Métodos Matemáticos de la Física

Oscar Reula

Acknowledgements

First of all, I want to thank Gloria Puente for supporting me throughout the time I wrote these notes, which took quite a bit of our shared hours, generally long nights. Secondly, to Bernardo González Kriegel who converted the first versions of these notes to LaTeX, doing a great job with much enthusiasm. Also to Bob Geroch, with whom I discussed several topics of the notes and from whom I was also inspired through his writings and books, not only in content but also in style. Finally, to several generations of students who stoically assimilated a large amount of the material from these courses in a very short time.

Contents

List of Figures

Preface

These notes, now turned into a book, originated as an attempt to condense in one place a large set of ideas, concepts, and mathematical tools that I consider basic for the understanding and daily work of a physicist today.

It usually happens that if a problem is formulated from a physical need, such as the description of some natural phenomenon, then it is well formulated, in the sense that a reasonable solution to it exists. This rule has generally been very fruitful and has particularly served as a guide to many mathematicians to make their way in unknown areas. But it has also served, particularly to many physicists, to work without worrying too much about formal aspects, whether analytical, algebraic, or geometric, and thus be able to concentrate on physical and/or computational aspects. Although this allows for the rapid development of some research, in the long run, it leads to stagnation because by proceeding in this way, one avoids facing problems that are very rich in terms of conceptualizing the phenomenon to be described. It is important to verify that the formulated problem has a mathematically and physically correct solution.

An example of this was the development, in the middle of the last century, of the modern theory of partial differential equations. Many of these equations arose because they describe physical phenomena: heat transmission, electromagnetic wave propagation, quantum waves, gravitation, etc. One of the first mathematical responses to the development of these areas was the Cauchy-Kowalevski theorem, which tells us that given a partial differential equation (under quite general circumstances), if an analytic function is given as data on a hypersurface (with certain characteristics), then there is a unique solution in a sufficiently small neighborhood of that hypersurface. It took a long time to realize that this theorem was not really relevant from the point of view of physical applications: there were equations admitted by the theorem such that if the data was not analytic, there was no solution! And in many cases, if they existed, they did not depend continuously on the given data, a small variation of the data produced a completely different solution. It was only in the middle of the last century that substantial progress was made on the problem, finding that there were different types of equations, hyperbolic, elliptic, parabolic, etc., that behaved differently and this reflected the different physical processes they simulated. Due to its relative novelty, this very important set of concepts is not part of the set of tools that many active physicists have, nor are they found in the textbooks usually used in undergraduate courses.

Like the previous one, there are many examples, particularly the theory of ordinary differential equations and geometry, without which it is impossible to understand many of the modern theories, such as relativity, elementary particle theories, and many phenomena of solid-state physics. As our understanding of the basic phenomena of nature advances, we realize that the most important tool for their description is geometry. This, among other things, allows us to handle a wide range of processes and theories with little in common with each other with a very reduced set of concepts, thus achieving a synthesis. These syntheses are what allow us to acquire new knowledge, since by adopting them we leave space in our minds to learn new concepts, which are in turn ordered more efficiently within our mental construction of the area.

These notes were originally intended for a four-month course. But in reality, they were more suited for an annual course or two semesters. Then, as more topics were incorporated into them, it became increasingly clear that they should be given in two semesters or one annual course. Basically, one course should contain the first chapters that include notions of topology, vector spaces, linear algebra, ending with the theory of ordinary differential equations. The task is considerably simplified if the students have previously had a good course in linear algebra. The correlation with physics subjects should be such that the course is prior to or concurrent with an advanced mechanics course. Emphasizing in it the fact that ultimately one is solving ordinary differential equations with a certain special structure. Using the concepts of linear algebra to find eigenmodes and the stability of equilibrium points. And finally using geometry to describe, albeit briefly, the underlying symplectic structure.

The second course consists of developing the tools to be able to discuss aspects of the theory of partial differential equations. It should be given before or concurrently with an advanced electromagnetism course, where emphasis should be placed on the type of equations that are solved (elliptic, hyperbolic), and the meaning of their initial or boundary conditions, as appropriate. Also using coherently the concept of distribution, which is far from being an abstract mathematical concept but is actually a concept that naturally appears in physics.

None of the content of these notes is original material, but some ways of presenting it are, for example, some simpler proofs than usual, or the way of integrating each content with the previous ones. Much of the material should be thought of as a first reading or an introduction to the topic and the interested reader should read the cited books, from which I have extracted much material, being these excellent and difficult to surpass.

Chapter 1

Basic Concepts of Topology

1.1 Introduction

The notion of a set, while telling us that certain objects –the elements that make it up– have something in common with each other, does not give us any idea of the *closeness* between these elements, whereas on the other hand, if we consider for example the real numbers, this notion is present. We know, for example, that the number 2 is much closer to 1 than 423 is. The concept of a topology in a set that we will define below tries to capture precisely this notion of closeness which, as we will see, admits many gradations.

Definition: A **topological space** consists of a pair $(X, \mathscr{T})$, where X is a set and $\mathscr{T}$ is a collection of subsets of X satisfying the following conditions:

1. The subsets $\emptyset$ and X of X are in $\mathscr{T}$.

2. Let O_λ, $\lambda \in I$, be a one-parameter family of subsets of X in $\mathscr{T}$, then $\bigcup_I O_\lambda$ is also in $\mathscr{T}$.

3. If O and O' are in $\mathscr{T}$, then $O \cap O'$ is also in $\mathscr{T}$.

The elements of $\mathscr{T}$, subsets of X, are called the **open subsets** of X, the set $\mathscr{T}$ itself is called a **topology** of X. Condition 2) tells us that infinite unions of elements of $\mathscr{T}$ are also in $\mathscr{T}$, while condition 3) tells us that in general only finite intersections remain in $\mathscr{T}$. The following examples illustrate why this asymmetry exists, they also illustrate how giving a topology is essentially giving a notion of closeness between the points of the set in question.

Example: a) Let $\mathscr{T} = \emptyset, X$, that is, the only open subsets of X are the empty subset and the subset X. It is clear that this collection of subsets is a topology, as it satisfies the three required conditions, this topology is called **indiscrete**. We can say that in this topology the points of X are arbitrarily close to each other, since if an open set contains one of them it contains all of them.

Example: b) Let $\mathscr{T} = \mathscr{P}(X)$, the collection of all subsets of X, clearly this collection also satisfies the conditions mentioned above and therefore it is also a topology of X, the so-called **discrete**. We can say that in this one all the points are arbitrarily separated from each other since, for example, given any point of X there is an open set that separates it from all the others, which consists of only the point in question.

Example: c) Let X be the set of real numbers, from now on, $\mathbb{R}$, and let
$\mathscr{T} = O| \, if \; r \in O, \exists \varepsilon > 0 \, such \; that \; if \; |r - r'| < \varepsilon, \; r' \in O$, that is, the collection of open sets in the usual sense. Let's see that this collection satisfies the conditions to be a topology. Clearly $\emptyset \in \mathscr{T}$, (since it has no r), as well as $\mathbb{R}$, (since it contains all r'), and thus condition 1) is satisfied. Let's see the second, let $r \in \bigcup_I O_\lambda$ then $r \in O_\lambda$ for some λ and therefore there will exist $\varepsilon > 0$ such that all r' with $|r - r'| < \varepsilon$ is also in O_λ, and therefore in $\bigcup_I O_\lambda$. Finally, let's see the third, let $r \in O \cap O'$ then r is in O and therefore there will exist $\varepsilon > 0$ such that all r' with $|r - r'| < \varepsilon$ will be in O, as r is also in O' there will exist $\varepsilon' > 0$ such that all r' with $|r - r'| < \varepsilon'$ will be in O'. Let $\varepsilon'' = min\{\varepsilon, \varepsilon'\}$ then all r' with $|r - r'| < \varepsilon''$ will be in O and in O' and therefore in $O \cap O'$, so we conclude that this last set is also in $\mathscr{T}$. $\mathbb{R}$ with this topology is called the **real line**.

Exercise: Find using the previous example an infinite intersection of open sets that is not open.

Example: d) Let $X = \mathbb{R} \times \mathbb{R} \equiv \mathbb{R}^2$, that is, the Cartesian product of $\mathbb{R}$with itself –the set of all pairs (x, y), with $x, y \in \mathbb{R}$– and let
$\mathscr{T} = O| \, if \; (x, y) \in O, \exists \varepsilon > 0 \, such \; that \; if \; |x - x'| + |y - y'| < \varepsilon, \; (x', y') \in O$. From the previous example, it can be seen that this is also a topological space and that this is the topology we usually use in $\mathbb{R}^2$

Definition: A **metric space**, (X, d) is a pair consisting of a set X and a map $d : X \times X \longrightarrow \mathbb{R}$, usually called distance, satisfying the following conditions:

1. $d(x, x') \geq 0, = 0 \Rightarrow x = x'$.
2. $d(x, x') = d(x', x)$.
3. $d(x, x') + d(x', x'') \geq d(x, x'')$.

Exercise: Prove that this space has a topology *induced* by its metric in a similar way to $\mathbb{R}$ in the previous example.

Exercise: See that $d(x, y) = 1$ if $x \neq y$, is a distance. What topology does this distance introduce?

Clearly a metric gives us a notion of closeness between points, as it gives us a numerical value of the distance between them. A topology, by not generally giving us any number, gives us a much vaguer notion of closeness, but still generally interesting.

1.1.1 Terminology

We now give a summary of the usual terminology in this area, which is a direct generalization of the commonly used one.

Definition: We will call the **complement**, O^c, of the subset O of X the subset of all elements of X that are not in O.

Definition: We will say that a subset O of X is **closed** if its complement O^c is open.

Definition: A subset N of X is called a **neighborhood of** $x \in X$ if there is an open set O_x, with $x \in O_x$, contained in N.

Definition: We will call the **interior** of $A \in X$ the subset $Int(A)$ of X formed by the union of all open sets contained in A.

Definition: We will call the **closure** of $A \in X$ the subset $Cl(A)$ of X formed by the intersection of all closed sets containing A.

Definition: We will call the **boundary** of $A \in X$ the subset ∂A of X formed by $Cl(A) - Int(A) \equiv Int(A)^c \cap Cl(A)$.

Exercise: Let (X, d) be a metric vector space, prove that: a) $C_x^1 = x' | d(x, x') \leq 1$ is closed and is a neighborhood of x. b) $N_x^\varepsilon = x' | d(x, x') < \varepsilon$, $\varepsilon > 0$ is also a neighborhood of x. c) $Int(N_x^\varepsilon) = N_x^\varepsilon$ d) $Cl(N_x^\varepsilon) = x' | d(x, x') \leq \varepsilon$ e) $\partial N_x^\varepsilon = x' | d(x, x') = \varepsilon$.

Exercise: Let $(X, \mathscr{T})$ be a topological space and A a subset of X. Prove that: a) A is open if and only if each $x \in A$ has a neighborhood contained in A. b) A is closed if and only if each x in A^c (that is, not belonging to A) has a neighborhood that does not intersect A.

Exercise: Let $(X, \mathscr{T})$ be a topological space, let $A \in X$ and $x \in X$. Prove that: a) $x \in Int(A)$ if and only if x has a neighborhood contained in A. b) $x \in Cl(A)$ if and only if every neighborhood of x intersects A. c) $x \in \partial A$ if and only if every neighborhood of x contains points in A and points in A^c.

1.2 Derived Concepts

In the previous sections, we have seen that the concept of a topology leads us to a generalization of a series of ideas and derived concepts that we handled in $\mathbb{R}^n$, which did not depend on the usual metric used in these spaces (the so-called Euclidean Metric). It is then worth asking if there are still other possible generalizations. In this and the next subsection, we will study two more of them, these in turn open up a vast area of mathematics, which we will not cover in this course but which is very important in what concerns modern physics.

The first of these is continuity.

1.2.1 Continuous Maps

Definition: Let $\varphi : X \to Y$ be a map between two topological spaces. (See box.) We will say that the map φ is **continuous** if given any open set O of Y, $\varphi^{-1}(O)$ is an open set of X.

Definition: A **map** $\phi : X \to Y$ between a set X and another Y is an assignment to *each* element of X of an element of Y.
This generalizes the usual concept of a function, note that the map is defined for every element of X, while in general its **image**, that is, the set $\phi(X) \equiv y \in Y; |; \exists x \in X; y; \phi(x) = y$, is not all of Y. In the case that it is, that is, $\phi(X) = Y$, we will say that the map is **surjective**. On the other hand, if it is fulfilled that $\phi(x) = \phi(\tilde{x}) \Longrightarrow x = \tilde{x}$ we will say that the map is **injective**. In such a case, there exists the inverse map to ϕ between the set $\phi(X) \subset Y$ and X. If the map is also surjective then its inverse is defined on all of Y and in this case, it is denoted by $\phi^{-1} : Y \to X$. It is also useful to consider the sets $\phi^{-1}(O) = x \in X; |; \phi(x) \in O$

Clearly, the previous definition only uses topological concepts. Does it have anything to do with the usual *epsilon-delta* used in $I\!R^n$? The answer is affirmative, as we will see below in our first theorem, but first, let's see some examples.

Example: a) Let X and Y be any sets and let the topology of X be the discrete one. Then any map between X and Y is continuous. Indeed, for any open set O in Y, $\varphi^{-1}(O)$ is some subset in X, but in the discrete topology, every subset of X is an open set.

Example: b) Let X and Y be any sets and let the topology of Y be the indiscrete one. Then any map between X and Y is also continuous. Indeed, the only open sets in Y are $\emptyset$ and Y, but $\varphi^{-1}(\emptyset) = \emptyset$, while $\varphi^{-1}(Y) = X$, but whatever the topology of X, $\emptyset$ and X are open sets.

From the previous examples, it might seem that our definition of continuity is not very interesting, that is because we have taken cases with the *extreme* topologies, in the intermediate topologies is where the definition becomes more useful.

Example: c) Let X and Y be real lines, and let $\varphi(x) = 1$ if $x \geq 0$, $\varphi(x) = -1$ if $x < 0$. This map is not continuous because, for example, $\varphi^{-1}((1/2, 3/2)) = x | x \geq 0$.

Theorem 1.1 *The map $\varphi : X \to Y$ is continuous if and only if it is fulfilled that: given any point $x \in X$ and any neighborhood M of $\varphi(x)$, there exists a neighborhood N of x such that $\varphi(N) \subset M$.*

This second definition is much closer to the intuitive concept of continuity.

Proof: Suppose φ is continuous. Let x be a point of X, and M a neighborhood of $\phi(x)$. Then there exists an open set O in Y contained in M and containing $\phi(x)$. By continuity $N = \varphi^{-1}(O)$ is an open set of X, and as it contains x, a neighborhood of x. It is then fulfilled that $\varphi(N) \subset O \subset M$. Now suppose that given any point $x \in X$ and any neighborhood M of $\varphi(x)$, there exists a neighborhood N of x such that $\varphi(N) \subset M$. Then let O be any open set of Y, we must now show that $\varphi^{-1}(O)$ is an open set of X. Let x be any point of $\varphi^{-1}(O)$, then $\varphi(x) \in O$ and therefore O is a neighborhood of $\varphi(x)$, therefore there exists a neighborhood N of x such that $\varphi(N) \subset O$ and therefore $N \subset \varphi^{-1}(O)$. But then $\varphi^{-1}(O)$ contains a neighborhood of each of its points and therefore it is open.

