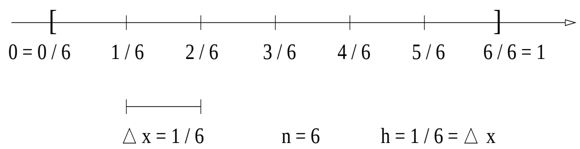

Figure 50. A uniform grid for $n = 6$ on the domain $[0, 1]$.

We use centered difference quotient given in Eqs. (7.12) and (7.16) to approximate the first and second derivatives of the function u, respectively. We choose $\Delta x = h$, and we denote the difference quotients evaluated at grid points x_i as follows,

$$d_c u(x_i) = d_c u_i, \qquad d_c^2 u(x_i) = d_c^2 u_i.$$

Therefore, we obtain the following formulas for the difference quotients,

$$d_c u_i = \frac{u_{i+1} - u_{i-1}}{2h}, \qquad d_c^2 u_i = \frac{u_{i+1} + u_{i-1} - 2u_i}{h^2}. \tag{7.17}$$

Now we state the approximate problem we will solve: Given the constants $\{f_i\}_{i=1}^{n-1}$, find the vector $\mathsf{u} = [u_i] \in \mathbb{R}^{n+1}$ solution of the $(n-1) \times (n-1)$ linear system and boundary conditions, respectively,

$$d_c^2 u_i + d_c u_i = f_i, \quad i = 1, \cdots, n-1, \tag{7.18}$$

$$u_0 = u_n = 0. \tag{7.19}$$

Eq. (7.18) is indeed a linear system for $\mathsf{u} = [u_i]$, since it is equivalent to the system

$$(2+h)u_{i+1} - 4u_i + (2-h)u_{i-1} = 2h^2 f_i, \qquad i = 1, \cdots, n-1.$$

When the boundary conditions in Eq. (7.19) are introduced in the equation above, we obtain an $(n-1) \times (n-1)$ linear system for the unknowns u_i, where $i = 1, \cdots, (n-1)$.

Part (b): In the case $n = 6$, we have $h = 1/6$, so we denote $a = 2 - 1/6$, $b = 2 + 1/6$ and $c = 2/36$. Also recall the boundary conditions $u_0 = u_6 = 0$. Then, the system above and its augmented matrix are given by, respectively,

$$\begin{aligned} -4u_1 + bu_2 &= c\,f_1, \\ au_1 - 4u_2 + bu_3 &= c\,f_2, \\ au_2 - 4u_3 + bu_4 &= c\,f_3, \\ au_3 - 4u_4 + bu_5 &= c\,f_4, \\ au_4 - 4u_5 &= c\,f_5, \end{aligned} \quad \Leftrightarrow \quad \left[\begin{array}{ccccc|c} -4 & b & 0 & 0 & 0 & c\,f_1 \\ a & -4 & b & 0 & 0 & c\,f_2 \\ 0 & a & -4 & b & 0 & c\,f_3 \\ 0 & 0 & a & -4 & b & c\,f_4 \\ 0 & 0 & 0 & a & -4 & c\,f_5 \end{array}\right].$$

We then conclude that the solution u of the boundary value problem in Eq. (7.7) can be approximated by the solution $\mathsf{u} = [u_i] \in \mathbb{R}^7$ of the 5×5 linear system above plus the two boundary conditions. The same type of approximate solution can be found for all $n \in \mathbb{N}$.

Part (c): The output of the finite difference method is a vector $\mathsf{u} = [u_i] \in \mathbb{R}^{(n+1)}$. An approximate solution to the ordinary differential equation in Eqs. (7.7) can be constructed from the vector u in many different ways. One way is polynomial interpolation, that is, to construct a polynomial of degree n whose graph contains all the points (x_i, u_i). Such polynomial is given by

$$p_n(x) = \sum_{i=0}^{n} u_i\, q_i(x), \qquad q_i(x) = \prod_{j\neq i} \frac{(x - x_j)}{(x_i - x_j)}.$$

It can be verified that the degree n polynomials q_i when evaluated at grid points satisfies

$$q_i(x_j) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

Therefore, the polynomial p_n has degree n and satisfies that $p_n(x_i) = u_i$. This polynomial function p_n approximates the solution u of the boundary value problem in Eqs. (7.7)-(7.8). ◁

EXAMPLE 7.3.2: Consider the boundary value problem for the partial differential equation given in Eqs. (7.9)-(7.11). Use the finite difference method to find an approximate solution of the function $u : D \to \mathbb{R}$ solution of that initial-boundary value problem.

(a) Use centered difference quotients to approximate the spatial derivatives and forward difference quotients to approximate time derivatives of the unknown function u.
(b) Repeat the calculations in part (a) now using backward difference quotients for the time derivatives of the unknown function u.

