Linear spaces

Joaquín Gómez

June 6, 2025

Contents

Symbol Glossary

Symbol	Meaning	Context / Notes
V	Linear space	The main set we're working in
O	Zero vector (element of V)	Use this consistently for the "zero of the space"
0	Scalar zero	Element of $\mathbb{R}$
I	Identity element	Scalar 1 in scalar multiplication or the identity matrix
$L(S)$	Linear span of a set S	Smallest subspace containing all linear combinations of S
$\dim V$	Dimension of V	Number of elements in a basis for V
$\\|x\\|$	Norm (length) of element x	Defined as $\sqrt{(x, x)}$ using the inner product
(x, y)	Inner product of x and y	Dot product in $\mathbb{R}^n$, or more generally defined
x_i	Component of vector x	Used in expressions like $x = \sum c_i e_i$
c_i	Scalar coefficient	Used in linear combinations
e_i	Basis vector	In an ordered basis of V
P	Projection matrix	Used for projecting onto subspaces
$\mathrm{proj}_u v$	Projection of v onto u	Defined as $\left(\frac{v \cdot u}{\\|u\\|^2}\right) u$

Chapter 1

Introduction

1.1 Definition of a linear space

Let V be a non-empty set of objects, called *elements*. The set V is called linear space if it satisfies the following ten axioms, which are stated in three groups.

Axioms of a linear space

Axiom 1. *Closure property of addition*: for every pair of elements $x, y \in V$, their sum is written as $z = x + y$ and $z \in V$.

Axiom 2. *Closure property of scalar product*: for any $x \in V$ and $a \in \mathbb{R}$, there is an element $z = ax \in V$.

Axioms for addition

There are four axioms of addition, we will use a number and a letter to refer to them. However if we are talking of addition properties we will simply use a letter to reference them. The same will be done for the axioms of scalar products.

Axiom 3.a. *Commutative law*: for any $x, y \in V$, we have $x + y = y + x$.

Axiom 4.b. *Associative law*: for any $x, y, z \in V$, we have $(x + y) + z = x + (y + z)$.

Axiom 5.c. *Existence of zero as an element*: there is a number in V, designated as O (big 'o'), that satisfies

$$x + O = x, \quad \forall x \in V$$

Axiom 6.d. *Opposite elements*: for all $x \in V$, the element $(-1)x$ has the property

$$x + (-1)x = O$$

Axioms for scalar product

Axiom 7.a. *Associative law*: for all $x \in V$ and every pair $a, b \in \mathbb{R}$, we have

$$a(bx) = (ab)x$$

Axiom 8.b. *Distributive law for addition in V*: for all $x, y \in V$ and $a \in \mathbb{R}$, it is true that

$$a(x + y) = ax + ay$$

Axiom 9.c. *Distributive law for addition in $\mathbb{R}$*: for any $x \in V$ and $a, b \in \mathbb{R}$, we have

$$(a + b)x = ax + bx$$

Axiom 10.d. *Existence of an identical element*: for all $x \in V$ theres an unique element I such that $Ix = x$ (commonly this element is 1. But, for example, the identical element in matrix spaces is called *identity matrix*, defined as $I = \text{diag}(1)$)

1.2 Examples of linear spaces

The following examples can be proven to be linear spaces

1. Real numbers
2. The vector space of real numbers $\mathbb{R}^n$
3. The set of all matrices
4. Polynomials P with $\deg P \leq n$ (in this case, if $\deg P = n$, we would have a problem with axioms of additions. We can't ensure the sum of two polynomials of degree n has degree n).
5. The set of all polynomials
6. The set of continuous functions on an interval $[a, b]$. This space is designated as $C(a, b)$.
7. The set of all integrable functions on an interval

8. The set of differentiable functions on an interval
9. A plane in $\mathbb{R}^3$ with the equation $ax+by+cz=0$. Note that this plane must always go through the origin to be a linear space.

There are plenty of examples for linear spaces. We can "create" a linear space if we define addition and multiplication for that space.

1.3 Consequences of the axioms

The following theorems are a consequence of the axioms of linear space.

Theorem 1.3.1 (Uniqueness of 'O')**.** *In any linear space there is one and only one zero element*

Proof. Axiom 5 ensures that there is at least one 'O' in V. Now, suppose there are two zeroes in V. Let $x = O_1$ and $O_2 = O$, thus $x + O = x + O_2 = x = O_1$, but as O_1 is zero, $O_1 + O_2 = O_2$, this means that $O_1 = O_2 = O$ □

Theorem 1.3.2 (Uniqueness of opposites)**.** *In any linear space each x has one and only one opposite y such that $x + y = O$*

Proof. Axiom 6 ensures there is at least one opposite of x in V. Let $y_1, y_2 \in V$ be two different opposite elements for x. Then $x+y_1 = O$ and $x+y_2 = O$, then

$$(x + y_1) + y_2 = y_2 + O = y_2$$

and

$$y_1 + (x + y_2) = y_1 + O = y_1$$

Thus $y_1 + (x + y_2) = y_1 + (x + y_1) = y_1 + O = O + y_1$, this proves that $y_1 = y_2$. □

1.4 Subspaces of a linear space

Let V be a linear space and let S be a subset of V, if S is also a linear space, then we say that "S is a subspace of V".

A subset of a linear space if a subspace only if it satisfies the axioms of closure.

Theorem 1.4.1. *Let V be a linear space, if $S \subset V$ and $S \neq \emptyset$ satisfies the ten axioms of closure then S is a subspace of V.*

The proof for this theorem is easy, and so I discarded it.

Definition 1.4.1. *Let $S \subset V$, and $S \neq \emptyset$, where V is a linear space. If $x \in V$ and*

$$x = \sum_{i=1}^{k} c_i x_i$$

where $x_1, x_2, \ldots, x_k \in S$ and $c_1, c_2, \ldots, c_k \in \mathbb{R}$, is called a linear combination of elements in S. The set of linear combinations of the elements of S satisfies the axioms of closure, so it is also a subspace of V. We say that this subspace is generated by S and we call it the linear span of S, designated by $L(S)$. If $S = \emptyset$, we define $L(S) = \{O\}$.

1.5 Dependent and independent subsets of a linear space

In this section we introduce the concept of independence, that is important when working with systems of linear equations, matrices, and other subjects in linear algebra.

Definition 1.5.1. *Let S be a set of elements of a linear space V. S is dependent if there exists a finite set of distinct elements in $x_1, x_2, \ldots, x_k \in S$, and a set of scalars $c_1, c_2, \ldots, c_k$ where not all of them are zero, that satisfies*

$$\sum_{i=1}^{k} c_i x_i = 0$$

A set is independent if it is not dependent. So the following

$$\sum_{i=1}^{k} c_i x_i = 0, \quad \text{implies } c_1 = c_2 = \cdots = c_k = 0$$

Independency and dependency are properties of sets of elements. However, we can apply the same concepts to the elements itself. For example, a set of vectors $v_1, v_2, \ldots, v_n \in \mathbb{R}^n$ is called independent if there is **not** a linear combination of these vectors that produce the zero vector.

Example 1.5.1. *Let $u_k(t) = t^k$ for $k = 1, 2, \ldots, n$ and $t \in \mathbb{R}$. The set $V = \{u_1, u_2, \ldots, u_n\}$ is independent except in the subset S where $t = -1$ and n is odd.*

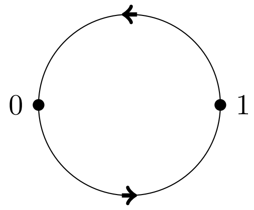
Proof. For S to be independent, there must be $c_1, c_2, \ldots, c_n$, where $c_1 = c_2 = \cdots = c_n = 0$ and

$$\sum_{k=0}^{n} c_k t^k = 0$$

To solve this, we set $c_0 = c_1 = \cdots = c_n$. If we define

$$f(t) = \sum_{k=0}^{n} c_k t^k,$$

note that $f(-1) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$ We can draw a picture for this problem. Imagine a circle, in which we have two points. We can travel the circumference counterclockwise starting from 1. We start from 1 because in the case $n = 0$, $f(-1) = 1$.



Now start performing counterclockwise turns, and count how many times you go from 0 to 1, and from 1 to 0.

If $C_{0\to 1}$ and $C_{1\to 0}$ are the counts of going from 0 to 1 and from 1 to 0 respectively, we have that if $C_{0\to 1} = C_{1\to 0}$, then $f(-1) = 1$. Otherwise, we must have $f(-1) = 0$.