Exercise: Let $\phi : X \to Y$ and $\psi : Y \to Z$ be continuous maps, prove that $\psi \circ \phi : X \to Z$ is also continuous. (Composition of maps preserves continuity.)

Induced Topology:

Let ϕ be a map between a set X and a topological space $Y, \mathcal{T}$. This map naturally provides, that is, without the help of any other structure, a topology on X, denoted by $\mathcal{T}_\phi$ and called the **topology induced** by ϕ on X. The set of its open sets is given by: $\mathcal{T}_\phi = O \subset X;|;O = \phi^{-1}(Q),;Q \in \mathcal{T}$, that is, O is an open set of X if there exists an open set Q of Y such that $O = \phi^{-1}(Q)$.

Exercise: Prove that this construction really defines a topology.
Not all topologies thus induced are of interest and in general, they depend strongly on the map, as shown by the following example:

Example:
a) Let $X = Y = I\!R$ with the usual topology and let $\phi : I\!R \to I\!R$ be the function $\phi(x) = 17\ \forall;x \in I\!R$. This function is clearly continuous with respect to the topologies of X and Y, those of the real line. However, $\mathcal{T}_\phi$, the topology induced on X by this map is the indiscrete one!
b) Let X and Y be as in a) and let $\phi(x)$ be an invertible map, then $\mathcal{T}_\phi$ coincides with the topology of the real line.

1.2.2 Compactness

The other generalization corresponds to the concept of **Compactness**. For this, we introduce the following definition: Let X be a set, A a subset of it, and $\{A_\lambda\}$ a collection of subsets of X parameterized by a continuous or discrete variable λ. We say that this collection **covers** A if $A \subset \cup_\lambda A_\lambda$.

Definition: We say that A is **compact** if given any collection $\{A_\lambda\}$ of *open sets* that cover it, there exists a *finite* number of these A_λ that also cover it.

Example: a) Let X be an infinite set of points with the discrete topology. Then a covering of X consists, for example, of all the points of it, each considered as a subset of it. But the topology of X is discrete, so this is a covering of open sets and no finite number of them will cover it, therefore X is not compact in this case. Clearly, if X had only a finite number of elements it would always be compact, regardless of its topology.

Example: b) Let X be any set with the indiscrete topology. Then X is compact. The only open sets of this set are $\emptyset$ and X, so any covering has X as one of its members and this alone is enough to cover X.

Thus, we see that this property strongly depends on the topology of the set. The relationship with the intuitive concept of compactness is clear from the following example and exercise.

Example: c) Let X be the real line and $A = (0, 1)$. This subset is not compact because, for example, the following is a covering of open sets of A such that any finite subset of it is not. $A_n = (\frac{1}{n}, \frac{n-1}{n})$

Exercise: Let X be the real line and $A = [0, 1]$. Prove that A is compact.
Proof:

Let $\{A_\lambda\}$ be a covering of $[0, 1]$ and $a \in [0, 1]$ the least upper bound of the $x \in (0, 1]$ such that $[0, x]$ is covered by a finite subcovering. a exists because 0 has an A that covers it. Let A_{λ_0} be an element of the covering such that $a \in A_{\lambda_0}$. Then there exists $b > a$ such that $b \in A_{\lambda_0}$ and b is already covered by a finite subcovering. Thus, a is in a finite subcovering and therefore, if $a \neq 1$ also some elements greater than it. This would constitute a contradiction. ♠

Now let's see the relationship between the two derived concepts of Topology, namely the continuity of maps between topological spaces and compactness. The fact that a map between topological spaces is continuous implies that this map is special, in the sense that *it carries or conveys information about the respective topologies and preserves the topological properties of the sets it associates.* This is seen in the following property, which –as derived from the following example– is very important.

Theorem 1.2 *Let X and Y be two topological spaces and ϕ a continuous map between them. Then if A is a compact subset of X, $\phi(A)$ is a compact subset of Y.*

Proof: Let O_λ be a collection of open sets in Y that cover $\phi(A)$. Then the collection $\phi^{-1}(O_\lambda)$ covers A, but A is compact and therefore there will be a finite subcollection $\phi^{-1}(O_{\tilde{\lambda}})$ of the former that also covers it. Therefore, the finite subcollection $O_{\tilde{\lambda}}$ will also cover $\phi(A)$. Since this is true for any collection of open sets covering $\phi(A)$ we conclude that it is compact.

Example: Let A be compact and let $\phi : A \to I\!R$ be continuous, that is, a map continuous between A and the real line. $\phi(A)$ is then a compact set in the real line and therefore a closed and bounded set, but then this set will have a maximum and a minimum, that is, the map ϕ reaches its maximum and minimum in A.

Finally, another theorem of fundamental importance about compact sets, which shows that they have another property that makes them very interesting. For this, we introduce the following definitions, which also only use topological concepts. A **sequence** in a set X $\{x_n\} = \{x_1, x_2, ...\}$, with $x_n \in X$, is a map from the integers to this set. Given a sequence $\{x_n\}$ in a topological space X, we say that $x \in X$ is a **limit point** of this sequence if given any open set O of X containing x there exists a number N such that for all $n > N$ $x_n \in O$. We say that $x \in X$ is an **accumulation point** of this sequence if given any open set O of X containing x, infinitely many elements of the sequence also belong to O.

Exercise: Find an example of a sequence in some topological space with different limit points.

Theorem 1.3 *Let A be compact. Then every sequence in A has an accumulation point.*

Proof: Suppose, –contrary to the theorem's assertion– that there exists a sequence $\{x_n\}$ without any accumulation point. That is, given any point x of A there exists a neighborhood O_x containing it and a number N_x such that if $n > N_x$ then $x_n \notin O_x$. Since this is valid for any x in A, the collection of sets $\{O_x | x \in A\}$ covers A, but A is compact and therefore there will exist a finite subcollection of these that also covers it. Let N be the maximum among the N_x of this finite subcollection. But then $x_n \notin A$ for all $n > N$ which is absurd.

Exercise: Prove that compact sets in the real line are the closed and bounded ones.

We can now ask the inverse question: If $A \subset X$ is such that every sequence has accumulation points, is it true then that A is compact? An affirmative answer would give us an alternative characterization of compactness, and this is affirmative for the case of the real line. In general, the answer is negative: there are topologies in which every sequence in a set has accumulation points in it, but this set is not compact. However, all the topologies we will see are **second countable** [See box] and in these the answer is affirmative.

In the real line, it is true that if $x \in \mathbb{R}$ is an accumulation point of a sequence $\{x_n\}$ then there exists a **subsequence**, $\{\tilde{x}_n\}$, that is, a restriction of the map defining the sequence to an infinite number of natural numbers, that will have x as a limit point. This is also not true in the generality of topological spaces, but it is if we consider only those that are **first countable** [See box]. All the spaces we will see in this course are.

***Countability of topological spaces.**

Definition: We say that a topological space $\{X, \mathcal{T}\}$ is first countable if for each $p \in X$ there exists a countable collection of open sets $\{O_n\}$ such that every open set containing p also contains at least one of these O_n.

Definition: We say that a topological space $\{X, \mathcal{T}\}$ is second countable if there is a countable collection of open sets such that any open set of X can be expressed as a union of sets from this collection.

Example:
a) X with the indiscrete topology is first countable.
b) X with the discrete topology is first countable. And second countable if and only if its elements are countable.

Exercise: Prove that the real line is first and second countable. Hint: For the first case, use the open sets $O_n = (p - \frac{1}{n}, p + \frac{1}{n})$ and for the second $O_{pqn} = (\frac{p}{q} - \frac{1}{n}, \frac{p}{q} + \frac{1}{n})$

***Separability of topological spaces.**

Definition: A topological space X is Hausdorff if given any pair of points of X, x and y, there exist neighborhoods O_x and O_y such that $O_x \cap O_y = \emptyset$.

Example:
a) X with the indiscrete topology is not Hausdorff.
b) X with the discrete topology is Hausdorff.

Exercise: Find a topology such that the integers are Hausdorff and compact.

Exercise: Prove that if a space is Hausdorff then if a sequence has a limit point, this is unique.

Exercise: Let X be compact, Y Hausdorff, and $\phi : X \to Y$ continuous. Prove that the images of closed sets are closed. Find a counterexample if Y is not Hausdorff.

Bibliography notes: This chapter is essentially a condensed version of chapters 26, 27, 28, 29, and 30 of [?], see also [?], [?], and [?]. Topology is one of the most fascinating branches of mathematics, if you delve deeper you will be captivated! Of particular interest in physics is the notion of Homotopy, a good place to understand these ideas is chapter 34 of [?].

Chapter 2

Linear Algebra

2.1 Vector Spaces

<u>Definition</u>: A **Real Vector Space** consists of three things: – i) A set, V, whose elements will be called **vectors**. ii) A rule that assigns to each pair of vectors, $\boldsymbol{v}$, $\boldsymbol{u}$, a third vector that we will denote by $\boldsymbol{v}+\boldsymbol{u}$ and that we will call their **sum**. iii) A rule that assigns to each vector, $\boldsymbol{v}$ and to each real number a, a vector that we will denote by $a\boldsymbol{v}$ and that we will call the **product** of $\boldsymbol{v}$ with a. – subject to the following conditions:

1.a) For any pair of vectors $\boldsymbol{u}, \boldsymbol{v} \in V$ it holds that,

$$\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u} \tag{2.1}$$

1.b) There exists in V a unique element called **zero** and denoted by $\mathbf{0}$, such that

$$\mathbf{0}+\boldsymbol{v}=\boldsymbol{v}\ \forall \boldsymbol{v}\in V. \tag{2.2}$$

1.c) For any vector $\boldsymbol{u}\in V$ there exists a unique vector in V, denoted $-\boldsymbol{u}$, such that,

$$\boldsymbol{u}+(-\boldsymbol{u})=\mathbf{0} \tag{2.3}$$

2.a) For any pair of real numbers a and a' and any vector $\boldsymbol{v}$ it holds that,

$$a(a'\boldsymbol{v})=(aa')\boldsymbol{v}.$$

2.b) For any vector $\boldsymbol{v}$ it holds that,

$$1\boldsymbol{v}=\boldsymbol{v}.$$

3.a) For any pair of real numbers a and a' and any vector $\boldsymbol{v}$ it holds that,

$$(a+a')\boldsymbol{v}=a\boldsymbol{v}+a'\boldsymbol{v}.$$

3.b) For any real number a and any pair of vectors v and v' it holds that,

$$a(v + v') = av + av'.$$

The first conditions involve only the rule of addition; these are actually the rules that define a group, a structure we will revisit in a later chapter, where addition represents the product between elements of the group. The next conditions involve only the rule of multiplication, while the last two deal with the relationship between these two operations. As we will see with examples later, real numbers can be replaced by any field, such as rationals, Q, integers, Z, complex numbers, C, and even finite fields.

Example: The set of all n-tuples of real numbers with the usual operations of addition and multiplication *tuple by tuple*. This space is denoted by $\mathbb{R}^n$.

Example: Let S be any set and let V be the set of all functions from S to the reals, $v : S \to \mathbb{R}$, with the following operations of addition and multiplication: The sum of the function v with the function v' is the function (element of V) that assigns to the element s of S the value $v(s) + v'(s)$. The product of $a \in \mathbb{R}$ with the function v is the function that assigns to $s \in S$ the value $av(s)$. This example will appear very often in the following chapters.

Definition: We will say that a set of vectors $\{e_i\}$ $i = 1, \ldots, n$ are **linearly independent** if $\sum_i a^i e_i = 0$ $a^i \in \mathbb{R}$, $i = 1, \ldots, n \Longrightarrow a^i = 0$, $i = 1, \ldots, n$, that is, if any non-trivial linear combination of these vectors gives us a non-trivial vector.

Definition: The $Span$ of a set of vectors $\{u_i\}$ is the set of all possible linear combinations of these elements. They generate a subspace of V, which we will denote by $Span\{u_i\} \subset V$.

If a finite number of linearly independent vectors, n, are sufficient to **span** V, [that is, if any vector in V can be obtained as a linear combination of these n vectors], then we will say that these form a **basis** of V and that the **dimension** of V is n, $\dim V = n$.

Exercise: Show that given a vector v and a basis, $\{e_i\}$ $i = 1, \ldots, n$, there exists a unique linear combination of elements of the basis that determines it. That is, if $v = \sum_i v^i e_i$ and $v = \sum_i \tilde{v}^i e_i$, then $v^i = \tilde{v}^i$ for all $i = 1, \ldots, n$.

Exercise: If we have two bases, $\{e_i\}$ and $\{\tilde{e}_i\}$, we can write the elements of one in terms of the other,

$$\tilde{e}_j = \sum_i P_j{}^i e_i, \quad e_i = \sum_l R_i{}^l e_l.$$

See that $R_i{}^l = P^{-1}_i{}^l$.

Exercise: Let S be a finite set, $S = \{s_1, s_2, \ldots, s_n\}$, find a basis of the vector space of all functions from S to $I\!R$. Find the dimension of this space.

Exercise: Show that the dimension of V is unique, that is, it does not depend on the basis used to define it.

Note that if in the previous example S consists of a finite number of elements, then the dimension of V is finite.[1] In the case that S had an infinite number of elements, we would say that the dimension of V would be infinite. In what follows in this chapter, we will only consider finite-dimensional vector spaces.

Let V be a vector space of dimension n and a basis of it, $\{e_i\}$ $i = 1, \ldots, n$. Given an n-tuple of real numbers, $\{c^i\}$ we then have determined an element of V, that is, the vector, $v = \sum_{i=1}^n c^i e_i$. On the other hand, we have seen in a previous exercise that given any vector, it determines a unique n-tuple of real numbers, the elements of v in that basis. We see that we then have an invertible map between V and the space of n-tuples, $I\!R^n$. This map is linear, assigning to the sum of two vectors v and $\tilde{v}$ the n-tuple sum of the respective n-tuples. This map depends on the basis, but it still tells us that finite-dimensional vector spaces do not hold many surprises; they are always copies of $I\!R^n$.

2.1.1 Covectors and Tensors

Let V be a vector space of dimension n. Associated with this vector space, consider the set, $V^* = \{$ the space of linear maps $\omega : V \to I\!R\}$. This is also a vector space, called the **dual space to** V , or **space of covectors**, with addition and multiplication given by:

$$(\omega + \alpha\tau)(v) = \omega(v) + \alpha\tau(v) \quad \forall\, v \in V$$

with $\omega, \tau \in V^*$, $\alpha \in I\!R$.

What is its dimension? Note that if $\{e_i\}$ $i = 1, \ldots, n$ is a basis of V, that is, a set of linearly independent vectors that span V, we can define n elements of V^* (called covectors) by the relation

$$\theta^i(e_j) = \delta^i{}_j. \tag{2.4}$$

That is, we define the action of θ^i on the $\{e_j\}$ as in the equation above and then extend its action to any element of V by writing this element in the basis $\{e_i\}$ and using the fact that the action must be linear.

It can be easily seen that any $\rho \in V^*$ can be obtained as a linear combination of the covectors $\{\theta^j\}$, $j = 1, \ldots, n$ and that these are linearly independent, therefore they form a basis and thus the dimension of V^* is also n.