SOLUTION:
Part (a): Introduce a grid in the domain $D = [0, \pi] \times [0, T]$ as follows: Fix the positive integers n_x, $n_t \in \mathbb{N}$, define the step sizes $h_x = \pi/n_x$ and $h_t = T/n_t$, and then introduce the uniform grids

$$\{x_i\}, \qquad x_i = ih_x, \qquad i = 0, 1, \cdots, n_x, \qquad \text{on } [0, \pi]$$
$$\{t_j\}, \qquad t_j = jh_t, \qquad j = 0, 1, \cdots, n_t, \qquad \text{on } [0, T].$$

A point of the form $(x_i, t_j) \in D$ is called a grid point. (See Fig. 51.)

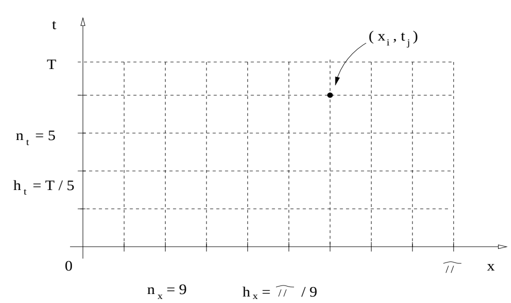

FIGURE 51. A rectangular grid with $n_x = 9$ and $n_t = 5$ on the domain $D = [0, \pi] \times [0, T]$.

We will compute the approximate solution values at grid points, and we use the notation $u_{i,j} = u(x_i, t_j)$ and $f_{i,j} = f(x_i, t_j)$ to denote unknown and source function values at grid points. We also denote $g_i = g(x_i)$ for the initial data function values at grid points. In this notation the boundary conditions in Eq. (7.10) have the form $u_{0,j} = 0$ and $u_{n_x,j} = 0$. We finally introduce the forward difference quotient in time d_{+t} and the second order centered difference quotient in space d^2_{cx} as follows,

$$d_{+t}u_{i,j} = \frac{u_{i,(j+1)} - u_{i,j}}{h_t}, \qquad d^2_{cx}u_{i,j} = \frac{u_{(i+1),j} + u_{(i-1),j} - 2u_{i,j}}{h_x^2}.$$

Now we state the approximate problem we will solve: Given the constants $f_{i,j}$ and g_i, for $i = 0, 1, \cdots, n_x$ and $j = 0, 1, \cdots, n_y$, find the constants $u_{i,j}$ solution of the linear equations and boundary conditions, respectively,

$$d_{+t}u_{i,j} - \kappa d^2_{cx}u_{i,j} = f_{i,j}, \tag{7.20}$$

$$u_{0,j} = u_{n_x,j} = 0, \tag{7.21}$$

$$u_{i,0} = g_i. \tag{7.22}$$

This system is simpler to solve than it looks at first sight. Let us rewrite it in as follows,

$$\frac{1}{h_t}\big(u_{i,(j+1)} - u_{i,j}\big) - \frac{\kappa}{h_x^2}\big(u_{(i+1),j} + u_{(i-1),j} - 2u_{i,j}\big) = f_{ij},$$

which is equivalent to the equation

$$u_{i,(j+1)} = r\big(u_{(i+1),j} + u_{(i-1),j}\big) + (1-2r)\,u_{i,j} + h_t f_{i,j}, \tag{7.23}$$

where $r = \kappa h_t/h^2_{cx}$. This last equation says that the solution at the time t_{j+1} can be computed if the solution at the previous time t_j is known. Since the solution at the initial time $t_0 = 0$ is known an equal to g_i, and since solution at the boundary of the domain $u_{0,j}$ and $u_{n_x,j}$ is known from the boundary conditions, then the solution $u_{i,j}$ can be computed from Eq. (7.23) time step after time step. For this reason the linear system in Eqs. (7.20)-(7.22) above is an example of an *explicit method*.

Part (b): We now repeat the calculations in part (a) using a backward difference quotient in time. We introduce the notation d_{-t} for the backward difference quotient in time, and we keep the notation d^2_{cx} for the second order centered difference quotient in space,

$$d_{-t}u_{i,j} = \frac{u_{i,j} - u_{i,(j-1)}}{h_t}, \qquad d^2_{cx}u_{i,j} = \frac{u_{(i+1),j} + u_{(i-1),j} - 2u_{i,j}}{h_x^2}.$$