But having $C_{0\to 1} = C_{1\to 0}$ and that the total count is $C = C_{0\to 1} + C_{1\to 0}$, means

$$C = 2C_{1\to 0} = 2C_{0\to 1}$$

Hence, C is an even number. To verify that S is dependent, we set $n = 2r - 1$ for any integer r and $t = -1$. With the results above, we can see that

$$\sum_{k=0}^{2r-1} c_k(-1)^k = 0$$

if $c_1 = c_2 = \cdots = c_{2r-1}$, but not necessarily zero. This proves that S is dependent. $\square$

Theorem 1.5.1. *Let $S = \{x_1, x_2, \ldots, x_k\}$ an independent set formed by k elements of a linear space V and let $L(S)$ be the linear span of S. Then, any set of $k + 1$ elements from $L(S)$ is dependent.*

What this theorem says is that, taking any set of vectors in the linear span of S, this is, formed by combining elements of S (this vectors can be of any nature), then, if we form a subset of the linear span of S and it has more elements than S itself, the set will be dependent. This is because we are not providing any new "dimension" to the new set. Say $S \in \mathbb{R}^{\nvDash}$, as we are restricted to be in $\mathbb{R}^{\nvDash}$, taking 4 vectors won't make any object in $R^{i>3}$

Proof. Let $T = \{y_1, y_2, \ldots, y_{n+1}\} \subset L(S)$, this means that each y_i is a linear combination of elements in S

$$y_i = \sum_{j=1}^{n} a_{ij} x_j, \quad \text{for } i = 1, 2, \ldots, n+1$$

For T to be dependent, there must be some scalar set $C = \{c_1, c_2, \ldots, c_{n+1}\}$, where not all of them are zero, that satisfies

$$\sum_{i=1}^{n+1} c_i y_i = 0$$

We now want to prove by induction that for $n-1$ elements of T, there is a linear combination that satisfies dependency. Thus, we can try to form an equation that represents T as a linear combination of $n-1$ elements. For this, we are going to take one element of T, multiply it by some scalar and subtract each element of T.

Take the 1^{st} element in T and multiply it by $c_i = \frac{a_{i1}}{a_{11}}$

$$c_i y_1 = a_{i1} x_1 + \sum_{j=2}^{n} c_i a_{1j} x_j$$

Now subtract y_1

$$\begin{aligned} c_i y_1 - y_i = a_{i1} x_1 + \sum_{j=2}^{n} c_i a_{1j} x_j - a_{i1} x1 + \sum_{j=2}^{n} a_{ij} x_j \\ = \sum_{j=2}^{n} c_i a_{1j} x_j - a_{ij} x_j \qquad (1.1) \\ = \sum_{j=2}^{n} \left(c_i a_{1j} - a_{ij}\right) x_j \end{aligned}$$

Equation (1.1) is indeed a linear combination of $n-1$ elements of S. By induction for n, we can prove that there are n scalars $t_2, t_3, \ldots, t_{n+1}$, that satisfy

$$\sum_{j=2}^{n+1} t_i \left(c_i y_1 - y_i\right) = 0 \qquad (1.2)$$

As each y_i is a linear combination of elements of S, we can write y_i in terms of y_1.

Equation (1.2) is solvable, because $y_i = c_i y_1$, this is true by the fact that $T \subset L(S)$. $\square$

1.6 Basis and dimension

Definition 1.6.1. *A finite set S of elements of a linear space V is called a finite basis of V is S is independent and spans V. V is of finite dimension if it has a finite basis. Otherwise, V has infinite dimension.*

Theorem 1.6.1. *Let V be a linear space of finite dimension. Then any finite basis of V has the same number of elements.*

Proof. This theorem can be proved with theorem 1.4, let S and T be two finite bases for V, with k and m elements respectively. If S generates V, then V must have k elements, we know that any set of $k+1$ elements of V is dependent. Thus, T must have $m \geq k$ elements. Applying the same reasoning vice-versa yields that $k = m$. $\square$

This does not mean that a set of $k+1$ elements of V can't span V. It states that, the number of elements for a finite basis of a linear space V of dimension k, must have the same number of elements.

Definition 1.6.2. *If a linear space V has a finite basis of n elements, we write $n = \dim V$.*

The following theorem will not be proven. However, it has an intuitive explanation.

Theorem 1.6.2. *Let V be a linear space of finite dimension, with $\dim V = n$. Then*

1. *If S is a finite basis for V, and T is a set of independent elements of V, then $T \subseteq S$.*

2. *Any set of n independent elements of V is a finite basis for V.*

1.7 Components

Let V be a linear space with $\dim V = n$, and consider an ordered basis

$$\{e_1, e_2, \ldots, e_n\}$$

This ordered basis is considered as an n-tuple $(e_1, e_2, \ldots, e_n)$.

Definition 1.7.1. *An ordered basis of a linear space V is a set of elements of V that form a basis and provides information about the order of its elements.*

If $x \in V$, we can express x as a linear combination of elements of the basis

$$x = \sum_{i=1}^{n} c_i e_i \tag{1.3}$$

This ensures that there is only one representation of x, take $x = \sum_{i=1}^{n} c_i e_i$, andintroductionintroduction $x = \sum_{i=1}^{n} d_i e_i$. Then

$$\sum_{i=1}^{n} c_i e_i = \sum_{i=1}^{n} d_i e_i$$

Then $\sum_{i=1}^{n} (c_i - d_i) e_i = O$, where O is the zero vector/element of V. This means that $c_i = d_i$ for $i = 1, 2, \ldots, n$. So there is only one representation of x in V.

Chapter 2

Euclidean spaces

We start the section by defining what is a Euclidean space.

Definition 2.0.1. *A Euclidean space is a finite-dimensional linear space that satisfies Euclidean geometry. They also are metric spaces, which are sets that have a notion of distance between its elements. Euclidean spaces are equipped with an inner product.*

Euclidean spaces have a set of properties, that were defined as axioms in *Euclid's Elements*, which are

1. If $a = b$ and $b = c$ then $a = c$ (the transitive property)
2. If $a = b$ then $a + c = b + c$ (the equal sum property)
3. If a line segment $\overline{AB}$ coincides in length and direction with $\overline{CD}$ then $\overline{AB} = \overline{CD}$.
4. The whole is greater than the part. This can be thought as: *let A and $B \subset A$ be two arbitrary sets, then A is "bigger" than B.*
5. Things that are double of the same thing are equal to each other. (This one is very obvious, consider two equal circles with radius r_1 and r_2, then we can say that $r_1 = r_2$).

2.1 Dot product and inner product

Definition 2.1.1. *The inner product is a function that maps two elements x and y from a linear space V to a real number. We write the inner product as (x, y).*

Any inner product satisfies the following properties:

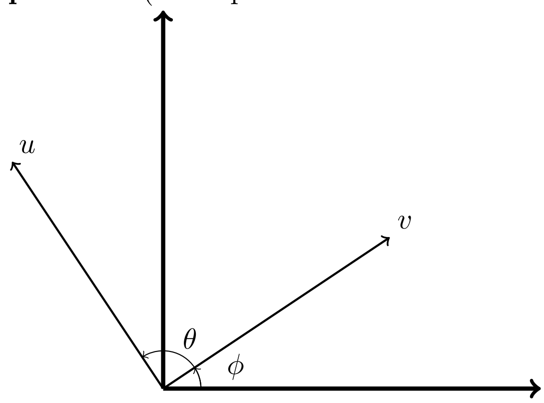
1. $(x, y) = \overline{(y, x)}$ *(hermitian symmetry)*
2. $(x, y + z) = (x, y) + (x, z)$ *(linearity)*

3. $c\,(x, y) = (cx, y)$ *(homogeneity)*

4. $(x, x) \geq 0$ *(positive definite)*

Remember: *a linear space with inner product is called a Euclidean space.*

Example 2.1.1 (Inner product of two vectors in $\mathbb{R}^2$)**.**



Now, we have $u = (u_x, u_y)$ *and* $v = (v_x, v_y)$*. If we define the inner product of two vectors as*

$$(u, v) = u \cdot v = \sum_{i=1}^{n} u_i v_i$$

For $n = 2$*, we have* $u \cdot v = u_x v_x + u_y v_y$*. We know the following relationships*

$$\begin{aligned} \cos\theta &= \frac{u_x}{|u|}, \qquad \sin\theta = \frac{u_y}{|u|} \\ \cos\phi &= \frac{v_x}{|v|}, \qquad \sin\phi = \frac{v_y}{|v|} \end{aligned} \tag{2.1}$$

If we solve for u *and* v *and substitute in the inner product formula, we get*

$$\begin{aligned} u \cdot v = |u| \cdot |v| \cos\theta \cos\phi + |u| \cdot |v| \sin\theta \sin\phi \\ = |u||v|(\cos\theta \cos\phi + \sin\theta \sin\phi) \\ = |u||v|(\cos{(\theta - \phi)}) \end{aligned} \tag{2.2}$$

This means that the dot product of two vectors is the product of their length times the cosine of the angle between them. The angle between u *and* v *is given by*

$$\theta - \phi = \arccos\left(\frac{u \cdot v}{|u||v|}\right) \tag{2.3}$$

Well, this rises a question. How can you derive the inner product for a real vector space? Well, there are various points to note, but let's imagine that we want to measure the length of a vector. How can we measure distance? But also, we want to measure the distance between two vectors. Let's go with an example to make things clearer.