Exercise: See that V^* is a vector space and that the $\{\theta^i\}$ really form a basis.

[1] That is, a finite number of linearly independent vectors span V.

Exercise: See that if $v = \sum_{i=1}^{n} v^i e_i$ then,

$$v^i = \theta(v).$$

Exercise: Let V be the space of functions from a set with a finite number of elements, n, to the reals. Let a basis be given by:

$$e_i(a) := \begin{cases} 1 & a \; is the i-th element \\ 0 & otherwise \end{cases}$$

Find the corresponding co-basis of its dual space.

Since V and V^* have the same dimension, they are, as vector spaces, the same thing, but since there is no map that identifies them, we have to consider them as different.

What happens if we now take the dual of V^*? Will we get more copies of V? The answer is no, since there is a natural identification between V and its double dual V^{**}.

Indeed, to each $v \in V$ we can associate an element X_v of V^{**}, that is, a linear functional from V^* to $I\!R$, in the following way: $X_v(\omega) := \omega(v) \;\; \forall \;\; \omega \in V^*$. That is, the element X_v of V^{**} associated with $v \in V$ is the one that, when acting on any covector ω, gives the number $\omega(v)$. Note that X_v acts linearly on the elements of V^* and therefore is an element of V^{**}. Are there elements of V^{**} that do not come from some vector in V? The answer is no, since the map $X_v : V \to V^{**}$ is injective $[X_v(\omega) = 0 \;\forall\; \omega \Longrightarrow v = 0]$ and therefore [2] $\dim X_V = \dim V$. On the other hand $\dim V^{**} = \dim V^*$ since V^{**} is the dual of V^* and thus $\dim V = \dim V^* = \dim V^{**}$, which indicates that the map in question is also surjective and therefore invertible. This allows us to identify V and V^{**} and conclude that by dualizing we will not be able to construct more interesting vector spaces. In the case where the dimension of the vector space is not finite, this is no longer true and there are cases – frequently used – where $X_V \subset V^{**}$ strictly.

Exercise: Given a basis in V, $\{e_i\}$ and the corresponding co-basis, $\{\theta^j\}$. Define the co-co-basis in V^{**}, $\{E_i\}$. Find the relationship between the components of a vector of the form X_v in the basis $\{E_i\}$ and those of the vector v in the basis $\{e_i\}$.

Exercise: See that indeed $\dim X_V = \dim V$.

However, nothing prevents us from also considering **multilinear maps** [3] from $\underbrace{V \times V \times \cdots \times V}_{k \text{ times}}$ to $I\!R$, or more generally,

$$\underbrace{V \times \cdots \times V}_{k \text{ times}} \times \underbrace{V^* \times \cdots \times V^*}_{l \text{ times}} \to I\!R.$$

[2] Denoting by X_V the image by $X_{(\cdot)}$ of V.

[3] That is, maps that are separately linear in each of their arguments.

The set of these maps (for each given pair (k,l)) is also a vector space, –with the obvious operations– and its elements are called **tensors of type** $\binom{l}{k}$.

Exercise: What is the dimension of these spaces as a function of the pair $\binom{l}{k}$?
Note: In finite dimensions, it can be shown that any tensor of type $\binom{l}{k}$ can be written as a linear combination of elements of the Cartesian product of k copies of V^* and l copies of V –where we have identified V with V^{**}–. For example, if $\boldsymbol{t}$ is of type $\binom{0}{2}$, –that is, a map that has as arguments two covectors–, then given a basis $\{\boldsymbol{e}_i\}$ of V, and the corresponding basis of V^{**}, $\{\boldsymbol{E}_i\}$ there will be $n\times n$ real numbers t^{ij}, $i=1,\dots,n$ such that

$$\boldsymbol{t}(\boldsymbol{\sigma},\boldsymbol{\omega})=\sum_{i,j=1}^{n} t^{ij}\boldsymbol{E}_i(\boldsymbol{\sigma})\boldsymbol{E}_j(\boldsymbol{\omega})=\sum_{i,j=1}^{n} t^{ij}\boldsymbol{\sigma}(\boldsymbol{e}_i)\boldsymbol{\omega}(\boldsymbol{e}_j), \quad \forall\boldsymbol{\sigma},\ \boldsymbol{\omega}\in V^*. \tag{2.5}$$

But the set of linear combinations of Cartesian products of k copies of V^* and l copies of V is also a vector space, it is called the **outer product** of k copies of V^* and l copies of V and is denoted by

$$\underbrace{V^*\otimes V^*\otimes\cdots\otimes V^*}_{k\text{ times}}\otimes\underbrace{V\otimes\cdots\otimes V}_{l\text{ times}}.$$

Therefore, tensors can also be considered as elements of these outer products.

Example: a) Let $\boldsymbol{t}$ be of type $\binom{0}{2}$, that is, $\boldsymbol{t}\in\ V^*\otimes V^*$. This is a bilinear map from $V\otimes V$ to $I\!R$, $\boldsymbol{t}(\boldsymbol{v},\boldsymbol{u})\in I\!R$. Let $\boldsymbol{t}$ be symmetric $[\boldsymbol{t}(\boldsymbol{v},\boldsymbol{u})=\boldsymbol{t}(\boldsymbol{u},\boldsymbol{v})]$ and non-degenerate $[\ \boldsymbol{t}(\boldsymbol{v},\ )=0\in V^*\Longrightarrow \boldsymbol{v}=0]$. Since $\boldsymbol{t}$ is non-degenerate, it defines an invertible map between V and its dual. Indeed, given $\boldsymbol{v}\in V$, $\boldsymbol{t}(\boldsymbol{v},\cdot)$ is an element of V^*. But if $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$ determine the same element of V^*, that is, if $\boldsymbol{t}(\boldsymbol{v},\boldsymbol{u})=\boldsymbol{t}(\tilde{\boldsymbol{v}},\boldsymbol{u});;\forall;\boldsymbol{u}\in V$ then $\boldsymbol{v}=\tilde{\boldsymbol{v}}$, which can be seen by taking $\boldsymbol{u}=\boldsymbol{v}-\tilde{\boldsymbol{v}}$ and using that $\boldsymbol{t}$ is non-degenerate. Since the dimensions of V and V^* are equal, the map thus defined is invertible.

Example: b) Let $\boldsymbol{\varepsilon}$ be an element of $\binom{0}{n}$ such that

$$\boldsymbol{\varepsilon}(\dots,\underbrace{\boldsymbol{v}}_{i},\dots,\underbrace{\boldsymbol{u}}_{j},\dots)=-\boldsymbol{\varepsilon}(\dots,\underbrace{\boldsymbol{u}}_{i},\dots,\underbrace{\boldsymbol{v}}_{j},\dots) \tag{2.6}$$

for any box i and j, that is, a totally antisymmetric tensor. Let $\boldsymbol{e}i$ be a basis of V and $\varepsilon 123\dots n:=\boldsymbol{\varepsilon}(\boldsymbol{e}_1,\boldsymbol{e}2,\dots,\boldsymbol{e}n)$, then any other component of $\boldsymbol{\varepsilon}$ in this basis will be either zero or $\varepsilon 123\dots n$ or $-\varepsilon 123\dots n$ depending on whether some $\boldsymbol{e}_i$ is repeated or is an even permutation of the above or an odd one. Indeed, for example,

$$\begin{aligned}
\varepsilon_{3124\dots n} &:= \boldsymbol{\varepsilon}(\boldsymbol{e}_3,\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_4,\dots,\boldsymbol{e}n)\\
&= -\boldsymbol{\varepsilon}(\boldsymbol{e}_1,\boldsymbol{e}_3,\boldsymbol{e}2,\boldsymbol{e}_4,\dots,\boldsymbol{e}n)\\
&= \boldsymbol{\varepsilon}(\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}3,\boldsymbol{e}_4,\dots,\boldsymbol{e}n)\\
&= \varepsilon 1234\dots n.
\end{aligned}$$

Therefore, given a basis, a single number, $\varepsilon_{123\ldots n}$, is enough to determine the tensor $\boldsymbol{\varepsilon}$ and given another tensor $\tilde{\boldsymbol{\varepsilon}}$ not identically zero with the properties mentioned above, there will exist a number α such that $\boldsymbol{\varepsilon} = \alpha\tilde{\boldsymbol{\varepsilon}}$. This last equality does not depend on the basis used and tells us that the dimension of the subspace of antisymmetric tensors of type $\binom{0}{n}$ is 1. Knowing one element is enough to generate the entire space by multiplying it by any real number.

Exercise: Let $\boldsymbol{\varepsilon}$ be not identically zero and $\boldsymbol{u}_i$ a set of $n = \dim V$ vectors of V. Show that these form a basis if and only if

$$\boldsymbol{\varepsilon}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n) \neq 0. \tag{2.7}$$

Example: c) Let A be an element of $\binom{1}{1}$,

$$\boldsymbol{u} \in V, ;; \boldsymbol{v}^{\in} V^* \to A(\boldsymbol{u}, \boldsymbol{v}^*) \in I\!R. \tag{2.8}$$

This implies that $A(\boldsymbol{u},\ )$ is also a vector (identifying V with V^{**}, the one that takes a form $\boldsymbol{\omega} \in V^*$ and gives the number $A(\boldsymbol{u}, \boldsymbol{\omega})$). Thus we have a **linear map** from V to V, that is, a linear operator on V.

Exercise: Let $\boldsymbol{u}_i$ be a basis of V and let $\boldsymbol{a}_i$ be the vectors $A(\boldsymbol{u}_i, ;;)$, then

$$\boldsymbol{\varepsilon}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = \boldsymbol{\varepsilon}(A(\boldsymbol{u}_1,\ ), \ldots, A(\boldsymbol{u}_n, ;;))$$

is totally antisymmetric in the u_i and therefore proportional to itself;

$$\boldsymbol{\varepsilon}(A(\boldsymbol{u}_1,\ ), \ldots, A(\boldsymbol{u}_n,\ )) \propto \boldsymbol{\varepsilon}(\boldsymbol{u}_1 \ldots, \boldsymbol{u}_n).$$

The proportionality constant is called the **determinant** of the operator A,

$$\boldsymbol{\varepsilon}(A(\boldsymbol{u}_1, ;;), \ldots, A(\boldsymbol{u}_n, ;;)) =: \det(A)\boldsymbol{\varepsilon}(\boldsymbol{u}_1 \ldots, \boldsymbol{u}_n).$$

Problem 2.1 *Show that this definition does not depend on the $\boldsymbol{\varepsilon}$ used nor on the basis and therefore is truly a function of the space of operators from V to $I\!R$.*

Exercise: If $\boldsymbol{A}$ and $\boldsymbol{B}$ are two operators on V, then $\boldsymbol{A} \cdot \boldsymbol{B}(\boldsymbol{v}) := A(\boldsymbol{B}(\boldsymbol{v}))$. Show that $\det(\boldsymbol{AB}) = \det(A) \cdot \det(\boldsymbol{B})$.

2.1.2 Complexification

Another way to obtain vector fields from a given one, say V, is by extending the field where the multiplication operation is defined, if this is possible. The most common case is the **complexification** of a real vector space; in this case, the product is simply extended to complex numbers, resulting in a vector field of the same dimension. One way to obtain it, for example, is by taking a set of linearly independent vectors from the initial space, that is, a basis, and considering all linear combinations with arbitrary complex coefficients. The space thus obtained is denoted by V^C. If the components of the vectors in V in the original basis were n-tuples of real numbers, they are now n-tuples of complex numbers. Since the basis is the same, the dimension is also the same. These extensions of vector spaces often appear, and we will see others later.

Multilinear maps must be extended in the same way, that is, for example, the dual of V will consist of all maps from V to C.

We can also take smaller fields, for example, Q^n or Z^n.

Example: Consider the vector space Q^n consisting of all n-tuples of rational numbers. In this space, the field is also the set of rationals. If we extend the field to the reals, we obtain $\mathbb{R}^n$.

Example: Consider a space V and any basis in it, $\boldsymbol{e}_i$. This choice characterizes a subspace of V, given by all elements of the form,

$$v = \sum_i^n m^i \boldsymbol{e}_i \quad m^i \in Z.$$

Exercise: Now consider the set of all linear maps from this subspace to Z. What form do their elements take?

2.1.3 Quotient Spaces

The last way we will see to obtain vector spaces from other vector spaces is by taking **quotients**. Let V be a vector space and let $W \subset V$ be a subspace of it. We will call the **quotient space** the set of equivalent classes in V, where we will say that two vectors in V are equivalent if their difference is a vector in W. This space is denoted as V/W.

Exercise: Prove that this is an equivalence relation. [At the end of the chapter there is a box with a discussion of the relevant definitions and properties of the central concept of equivalence relations.]

Let's see that this is a vector space. The elements of V/W are equivalent classes; two elements of V, v and v', belong to the same equivalent class if $v - v' \in W$. Let ζ and ζ' be two elements of V/W, that is, two equivalent classes of elements of V. We will define the operations proper to vector spaces as follows: $\zeta + \alpha\zeta'$ will be the equivalent class corresponding to an element $\tilde{v}$ obtained by taking an element of V in ζ, say v, another in ζ', say v', and defining $\tilde{v} := v + \alpha v'$, we have $\tilde{\zeta} = \zeta + \alpha\zeta'$, where $\tilde{\zeta}$ is the equivalent class containing the element $\tilde{v} = v + \alpha v'$. To facilitate notation, the equivalent class containing a given element, say v, is usually represented as v. In this case, we have,

$$v + \alpha v' = v + \alpha v'.$$

Exercise: See that this definition does not depend on the choice of elements in the equivalent class taken to perform the operation. That is, consider two other elements in ζ and ζ', say $\hat{v}$ and $\hat{v}'$, and see that with them you obtain an element in the same class as $\tilde{v} = v + \alpha v'$.

Example: Let $V = I\!R^2$, that is, the space of 2-tuples of real numbers. Let v be any element. This element generates the one-dimensional space W_v consisting of all vectors of the form αv, for $\alpha \in I\!R$. The quotient space V/W_v is the space composed of lines parallel to v. That is, each line is an element of the quotient space, and there is a notion of addition and scalar multiplication among them.

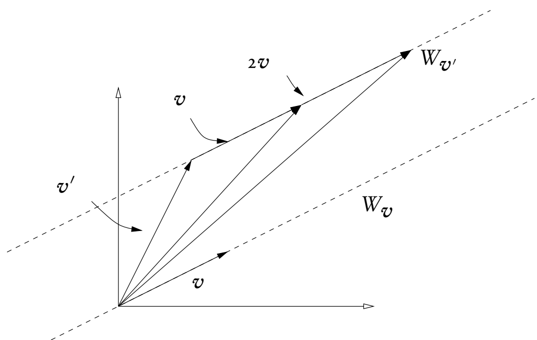

Figure 2.1: Geometric interpretation of the quotient space.

Exercise: Let V be the set of continuous functions from $I\!R$ to $I\!R$ and let W be the subset of those that vanish in the interval $[0,1]$. See that this is a vector subspace. Consider the space V/W. What space can you associate with this?