Now we state the approximate problem we will solve: Given the constants $f_{i,j}$ and g_i, for $i = 0, 1, \cdots, n_x$ and $j = 0, 1, \cdots, n_y$, find the constants $u_{i,j}$ solution of the linear equations and boundary conditions, respectively,

$$d_{-t}u_{i,j} - \kappa d^2_{cx}u_{i,j} = f_{i,j}, \tag{7.24}$$

$$u_{0,j} = u_{n_x,j} = 0, \tag{7.25}$$

$$u_{i,0} = g_i. \tag{7.26}$$

This system is simpler to solve than it looks at first sight. Let us rewrite it in as follows,

$$\frac{1}{h_t}\big(u_{i,j} - u_{i,(j-1)}\big) - \frac{\kappa}{h_x^2}\big(u_{(i+1),j} + u_{(i-1),j} - 2u_{i,j}\big) = f_{ij},$$

which is equivalent to the equation

$$(+2r)u_{i,j} - r\big(u_{(i+1),j} + u_{(i-1),j}\big) = u_{i,(j-1)} + h_t f_{i,j}, \tag{7.27}$$

where $r = \kappa h_t/h^2_{cx}$, as above. This last equation says that the solution at the time t_{j+1} can be computed if the solution at the previous time t_j is known. However, in this case we need to solve an $(n_x - 1) \times (n_x - 1)$ linear system at each time step. Such system is similar to the one that appeared in Example 7.3.1. Since the solution at the initial time $t_0 = 0$ is known an equal to g_i, and since solution at the boundary of the domain $u_{0,j}$ and $u_{n_x,j}$ is known from the boundary conditions, then the solution $u_{i,j}$ can be computed from Eq. (7.27) time step after time step. We emphasize that the solution is computed by solving an $(n_x - 1) \times (n_x - 1)$ system at every time step. For this reason the linear system in Eqs. (7.24)-(7.26) above is an example of an *implicit method*. ◁

What we have seen in this Section is just the first part of the story. We have seen how we can use linear algebra to obtain approximate solutions of few problems involving differential equations. The second part is to study how the solutions of the approximate problems approaches the solution of the original problem as the grid step size approaches zero. Consider the approximate solution $\mathsf{u} \in \mathbb{R}^{n+1}$ found in Example 7.3.1. Does the interpolation polynomial p_n constructed with the components of u approximate the function $u : [0, 1] \to \mathbb{R}$ solution to the boundary value problem in Eq. (7.7) in the limit $n \to \infty$? A similar question can be asked for the solutions $\{u_{i,j}\}$ obtained in parts (b) and (c) in Example 7.3.2. We will study the answers to these questions in the following Chapters.

One last remark is the following. Comparing parts (a) and (b) in Example 7.3.2 we see that explicit methods are simpler to solve than implicit methods. A matrix must be inverted to solve an implicit method, while this is not needed to solve an explicit method. So, why are implicit methods studied at all? The reason is that in the limit $n \to \infty$ the approximate solutions of explicit methods do not approximate the solution of the original differential equation as good as a solution of an implicit method. Moreover, the solution of an explicit method may not converge at all, while the solutions of implicit methods always converges.

Further reading. See Section 1.4 in Meyer's book [3] for a detailed discussion on discretizations of two-point boundary values problems.

7.3.4. **Exercises.**

7.3.1.- Consider the boundary value problem for the function u given by

$$\frac{d^2u}{dx^2}(x) = 25\,x,$$
$$u(0) = 0, \quad u(1) = 0,$$
$$x \in [0,1].$$

Divide the interval $[0,1]$ into five equal subintervals and use the finite difference method to find an approximate solution vector $\mathsf{u} = [u_i]$ to the boundary value problem above, where $i = 0, \cdots, 5$. Use **centered** difference quotients given in Eq. (7.17) to approximate the derivatives of function u.

7.3.2.- Given an infinite differentiable function $u : \mathbb{R} \to \mathbb{R}$. apply twice the forward difference quotient d_+ to show that the second order forward difference quotient has the form

$$d_+^2 u(x) = \frac{u(x+2\Delta x) - 2u(x+\Delta x) + u(x)}{(\Delta x)^2}.$$

7.3.3.- Given an infinite differentiable function $u : \mathbb{R} \to \mathbb{R}$. apply twice the backward difference quotient d_+ to show that the second order backward difference quotient has the form

$$d_-^2 u(x) = \frac{u(x) - 2u(x-\Delta x) + u(x-2\Delta x)}{(\Delta x)^2}.$$

7.3.4.- Consider the boundary value problem given in the Exercise **7.3.1**. Divide again the interval $[0,1]$ into five equal subintervals and find the 4×4 linear system that approximates the original problem using **forward** difference quotients to approximate derivatives of function u. You do not need to solve the linear system.

7.3.5.- Consider the boundary value problem given Problem **7.3.1**. Divide again the interval $[0,1]$ into five equal subintervals and find the 4×4 linear system that approximates the original problem using **backward** difference quotients to approximate derivatives of function u. You do not need to solve the linear system.