Example 2.1.2 (Distance of two vectors in $\mathbb{R}^n$)**.** *We first define two vectors* $u = (u_1, u_2, \dots, u_n)$ *and* $v = (v_1, v - 2, \dots, v_n)$

Now, a third vector, we call it $w = u - v$ *has squared length*

$$|w|^2 = (u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2 \tag{2.4}$$

Expanding one of the right-hand side terms we get $(u_i - v_i)^2 = u_i^2 - 2v_iu_i + v_i^2$*. Grouping the terms in (2.4) results in*

$$\sum_{i=1}^{n} \left(w_i^2\right) = \sum_{i=1}^{n} \left(u_i^2\right) + \sum_{i=1}^{n} \left(v_i^2\right) - 2\sum_{i=1}^{n} u_iv_i \tag{2.5}$$

Note that in (2.5) the dot product appears in the last term of the right-hand side. We can rewrite the equation as

$$w \cdot w = v \cdot v + u \cdot u - 2u \cdot v \tag{2.6}$$

And using formula (2.2)

$$|w|^2 = |v|^2 + |u|^2 - 2|u||v|\cos\theta \tag{2.7}$$

Where θ *is the angle between* u *and* v*. Note that, because the angle between a vector and itself is* $\theta = 0$*,* $\cos\theta = 1$*.*

Equation (2.7) is nothing more than the *Law of Cosines*. Now, the dot product does not follow a "natural pattern" as one would call it. Think of the exponential function, it has a very natural reasoning, for example, in the growth of populations or in differential equations. However, the dot product is present when we measure elements in Euclidean spaces, like segments or vectors.

The dot product is not "derived" in a way most things are. Instead, it is useful because it simply "appears" in measurements.

The **inner product** is a generalization of the dot product in more general spaces. Each space can have a different definition for its inner product. For example

[Inner product of a functional space $C(a, b)$]

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be continuous functions in an interval $[a, b]$, the inner product is defined as

$$(f, g) = \int_a^b f(x)g(x)dx \tag{2.8}$$

2.2 Norms and length

The norm of an element x in a linear space is written as $\|x\|$ and has the following properties:

1. $\|x\| > 0$ if $x \neq 0$
2. $\|x\| = 0$ if $x = 0$
3. For a scalar a, $\|ax\| = |a| \cdot \|x\|$
4. For two elements y and x in a linear space, $\|x + y\| \leq \|x\| + \|y\|$

The 4^{th} property is the triangle inequality.

Definition 2.2.1. *Let x be an element of a linear space V. The norm of x is defined as $(x, x)^{1/2}$, this is, the square root of the inner product of x with itself.*

Definition (2.3) satisfies the properties of a norm.

Example 2.2.1. *Let $x \in \mathbb{R}^n$, the norm of a vector is given by the pythagorean theorem*

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Finally, we can see that $\|x\| = \sqrt{(x, x)}$, we defined the inner product of a real vector space as the dot product $(\cdot, \cdot) : V \times V \to \mathbb{R}$. Here V denotes the linear space, in our current example $V = \mathbb{R}^n$.

Example 2.2.2. *Let V be the functional space $C(a, b)$, the norm of a function f the interval $[a, b]$ is*

$$\|f\| = \sqrt{\int_a^b [f(\psi)]^2 \, d\psi}$$

This measure in functional spaces are useful when dealing with negative values on the integral. For example, the $\sin(\psi)$ function is zero when integrated in its period. However we can use the norm to measure it:

$$
\begin{aligned}
\|\sin(\psi)\| &= (\sin(\psi), \sin(\psi)) \\
&= \sqrt{\int_{\theta_1}^{\theta_2} [\sin(\psi)]^2 \, d\psi} \\
&= \sqrt{\left[\frac{\psi}{2} - \frac{\sin(2\psi)}{4}\right]_{\theta_1}^{\theta_2}} \\
&= \sqrt{\frac{\theta_1 + \theta_2}{2} - \frac{\sin(\theta_2)\cos(\theta_2) - \sin(\theta_1)\cos(\theta_1)}{2}}
\end{aligned}
$$

If $\theta_1 = 0$ and $\theta_2 = 2\pi$ we have $\|\sin(\psi)\| = \sqrt{\pi}$.

Theorem 2.2.1. *In the Euclidean space V, every inner product satisfies the Cauchy-Schwarz inequality:*

$$
|(x, y)|^2 \leq (x, x)(y, y) \tag{2.9}
$$

Proof. Let $z = ax + by$. If $x = 0$ and $y = 0$ the problem is trivial and the equality holds. Else, we can use the fact that $(z, z) \geq 0$ and by using the properties of the inner product:

$$
\begin{aligned}
&(z, z) = (ax + by, ax + by) \\
&= (ax, ax) + (ax, by) + (by, ax) + (by, by) \\
&= a\bar{a}\,(x, x) + a\bar{b}\,(x, y) + b\bar{a}\,(y, x) + b\bar{b}\,(y, y) \\
&\geq 0
\end{aligned}
$$

Now, let $a = \bar{a} = (y, y)$

$$
(y, y)(y, y)(x, x) + \bar{b}\,(y, y)(x, y) + b\,(y, y)(y, x) + b\bar{b}\,(y, y) \geq 0
$$

Dividing everything by (y, y) leaves

$$
(y, y)(x, x) + \bar{b}\,(x, y) + b\,(y, x) + b\bar{b} \geq 0
$$

If we let $b = -(x, y)$, such that its conjugate $\bar{b} = -(y, x)$, we obtain

$$
\begin{aligned}
&(x, x)(y, y) - (x, y)(y, x) - (x, y)(y, x) + (x, y)(y, x) \\
&(x, x)(y, y) - (x, y)(y, x) \geq 0
\end{aligned}
$$

If we reorder the terms of this equation, we are left with the Cauchy-Schwarz inequality

$$
(x, y)(x, y) \leq (x, x)(y, y)
$$

□

Example 2.2.3. *Applying theorem (2.1) in $C(a,b)$, with inner product $(f,g) = \int_a^b f(t)g(t)dt$, results in*

$$\left(\int_a^b f(t)g(t)dt\right)^2 \leq \left(\int_a^b [f(t)]^2 dt\right)\left(\int_a^b [g(t)]^2 dt\right)$$

The triangle inequality is a direct consequence of the Cauchy-Schwarz inequality, see

The triangle inequality. Let $x, v \in V$ where V is an Euclidean space. With the properties of the inner product we can see that

$$\begin{aligned} \|x+y\|^2 &= (x+y, x+y) \\ &= (x,x) + (x,y) + (y,x) + (y,y) \\ &= (x,x) + (x,y) + \overline{(x,y)} + (y,y) \end{aligned} \tag{2.10}$$

□

The sum $(x,y) + \overline{(x,y)}$ is real, see

$$z = a + bi \longrightarrow \bar{z} = a - bi, \text{ such that } z + \bar{z} = 2a$$

So, we can use the Cauchy-Schwarz inequality. See that $\|(x,y)\|^2 \leq \|x\|\|y\|$ and $\|(y,x)\|^2 = \|\overline{(x,y)}\|^2 \leq \|x\|\|y\|$. Transforming (2.10) into an inequality holds

$$\begin{aligned} \|x+y\|^2 &= (x,x) + (x,y) + \overline{(x,y)} + (y,y) \\ &= \|x\|^2 + \|y\|^2 + (x,y) + \overline{(x,y)} \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \end{aligned}$$

You can easily see that $\|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2$. We get

$$\|x+y\|^2 \leq (\|x\| + \|y\|)^2 \tag{2.11}$$

This proves the triangular inequality

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

Definition 2.2.2. *In a real Euclidean space V, the angle between two non-null elements x and y is defined as the number θ in the interval $[0, \pi]$. This number satisfies the following equation*

$$\cos\theta = \frac{(x,y)}{\|x\|\|y\|} \tag{2.12}$$

By using the Cauchy-Schwarz inequality in (2.4) we can prove that

$$\|(x,y)\|^2 = \|x\|^2\|y\|^2\cos^2\theta \leq \|x\|\|y\|$$

Such that $\|x\|\|y\|\cos^2\theta \leq 1$. But we know that $\|x\| \geq 0$ and $\|y\| \geq 0$, so

$$0 \leq \|x\|\|y\|\cos^2\theta \leq 1$$

This is the same as

$$0 \leq \|(x,y)\|^2 \leq 1$$

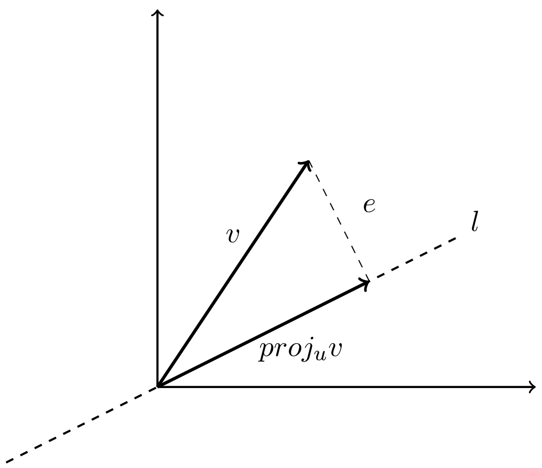
Thus, $-1 \leq (x,y) \leq 1$. This proves that the quotient in the right-hand side of (2.12) is in $[-1,1]$, so $\cos\theta$ will go from $[0,\pi]$.

Chapter 3

Projections onto subspaces

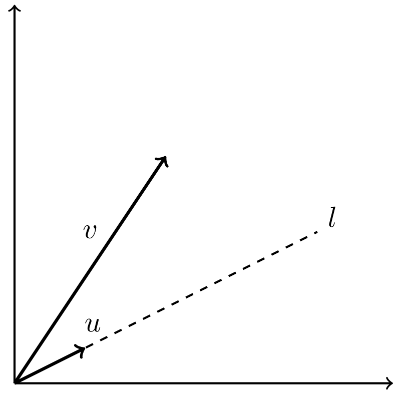
A projection is an idempotent mapping of a set (or any structure) into a subset (or sub-structure). Idempotent means that, projecting once is the same as projecting n-times.