2.2 Norms

Definition: A **norm** in a vector space V is a map $|x| : V \to I\!R^+$, satisfying for all $x, y \in V$, $\alpha \in I\!R$,,

$i)$ $|x| \geq 0 \quad (= \longleftrightarrow ; x = 0)$

$ii)$ $|\alpha x| = |\alpha| ; |x|$

$iii)$ $|x + y| \leq |x| + |y|$

Examples: In $I\!R^2$:

a) $|(x,y)| := max|x|,|y|;$

b) $|(x,y)|_2 := \sqrt{x^2 + y^2}$, Euclidean norm;

c) $|(x,y)|_1 := |x| + |y|;$

d) $|(x,y)|_p := (|x|^p + |y|^p)^{\frac{1}{p}} ;;;;; p \geq 1;$

e) In V let $\boldsymbol{t}$ be a positive definite symmetric tensor of type $\binom{0}{2}$, that is $\boldsymbol{t}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{t}(\boldsymbol{v},\boldsymbol{u})$, $\boldsymbol{t}(\boldsymbol{u},\boldsymbol{u}) \geq 0\ (= \longleftrightarrow; \boldsymbol{u} = 0)$. The function $|\boldsymbol{u}|_{\boldsymbol{t}} = \sqrt{\boldsymbol{t}(\boldsymbol{u},\boldsymbol{u})}$ is a norm. Each tensor of this type generates a norm, but there are many norms that do not come from any tensor of this type. Give an example.

Exercise: Prove that $|\boldsymbol{t}(\boldsymbol{u},\boldsymbol{v})|^2 \leq ||\boldsymbol{u}||\boldsymbol{t}||\boldsymbol{v}||\boldsymbol{t}$. Hint: Consider the polynomial: $P(\lambda) := \boldsymbol{t}(\boldsymbol{u}+\lambda\boldsymbol{v},\boldsymbol{u}+\lambda\boldsymbol{v})$.

Exercise: Prove that the given examples are indeed norms. Graph the level curves of the first four norms, that is, the sets $S_a = (x,y) \in I\!R^2;/;|(x,y)| =,a$ and the "balls of radius a", that is $B_a = (x,y) \in I\!R^2/|(x,y)| \leq a$.

Exercise: Prove that the map $d : V \times V \to I\!R^+$ given by $d(\boldsymbol{x},\boldsymbol{y}) = |\boldsymbol{x}-\boldsymbol{y}|$ is a metric.

What is a norm geometrically? Given a vector $\boldsymbol{x} \neq 0$ of V and any positive number, a, there is a unique number $\alpha > 0$ such that $|\alpha\boldsymbol{x}| = a$. This indicates that the level surfaces of the norm, that is, the hypersurfaces $S_a = \boldsymbol{x} \in V/|\boldsymbol{x}| = a$, $a > 0$ form a smooth family of layers one inside the other, and each of them divides V into three disjoint sets, the *interior* of S_a –containing the element $\boldsymbol{x} = 0$– , S_a and the *exterior* of S_a. The *interior* of S_a is a convex set, that is, if $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to the interior of S_a, then $\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y},;\alpha \in [0,1]$ also belongs [since $|\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}| \leq \alpha|\boldsymbol{x}| + (1-\alpha)|\boldsymbol{y}| \leq \alpha a + (1-\alpha)a = a$]. A level curve completely characterizes a norm in the sense that if we give a subset N of V, such that N: **a)** has the radial property, that is, given $\boldsymbol{x} \neq 0$ there is a unique $\alpha > 0$ such that $\alpha,\boldsymbol{x} \in N$ and $-\alpha,\boldsymbol{x} \in N$ and **b)** is convex, then there is a unique norm such that N is the level surface S_1. This norm is defined as follows: given $\boldsymbol{x}$ we know that there will be a unique $\alpha > 0$ such that $\alpha\boldsymbol{x} \in N$ and then the norm of $\boldsymbol{x}$ will be $|\boldsymbol{x}| := \frac{1}{\alpha}$.

Exercise: Prove that this is a norm. Hint, given any two vectors $\boldsymbol{x},;;\boldsymbol{y};\in V$, then $\frac{\boldsymbol{x}}{||\boldsymbol{x}||}$ and $\frac{\boldsymbol{y}}{||\boldsymbol{y}||}$ are unitary and therefore are in N. But then we have that $||\lambda\frac{\boldsymbol{x}}{||\boldsymbol{x}||} - (1-\lambda)\frac{\boldsymbol{y}}{||\boldsymbol{y}||}|| \leq 1;;;\forall;\lambda \in [0,1]$. Now find a convenient value for λ.

From this image, we see that given two norms of a finite-dimensional vector space and a level surface of one, there will be level surfaces of the other that will have the first one inside or outside it. In the norms a) and b) of the previous example, we see that given a square containing zero, there will be two circles containing zero, one containing the square and the other contained by it. This leads us to the following theorem.

Theorem 2.1 *Let V be a finite-dimensional vector space. Then all its norms are equivalent to each other, in the sense that given $|\cdot|$ and $|\cdot|'$ there are positive constants M_1 and M_2 such that for all $\boldsymbol{x} \in V$ it holds that $M_1,|\boldsymbol{x}| \leq |\boldsymbol{x}|' \leq M_2|\boldsymbol{x}|$.*

Proof: We will show that all are equivalent to the norm $|\boldsymbol{x}|1 = \sum i = m^n |a^i|$, where the a^i are the components of $\boldsymbol{x}$, with respect to a given basis $\boldsymbol{e}_i$, $\boldsymbol{x} = a^i, \boldsymbol{e}_i$.

Let any other norm, then

$$\begin{aligned} |,|x|,| &\leq |x| = |\sum_i^n a^i, e_i| \leq \sum_i^n |a^i, ei| \\ &\leq \sum_i^n |a^i|, |ei| \leq (max j = 1, n|ej|) \sum i = 1^n |a^i| \\ &= (max j = 1, n|e_j|); |x|_1. \end{aligned} \tag{2.9}$$

And we have easily obtained the upper bound. Now let's see the lower bound. To do this, we must prove that the norm $|\cdot|$ is, as a function from V to $I\!R^+$, a continuous function. This easily follows from the already found bound, indeed, let any other vector, $y = b^i, ei$, then

$$\begin{aligned} |,|x| - ||y|,| &\leq |x - y| \\ &= (max j = 1, n|e_j|); |x - y|_1 \end{aligned} \tag{2.10}$$

This shows that the norm $|\cdot|$ is a continuous function with respect to the norm $|\cdot|_1$. Let S_1 be the level surface of radius 1 with respect to the metric $|\cdot|_1$. S_1 is a closed and bounded set and therefore compact. Therefore, by continuity, $|\cdot|$ has a maximum value, M_2, and a minimum, M_1, which give the sought inequality. The maximum value we have already found, the minimum is what allows us to bound the norm from below and conclude the theorem.

Notes:

i) In this proof, it is crucial that S_1 is compact. If V is infinite-dimensional, this is not the case, and there are many non-equivalent norms.

ii) For our purposes, any norm is sufficient –since if, for example, $f : V \to I\!R$ is continuous with respect to one norm, it is also continuous with respect to any other equivalent to it– and for simplicity, from now on, we will use the Euclidean norm.

iii) In this sense, the norms of finite-dimensional vector spaces are equivalent to the one generated by any positive-symmetric element of the exterior product of its dual with itself.

iv) Since equivalent norms generate the same topology, we see that in finite-dimensional vector spaces, there is a unique topology associated with all its possible norms. This is usually called the **strong topology**.

Exercise: Prove that the above is indeed an equivalence relation.

2.2.1 Induced Norms in $V^\star$

The norms defined in V naturally induce norms in its dual, $V^\star$. This is given by:

$$|\omega| := max_{|v|=1} |\omega(v)|. \tag{2.11}$$

Exercise: Show that this is a norm and that $|\omega(v)| \leq |\omega||v|$.

Exercise: Consider $V = I\!R^2$ with the norm: $|(x,y)| := max|x|, |y|$. What is the induced norm in $V^\star$?

2.3 Linear Operator Theory

A **linear operator** A on a vector space V is a continuous map [4] from V to V such that $\forall\, x, y \in V$, $\alpha \in \mathbb{R}$, $A(\alpha x + y) = \alpha A(x) + A(y)$, that is, a tensor of type $\binom{1}{1}$.

The set of linear operators $\mathscr{L}$ is an algebra, that is, a vector space with a product. Indeed, if A, $B \in \mathscr{L}$, $\alpha \in \mathbb{R}$, then $A + \alpha B \in \mathscr{L}$ and also $A \cdot B$ (the operator that sends $x \in V$ to $A(B(x)) \in V$) also belongs to $\mathscr{L}$. Due to this property, we can also define non-linear functions from $\mathscr{L}$ to $\mathbb{R}$ and maps from $\mathscr{L}$ to $\mathscr{L}$. To study the continuity and differentiability of these maps, we introduce a norm in $\mathscr{L}$, the most convenient being the following norm induced by the one used in V,

$$||A||_{\mathscr{L}} = \max_{||x||_V = 1} ||A(x)||_V. \tag{2.12}$$

If V is finite-dimensional (which we will assume from now on), the vector space $\mathscr{L}$ is also finite-dimensional and therefore all its norms are equivalent. The fact that the norm of A is finite again uses that $A : V \to V$ is continuous and that $\{x \in V / ||x||_V = 1\}$ is compact, in the infinite-dimensional case, neither of these things is necessarily true, and within $\mathscr{L}$ we only have a subspace of linear operators with finite (bounded) norms.

Exercise: Show that

$$||A(v)|| \leq ||A||_{\mathscr{L}} ||v||.$$

Exercise: Using the result of the previous exercise, show that

$$||AB||_{\mathscr{L}} \leq ||A||_{\mathscr{L}} ||B||_{\mathscr{L}}.$$

Exercise: Show that $|| \, ||_{\mathscr{L}} : \mathscr{L} \to \mathbb{R}^+$ is a norm.

Next, we study various functions in the space of operators.

The determinant of an operator, introduced in the previous section, is a polynomial of degree $n = \dim V$ in A and therefore differentiable. Using the chain rule, we see that $det(I + \varepsilon A)$ is differentiable in ε, and indeed a polynomial of degree n in ε. Each of the coefficients of this polynomial is a function of A. Of importance in what follows is the linear coefficient in A, which is obtained using the formula

$$\frac{d}{d\varepsilon} det(I + \varepsilon A)|_{\varepsilon=0} = \frac{\varepsilon(A(u_1), u_2, \ldots, u_n) + \cdots + \varepsilon(u_1, \ldots, A(u_n))}{\varepsilon(u_1, \ldots, u_n)} \tag{2.13}$$

this function is called the **trace** of A and is denoted $tr(A)$.

[4] With respect to the topology induced by any of the equivalent norms of V.

Among the maps from $\mathscr{L}$ to $\mathscr{L}$, consider the exponential map, defined as,

$$e^{A} = \sum_{i=0}^{\infty} \frac{A^i}{i!} = I + A + \frac{A^2}{2} + \cdots \tag{2.14}$$

Theorem 2.2 *$e^{A} \in \mathscr{L}$ if $A \in \mathscr{L}$ and e^{tA} is infinitely differentiable with respect to t.*

Proof: Consider the Cauchy sequence $\{e^{A}_n\}$, where $e^{A}_n \equiv \sum_{i=0}^{n} \frac{A^i}{i!}$. This sequence is Cauchy because, taking $m > n$, we have

$$\begin{aligned}
||e^{A}_m - e^{A}_n||_{\mathscr{L}} &= ||\frac{A^m}{(m)!} + \frac{A^{m-1}}{(m-1)!} + \frac{A^{m-2}}{(m-2)!} + \cdots + \frac{A^{n+1}}{(n+1)!}||_{\mathscr{L}} \\
&\leq ||\frac{A^m}{(m)!}||_{\mathscr{L}} + ||\frac{A^{m-1}}{(m-1)!}||_{\mathscr{L}} + ||\frac{A^{m-2}}{(m-2)!}||_{\mathscr{L}} + \cdots + ||\frac{A^{n+1}}{(n+1)!}||_{\mathscr{L}} \\
&\leq \frac{||A||^m_{\mathscr{L}}}{(m)!} + \frac{||A||^{m-1}_{\mathscr{L}}}{(m-1)!} + \frac{||A||^{m-2}_{\mathscr{L}}}{(m-2)!} + \cdots + \frac{||A||^{n+1}_{\mathscr{L}}}{(n+1)!} \\
&= |e_m^{||A||_{\mathscr{L}}} - e_n^{||A||_{\mathscr{L}}}| \to 0.
\end{aligned} \tag{2.15}$$

Where $e_n^{||A||_{\mathscr{L}}} \equiv \sum_{i=0}^{n} \frac{||A||^i_{\mathscr{L}}}{i!}$ and the last implication follows from the fact that the numerical series $e^{||A||_{\mathscr{L}}}$ converges. But by completeness [5] of $\mathscr{L}$, every Cauchy sequence converges to some element of $\mathscr{L}$ that we will call e^{A}. The differentiability of e^{tA} follows from the fact that if a series $\sum_{i=0}^{\infty} f_i(t)$ is convergent and $\sum_{i=0}^{\infty} \frac{df_i}{dt}$ is uniformly convergent, then $\frac{d}{dt}\sum_{i=0}^{\infty} f_i(t) = \sum_{i=0}^{\infty} \frac{d}{dt} f_i(t)$ ♠

Exercise: Show that
a) $e^{(t+s)A} = e^{tA} \cdot e^{sA}$,
b) If A and B commute, that is if $AB = BA$, then $e^{A+B} = e^{A}\, e^{B}$.
c) $det(e^{A}) = e^{tr(A)}$
d) $\frac{d}{dt} e^{tA} = A e^{tA}$.
Hint: For point c) use that e^{A} can also be defined as,

$$e^{A} = \lim_{m \to \infty} (I + \frac{A}{m})^m.$$

[5] Every finite-dimensional real vector space is complete.

2.3.1 Matrix Representation

To describe certain aspects of linear operators, it is convenient to introduce the following matrix representation.

Let $\{\boldsymbol{u}_i\}$, $i = 1, \ldots, n$ be a basis of V, that is, a set of linearly independent vectors of V $[\sum_{i=1}^n c^i \boldsymbol{u}_i = 0 \Longrightarrow c^i = 0]$ that span it [if $\boldsymbol{v} \in V$, there exist numbers $\{v^i\}$, $i = 1, \ldots, n$ such that $\boldsymbol{v} = \sum_{i=1}^n v^i \boldsymbol{u}_i)$]. Applying the operator $\boldsymbol{A}$ to a member of the basis $\boldsymbol{u}_i$ we obtain a vector $\boldsymbol{A}(\boldsymbol{u}_i)$ which in turn can be expanded in the basis, $\boldsymbol{A}(\boldsymbol{u}_i) = \sum_{j=1}^n A^j{}_i \boldsymbol{u}_j$. The matrix thus constructed, $A^j{}_i$, is a representation of the operator $\boldsymbol{A}$ in that basis. In this language, we see that the matrix $A^j{}_i$ transforms the vector of components $\{v^i\}$ into the vector of components $\{A^j{}_i v^i\}$. Given a basis, $\{\boldsymbol{u}_i\}$, and a matrix $A^j{}_i$, we can construct a linear operator as follows: Given the basis, we define its co-basis, that is, a basis in V^* as $\{\theta^i\}$, $i = 1, \ldots, n$, such that $\theta^i(\boldsymbol{u}_j) = \delta^i_j$, then $\boldsymbol{A} = \sum_{i,j=1}^n A^j{}_i \boldsymbol{u}_j \theta^i$.