7.3.6.- Consider the boundary value problem for the function u given by

$$\frac{d^2u}{dx^2}(x) + 2\frac{du}{dx}(x) = 25\,x,$$
$$u(0) = 0, \quad u(1) = 0,$$
$$x \in [0,1].$$

Divide the interval $[0,1]$ into five equal subintervals and use the finite difference method to find an algebraic linear system that approximates the original boundary value problem above. Use centered difference quotients given in Eq. (7.17) to approximate derivatives of function u. You do not need to solve the linear system.

7.4. Finite element method

The finite element method permits the computation of approximate solutions to differential equations. Boundary value problems involving a differential equations are first transformed into integral equations. The boundary conditions are included into the integral equation by performing integration by parts. The original problem is transformed into inverting a bilinear form on a vector space. The approximate problem is obtained when the integral equation is solved not on the whole vector space but on a finite dimensional subspace. By a careful choice of the subspace, the calculations needed to obtain the approximate solution can be simplified. In this Section we study the same differential equations we have seen in Sect. 7.3 when we described the finite difference methods. Finite difference methods are different form finite element methods. The former method approximates derivatives by difference quotients in order to obtain the approximate problem involving an algebraic linear system. The latter method produces an algebraic linear system by restricting an integral version of the original differential equation onto a finite dimensional subspace of the original vector space where the integral equation is defined.

7.4.1. **Differential equations.**

We now recall the boundary value problem we are interested to study. This is the first problem we studied in Sect. 7.3, that is, to find an approximate solution to a boundary value problem for an ordinary differential equation: Given an infinitely differentiable function $f : [0,1] \to \mathbb{R}$, find a function $u : [0,1] \to \mathbb{R}$ solution of the boundary value problem

$$\frac{d^2u}{dx^2}(x) + \frac{du}{dx}(x) = f(x), \tag{7.28}$$

$$u(0) = u(1) = 0. \tag{7.29}$$

Recall that finding the function u involves doing two integrations that introduce two integration constants. These constants are determined by the two conditions on $x = 0$ and $x = 1$ above, called boundary conditions, since they are conditions on the boundaries of the interval $[0,1]$. Because of this extra condition on the ordinary differential equation this problem is called a boundary value problem.

This problem can be expressed using linear transformations on vector spaces. Consider the inner product spaces $\big(V, \langle\,,\rangle\big)$ and $\big(W, \langle\,,\rangle\big)$, where $V = C_0^\infty([0,1],\mathbb{R})$ is the space of infinitely many differentiable functions that vanish at $x = 0$ and $x = 1$, $W = C^\infty([0,1],\mathbb{R})$ is the space of infinitely many differentiable functions, while the inner product is defined as

$$\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \int_0^1 \boldsymbol{f}(x)\boldsymbol{g}(x)\,dx.$$

Introduce the linear transformation $L : V \to W$ defined by

$$L(\boldsymbol{v}) = \frac{d^2\boldsymbol{v}}{dx^2} + \frac{d\boldsymbol{v}}{dx}$$

Then, the boundary value problem in Eqs. (7.28)-(7.29) can be expressed in the following way: Given a vector $\boldsymbol{f} \in W$, find a vector $\boldsymbol{u} \in V$ solution of the equation

$$L(\boldsymbol{u}) = \boldsymbol{f}. \tag{7.30}$$

Notice that the boundary conditions in Eq. (7.29) have been included into the definition of the vector space V. The problem defined by Eqs. (7.28)-(7.29), which is equivalently defined by Eq. (7.30), is called the **strong formulation** of the problem.

So, we have expressed a boundary value problem for a differential equation in terms of linear transformations between appropriate infinite dimensional vector spaces. The next step is to use the inner product defined on V and W to transform the differential equation in (7.30) into an integro-differential equation, which will be called the weak formulation of the problem. This idea is summarized below.

7.4.2. **The Galerkin method.** The Galerkin method refers to a collection of ideas to transform a problem involving a linear transformation between infinite dimensional inner product spaces into a problem involving a matrix as a function between finite dimensional subspaces. We describe in this Section the original idea, introduced by Boris Galerkin around 1915. Galerkin worked only with partial differential equations, but we now know that his idea works in the more general context of a linear transformation between infinite dimensional inner product spaces. For this reason we describe Galerkin's idea in this more general context. The Galerkin method is to transform the strong formulation of the problem in (7.30) into what is called the weak formulation of the problem. This transformation is done using the inner product defined on the infinite dimensional vector spaces. Before we describe this transformation we need few definitions.

7.4.3. **Finite element method.**

7.4.4. **Exercises.**

7.4.1.- .

7.4.2.- .