We will see this as a vector being projected in a line, a line is a subspace of $\mathbb{R}$, and our vector is in $\mathbb{R}^2$.



This is the most basic example of projection. Let's deduce a formula.

Example 3.0.1. *Let $v \in \mathbb{R}^2$ and let l be a line in $\mathbb{R}^2$ parametrized by a vector u as $f(t) = ut$.*



What we want to find is some constant k such that ku is perpendicular to l. We know from the previous image that $e = proj_u v - v = uk - v$

$$(ku - v) \cdot u = k\|u\|^2 - v \cdot u = 0$$

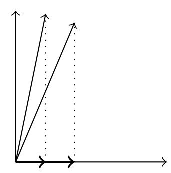
Then $k = \frac{v \cdot u}{\|u\|^2}$. The formula for projection becomes

$$proj_u v = ku = \frac{v \cdot u}{\|u\|^2} \cdot u \tag{3.1}$$

We call this projection in terms of u, because l is parametrized by u.

Now, we want to extend this idea to any linear space. We can see that a projection is a multiplication of the "projection basis" by some scalar.

This gives us an idea of what the inner product is: we can think of the inner product as "how much of some element is onto other element". Think of projecting a vector v onto u when $v \perp u$.



You can see that the projection tends to a vector with no length. This can be proven easily with (2.2):

$$\begin{aligned} proj_u v &= u \cdot \frac{u \cdot v}{\|u\|^2} \\ &= u \cdot \frac{\|v\|}{\|u\|} \cos\psi \end{aligned} \tag{3.2}$$

The $\frac{\|v\|}{\|u\|}$ part is the amount of times u fits in v, and if $\psi = \frac{\pi}{2}$ then the projection is the O vector.

Example 3.0.2 (Projections in $\mathbb{R}^n$)**.** *Let $x \in \mathbb{R}^n$ be a vector in a plane, this plane must be of dimension $n-1$. Thus, it has a basis with $n-1$ elements given by*

$$W = (e_1, e_2, \ldots, e_{n-1})$$

Where $e_1, e_2, \ldots, e_{n-1} \in \mathbb{R}^n$. These vectors parametrize the plane, such that x can be written as

$$x = \sum_{i=1}^{n-1} c_i e_i \tag{3.3}$$

for any scalars $c_1, c_2, \ldots, c_{n-1}$. Now, suppose that we have a vector $v \in \mathbb{R}^n$ that preferably, does not lie in this plane.

To find the projection, we can set a set of equation that are analogous to the example in $\mathbb{R}^2$:

$$\begin{cases} e_1(v - kx) = 0 \\ e_2(v - kx) = 0 \\ \cdots \\ e_{n-1}(v - kx) = 0 \end{cases} \tag{3.4}$$

Where $k \in \mathbb{R}$. What we want is to make each basis vector orthogonal to a vector x that connects v to the plane. In other words, the vector $(v - kx)$ will be perpendicular to the plane.

We will collect each basis vector into a matrix

$$A = \begin{bmatrix} \vdots & \vdots & & \vdots \\ e_1 & e_2 & \cdots & e_{n-1} \\ \vdots & \vdots & & \vdots \end{bmatrix} \tag{3.5}$$

Here, A has dimension $n \times (n-1)$. The system of equations in (3.4) becomes

$$A^T (v - kx) = 0 \tag{3.6}$$

But we can rewrite x as a vector c times A, so that

$$x = Ac = \begin{bmatrix} \vdots & \vdots & & \vdots \\ e_1 & e_2 & \cdots & e_{n-1} \\ \vdots & \vdots & & \vdots \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix} \tag{3.7}$$

You can verify that this is the same as writing x in the form (3.3). Rewriting equation (3.6) leaves

$$A^T (v - Ac) = 0$$

We omit the term k since we will assume that the vector c scales x accordingly. Expanding the equation

$$A^T(v - Ac) = A^T v - A^T Ac = 0 \tag{3.8}$$

Now, we solve for c

$$c = (A^T A)^{-1} A^T v \tag{3.9}$$

Now, as the projection is given by scaling all the basis vectors in A by c (think of c as the scalar term in the 2-dimensional example), we can write our projection vector p as

$$p = Ac = A(A^T A)^{-1} A^T v = Pv \tag{3.10}$$

We define P to be the ***projection matrix****, and v is the vector projected onto the plane A.*

To figure out the equation of the plane, we think of the definition of a plane: each vector in the plane must be perpendicular to a normal vector. This means that we always have $A \cdot n = O$, where n is the normal vector.

We can write the equation for an hyperplane in $\mathbb{R}^{n-1}$ as

$$n^T(x - x_0) = 0 \tag{3.11}$$

where:

- *$n \in \mathbb{R}^n$ is the* ***normal vector*** *to the hyperplane,*
- *$x \in \mathbb{R}^n$ is any point* ***on*** *the hyperplane,*
- *$x_0 \in \mathbb{R}^n$ is a fixed point in the hyperplane (used to position or "anchor" the hyperplane in space).*

This equation states that the vector from x_0 to any point x on the hyperplane is ***orthogonal*** *to the normal vector n, ensuring all such points lie in the same flat $(n-1)$-dimensional space.*

In this definition we didn't use the basis vectors in W, however as x is a combination of basis vectors, it is not necessary to explicitly write them.

Some important applications of projections are **linear regression**, or the **Fourier series** we will see how to derive them later.

Example 3.0.3 (Linear Regression)**.** *Let S be a set of points in $\mathbb{R}^{n-1}$, suppose we can map the points with a scalar field given by $f : \mathbb{R}^{n-1} \to \mathbb{R}$, this means that S can be written as:*

$$S = \left\{ \left(x_{11}, x_{12}, \ldots, x_{1(n-1)}, z_1\right), \ldots, \left(x_{m1}, x_{m2}, \ldots, x_{m(n-1)}, z_m\right) \right\} \tag{3.12}$$

We can set a system of equations that uses a coefficient vector u and a bias:

$$\begin{cases} x_{11}u_1 + x_{12}u_2 + \cdots + x_{1(n-1)}u_{n-1} + b = z_1 \\ x_{21}u_1 + x_{22}u_2 + \cdots + x_{2(n-1)}u_{n-1} + b = z_2 \\ \vdots \\ x_{m1}u_1 + x_{m2}u_2 + \cdots + x_{m(n-1)}u_{n-1} + b = z_m \end{cases} \tag{3.13}$$

Each z is a combination of $n-1$ independent variables, plus a bias term. We can rewrite the system as an $m \times n$ matrix multiplied by an $n \times 1$ vector

$$\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1(n-1)} \\ 1 & x_{21} & x_{22} & \cdots & x_{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{m(n-1)} \end{bmatrix} \cdot \begin{bmatrix} b \\ u_1 \\ \vdots \\ u_{n-1} \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} \tag{3.14}$$

$$X \cdot u = z$$

We want to fit the equation to an approximation vector û. The vector z lies in a hyperplane. û is essentially a vector that gives the direction of a line that fits these points at best. This works because, the projection gives the minimum distance to z.

$$\hat{u} = \left(X^T X\right)^{-1} X^T z \tag{3.15}$$

But this approximation only has one dimension for the output. If you let z be a matrix, so that each output of the system of equations is a vector, then define u to be a $n \times p$ matrix and z to be $m \times p$.

$$\begin{bmatrix} 1 & x_{11} & \cdots & x_{1(n-1)} \\ 1 & x_{21} & \cdots & x_{2(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & \cdots & x_{m(n-1)} \end{bmatrix} \cdot \begin{bmatrix} b_1 & \cdots & b_p \\ u_{11} & \cdots & u_{1p} \\ \vdots & \ddots & \vdots \\ u_{(n-1)1} & \cdots & u_{(n-1)p} \end{bmatrix} = \begin{bmatrix} z_{11} & \cdots & z_{1p} \\ z_{21} & \cdots & z_{2p} \\ \vdots & \ddots & \vdots \\ z_{m1} & \cdots & z_{mp} \end{bmatrix} \tag{3.16}$$

Then û has dimension $n \times p$, defining a hyperplane that best fits the data in X across all output dimensions.

While û provides an approximation for each point in X, the hyperplane must fit all points. In an n-dimensional space, this requires only n coefficients.

For that, we can take the mean of each row (as we have n rows) and approximate a plane with the following equation:

$$P(x_1, x_2, \ldots, x_{n-1}) = \bar{b} + \bar{u}_1 x_1 + \cdots + \bar{u}_{n-1} x_{n-1} \tag{3.17}$$

Where

- $\bar{b} = \frac{1}{p}\sum_{i=1}^{p} b_i$
- $\bar{u}_i = \frac{1}{p}\sum_{j=1}^{p} u_{ij}$ *for* $i \in \{1, 2, \ldots, n-1\}$

Worked example: *Suppose that we have a set of points given in the table below:*

Point	x_1	x_2	$z^{(1)}$	$z^{(2)}$
1	*2*	*3*	*10*	*11*
2	*6*	*7*	*9*	*8*
3	*4*	*2*	*7*	*9*

Note that we have two different outputs, namely $z^{(1)}$ *and* $z^{(2)}$. *This is because we can have multiple readings for the same inputs* x_1, x_2.