If we change the basis, the matrices representing the operators will change. Indeed, if we take another basis $\{\hat{\boldsymbol{u}}_i\}$ and write its components with respect to the previous basis as $\hat{\boldsymbol{u}}_i = P^k{}_i \boldsymbol{u}_k$, and therefore, $\hat{\theta}^j = (P^{-1})^j{}_l \theta^l$, then the relationship between the components of the operator $\boldsymbol{A}$ in both bases is given by

$$\hat{A}^j{}_i = \boldsymbol{A}(\hat{\theta}^j, \hat{\boldsymbol{u}}_i) = (P^{-1})^j{}_l A^l{}_k P^k{}_i \quad \text{o} \quad \hat{A} = P^{-1} A P \tag{2.16}$$

that is, $\hat{A}$ and A are **similar matrices**.

<u>Exercise</u>: See, from its definition (equation (??)), that in a basis we have, $trA = \sum_{i=1}^n A^i{}_i$.

<u>Exercise</u>: See with an example in two dimensions that the definition of determinant conforms with the usual one when we use a basis.

2.3.2 Invariant Subspaces

<u>Definition</u>: Let $A : V \to V$ be an operator and let W be a subspace of V. We will say that W is an **invariant subspace** of A if $AW \subseteq W$.

The invariant subspaces of an operator are important because they allow us to understand its action. Note that given any operator $\boldsymbol{A}$, there are always at least two invariant spaces, V and $\{0\}$, and in reality, many more. For example, as we will see later, given any number between 1 and n (=dim V), there exists an invariant subspace with that number as its dimension. The ones that truly encode the action of the operator are its **irreducible** invariant subspaces, that is, those that cannot be further decomposed into invariant subspaces such that their direct sum is the whole V [6]

[6] A vector space is said to be the direct sum of two of its subspaces, W_1 and W_2, and is denoted by

Example: Let V with $\dim(V) = 2$ and $A : V \to V$ given by, $A(\boldsymbol{u}_1) = \lambda_1 \boldsymbol{u}_1$, $A(\boldsymbol{u}_2) = \lambda_2 \boldsymbol{u}_2$, where $(\boldsymbol{u}_1, \boldsymbol{u}_2)$ are linearly independent (and therefore a basis). Note that this completely defines the operator, since given any $\boldsymbol{v} \in V$, we can uniquely write it as $\boldsymbol{v} = v^1 \boldsymbol{u}_1 + v^2 \boldsymbol{u}_2$ and therefore,

$$A(\boldsymbol{v}) = \lambda_1 v^1 \boldsymbol{u}_1 + \lambda_2 v^2 \boldsymbol{u}_2.$$

In this case, the invariant subspaces are $Span\{\boldsymbol{u}_1\}$ and $Span\{\boldsymbol{u}_2\}$, and clearly, we have $V = Span\{\boldsymbol{u}_1\} \oplus Span\{\boldsymbol{u}_2\}$. Since each of these invariant subspaces is one-dimensional, the action of the operator on them is simply a dilation, that is, the multiplication of their elements by a number. Note that in the case where $\lambda_1 = \lambda_2$, the operator is proportional to the identity and therefore we have infinite invariant spaces.

Exercise: Let V be the space from the previous example and let A be given by $A(\boldsymbol{u}_1) = 0$, $A(\boldsymbol{u}_2) = \boldsymbol{u}_1$. Find its irreducible invariant subspaces. Do the same for the operator given by $A(\boldsymbol{u}_1) = \boldsymbol{u}_1$, $A(\boldsymbol{u}_2) = a\boldsymbol{u}_2 + \boldsymbol{u}_1$. What happens when $a = 1$?

We will study in detail the one-dimensional invariant subspaces, note that they are irreducible. To study the invariant spaces, it is convenient to study the invariant subspaces of the operator when its action is extended to V^C, that is, the **complexification** of V.

Let's see that an operator always has at least one one-dimensional invariant subspace (and therefore always has a non-trivial irreducible invariant subspace).

Lemma 2.1 *Given $A : V^C \to V^C$, where V^C is finite-dimensional, there always exists a $\boldsymbol{u} \in V^C$ and a $\lambda \in C$ such that,*

$$(A - \lambda \boldsymbol{I})\boldsymbol{u} = 0 \tag{2.17}$$

Proof:

A solution to this equation consists of a scalar λ, called the **eigenvalue** of the operator A, and a vector $\boldsymbol{u}$, called the **eigenvector** of the operator A. The subspace of V^C given by $\{\alpha \boldsymbol{u} \mid \alpha \in C\}$ is the invariant subspace sought.

It is clear that the system has a solution if and only if $\det(A - \lambda \boldsymbol{I}) = 0$. But this is a polynomial in λ of order equal to the dimension of V and therefore, by the Fundamental Theorem of Algebra, it has at least one solution or root, (generally complex), λ_1, and therefore there will be, associated with it, at least one $\boldsymbol{u}_1$ solution of (**??**) with $\lambda = \lambda_1$ ♠

The need to consider all these solutions is what leads us to treat the problem for complex vector spaces.

Exercise: In the infinite-dimensional case, the theorem is no longer true. Find an example of an operator on the set of infinite tuples without any eigenvector. Find it

$V = W_1 \oplus W_2$ if each element of V can be uniquely written as the sum of two elements, one from each of these subspaces.

by looking for infinite matrices constructed in such a way that they have only one eigenvector and consider the limit to infinite components.

Application: Schur's Triangulation Lemma

Definition: An $n \times n$ matrix $A^j{}_i$ has an upper triangular form if $A^j{}_i = 0 \;\; \forall j > i,\ j, i = 1, \ldots, n$. That is, it is a matrix of the form,

$$A = \begin{pmatrix} A^1{}_1 & A^1{}_2 & \cdots & \cdots & A^1{}_n \\ 0 & A^2{}_2 & \cdots & \cdots & A^2{}_n \\ 0 & 0 & \ddots & \ddots & A^3{}_n \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & A^n{}_n \end{pmatrix}. \tag{2.18}$$

As we will see later, in chapter ??, this is a very convenient form to understand the solutions to systems of ordinary differential equations. And the most attractive thing about it is that any operator has a matrix representation with an upper triangular form! Moreover, if an inner product is present, the basis for this representation can be chosen to be orthonormal.

Lemma 2.2 (Schur) *Let $A : V \to V$ be a linear operator acting on a finite-dimensional complex vector space V, n, and let $(\cdot,\cdot)$ be an inner product on V. Then, there exists an orthonormal basis $\{\boldsymbol{u}_i\}$, $i = 1, \ldots, n$ with respect to which the matrix representation of A is upper triangular.*

Proof: Consider the eigenvalue-eigenvector problem for A,

$$(A - \lambda \boldsymbol{I})(\boldsymbol{u}) = 0. \tag{2.19}$$

As we have already seen, this problem always has at least one non-trivial solution, and therefore we have a pair $(\lambda_1, \boldsymbol{u}_1)$ solution of the problem. We take $\boldsymbol{u}_1$ of unit norm as the first element of the basis to be determined. We then have

$$A^j{}_1 := \theta^j(A(\boldsymbol{u}_1)) = \theta^j(\lambda_1 \boldsymbol{u}_1) = \lambda_1 \delta^j{}_1,$$

which gives us the result for the first column of the matrix.

Now consider the space

$$V_{n-1} = Span\{\boldsymbol{u}_1\}^{\perp} := \{\boldsymbol{u} \in V | (\boldsymbol{u}_1, \boldsymbol{u}) = 0\}$$

and the operator from $V_{n-1} \to V_{n-1}$ given by $A_1 := (\boldsymbol{I} - \boldsymbol{u}_1 \theta^1)A$. Note that as we form an orthonormal basis, we already know the first member of the co-basis, $\theta^1 = (\boldsymbol{u}_1, \cdot)$. The operator $P_1 := \boldsymbol{I} - \boldsymbol{u}_1 \theta^1$ satisfies $P_1(\boldsymbol{u}_1) = 0$, $(P_1(v), \boldsymbol{u}_1) = 0$ and $P_1 \cdot P_1 = P_1$, that is, it is a projection operator in the subspace V_{n-1}.

We then have that $A_1 : V_{n-1} \to V_{n-1}$. Therefore, in this space, we can also pose the eigenvalue-eigenvector equation,

$$(A_1 - \lambda I)u = ((I - u_1\theta^1)A - \lambda I)u = 0. \tag{2.20}$$

We thus obtain a new pair (λ_2, u_2), with $u_2 \in V_{n-1}$, and therefore perpendicular to u_1 and also, $Au_2 = \lambda_2 u_2 + u_1\theta^1(A(u_2))$. Therefore

$$A^j{}_2 = \theta^j(A(u_2)) = \theta^j(\lambda_2 u_2 + u_1\theta^1(A(u_2))) = \lambda_2\delta^j{}_2 + \delta^j{}_1 A^1{}_2.$$

We thus see that with this choice of basis element, the second column of A satisfies the condition of the Lemma. The next step is to consider the subspace,

$$V_{n-2} = Span\{u_1, u_2\}^{\perp} = \{u \in V | (u_1, u) = (u_2, u) = 0\},$$

and there the eigenvalue-eigenvector equation for the operator $(I - u_1\theta^1 - u_2\theta^2)A$. Proceeding in this way, we generate the entire basis ♠

The previous proof used an inner product that is actually exogenous to the property itself. We did it this way because the proof is conceptually simple and useful in the case that we are in the presence of a given inner product. However, the previous theorem can be proven using the notion of quotient space and thus dispensing with the inner product. This proof can be obtained by doing the following exercises:

Exercise: Let $A : V \to V$ be a linear operator, let $W \subset V$ be invariant under A, that is, $A[W] \subset W$. This operator induces an operator in the quotient space V/W as follows: $\hat{A}\zeta = \tilde{\zeta}$ where $\tilde{\zeta}$ is the equivalent class to which $A(u)$ belongs when u belongs to ζ. See that this definition is consistent, that is, if we choose another element $v \in \zeta$, we obtain that $A(v) \in \tilde{\zeta}$.

Exercise: The induced operator will have at least one eigenvalue-eigenvector pair in V/W, that is, there will be a pair $(\hat{\lambda}, \zeta)$ such that $\hat{A}\zeta = \hat{\lambda}\zeta$. What does this equation mean in terms of the operator A in V?

Exercise: Specialize the previous case when W is the subspace of V generated by an eigenvector of A. $Au_1 = \lambda_1 u_1$, $W_1 = Span\{u_1\}$. What does it mean, in terms of the space V and the operator A, that $\hat{A} : V/W_1 \to V/W_1$ has an eigenvalue-eigenvector pair?

Exercise: Iterate the previous procedure, taking at each step an element from each equivalent class of generalized eigenvectors to form a basis where the matrix representation of A is upper triangular.

Continuing now with the study of invariant subspaces. If $\det(A - \lambda I)$ has $1 \leq m \leq n$ distinct roots, $\{\lambda_i\}$, $i = 1, \ldots, m$, then there will be at least one complex eigenvector $\boldsymbol{u}_i$ associated with each of them. Let's see that these form distinct invariant subspaces.

Lemma 2.3 *Let* $\{(\lambda_i, \boldsymbol{u}_i)\}$ $i = 1 \ldots m$ *be a set of eigenvalue-eigenvector pairs. If* $\lambda_i \neq \lambda_j$ $\forall\, i \neq j$, $i, j = 1 \ldots m$, *then these eigenvectors are linearly independent.*

Proof: Suppose by contradiction that they are not, and therefore there exist constants $c^i \in \boldsymbol{C}$, $i = 1, \ldots, m-1$, such that

$$\boldsymbol{u}_m = \sum_{i=1}^{m-1} c^i\, \boldsymbol{u}_i \tag{2.21}$$

Applying $\boldsymbol{A}$ to both sides, we get

$$\boldsymbol{A}\boldsymbol{u}_m = \lambda_m\, \boldsymbol{u}_m = \sum_{i=1}^{m-1} c^i\, \lambda_i\, \boldsymbol{u}_i \tag{2.22}$$

or,

$$0 = \sum_{i=1}^{m-1} c^i\, (\lambda_m - \lambda_i)\, \boldsymbol{u}_i. \tag{2.23}$$

We conclude that $\{\boldsymbol{u}_i\}$ $i = 1, \ldots, m-1$ are linearly dependent. Due to **??** and the hypothesis that the eigenvalues are distinct, at least one of the coefficients must be non-zero, and therefore we can solve for one of the remaining eigenvectors in terms of the others $m-2$. Repeating this procedure $(m-1)$ times, we arrive at the conclusion that $\boldsymbol{u}_1 = 0$, which is a contradiction since, as we have seen, the eigenvector equation always has a non-trivial solution for each distinct eigenvalue ♠

If for each eigenvalue there is more than one eigenvector, then these form a higher-dimensional invariant vector subspace (reducible). Within each of these subspaces, we can take a basis composed of eigenvectors. The previous lemma ensures that all these eigenvectors thus chosen, for all eigenvalues, form a large linearly independent set.

Exercise: Convince yourself that the set of eigenvectors with the same eigenvalue forms a vector subspace.

If a given operator $\boldsymbol{A}$ has all its eigenvalues distinct, then the corresponding eigenvectors are linearly independent and equal in number to the dimension of V, that is, they generate a basis of V^C. In that basis, the matrix representation of $\boldsymbol{A}$ is diagonal, that is, $A^j{}_i = \delta^j{}_i \lambda_i$. Each of its eigenvectors generates an irreducible invariant subspace, and together they generate V^C. In each of them, the operator $\boldsymbol{A}$ acts merely by multiplication by λ_i. Note that the λ_i are generally complex, and therefore such multiplication is actually a rotation plus a dilation. Note that, unlike the basis of

Schur's triangulation lemma, this is not generally orthogonal with respect to any given inner product. [7]

Example: Let V be a vector space with $\dim(V) = 2$ and let A be given by $A(\boldsymbol{e}_1) = \boldsymbol{e}_2$, $A(\boldsymbol{e}_2) = -\boldsymbol{e}_1$, where $(\boldsymbol{e}_1, \boldsymbol{e}_2)$ are any two linearly independent vectors. If we interpret them as two orthonormal vectors, then A is a rotation by $\pi/2$ in the plane.

Now let's calculate the determinant of $\boldsymbol{A} - \lambda \boldsymbol{I}$,

$$\begin{aligned} \det(\boldsymbol{A} - \lambda \boldsymbol{I}) &= \varepsilon((\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{e}_1, (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{e}_2)/\varepsilon(\boldsymbol{e}_1, \boldsymbol{e}_2) \\ &= \varepsilon(\boldsymbol{e}_2 - \lambda \boldsymbol{e}_1, -\boldsymbol{e}_1 - \lambda \boldsymbol{e}_2)/\varepsilon(\boldsymbol{e}_1, \boldsymbol{e}_2) \\ &= 1 + \lambda^2. \end{aligned} \tag{2.24}$$

and therefore the eigenvalues are $\lambda_1 = \imath$, and $\lambda_2 = -\imath$. The eigenvectors are $\boldsymbol{u}_1 = \boldsymbol{e}_1 + \imath \boldsymbol{e}_2$ and $\boldsymbol{u}_2 = \boldsymbol{e}_1 - \imath \boldsymbol{e}_2 = \bar{\boldsymbol{u}}_1$. We see then that the action of $\boldsymbol{A}$ in these subspaces is multiplication by $\pm\imath$ and that both invariant subspaces are genuinely complex. In this new basis, the space V is generated by all linear combinations of the form $z\boldsymbol{u}_1 + \bar{z}\boldsymbol{u}_2$, and the action of $\boldsymbol{A}$ is simply multiplication by $\imath$ of z.

If the multiplicity of any of the roots $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$ is greater than one, there will be fewer eigenvalues than the dimension of the space, and therefore we will not be guaranteed to have enough eigenvectors to form a basis, as we can only guarantee the existence of one for each eigenvalue.

Example: Let V be the set of 2-tuples of real numbers with a generic element (a, b) and let A be given by $A(a, b) = (\lambda a + \epsilon b, \lambda b)$. Taking a basis, $\boldsymbol{e}_1 = (1, 0)$, $\boldsymbol{e}_2 = (0, 1)$, we see that its matrix representation is:

$$\begin{pmatrix} \lambda & \epsilon \\ 0 & \lambda \end{pmatrix} \tag{2.25}$$

We see that this operator has only one eigenvalue with multiplicity 2. But it has only one eigenvector, proportional to $\boldsymbol{e}_1 = (1, 0)$. For future use, note that if we define $\boldsymbol{\Delta} = \boldsymbol{A} - \lambda \boldsymbol{I}$, then $\boldsymbol{e}_1 = \frac{1}{\epsilon}\boldsymbol{\Delta}\boldsymbol{e}_2$, and therefore, in the basis $\{\tilde{\boldsymbol{e}}_1 = \boldsymbol{e}_1, \tilde{\boldsymbol{e}}_2 = \frac{1}{\epsilon}\boldsymbol{e}_2\}$, the operator is represented by the matrix,

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \tag{2.26}$$

We must therefore analyze what happens in these cases. To do this, let us define, given λ_i an eigenvalue of $\boldsymbol{A}$, the following subspaces:

[7] Note, however, that it can be *declared* orthogonal by defining the inner product as $(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i=1}^{n} \theta^i(\bar{\boldsymbol{u}})\theta^i(\boldsymbol{v})$

$$W_{\lambda_i p} = \{\boldsymbol{u} \in V | (\boldsymbol{A} - \lambda_i \boldsymbol{I})^p \boldsymbol{u} = 0\} \tag{2.27}$$

Note that these are invariant spaces: $\boldsymbol{A} W_{\lambda_i p} \subset W_{\lambda_i p}$. Moreover, $W_{\lambda_i p} \subset W_{\lambda_i p+1}$, and therefore, for a sufficiently large p ($p \leq n$), we will have that $W_{\lambda_i p} = W_{\lambda_i p+1}$, taking the minimum among the p's where this occurs, we define $W_{\lambda_i} := W_{\lambda_i p}$. Note that if for some λ_i, $p = 1$, then the subspace W_{λ_i} is composed of eigenvectors. These are the maximum invariant spaces associated with the eigenvalue λ_i, indeed we have:

Lemma 2.4 *The only eigenvalue of A in W_{λ_i} is λ_i.*

Proof: Let λ be an eigenvalue of $\boldsymbol{A}$ in W_{λ_i}. Let's see that $\lambda = \lambda_i$. As we have already seen, there will be an eigenvector $\zeta \in W_{\lambda_i}$ with λ as the eigenvalue. Since it is in W_{λ_i}, there will be some $p \geq 1$ such that $(\boldsymbol{A} - \lambda_i \boldsymbol{I})^p \zeta = 0$, but since it is an eigenvector, we have that $(\lambda - \lambda_i)^p \zeta = 0$, and therefore $\lambda = \lambda_i$ ♠

Now let's see that these subspaces are linearly independent and generate all of V^C. We will prove this theorem again in Chapter ??.

Theorem 2.3 (See Chapter ??, Theorem ??) *Given an operator $A : V \to V$, with eigenvectors $\{\lambda_i\}$, $i = 1...m$, the space V^C admits a direct decomposition into invariant subspaces W_{λ_i}, where in each of them A has only λ_i as an eigenvalue.*

Proof:

The W_{λ_i} are independent. Let $\boldsymbol{v}_1 + \cdots + \boldsymbol{v}_s = 0$, with $\boldsymbol{v}_i \in W_{\lambda_i}$, then we must prove that each $\boldsymbol{v}_i = 0$. Applying $(\boldsymbol{A} - \lambda_2 \boldsymbol{I})^{p_2} \ldots (\boldsymbol{A} - \lambda_s \boldsymbol{I})^{p_s}$ to the previous sum, we get, $(\boldsymbol{A} - \lambda_2 \boldsymbol{I})^{p_2} \ldots (\boldsymbol{A} - \lambda_s \boldsymbol{I})^{p_s} \boldsymbol{v}_1 = 0$, but since λ_i, $i \neq 1$, is not an eigenvalue of $\boldsymbol{A}$ in W_{λ_1}, the operator $(\boldsymbol{A} - \lambda_2 \boldsymbol{I})^{p_2} \ldots (\boldsymbol{A} - \lambda_s \boldsymbol{I})^{p_s}$ is invertible [8] in that subspace, and therefore $\boldsymbol{v}_1 = 0$. Continuing in this way, we see that all the $\boldsymbol{v}_i$ must be zero.

The W_{λ_i} generate all of V^C. Suppose by contradiction that this is not the case, and consider V^C/W, where W is the space generated by all the W_{λ_i}, that is, the space of all linear combinations of elements in the W_{λ_i}. The operator $\boldsymbol{A}$ acts on V^C/W $[\boldsymbol{A}\{\boldsymbol{u}\} = \{\boldsymbol{A}\boldsymbol{u}\}]$, and therefore it has an eigenvalue-eigenvector pair there. This implies that for some element ζ of V^C in some equivalent class of V^C/W, we have:

$$A\zeta = \lambda\zeta + \boldsymbol{u}_1 + \cdots + \boldsymbol{u}_s \tag{2.28}$$

where the $\boldsymbol{u}_i$ belong to each W_{λ_i}. Now suppose that $\lambda \neq \lambda_i \ \forall\, i = 1..s$, then $\boldsymbol{A} - \lambda \boldsymbol{I}$ is invertible in each W_{λ_i}, and therefore there exist vectors $\zeta_i = (\boldsymbol{A} - \lambda \boldsymbol{I})^{-1} \boldsymbol{u}_i \in W_{\lambda_i}$.

[8]Note that $(\boldsymbol{A} - \lambda_i \boldsymbol{I})^s|_{W_{\lambda_j}}$ is invertible if its determinant is non-zero. But $\det(\boldsymbol{A} - \lambda_i \boldsymbol{I})^s = (\det(\boldsymbol{A} - \lambda_i \boldsymbol{I}))^s = (\lambda_j - \lambda_i)^{s \dim(W_{\lambda_j})} \neq 0$

But then $\tilde{\zeta} := \zeta - \zeta_1 - \cdots - \zeta_s$ is an eigenvalue of A! This is impossible since λ is not a root of the characteristic polynomial, nor is $\tilde{\zeta} = 0$, since belonging to ζ to V^C/W is not a linear combination of elements in the W_{λ_i}. We thus have a contradiction. Now suppose that $\lambda = \lambda_j$ for some $j \in \{1..s\}$. We can still define the vectors $\zeta_i = (A - \lambda_j I)^{-1} u_i$ for all $i \neq j$, and $\tilde{\zeta}$, where we only subtract from ζ all the ζ_i with $i \neq j$, therefore we have that

$$(A - \lambda_j I)\tilde{\zeta} = u_i \tag{2.29}$$

But applying $(A - \lambda_j I)^{p_j}$ to this equation, with p_j the minimum value for which $W_{\lambda_j p_j} = W_{\lambda_j p_j + 1}$, we get that $\tilde{\zeta} \in W_{\lambda_j}$, and thus another contradiction, therefore it can only be that V^C/W is the trivial space, and the W_{λ_i} generate all of V^C ♠

We see that we only need to study each of these subspaces W_{λ_i} to find all their irreducible parts (from now on we will suppress the subscript i). But in these subspaces, the operator A acts very simply!

Indeed, let $\Delta_\lambda : W_\lambda \to W_\lambda$ be defined by $\Delta_\lambda := A|_{W_\lambda} - \lambda I|_{W_\lambda}$, then Δ has only 0 as an eigenvalue, and therefore it is **nilpotent** , that is, there exists an integer $m \leq n$ such that $\Delta_\lambda^m = 0$.

Lemma 2.5 *Let $\Delta : W \to W$ be such that its only eigenvalue is 0, then Δ is nilpotent.*

Proof: Let $W^p := \Delta^p[W]$, then we have that $W^p \subseteq W^q$ if $p \geq q$. Indeed, $W^p = \Delta^p[W] = \Delta^q[\Delta^{p-q}[W]] \subset \Delta^q[W]]$. Since the dimension of W is finite, it must happen that for some integer p, we will have that $W^p = W^{p+1}$, we see then that Δ^p acts injectively on W^p and therefore cannot have 0 as an eigenvalue. But we have seen that every operator has some eigenvalue, and therefore we have a contradiction unless $W^p = \{0\}$. That is, $\Delta^p = 0$ ♠

Nilpotent operators have the important property of generating a basis of the space in which they act from their repeated application on a smaller set of linearly independent vectorsContinuing now with the study of invariant subspaces. If $\det(A - \lambda I)$ has $1 \leq m \leq n$ distinct roots, $\{\lambda_i\}$, $i = 1, \ldots, m$, then there will be at least one complex eigenvector u_i associated with each of them. Let's see that these form distinct invariant subspaces.

Lemma 2.6 *Let $\{(\lambda_i, u_i)\}$ $i = 1 \ldots m$ be a set of eigenvalue-eigenvector pairs. If $\lambda_i \neq \lambda_j \quad \forall\, i \neq j$, $i, j = 1 \ldots m$, then these eigenvectors are linearly independent.*

Proof: Suppose by contradiction that they are not, and therefore there exist constants

$c^i \in \mathbf{C}$, $i = 1, \ldots, m-1$, such that

$$\boldsymbol{u}_m = \sum_{i=1}^{m-1} c^i \, \boldsymbol{u}_i \tag{2.30}$$

Applying $\boldsymbol{A}$ to both sides, we get

$$\boldsymbol{A}\boldsymbol{u}_m = \lambda_m \, \boldsymbol{u}_m = \sum_{i=1}^{m-1} c^i \, \lambda_i \, \boldsymbol{u}_i \tag{2.31}$$

or,

$$0 = \sum_{i=1}^{m-1} c^i \, (\lambda_m - \lambda_i) \, \boldsymbol{u}_i. \tag{2.32}$$

We conclude that $\{\boldsymbol{u}_i\}$ $i = 1, \ldots, m-1$ are linearly dependent. Due to ?? and the hypothesis that the eigenvalues are distinct, at least one of the coefficients must be non-zero, and therefore we can solve for one of the remaining eigenvectors in terms of the others $m-2$. Repeating this procedure $(m-1)$ times, we arrive at the conclusion that $\boldsymbol{u}_1 = 0$, which is a contradiction since, as we have seen, the eigenvector equation always has a non-trivial solution for each distinct eigenvalue ♠

If for each eigenvalue there is more than one eigenvector, then these form a higher-dimensional invariant vector subspace (reducible). Within each of these subspaces, we can take a basis composed of eigenvectors. The previous lemma ensures that all these eigenvectors thus chosen, for all eigenvalues, form a large linearly independent set.

Exercise: Convince yourself that the set of eigenvectors with the same eigenvalue forms a vector subspace.

If a given operator $\boldsymbol{A}$ has all its eigenvalues distinct, then the corresponding eigenvectors are linearly independent and equal in number to the dimension of V, that is, they generate a basis of V^C. In that basis, the matrix representation of $\boldsymbol{A}$ is diagonal, that is, $A^j{}_i = \delta^j{}_i \lambda_i$. Each of its eigenvectors generates an irreducible invariant subspace, and together they generate V^C. In each of them, the operator $\boldsymbol{A}$ acts merely by multiplication by λ_i. Note that the λ_i are generally complex, and therefore such multiplication is actually a rotation plus a dilation. Note that, unlike the basis of Schur's triangulation lemma, this is not generally orthogonal with respect to any given inner product. [9]

Example: Let V be a vector space with $\dim(V) = 2$ and let $\boldsymbol{A}$ be given by $\boldsymbol{A}(\boldsymbol{e}_1) = \boldsymbol{e}_2$, $\boldsymbol{A}(\boldsymbol{e}_2) = -\boldsymbol{e}_1$, where $(\boldsymbol{e}_1, \boldsymbol{e}_2)$ are any two linearly independent vectors. If we interpret them as two orthonormal vectors, then A is a rotation by $\pi/2$ in the plane.

[9] Note, however, that it can be *declared* orthogonal by defining the inner product as $(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i=1}^{n} \theta^i(\bar{\boldsymbol{u}})\theta^i(\boldsymbol{v})$

Now let's calculate the determinant of $\boldsymbol{A}-\lambda\boldsymbol{I}$,

$$\begin{aligned}\det(\boldsymbol{A}-\lambda\boldsymbol{I}) &= \varepsilon((\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{e}_1,(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{e}_2)/\varepsilon(\boldsymbol{e}_1,\boldsymbol{e}_2) \\ &= \varepsilon(\boldsymbol{e}_2-\lambda\boldsymbol{e}_1,-\boldsymbol{e}_1-\lambda\boldsymbol{e}_2)/\varepsilon(\boldsymbol{e}_1,\boldsymbol{e}_2) \\ &= 1+\lambda^2. \end{aligned} \tag{2.33}$$

and therefore the eigenvalues are $\lambda_1 = \imath$, and $\lambda_2 = -\imath$. The eigenvectors are $\boldsymbol{u}_1 = \boldsymbol{e}_1 + \imath\boldsymbol{e}_2$ and $\boldsymbol{u}_2 = \boldsymbol{e}_1 - \imath\boldsymbol{e}_2 = \bar{\boldsymbol{u}}_1$. We see then that the action of A in these subspaces is multiplication by $\pm\imath$ and that both invariant subspaces are genuinely complex. In this new basis, the space V is generated by all linear combinations of the form $z\boldsymbol{u}_1 + \bar{z}\boldsymbol{u}_2$, and the action of A is simply multiplication by $\imath$ of z.

If the multiplicity of any of the roots $\det(\boldsymbol{A}-\lambda\boldsymbol{I}) = 0$ is greater than one, there will be fewer eigenvalues than the dimension of the space, and therefore we will not be guaranteed to have enough eigenvectors to form a basis, as we can only guarantee the existence of one for each eigenvalue.