Now, using the method described above we can set up our system of equations

$$X \cdot u = z$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 6 & 7 \\ 1 & 4 & 2 \end{bmatrix} \cdot \begin{bmatrix} b_1 & b_2 \\ u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \begin{bmatrix} 10 & 11 \\ 9 & 8 \\ 7 & 9 \end{bmatrix} \tag{3.18}$$

Solving for $\hat{u} = \left(X^T X\right)^{-1} X^T z$:

$$\hat{u} = \begin{bmatrix} 9.6667 & 12.3333 \\ -1.0833 & -0.9167 \\ 0.8333 & 0.1667 \end{bmatrix}$$

Now, we take the mean of each row and convert it to a column vector $\bar{u}$

$$\bar{u} = \begin{bmatrix} \bar{b} \\ \bar{u}_1 \\ \bar{u}_2 \end{bmatrix} = \begin{bmatrix} 11 \\ -1 \\ 0.5 \end{bmatrix}$$

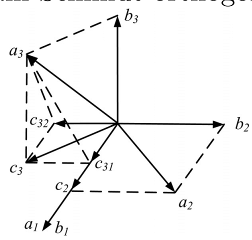
This gives a final approximation

$$P(x, y) = 11 - x + \frac{1}{2}y$$

3.1 Orthogonalization

Formally, orthogonalization is the process of taking a set of independent vectors in some linear space V and transforming them into an orthogonal basis. We define an orthogonal basis as

Figure 3.1: Gram-Schmidt orthogonalization in $\mathbb{R}^3$



Definition 3.1.1. *An orthogonal basis of a linear space V is a set of $n = \dim V$ elements, called $S = \{v_1, v_2, \ldots, v_n\}$ such that the inner product*

$$(v_i, v_j) = 0, \text{ if } i \neq j \tag{3.19}$$

Example 3.1.1. *We commonly define the basis of vector spaces like $\mathbb{R}^2$, $\mathbb{R}^3$, ..., $\mathbb{R}^n$ to be orthogonal. We denote their elements as $e_1, e_2, \ldots, e_n$.*

The **Gram-Schmidt** process converts a set of n independent elements to an orthogonal basis for the space that they span.

3.2 Gram-Schmidt orthogonalization

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of independent elements of a linear space V. We want to build an orthogonal basis $Y = \{y_1, y_2, \ldots, y_n\}$ in V.

To do so, we start by taking the first element of X. Let $y_1 = x_1$ be the first element of the basis. The rest of the elements will be built around this one.

If we take the second element, namely x_2, and we project x_2 onto y_1, you can easily check (by drawing it) that the vector $x_2 - \text{proj}_{y_1} x_2$ is orthogonal to x_1. You can see the projection formula derivation in the previous section. With this, we define $y_2 = x_2 - \frac{(x_2,y_1)}{(y_1,y_1)} y_1$.

Now, we have a basis of two vectors. Geometrically, it represents a plane. The third vector will not be coplanar to these two, because of their independency. Thus, we can project the vector onto the plane by combining both projections onto y_1 and y_2. So $y_3 = x_3 - \left(\text{proj}_{y_2} x_3 + \text{proj}_{y_1} x_3\right)$.

In figure 3.1 you can see a graphical description. The orthogonal basis is b_1, b_2, b_3. Take a closer look at a_3, this is the third independent vector, c_3 is a linear combination of c_{32} and c_{31}, who are the projections of a_3 onto b_1 and b_2. You can think of these projections as the "components" of a_3 in the direction of b_1 and b_2. Thus, the sum of those components form c_3, and $a_3 - c_3$ is a vector orthogonal to b_1 and b_2.

To find all the vectors, we can write the following equations. For $r = 1, 2, \ldots, n-1$ we have

$$y_1 = x_1 \tag{3.20}$$

$$y_{r+1} = x_{r+1} - \sum_{i=1}^{r} \frac{(x_{r+1}, y_i)}{(y_i, y_i)} x_{r+1} \tag{3.21}$$

Or, alternatively

$$y_{r+1} = x_{r+1} - \sum_{i=1}^{r} \text{proj}_{y_i} x_{r+1} \tag{3.22}$$

These are the formulas for **Gram-Schmidt** orthogonalization.

Example 3.2.1 (Fourier Series)**.** *To approximate continuous functions in an interval* $[0, 2\pi]$*, Fourier discovered that he could combine trigonometric polynomials. Let* $V = C(0, 2\pi)$ *the linear space of all real continuous functions in the interval* $[0, 2\pi]$*, we define the inner product as* $(f, g) = \int_a^b f(x)g(x)dx$*.*

Now, we try to find a normal set of trigonometric functions, this are the basis elements of our linear space. First, consider the functions $\cos kx$ *and* $\sin kx$*. In a previous section we've seen that the norm of these functions is* $\sqrt{\pi}$*. Now, define a finite set of n elements consisting of these trigonometric functions, and define them as* $\psi_0, \psi_1, \ldots$ *where*

$$\psi_0(x) = \frac{1}{\sqrt{2\pi}}, \quad \psi_{2k-1}(x) = \frac{\cos kx}{\sqrt{\pi}}, \quad \psi_{2k}(x) = \frac{\sin kx}{\sqrt{\pi}} \tag{3.23}$$

for $k \geq 1$*. You can check that* $(\psi_{2k-1}, \psi_{2k}) = \frac{1}{\pi} \int_0^{2\pi} \cos(kx) \sin(kx) dx = 0$ *for all k. Also, for two elements* $p, q \leq \dim V$ *where* $p \neq q$*, we have that*

$$\begin{aligned} \frac{1}{\pi} \int_0^{2\pi} \cos(px)\cos(qx)dx = 0 \\ \frac{1}{\pi} \int_0^{2\pi} \sin(px)\sin(qx)dx = 0 \end{aligned} \tag{3.24}$$

Thus, the set $\Psi = \{\psi_0, \psi_1, \ldots, \psi_{2k-1}, \psi_{2k}, \ldots\}$ *is a normalized orthogonal basis of* V*.*

These elements span a subspace S*, so* $S = L(\psi_0, \psi_1, \ldots)$*. If we want to approximate a function* f*, we must take the projection of* f *onto* S*. Let* f_n *be the projection onto an n dimensional subspace* S *spanned by the elements of* Ψ*:*

$$f_n = \sum_{k=1}^{n} (f, \psi_k)\psi_k, \quad \text{where } (f, \psi_k) = \int_0^{2\pi} f(x)\psi_k(x)\, dx \tag{3.25}$$

Note that we omit the coefficient $\frac{1}{(\psi_k, \psi_k)}$*, as* ψ_k *is normalized it has norm* 1*.*

Using the formulas in (3.23) we can write (3.25) as

$$f_n(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx) \tag{3.26}$$

where

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_0^{2\pi} f(x)\, dx \\ a_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx)\, dx \\ b_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx)\, dx \end{aligned} \tag{3.27}$$

Fourier series *can be written as*

$$f(x) = a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \tag{3.28}$$

Chapter 4

Linear Transformations

4.1 Definition

A transformation is a function whose domain and image are subspaces of a linear space. Let T be a transformation, we denote

$$T : V \to W \tag{4.1}$$

to say that T maps elements from its domain, V, to its image, W. Remember, V is any linear space and W is a subset of that linear space.

A linear transformation is a kind of transformation that follows certain properties. Let $T : V \to W$ be a linear transformation, then

1. T(x+y) = T(x)+T(y)
2. T(cx) = cT(x)

for $x, y \in V$ and $c \in \mathbb{R}$.

Both properties can be combined to get

$$T(ax + by) = aT(x) + bT(y),$$

which can be generalized as

$$T\left(\sum_{i=1}^{n} k_i x_i\right) = \sum_{i=1}^{n} k_i T(x_i) \tag{4.2}$$

for $x_1, x_2, \ldots, x_n \in V$ and $k_1, k_2, \ldots, k_n \in \mathbb{R}$.

4.2 Kernel and range

We must recap some concepts from the previous chapters. We know that any linear space must have a *zero element*, commonly we will denote this element as O.

Linear transformations map from a subspace to another subspace of the same linear space. The definition of a subspace given in theorem 1.4.1 says that any subspace S of a linear space V must satisfy the ten axioms of closure. Axiom 5 says that there must be a zero element, so any transformation must map the zero element.

There will be a subset in the domain of T that will map this element. This is called **kernel**.

Definition 4.2.1. *The **kernel** of a linear transformation $T : V \to W$ is a subset $K \subseteq V$ such that*

$$K(T) = \{x : x \in V \wedge T(x) = 0\} \tag{4.3}$$

As any subspace, $T(V)$ spans another subspace. This is called the **range** of T.

Theorem 4.2.1. *The set $T(V)$ (range of T) is a subspace of W. Furthermore, T maps the zero element on V to the zero element of W.*

Proof. Let $e_1, e_2, \ldots, e_n \in V$ be independent, and $k_1, k_2, \ldots, k_n \in \mathbb{R}$. We can form every element in V as a combination of these elements. Let y be any element in V given by

$$y = \sum_{i=1}^{n} k_i e_i,$$

then we have $T(y) \in W$. If this applies for any y, then it does for any element in V.

Now, to prove the second sentence, let $x \in V$ and $c \in \mathbb{R}$. The transformation $T(cx) = 0$, then $cT(x) = 0$ if $c = 0$ for any x $(0x = O)$. □

4.2.1 Dimension of the kernel and range of a transformation

We are interested in the relationship between the dimension of $K(T)$ and $T(V)$. For this, we present the following theorem.

Theorem 4.2.2 (Rank-Nullity Theorem). *If V is a linear space of finite dimension. Let T be a linear transformation, then $T(V)$ must be finite and we have*

$$\dim N(T) + \dim K(T) = \dim V \tag{4.4}$$

Proof. Let $n = \dim V$, and $e_1, e_2, \ldots, e_k$ be a basis for $K(T)$, where $k = \dim K(T)$. A basis of V can be

$$e_1, \ldots, e_k, e_{k+1}, \ldots, e_{k+r}$$

where $k + r = n$.

If we take an element y from V, we can write it as

$$y = \sum_{i=1}^{k+r} c_i e_i, \quad \text{for any scalars } c_1, c_2, \ldots, c_{k+r}$$

If we apply the transformation, we get

$$T(y) = T\left(\sum_{i=1}^{k+r} c_i e_i\right) = \sum_{i=k+1}^{k+r} c_i T(e_i)$$

Note that we have changed i to be $k+1$, because $T(e_1) = \cdots = T(e_k) = O$.

This proves that any element in $T(V)$ is a combination of r independent elements, so $r = \dim T(V)$, which proves the theorem. □

Now, we are going to give an example with infinite linear spaces.

Theorem 4.2.3. *Let V be a linear space of* ***infinite dimension****. Let T be a linear transformation $T : V \to W$. Then at least $T(V)$ (image of T) or $K(T)$ (kernel of T) must have infinite dimension.*

Proof. Suppose that

$$\begin{cases} \dim T(V) = r \\ \dim K(T) = k \end{cases}$$

Now, let $e_1, \ldots, e_k$ be a base for $K(T)$. And let a base for V be

$$e_1, \ldots, e_k, e_{k+1}, \ldots, e_{k+n},$$

where $n > r$. By contradiction, we want to prove that this is not possible. Any element $y \in V$ can be expressed as

$$y = \sum_{i=1}^{k+n} c_i e_i, \quad \text{for scalars } c_1, \ldots, c_{k+n}.$$

Applying the transformation on y leaves

$$T(y) = \sum_{i=1}^{k+n} c_i T(e_i),$$

but because $T(e_1) = \cdots = T(e_k) = 0$, we can express it as

$$T(y) = \sum_{i=k+1}^{k+n} c_i T(e_i)$$

The contradiction arises because if $\dim T(V) = r$, then the set $T(V)$ will be linearly dependent by theorem 1.5.1 (because $n > r$). This violates the Rank-Nullity Theorem (4.2.2). For the theorem to hold, at least the range $T(V)$ or the kernel $K(T)$ must have infinite dimension. □

4.3 Inverses

Let $T : V \to W$ be a function. An **inverse** is a function that, multiplied by T, maps from W to V.

Definition 4.3.1. *Let $T : V \to W$. A left inverse I_l is a function such that*

$$I_l \circ T(x) = x, \quad x \in V \tag{4.5}$$

A right inverse I_r satisfies

$$T(x) \circ I_r = x \tag{4.6}$$

If $I = I_r = I_l$, we call I the inverse of T.

(We are doing a composition of functions, defined as: *for two functions $f : V \to W$ and $g : S \to V$, and $x \in S$*

$$f \circ g = (f \circ g)(x) = f\left[g(x)\right] \tag{4.7}$$

note that the domain of f is the image of g)

The condition for a function to be invertible is that it must be a one-to-one mapping from V to W.

Definition 4.3.2 (One-to-one mapping)**.** *A one-to-one mapping from a set V to a set W is a function $T : V \to W$ that satisfies the following property. Let $x_1, x_2 \in V$, then*

$$T(x_1) = T(x_2), \text{ implies } x_1 = x_2 \tag{4.8}$$

In other words, a one-to-one mapping ensures that for each element in V, there is only one equivalent element in W.

With this criteria, there is a theorem that ensures that every one-to-one mapping has left and right inverses. The inverse equivalence is not always the case.

Theorem 4.3.1. *If $T : V \to W$ is a one-to-one mapping, then it has a left inverse and a right inverse. Also, T only has one left inverse, and it is the same as the right inverse.*

Proof. The one-to-one mapping definition says that for two elements $x_1, x_2 \in V$, the transformation satisfies $T(x_1) = T(x_2)$ if and only if $x_1 = x_2$. This ensures that there is at least one left inverse, let $y \in W$ and $S(y) = x$ be a left inverse

$$T\left[S(y)\right] = y$$

Now, let $S'(y)$ be another left inverse, as $T\left[S(y)\right] = y$ and $T\left[S'(y)\right] = y$, there is only one left inverse.

Now, if $y = T(x)$, we have

$$x = S[T(x)] = S(y) \tag{4.9}$$

Applying T again, we are left with $T(x)$, so indeed, the left and right inverses are the same. $\square$

Let's recap and give a complete definition

Definition 4.3.3. *Let $T : V \to W$ be one-to-one in V. T has only one unique inverse (left and right), and we designate it by T^{-1}. We say that T is **invertible** and we name T^{-1} the inverse of T.*

4.4 One to one mappings

We have already defined a one-to-one mapping as a **function** f with domain in V, such that

$$f(x_1) = f(x_2), \text{ if and only if } x_1 = x_2$$

The following theorem proves that any linear transformation is a one-to-one mapping.

Theorem 4.4.1. *Let $T : V \to W$ be a linear transformation. Then the following are equivalent:*

1. *T is one-to-one (injective)*
2. *T is invertible and T^{-1} is a linear transformation*
3. *For all $x \in V$, if $T(x) = O$, then $x = O$*

Proof. We will prove that (1) $\implies$ (2) $\implies$ (3) $\implies$ (1) (1 implies 2, and 2 implies 3, so 1 implies 3).

- First, if T is one-to-one, by theorem 4.3.1 we can prove (2). So (1) implies (2). If T has an inverse T^{-1}, then it must be one-to-one. So (2) implies (1).
- Suppose that $x \in V$ with $\dim V = n$ and an independent basis $e_1, \ldots, e_n$, we can write x as

  $$x = \sum_{i=1}^{n} c_i e_i, \quad \text{for a set of scalars } c_1, \ldots, c_n$$

  if $x = O$, it implies $c_1 = \cdots = c_n = 0$, because the basis of V is independent. Then, applying T leads to

  $$T(x) = \sum_{i=1}^{n} c_i T(e_i)$$

All the scalars are zero, so $T(x) = O$.

As T is linear, suppose that $T(x) = T(y)$, so $T(x) - T(y) = T(x - y) = O$, this means that $x - y = O$.

The inverse $T^{-1}(x) = T^{-1}(y)$ must by linear by (2), so $T^{-1}(x) - T^{-1}(y) = T^{-1}(x - y) = O$. This proves (3).

As all implications are true, the theorem is proven.

□

Theorem 4.4.2. *Let $T : V \to W$ be a linear transformation, suppose that V is a finite set with* $\dim V = n$. *Then, the following propositions are equivalent:*

1. *T is one-to-one in V.*
2. *If $e_1, \ldots, e_p$ are independent elements of V, then $T(e_1), \ldots, T(e_p)$ are independent elements of $T(V)$.*
3. $\dim T(V) = n$.
4. *If $\{e_1, \ldots, e_n\}$ is a basis for V, $\{T(e_1), \ldots, T(e_n)\}$ is a basis for $T(V)$.*

Proof. If V has independent elements $e_1, \ldots, e_p$, and T is a linear transformation. With the previous theorem we have that

$$T(e_i) - T(e_j) = T(e_i - e_j) = O \implies e_i = e_j \tag{4.10}$$

For any combination of the basis, let $x_i = \sum_{j=1}^{n} c_{ij} e_j$ for a set of scalars $c_{ij}, \ldots, c_{in}$, then

$$T(x_i) = \sum_{j=1}^{n} c_{ij} T(e_j) \tag{4.11}$$

So for two transformations to be equal, we have

$$T(x_k) - T(x_p) = \sum_{j=1}^{n} T(e_j)(c_{kj} - c_{pj}) = 0 \tag{4.12}$$

So $c_{kj} = c_{pj}$ for $j = 1, 2, \ldots, n$.Thus, $x_k = x_p$. This proves (1).

To prove (2), suppose (3) is true, so $T(V)$ has dimension n and $p \leq n$

$$\sum_{i=1}^{p} c_i T(e_p) = 0 \tag{4.13}$$

Such that $T\left(\sum_{i=1}^{p} c_i e_p\right) = 0$, this proves that $c_1 = \cdots = c_p = 0$, so the subset $\{T(e_1), \ldots, T(e_p)\}$ is independent. For $p = n$ this is true, so this proves (4), because any set of independent elements of a linear space is a basis; the second statement of (4) is true because $\{T(e_1), \ldots, T(e_n)\}$ is independent. □

4.5 Matrix representation

Any linear transformation has a matrix representation. But first, we will define how transformations are written, as we need a more clear syntax.