Example: Let V be the set of 2-tuples of real numbers with a generic element (a,b) and let A be given by $A(a,b) = (\lambda a + \epsilon b, \lambda b)$. Taking a basis, $\boldsymbol{e}_1 = (1,0)$, $\boldsymbol{e}_2 = (0,1)$, we see that its matrix representation is:

$$\begin{pmatrix} \lambda & \epsilon \\ 0 & \lambda \end{pmatrix} \tag{2.34}$$

We see that this operator has only one eigenvalue with multiplicity 2. But it has only one eigenvector, proportional to $\boldsymbol{e}_1 = (1,0)$. For future use, note that if we define $\boldsymbol{\Delta} = \boldsymbol{A}-\lambda\boldsymbol{I}$, then $\boldsymbol{e}_1 = \frac{1}{\epsilon}\boldsymbol{\Delta}\boldsymbol{e}_2$, and therefore, in the basis $\{\tilde{\boldsymbol{e}}_1 = \boldsymbol{e}_1, \tilde{\boldsymbol{e}}_2 = \frac{1}{\epsilon}\boldsymbol{e}_2\}$, the operator is represented by the matrix,

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \tag{2.35}$$

We must therefore analyze what happens in these cases. To do this, let us define, given λ_i an eigenvalue of A, the following subspaces:

$$W_{\lambda_i p} = \{\boldsymbol{u} \in V \,|\, (\boldsymbol{A}-\lambda_i\boldsymbol{I})^p\boldsymbol{u} = 0\} \tag{2.36}$$

Note that these are invariant spaces: $\boldsymbol{A}W_{\lambda_i p} \subset W_{\lambda_i p}$. Moreover, $W_{\lambda_i p} \subset W_{\lambda_i p+1}$, and therefore, for a sufficiently large p ($p \leq n$), we will have that $W_{\lambda_i p} = W_{\lambda_i p+1}$, taking the minimum among the p's where this occurs, we define $W_{\lambda_i} := W_{\lambda_i p}$. Note that if for some λ_i, $p = 1$, then the subspace W_{λ_i} is composed of eigenvectors. These are the maximum invariant spaces associated with the eigenvalue λ_i, indeed we have:

Lemma 2.7 *The only eigenvalue of A in W_{λ_i} is λ_i.*

Proof: Let λ be an eigenvalue of A in W_{λ_i}. Let's see that $\lambda = \lambda_i$. As we have already seen, there will be an eigenvector $\zeta \in W_{\lambda_i}$ with λ as the eigenvalue. Since it is in W_{λ_i}, there will be some $p \geq 1$ such that $(\boldsymbol{A} - \lambda_i \boldsymbol{I})^p \zeta = 0$, but since it is an eigenvector, we have that $(\lambda - \lambda_i)^p \zeta = 0$, and therefore $\lambda = \lambda_i$ ♠

Now let's see that these subspaces are linearly independent and generate all of V^C. We will prove this theorem again in Chapter ??.

Theorem 2.4 (See Chapter ??, Theorem ??) *Given an operator $A : V \to V$, with eigenvectors $\{\lambda_i\}$, $i = 1...m$, the space V^C admits a direct decomposition into invariant subspaces W_{λ_i}, where in each of them A has only λ_i as an eigenvalue.*

Proof:

The W_{λ_i} are independent. Let $\boldsymbol{v}_1 + \cdots + \boldsymbol{v}_s = 0$, with $\boldsymbol{v}_i \in W_{\lambda_i}$, then we must prove that each $\boldsymbol{v}_i = 0$. Applying $(\boldsymbol{A} - \lambda_2 \boldsymbol{I})^{p_2} \dots (\boldsymbol{A} - \lambda_s \boldsymbol{I})^{p_s}$ to the previous sum, we get, $(\boldsymbol{A} - \lambda_2 \boldsymbol{I})^{p_2} \dots (\boldsymbol{A} - \lambda_s \boldsymbol{I})^{p_s} \boldsymbol{v}_1 = 0$, but since λ_i, $i \neq 1$, is not an eigenvalue of $\boldsymbol{A}$ in W_{λ_1}, the operator $(\boldsymbol{A} - \lambda_2 \boldsymbol{I})^{p_2} \dots (\boldsymbol{A} - \lambda_s \boldsymbol{I})^{p_s}$ is invertible [10] in that subspace, and therefore $\boldsymbol{v}_1 = 0$. Continuing in this way, we see that all the $\boldsymbol{v}_i$ must be zero.

The W_{λ_i} generate all of V^C. Suppose by contradiction that this is not the case, and consider V^C/W, where W is the space generated by all the W_{λ_i}, that is, the space of all linear combinations of elements in the W_{λ_i}. The operator $\boldsymbol{A}$ acts on V^C/W $[A\{\boldsymbol{u}\} = \{A\boldsymbol{u}\}]$, and therefore it has an eigenvalue-eigenvector pair there. This implies that for some element ζ of V^C in some equivalent class of V^C/W, we have:

$$A\zeta = \lambda\zeta + \boldsymbol{u}_1 + \cdots + \boldsymbol{u}_s \tag{2.37}$$

where the $\boldsymbol{u}_i$ belong to each W_{λ_i}. Now suppose that $\lambda \neq \lambda_i \ \forall\, i = 1..s$, then $\boldsymbol{A} - \lambda \boldsymbol{I}$ is invertible in each W_{λ_i}, and therefore there exist vectors $\zeta_i = (\boldsymbol{A} - \lambda \boldsymbol{I})^{-1} \boldsymbol{u}_i \in W_{\lambda_i}$. But then $\tilde{\zeta} := \zeta - \zeta_1 - \cdots - \zeta_s$ is an eigenvalue of $\boldsymbol{A}$! This is impossible since λ is not a root of the characteristic polynomial, nor is $\tilde{\zeta} = 0$, since belonging to ζ to V^C/W is not a linear combination of elements in the W_{λ_i}. We thus have a contradiction. Now suppose that $\lambda = \lambda_j$ for some $j \in \{1..s\}$. We can still define the vectors $\zeta_i = (\boldsymbol{A} - \lambda_j \boldsymbol{I})^{-1} \boldsymbol{u}_i$ for all $i \neq j$, and $\tilde{\zeta}$, where we only subtract from ζ all the ζ_i with $i \neq j$, therefore we have that

$$(\boldsymbol{A} - \lambda_j \boldsymbol{I})\tilde{\zeta} = \boldsymbol{u}_i \tag{2.38}$$

But applying $(\boldsymbol{A} - \lambda_j \boldsymbol{I})^{p_j}$ to this equation, with p_j the minimum value for which $W_{\lambda_j p_j} = W_{\lambda_j p_j + 1}$, we get that $\tilde{\zeta} \in W_{\lambda_j}$, and thus another contradiction, therefore it can only be that V^C/W is the trivial space, and the W_{λ_i} generate all of V^C ♠

[10] Note that $(\boldsymbol{A} - \lambda_i \boldsymbol{I})^s|_{W_{\lambda_j}}$ is invertible if its determinant is non-zero. But $\det(\boldsymbol{A} - \lambda_i \boldsymbol{I})^s = (\det(\boldsymbol{A} - \lambda_i \boldsymbol{I}))^s = (\lambda_j - \lambda_i)^{s \dim(W_{\lambda_j})} \neq 0$

We see that we only need to study each of these subspaces W_{λ_i} to find all their irreducible parts (from now on we will suppress the subscript i). But in these subspaces, the operator A acts very simply!

Indeed, let $\mathbf{\Delta}_\lambda : W_\lambda \to W_\lambda$ be defined by $\mathbf{\Delta}_\lambda := A|_{W_\lambda} - \lambda I|_{W_\lambda}$, then Δ has only 0 as an eigenvalue, and therefore it is **nilpotent** , that is, there exists an integer $m \leq n$ such that $\mathbf{\Delta}_\lambda^m = 0$.

Lemma 2.8 *Let $\mathbf{\Delta} : W \to W$ be such that its only eigenvalue is 0, then $\mathbf{\Delta}$ is nilpotent.*

Proof: Let $W^p := \mathbf{\Delta}^p[W]$, then we have that $W^p \subseteq W^q$ if $p \geq q$. Indeed, $W^p = \mathbf{\Delta}^p[W] = \mathbf{\Delta}^q[\mathbf{\Delta}^{p-q}[W]] \subset \mathbf{\Delta}^q[W]]$. Since the dimension of W is finite, it must happen that for some integer p, we will have that $W^p = W^{p+1}$, we see then that $\mathbf{\Delta}^p$ acts injectively on W^p and therefore cannot have 0 as an eigenvalue. But we have seen that every operator has some eigenvalue, and therefore we have a contradiction unless $W^p = \{0\}$. That is, $\mathbf{\Delta}^p = \mathbf{0}$ ♠

Nilpotent operators have the important property of generating a basis of the space in which they act by repeatedly applying them to a smaller set of linearly independent vectors.

Lemma 2.9 *Let $\mathbf{\Delta} : W \to W$ be nilpotent, then there exists a basis of W consisting of elements of the form:*

$$\{\{\boldsymbol{v}_1, \mathbf{\Delta}\boldsymbol{v}_1, \ldots, \mathbf{\Delta}^{p_1}\boldsymbol{v}_1\}, \ldots, \{\{\boldsymbol{v}_d, \mathbf{\Delta}\boldsymbol{v}_d, \ldots, \mathbf{\Delta}^{p_d}\boldsymbol{v}_d\}\}$$

where p_i is such that $\mathbf{\Delta}^{p_i+1}\boldsymbol{v}_i = 0$

Note that if $n = \dim W$ then $n = \sum_{i=1}^d p_i$. Each of these sets formed by repeated applications of an operator is called a **cycle** . In this case, the basis is formed by the elements of d cycles. Note that cycles are not necessarily unique entities; indeed, if we have two cycles with the same number of elements, then any linear combination of them will also be a cycle. Note that each cycle contains only one eigenvector.

Example: Consider the matrix,

$$\mathbf{\Delta} := \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.39}$$

its powers are,

$$\mathbf{\Delta}^2 := \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{\Delta}^3 := \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{\Delta}^4 := \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.40}$$

In this case, we have a single cycle, corresponding to the eigenvector $\mathbf{e}_1 = (1, 0, 0, 0)$, which is the vector whose span is $\mathbf{\Delta}^3[\mathbb{R}^4]$, the cycle is,

$$\begin{aligned} \mathbf{e}_4 &= (0, 0, 0, 1) \\ \mathbf{e}_3 &= (0, 0, 1, 0) = \mathbf{\Delta}\mathbf{e}_4 \\ \mathbf{e}_2 &= (0, 1, 0, 0) = \mathbf{\Delta}\mathbf{e}_3 = \mathbf{\Delta}^2\mathbf{e}_4 \\ \mathbf{e}_1 &= (1, 0, 0, 0) = \mathbf{\Delta}\mathbf{e}_2 = \mathbf{\Delta}^2\mathbf{e}_3 = \mathbf{\Delta}^3\mathbf{e}_4. \end{aligned}$$

<u>Example</u>: Consider the matrix,

$$\mathbf{\Delta} := \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.41}$$

its non-trivial powers are,

$$\mathbf{\Delta}^2 := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{\Delta}^3 := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In this case, we have two cycles, corresponding to the two eigenvectors, $\mathbf{e}_1$ and $\mathbf{e}_3$.

<u>Proof</u>: We will prove it by induction on the dimension of W. If $n = 1$, we take any vector to generate the basis, since in this case $\mathbf{\Delta} = 0$. We now assume it is true for any dimension less than n. In particular, since $\mathbf{\Delta}$ has a zero eigenvalue, dim(ker $\mathbf{\Delta}) \geq 1$, and therefore we have that W' $(= \mathbf{\Delta}(W))$ has a dimension less than n, say n', and by the inductive hypothesis, a basis of the form

$$\{\{\mathbf{v}'_1, \mathbf{\Delta}\mathbf{v}'_1, \ldots, \mathbf{\Delta}^{p'_1}\mathbf{v}'_1\}, \ldots, \{\{\mathbf{v}'_{d'}, \mathbf{\Delta}\mathbf{v}'_{d'}, \ldots, \mathbf{\Delta}^{p'_{d'}}\mathbf{v}'_{d'}\}\}.$$

To form a basis of W, we will add to these vectors d' vectors $\mathbf{v}_i$ such that $\mathbf{\Delta}\mathbf{v}_i = \mathbf{v}'_i,\ i = 1, \ldots, d'$. This can always be done since $\mathbf{v}'_i \in W' = \mathbf{\Delta}W$. We thus see that we have increased the set of vectors to

$$\{\{\mathbf{v}_1, \mathbf{\Delta}\mathbf{v}_1, \ldots, \mathbf{\Delta}^{p'_1+1}\mathbf{v}_1\}, \ldots, \{\{\mathbf{v}_{d'}, \mathbf{\Delta}\mathbf{v}_{d'}, \ldots, \mathbf{\Delta}^{p'_{d'}+1}\mathbf{v}_{d'}\}\},$$

that is, we now have $r = \sum_{i=1}^{d'}(p'_i + 1) = n' + d'$ vectors. To obtain a basis, we must then increase this set with $n - n' - d'$ vectors. Note that this number is non-negative; indeed, $n - n' = \dim(\ker \Delta) \geq \dim(\ker \Delta \cap W') = d'$, and it is precisely the dimension of the subspace of $\ker \Delta$ that is not in W'. We then complete the proposed basis for W by incorporating into the already obtained set $n - n' - d'$ vectors $\{z_i\}, i = 1, ..n-n'-d'$ from the null space of Δ that are linearly independent among themselves and with the other elements of $\ker \Delta$ in W and that are also not in W'. We have thus obtained a set of $d = d' + n - n' - d' = n - n'$ cycles. Let's see that the set thus obtained is a basis. Since they are n in number, we only need to see that they are linearly independent. We must then prove that if we have constants $\{C_{i,j}\},\ i = 1..d,\ j = 0..p'_i + 1$ such that

$$0 = \sum_{i=1}^{d}\sum_{j=0}^{p'_i+1} C_{ij}\Delta^{p_i} v_i \tag{2.42}$$

then $C_{ij} = 0$. Applying Δ to this relation, we get,

$$\begin{aligned} 0 &= \Delta\sum_{i=1}^{d}\sum_{j=0}^{p'_i+1} C_{ij}\Delta^{p_i} v_i \\ &= \sum_{i=1}^{d'}\sum_{j=0}^{p'_i} C_{ij}\Delta^{p'_i} v'_i, \end{aligned} \tag{2.43}$$

where we have used that $\Delta^{p'_i+1}v'_i = 0$. But this is the orthogonality relation of the basis of W', and therefore we conclude that $C_{ij} = 0\ \forall i \leq d',\ \ j \leq p'_i$. The initial relation is then reduced to

$$\begin{aligned} 0 &= \sum_{i=1}^{d} C_{i\,p'_i+1}\Delta^{p'_i+1} v_i \\ &= \sum_{i=1}^{d'} C_{i\,p'_i}\Delta^{p'_i} v'_i + \sum_{i=d'+1}^{d} C_{i\,1} z_i, \end{aligned} \tag{2.44}$$

but the members of the first summation are part of the basis of W' and therefore linearly independent among themselves, while those of the second are a set of elements outside W' chosen to be linearly independent among themselves and with those of the first summation, and therefore we conclude that all C_{ij} are zero ♠

Alternative Proof: Alternatively, the previous lemma can be proven constructively. Indeed, if $m + 1$ is the power for which Δ is nullified, we can take the space $W^m = \Delta^m[W]$ and a basis $\{v_i^m\}$ of it. Note that all elements of W^m are eigenvectors of A, and therefore the elements of the basis. Then we consider the space $W^{m-1} = \Delta^{m-1}[W]$. Note that $W^m = \Delta^m[W] = \Delta^{m-1}[\Delta[W]]$ and $\Delta[W] \subset W$, therefore $W^m \subset W^{m-1}$. Since $W^m = \Delta[W^{m-1}]$, for each vector v_i^m of the basis $\{v_i^m\}$ of

W^m there will be a vector $\boldsymbol{v}_i^{m-1}$ such that $\boldsymbol{\Delta}\boldsymbol{v}_i^{m-1} = \boldsymbol{v}_i^m$. Since $W^m \subset W^{m-1}$, the set $\{\boldsymbol{v}_i^m\} \cup \{\boldsymbol{v}_i^{m-1}\}$ is contained in W^{m-1}. Note that $\dim W^m = \dim(\boldsymbol{\Delta}[W^{m-1}]) \leq \dim(\ker \boldsymbol{\Delta} \cap W^{m-1})$, since $W^m \subset W^{m-1}$ and all its elements belong to $\ker \boldsymbol{\Delta} \cap W^{m-1}$.