The following theorem defines how a linear transformation maps a linear space to another

Theorem 4.5.1. *Let $e_1, e_2, \dots, e_n$ be a basis for an n-dimensional linear space V. Let $u_1, u_2, \dots, u_n$ be a basis for a linear space W. Then, there exists one and only one linear transformation such that*

$$T(e_k) = u_k \quad \text{for } k = 1, 2, \dots, n \tag{4.14}$$

This transformation applies any element $x \in V$ like

$$\text{if } x = \sum_{k=1}^{n} x_k e_k \text{ then } T(x) = \sum_{k=1}^{n} x_k u_k \tag{4.15}$$

We will ignore the proof for this theorem because is trivial. However, you can prove it by yourself by applying the previous concepts.

To find a matrix representation, we can group the scalars $x_1, x_2, \dots, x_n$ from (4.15) to form a vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{4.16}$$

We have already defined the dot product of two vectors. We can see that applying T on x is the same as

$$T(x) = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \cdot \begin{bmatrix} T(e_1) \\ T(e_2) \\ \vdots \\ T(e_n) \end{bmatrix} \tag{4.17}$$

Using theorem (4.5.1) we have

$$T(x) = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \sum_{i=1}^{n} x_k u_k \tag{4.18}$$

Now, suppose that the elements $u_1, \dots, u_n$ can be expressed as a combination of m independent elements $w_1, \dots, w_m \in W$, then, we have

$$T(e_k) = u_k = \sum_{i=1}^{m} t_{ik} w_i \quad \text{for } k = 1, 2, \dots, n \tag{4.19}$$

So each u_k has a vector representation

$$\begin{bmatrix} t_{1k} & t_{2k} & \cdots & t_{mk} \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}$$

We ignore the basis vector $\begin{bmatrix} w_1 & w_2 & \cdots & w_m \end{bmatrix}$ And just write

$$t_k = \begin{bmatrix} t_{k1} \\ t_{2k} \\ \cdots \\ t_{mk} \end{bmatrix} \tag{4.20}$$

We can represent the whole transformation as an $m \times n$ matrix

$$M = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{bmatrix} \tag{4.21}$$

Each column corresponds to a basis element from W. To apply the transformation we can multiply this matrix by a column vector in V.

$$T(x) = M^T x = \begin{bmatrix} t_{11} & t_{21} & \cdots & t_{n1} \\ t_{12} & t_{22} & \cdots & t_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ t_{1m} & t_{2m} & \cdots & t_{nm} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{4.22}$$

Here T is the transpose operator. In linear algebra we multiply "rows by columns", so, to follow that way of thinking, we transpose the matrix. The transpose operator affects a matrix A of size $m \times n$ in the following way: *if* $A = (a_{ij})_{i,j=1}^{m,n}$, then

$$A^T = A = (a_{ji})_{i,j=1}^{m,n} \tag{4.23}$$

Pay close attention to the change of indexes, from ij to ji.

So we have successfully related matrices with transformation. If you know linear algebra, the notions of kernel and linear span apply equally. We could work on other concepts of matrix algebra, but will skip it. However, I will use matrix representation when necessary.

4.6 Matrix multiplication

In regard to linear transformations, a matrix multiplication is a composition of two transformations.

Definition 4.6.1. *Let $T : U \to V$ and $S : V \to W$ be two linear transformations. Then a transformation $ST : U \to W$ is a composition given by $S[T(x)]$ for $x \in U$.*

Now, we will conveniently define a "function" to get the matrix representation of a transformation.

Definition 4.6.2. *If $T : V \to W$ is any linear transformation, it has a matrix representation given by*

$$m(T) = (t_{ij})_{i,j=1}^{m,n} \tag{4.24}$$

Where $\dim V = n$ *and* $\dim W = m$.

This new definition will be useful for notation purposes, so we don't have to give each matrix a new letter. Now, we want to give definition (4.6.2) a matrix notation. We know that we can apply both transformations the following way

Theorem 4.6.1. *Given two transformations $T : U \to V$ and $S : V \to W$, for linear spaces U, V and W, such that*

$$\dim U = n \quad \dim V = p \quad \dim W = m$$

There is a matrix representation of $ST = S[T(x)]$ for any $x \in U$ given by

$$S[T(x)] = [m(S) \cdot m(T)] \cdot x$$

Proof. If we have a basis for U given by $e_1, e_2, \ldots, e_n$ that can be written in terms of an independent set of elements $u_1, u_2, \ldots, u_p \in V$, then

$$T(e_k) = \sum_{i=1}^{p} t_{ik} u_i \tag{4.25}$$

Then, if the set $u_1, u_2, \ldots, u_p$ is independent, it is a basis for V, if it can be written in terms of independent elements $w_1, w_2, \ldots, w_m \in W$, then

$$S(u_k) = \sum_{i=1}^{m} s_{ik} w_i \tag{4.26}$$

Now, if we calculate ST we have, for each $e_1, e_2, \ldots, e_n$

$$S[T(e_k)] = S\left[\sum_{i=1}^{p} t_{ik} u_i\right] = \sum_{i=1}^{p} t_{ik} S(u_i) \tag{4.27}$$

Applying (4.26) we are left with

$$S[T(e_k)] = \sum_{i=1}^{p} t_{ik} \left(\sum_{j=1}^{m} s_{ji} w_j\right) \tag{4.28}$$

The matrix representation for this transformation is, for each $j = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, n$

$$\begin{bmatrix} s_{j1} & s_{j2} & \cdots & s_{jp} \end{bmatrix} \begin{bmatrix} t_{1k} \\ t_{2k} \\ \vdots \\ t_{pk} \end{bmatrix}$$

If we want to represent the whole sum in terms of matrices, we would write

$$\left(\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \end{bmatrix} + \cdots + \begin{bmatrix} s_{m1} & s_{m2} & \cdots & s_{mp} \end{bmatrix} \right) \cdot \begin{bmatrix} t_{1k} \\ t_{2k} \\ \cdots \\ t_{pk} \end{bmatrix} \tag{4.29}$$

For any $x \in U$, applying $S\left[T(x)\right]$ results in

$$S\left[T(x)\right] = S\left[T\left(\sum_{j=1}^{n} x_j e_j\right)\right] = \sum_{j=1}^{n} x_j S\left[T(e_j)\right] \tag{4.30}$$

The result (4.28) is useful now

$$S\left[T(x)\right] = \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{p} t_{ij} S(u_i)\right) = \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{p} t_{ij} \left\{\sum_{k=1}^{m} s_{ki} w_k\right\}\right) \tag{4.31}$$

Reordering the sums, we have

$$S\left[T(x)\right] = \sum_{i=1}^{p} \sum_{j=1}^{n} \sum_{k=1}^{m} x_j t_{ij} s_{ki} w_k \tag{4.32}$$

This is super confusing! Although it seems complicated, it is not at all. If we let k fixed, we can represent the sum in matrix form as

$$\begin{bmatrix} s_{k1} & s_{k2} & \cdots & s_{kp} \end{bmatrix} \left(\begin{bmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{p1} \end{bmatrix} + \cdots + \begin{bmatrix} t_{1n} \\ t_{2n} \\ \vdots \\ t_{pn} \end{bmatrix} \right) \tag{4.33}$$

Now, if we iterate for $k = 1, 2, \ldots, m$, we would have a matrix product, so

$$S\left[T(x)\right] = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mp} \end{bmatrix} \cdot \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{p1} & t_{p2} & \cdots & t_{pn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{4.34}$$

Which is equal to writing $S\left[T(x)\right] = \left[m(S) \cdot m(T)\right] \cdot x$. □

The result of $m(S) \cdot m(T)$ is an $m \times n$ matrix, because $m(S)$ is of dimension $m \times p$ and $m(T)$ is a $p \times n$ matrix.

4.6.1 Matrix multiplication definition

Before defining matrix multiplication, let's state the following notation to make a clear interpretation of each component of the matrix

Definition 4.6.3. *Given a matrix A, we denote by $A_{j=k}$ the k^{th} column of A, and by $A_{i=k}$ the k^{th} row of A.*

Now, we define matrix multiplication

Definition 4.6.4 (Matrix multiplication)**.** *Given two linear transformations $T : U \to V$ and $S : V \to W$ such that*

$$\dim U = n \quad \dim V = p \quad \dim W = m$$

We let matrix multiplication be a function

$$\lambda : \mathbb{R}^{m\times p} \times \mathbb{R}^{p\times n} \to \mathbb{R}^{m\times n}$$

Which is defined as

$$m(S) \cdot m(T) = m(ST) = \begin{bmatrix} \sum_{k=1}^{p} s_{1k}t_{k1} & \cdots & \sum_{k=1}^{p} s_{1k}t_{kn} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{p} s_{mk}t_{k1} & \cdots & \sum_{k=1}^{p} s_{mk}t_{kn} \end{bmatrix}$$

So that

$$m(S) \cdot m(T) = m(ST) = \begin{bmatrix} m(S)_{i=1} \cdot m(T)_{j=1} & \cdots & m(S)_{i=1} \cdot m(T)_{j=n} \\ \vdots & \ddots & \vdots \\ m(S)_{i=m} \cdot m(T)_{j=1} & \cdots & m(S)_{i=m} \cdot m(T)_{j=n} \end{bmatrix}$$

Although this definition is confusing at first, it is useful because it emphasizes the relationship between linear transformation composition and matrix multiplication. However, we now provide a second, more computationally convenient definition.

Definition 4.6.5 (Matrix multiplication, second definition)**.** *If $A = (a_{ij})_{i,j=1}^{p,n}$, $B = (b_{ij})_{i,j=1}^{m,p}$, then*

$$C = B \cdot A$$

defines a matrix $C = (c_{ij})_{i,j=1}^{m,n}$ where

$$c_{ij} = \sum_{k=1}^{p} b_{ik}a_{kj}$$

This makes clear that each entry in the matrix is the dot product between the i^{th} row vector of the left-hand matrix and the j^{th} column vector of the right-hand matrix.

Example 4.6.1. *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ and $S : \mathbb{R}^2 \to \mathbb{R}^2$ be two linear transformations defined by*

$$\begin{cases} T(e_1) = e_1 \cos\left(\dfrac{3\pi}{2}\right) - e_2 \sin\left(\dfrac{3\pi}{2}\right) \\ T(e_2) = e_1 \sin\left(\dfrac{3\pi}{2}\right) + e_2 \cos\left(\dfrac{3\pi}{2}\right) \end{cases}$$

And

$$\begin{cases} S(e_1) = e_1 \cos\left(\dfrac{4\pi}{5}\right) - e_2 \sin\left(\dfrac{4\pi}{5}\right) \\ S(e_2) = e_1 \sin\left(\dfrac{4\pi}{5}\right) + e_2 \cos\left(\dfrac{4\pi}{5}\right) \end{cases}$$

These transformations have a matrix representation given by

$$m(T) = \begin{bmatrix} \cos\left(\dfrac{3\pi}{2}\right) & -\sin\left(\dfrac{3\pi}{2}\right) \\ \sin\left(\dfrac{3\pi}{2}\right) & \cos\left(\dfrac{3\pi}{2}\right) \end{bmatrix}, \quad m(S) = \begin{bmatrix} \cos\left(\dfrac{4\pi}{5}\right) & -\sin\left(\dfrac{4\pi}{5}\right) \\ \sin\left(\dfrac{4\pi}{5}\right) & \cos\left(\dfrac{4\pi}{5}\right) \end{bmatrix}$$

These matrices are known as ***rotation matrices****. When applied on some vector $x = (x_1, x_2) \in \mathbb{R}^2$ they perform a counterclockwise turn on it, changing the angle with respect to the x axis.*

The composite transformation $ST = S[T(x)]$ will perform a counterclockwise rotation of $\dfrac{3\pi}{2} = 270\,\mathrm{deg}$ on the vector, and then a $\dfrac{4\pi}{5} = 144\,\mathrm{deg}$ rotation, also counterclockwise. Resulting in a $54\,\mathrm{deg}$ counterclockwise rotation.

$$S[T(x)] = m(S) \cdot m(T) \cdot x = \begin{bmatrix} \cos\left(\dfrac{3\pi}{2} + \dfrac{4\pi}{5}\right) & -\sin\left(\dfrac{3\pi}{2} + \dfrac{4\pi}{5}\right) \\ \sin\left(\dfrac{3\pi}{2} + \dfrac{4\pi}{5}\right) & \cos\left(\dfrac{3\pi}{2} + \dfrac{4\pi}{5}\right) \end{bmatrix}$$

You can verify this result by using trigonometric identities and the matrix multiplication rule.

Chapter 5

Differentiation

In this chapter we will explore differentiation. However, I will assume the reader has knowledge of one-dimensional differentiation. We'll explore multi-dimensional derivatives in scalar and vector fields.

Definition 5.0.1. *(Scalar field) A scalar field is a mapping* $f : \mathbb{R}^n \to \mathbb{R}$ *for* $n \in \mathbb{N}$.

Definition 5.0.2. *(Vector field) A vector field is a mapping* $f : \mathbb{R}^n \to \mathbb{R}^m$ *for* $n, m \in \mathbb{N}$.

Let's first introduce the concept of "derivative" for a vector field. Let $f : \mathbb{R}^n \to \mathbb{R}^m$, we would need a set of functions $f_1, f_2, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ to describe the derivative in terms of scalar fields. For each f_i, we can form a vector of partial derivatives

$$\nabla f_i = \begin{bmatrix} \frac{\partial f_i}{\partial x_1} & \cdots & \frac{\partial f_i}{\partial x_n} \end{bmatrix}$$

This vector is called the gradient of f. The gradient has various properties that we will not describe or talk about right now. However, it is important to note that it is a vector that **points to the steepest direction** on the scalar field, and always points upwards[1].

In other words, the value of this vector at a certain point p gives the direction and the rate of fastest increase.

Also, we can demonstrate (we will not for now, because we need the chain rule to do so) that the gradient is perpendicular to the level curves of the scalar field. Let $f : \mathbb{R}^2 \to \mathbb{R}$, we can only provide this example in three dimensions, in other way it would be impossible to imagine. A level curve for f will be, for example

$$f(x, y) = k \quad \text{for some } k \in \mathbb{R} \tag{5.1}$$

[1] This has to do with the definition of the derivative, visualize in 3-dimensions a surface and put a tangent line on it. What can you say about the slope of the tangent? We define it is always pointing upwards, so the tangent line has positive slope.

Then, the gradient $\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}$ will satisfy $\nabla f \cdot T = 0$, where T is the tangent vector to the level curve. This vector is obtained by parametrisation on f. We can let $r(t)$ defined in $[a, b]$, and let $f(x, y) = f\left[X(t), Y(t)\right] = f\left[r(t)\right] = g(t)$, by the chain rule we have

$$g'(t) = f'\left[r(t)\right] \cdot r'(t) \tag{5.2}$$

and we know that $r'(t)$ is tangent to the path of $r(t)$. Given that the path is a level curve of f, we would have $g(t) = k$, because $f(x, y) = k$, we get

$$f'\left[r(t)\right] \cdot r'(t) = 0 \tag{5.3}$$

which implies that $f'\left[r(t)\right] \perp r'(t)$. Now, we have to prove that

$$f'\left[r(t)\right] = \nabla f[r(t)] \tag{5.4}$$

5.1 Directional derivative

The derivative in a scalar field depends from the direction that we choose. We define the *directional derivative* of $f(x)$ with respect to the vector a as

$$f'(x; a) = \lim_{h \to 0} \frac{f(x + ha) - f(x)}{h} \tag{5.5}$$

such that, for sufficiently small h we have

$$f(x) \approx f(x + ha) - hf(x; a) \tag{5.6}$$

The directional derivatives in the direction of the basis vectors is a set $S = \{f(x; e_1), f(x; e_2), \ldots, f(x; e_n)\}$. These derivatives are the partial derivatives of f, so

$$f'(x; e_k) = \frac{\partial f}{\partial x_k} \tag{5.7}$$

We can express $f(x; a)$ using linearity

$$f'(x; a) = f'\left(x; \sum_{i=1}^{n} a_i e_i\right) = \sum_{i=1}^{n} a_i f'\left(x; e_i\right) = \sum_{i=1}^{n} a_i \frac{\partial f}{\partial x_k} \tag{5.8}$$

this is equal to the dot product

$$f'(x; a) = \nabla f(x) \cdot a \tag{5.9}$$

We can now rewrite (5.7) as

$$f(x) \approx f(x + ha) - h\nabla f(x) \cdot a \tag{5.10}$$

we can also set a to be a very small vector, call it δx, which leads to

$$f(x) \approx f(x + \delta x) - \nabla f(x) \cdot \delta x \tag{5.11}$$

This is called a *first order Taylor's approximation.*

$$f(x) \approx f(x + \delta x) - \sum_{k=1}^{n} \frac{\partial f}{\partial x_k} \cdot \delta x \tag{5.12}$$

5.2 Differentials in vector fields

We have already defined a vector field, let $f : \mathbb{R}^n \to \mathbb{R}^m$. We can represent this transformation as a system. Also, let $x = (x_1, \ldots, x_n)$, and

$$f(x) = \begin{cases} f_1(x_1, \ldots, x_n) \\ f_2(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{cases} \tag{5.13}$$

so we can define $f(x) = [f_1(x), \ldots, f_m(x)]$ In terms of scalar fields, we have

$$\nabla f_i(x) = \left[\frac{\partial f_i}{\partial x_1} \cdots \frac{\partial f_1}{\partial x_n} \right] \tag{5.14}$$

We can use a first order Taylor's approximation to approximate $f(x)$, using (5.11) we have

$$f(x) \approx f(x + \delta x) - \nabla f(x) \cdot \delta x \tag{5.15}$$

We can write the full approximation as a sum of two matrices

$$\begin{aligned} f(x) &\approx \begin{bmatrix} f_1(x + \delta x) \\ f_2(x + \delta x) \\ \vdots \\ f_m(x + \delta x) \end{bmatrix} + \begin{bmatrix} \nabla f_1(x) \cdot \delta x \\ \nabla f_2(x) \cdot \delta x \\ \vdots \\ \nabla f_m(x) \cdot \delta x \end{bmatrix} \\ &= f(x + \delta x) + D[f(x)] \end{aligned} \tag{5.16}$$

where $D[f(x)]$ is called the **Jacobian matrix** of $f(x)$

$$D[f(x)] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{5.17}$$