Adding to the previous set a set $\{\boldsymbol{z}_i\}$ of $\dim(\ker \boldsymbol{\Delta} \cap W^{m-1}) - \dim(\ker \boldsymbol{\Delta} \cap W^m)$ vectors from the null space of $\boldsymbol{\Delta}$ in W^{m-1}, such that they are linearly independent among themselves and with the elements of the basis of W^m, we obtain a set of $\dim W^{m-1}$ vectors. Note that the mentioned choice of elements $\{\boldsymbol{z}_i\}$ can be made since it is merely an extension of the basis $\{\boldsymbol{v}_i^m\}$ of W^m to a basis of $\ker \boldsymbol{\Delta} \cap W^{m-1}$. Now let's prove that they are linearly independent and therefore form a basis of W^{m-1}. To do this, we need to prove that if

$$0 = \sum_i C_i^m \boldsymbol{v}_i^m + \sum_i C_i^{m-1} \boldsymbol{v}_i^{m-1} + \sum_j C_j^z \boldsymbol{z}_j \tag{2.45}$$

then each of the coefficients C_i^m, C_i^{m-1}, C_j must be zero. Multiplying the previous expression by $\boldsymbol{\Delta}$, we get,

$$0 = \sum_i C_i^{m-1} \boldsymbol{\Delta} \boldsymbol{v}_i^{m-1} = \sum_i C_i^{m-1} \boldsymbol{v}_i^m, \tag{2.46}$$

but then the linear independence of the basis of W^m ensures that the $\{C_i^{m-1}\}$ are all zero. We then have,

$$0 = \sum_i C_i^m \boldsymbol{v}_i^m + \sum_j C_j^z \boldsymbol{z}_j. \tag{2.47}$$

But these vectors were chosen to be linearly independent among themselves and therefore all the coefficients in this sum must be zero. We thus see that the set $\{\boldsymbol{v}_i^m\} \cup \{\boldsymbol{v}_i^{m-1}\} \cup \{\boldsymbol{z}_i\}$ forms a cyclic basis of W^{m-1}. Continuing with W^{m-2} and so on, we obtain a cyclic basis for all of W ♠

We thus see that the irreducible invariant subspaces of an operator are constituted by cycles within invariant subspaces associated with a given eigenvector. Each cycle contains a unique eigenvalue of the operator. We will denote the subspaces generated by these cycles (and usually also called cycles) by $C_{\lambda_i}^k$, where the lower index refers to the eigenvalue of the cycle and the upper index indexes the different cycles within each W_{λ_i}.

Exercise: Show that the obtained cycles are irreducible invariant subspaces of A.

2.3.3 Jordan Canonical Form

Definition: *Let $A : V \to V$ be a linear operator. We will say that A is of Jordan type with eigenvalue λ if there exists a basis $\{\boldsymbol{u}_i\}$ of V such that* [11]

[11] In the sense that $A(\boldsymbol{v}) = \lambda \sum_{i=1}^n \boldsymbol{u}_i \theta^i(\boldsymbol{v}) + \sum_{i=2}^n \boldsymbol{u}_i \theta^{i-1}(\boldsymbol{v})\ \forall \boldsymbol{v} \in V$

$$A = \lambda \sum_{i=1}^{n} \boldsymbol{u}_i \theta^i + \sum_{i=2}^{n} \boldsymbol{u}_i \theta^{i-1} \equiv \lambda I + \Delta \tag{2.48}$$

where $\{\theta^i\}$ is the co-basis of the basis $\{\boldsymbol{u}_i\}$.

That is, in this basis, the components of A form a matrix $A^j{}_i$ given by

$$A = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ & & \lambda & 1 & 0 \\ & 0 & & \lambda & 1 \\ & & & & \lambda \end{pmatrix} \tag{2.49}$$

Note that the matrix Δ is n-nilpotent, that is $\Delta^n = 0$.

Not every operator is of Jordan type –find one that is not– but it is clear that the restriction of any operator to one of its irreducible invariant subspaces (cycles) is. This can be seen by conveniently numbering the elements of the basis generated by the cycle.

Exercise: Find such an ordering of the elements of the basis.

Therefore, we can summarize the results found earlier in the following theorem about the matrix representations of any operator acting in a finite-dimensional space.

Theorem 2.5 *de Jordan Let $\boldsymbol{A} : V \to V$ be a linear operator acting in a complex vector space V. Then, there exists a unique decomposition into a direct sum [12] of V into subspaces $C^k_{\lambda_i}$, $V = C^1_{\lambda_1} \oplus \cdots C^{k_i}_{\lambda_1} \oplus \cdots \oplus C^1_{\lambda_d} \oplus \cdots C^{k_d}_{\lambda_d}$, $d \leq n$ such that*

i) The $C^k_{\lambda_i}$ are invariant under the action of $\boldsymbol{A}$, that is, $\boldsymbol{A}\, C^k_{\lambda_i} \subseteq C^k_{\lambda_i}$

ii) The $C^k_{\lambda_i}$ are irreducible, that is, there are no invariant subspaces of $C^k_{\lambda_i}$ such that their sums are the entire $C^k_{\lambda_i}$.

iii) Due to property i), the operator $\boldsymbol{A}$ induces in each $C^k_{\lambda_i}$ an operator $\boldsymbol{A}_i : C^k_{\lambda_i} \to C^k_{\lambda_i}$, which is of Jordan type with λ_i being one of the roots of the polynomial of degree n_i,

$$det(\boldsymbol{A}_i - \lambda_i \boldsymbol{I}) = 0. \tag{2.50}$$

This theorem tells us that given A, there exists a basis, generally complex, such that the matrix of its components has the form of square diagonal blocks of $n_i \times n_i$, where n_i is the dimension of the subspace $C^k_{\lambda_i}$, each with the form given in **??**. This form of the matrix is called the **Jordan canonical form**.

[12] Recall that a vector space V is said to be the direct sum of two vector spaces W and Z, and we denote it as $V = W \oplus Z$ if each vector in V can be obtained in a unique way as the sum of an element in W and another in Z.

Exercise: Show that the roots, λ_i, that appear in the operators A_i are invariant under similarity transformations, that is, $\lambda_i(A) = \lambda_i(PAP^{-1})$.

Example: Let $A : C^3 \to C^3$, then $det(A - \lambda I)$ is a polynomial of degree 3, and therefore has three roots. If these are distinct, there will be at least three invariant and irreducible subspaces of C^3, but $dim C^3 = 3$ and therefore each of them has $n_i = 1$. The Jordan canonical form is then,
If two of them coincide, we have two possibilities: either we have three subspaces, in which case the Jordan canonical form will be

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \tag{2.51}$$

or, we have two subspaces, one necessarily of dimension 2, the Jordan canonical form will be

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \tag{2.52}$$

If all three roots coincide, then there will be three possibilities,

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \; and \; \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \tag{2.53}$$

Example: We will now illustrate the case of coincident eigenvalues in two dimensions. This case is not generic, in the sense that any perturbation in the system –that is, any minimal change in the equations– separates the roots, making them distinct.

Let $A : C^2 \to C^2$. In this case, the characteristic polynomial $det(A - \lambda I)$ has only two roots, which we will assume are coincident, $\lambda_1 = \lambda_2 = \lambda$. We find ourselves with two possibilities: either there are two linearly independent eigenvectors u_1 and u_2, in which case $V = B_1 \oplus B_2$ and A is diagonalizable ($A = diag(\lambda, \lambda) = I\lambda$), or there is only one eigenvector $\tilde{u}_1$. In this case, let $\tilde{u}_2$ be any vector linearly independent of $\tilde{u}_1$, then $A\tilde{u}_2 = c^1\tilde{u}_1 + c^2\tilde{u}_2$ for some scalars c^1 and c^2 in C. Calculating the determinant of $A - \tilde{\lambda} I$ in this basis, we get $(\lambda - \tilde{\lambda})(c^2 - \tilde{\lambda})$, but λ is a double root and therefore $c^2 = \lambda$.

Reordering and rescaling the bases $u_1 = \tilde{u}_1$, $u_2 = c^1\tilde{u}_2$, we obtain

$$\begin{aligned} Au_1 &= \lambda u_1 \\ Au_2 &= \lambda u_2 + u_1, \end{aligned} \tag{2.54}$$

and therefore

$$A = \lambda(u_1 \oplus \theta^1 + u_2 \oplus \theta^2) + u_1 \oplus \theta^2, \tag{2.55}$$

where $\{\theta^i\}$ is the co-basis of the basis $\{\boldsymbol{u}_i\}$.

Note that $(\boldsymbol{A}-\lambda \boldsymbol{I})\boldsymbol{u}_2 = \boldsymbol{u}_1$ and $(\boldsymbol{A}-\lambda \boldsymbol{I})\boldsymbol{u}_1 = 0$, that is, $\Delta^2 = (\boldsymbol{A}-\lambda \boldsymbol{I})^2 = 0$.

As we will see later in physical applications, the invariant subspaces have a clear physical meaning, they are called normal modes –one-dimensional case– and cycles –in other cases.

2.3.4 Similarity Relation

In physics applications, the following equivalence relation is common: [See box at the end of the chapter.] We will say that the operator $\boldsymbol{A}$ is similar to the operator $\boldsymbol{B}$ if there exists another operator $\boldsymbol{P}$, invertible, such that

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{B}\boldsymbol{P}^{-1}. \tag{2.56}$$

That is, if we *rotate* V with an invertible operator $\boldsymbol{P}$ and then apply $\boldsymbol{A}$, we obtain the same action as if we first apply $\boldsymbol{B}$ and then *rotate* with $\boldsymbol{P}$.

<u>Exercise</u>:
a) Prove that similarity is an equivalence relation.
b) Prove that the functions and maps defined above are the same in the different equivalent classes, that is,

$$\begin{array}{rcl} det(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}) & = & det(\boldsymbol{A}) \\ tr(\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}) & = & tr(\boldsymbol{A}) \\ e^{\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}} & = & \boldsymbol{P}e^{\boldsymbol{A}}\boldsymbol{P}^{-1}. \end{array} \tag{2.57}$$

2.4 Adjoint Operators

Let A be a linear operator between two vector spaces, $A : V \to W$, that is, $A \in \mathcal{L}(V, W)$. Since V and W have dual spaces, this operator naturally induces a linear operator from W' to V', called its **dual** ,

$$A'(\omega)(v) := \omega(A(v)) \;\; \forall \;\, \omega \in W', \;\; v \in V. \tag{2.58}$$

That is, the operator that when applied to an element $\omega \in W'$ gives us the element $A'(\omega)$ of V' which, when acting on $v \in V$, gives the number $\omega(A(v))$. See figure.

$$\begin{tikzcd} & & \omega(A(v)) \\ V \arrow[r, "A"] \arrow[d] & W & \\ V' & W' \arrow[l, "A'"'] \arrow[u] & \\ A'(\omega)(v) & & \end{tikzcd}$$

Figure 2.2: Diagram of the dual operator.

Note that this is a linear operator since,

$$\begin{aligned} A'(\alpha\omega + \sigma)(v) &= (\alpha\omega + \sigma)(A(v)) \\ &= \alpha\omega(A(v)) + \sigma(A(v)) \\ &= \alpha A'(\omega)(v) + A'(\sigma)(v). \end{aligned} \tag{2.59}$$

In the matrix representation, this operator is represented merely by the same matrix as the original, but now acting on the left, $A'(w)_i = w_j A^j{}_i$, that is, $A'^{\,j}_i = A^j{}_i$.

If there are norms defined in V and W and we define the norm of $A : V \to W$ in the usual way,

$$||A|| := \sup_{||v||_V = 1} \{||A(v)||_W\} \tag{2.60}$$

Then we see that

$$
\begin{aligned}
||A'|| &:= \sup_{||\omega||_{W'}=1} \{||A'(\omega)||_{V'}\} \\
&= \sup_{||\omega||_{W'}=1} \{ \sup_{||v||_V=1} \{|A'(\omega)(v)|\}\} \\
&= \sup_{||\omega||_{W'}=1} \{ \sup_{||v||_V=1} \{|\omega(A(v))|\}\} \\
&\leq \sup_{||\omega||_{W'}=1} \{ \sup_{||v||_V=1} \{||\omega||_{W'}||(A(v))||_W\}\} \\
&= \sup_{||v||_V=1} \{||(A(v))||_W\} \\
&= ||A||.
\end{aligned} \tag{2.61}
$$

Thus we see that if an operator is bounded, then its dual is also bounded. In fact, the equality of the norms can be proven, but this would require introducing new tools (in the most general case, the Hahn-Banach theorem) that we do not wish to incorporate in this text.

Let's see what the components of the dual of an operator are in terms of the components of the original operator. Let $\{e_i\}$, $\{\theta^i\}$, $i = 1,..n$ be a basis and respectively a co-basis of V, and let $\{\hat{e}_i\}$, $\{\hat{\theta}^i\}$, $i = 1,..m$ be a pair of basis and co-basis of W. We then have that the components of A with respect to these bases are: $A^i{}_j := \hat{\theta}^i(A(e_j))$, that is, $A(v) = \sum_{i=1}^m \sum_{j=1}^n A^i{}_j \hat{e}_i \theta^j(v)$.

Therefore, if v has components $(v^1, v^2, \ldots, v^n)$ in the basis $\{e_i\}$, $A(v)$ has components

$$(\sum_{i=1}^n A^1{}_i v^i, \sum_{i=1}^n A^2{}_i v^i, \ldots, \sum_{i=1}^n A^m{}_i v^i)$$

in the basis $\{\hat{e}_i\}$

Now let's see the components of A'. By definition we have,

$$A'^i{}_j := A'(\hat{\theta}^i)(e_j) = \hat{\theta}^i(A(e_j)) = A^i{}_j.$$

That is, the same components, but now the matrix acts on the left on the components $(\omega_1, \omega_2, \ldots, \omega_m)$ of an element ω of W' in the co-basis $\{\hat{\theta}^i\}$. The components of $A'(\omega)$ in the co-basis $\{\theta^i\}$ are,

$$(\sum_{i=1}^m A^i{}_1 \omega_i, \sum_{i=1}^m A^i{}_2 \omega_i, \ldots, \sum_{i=1}^m A^i{}_n \omega_i).$$

A particularly interesting case of this construction is when $W = V$ and this is a space with an inner product, that is, a Hilbert space. In this case, the inner product gives us a canonical map between V and its dual V':

$$\phi : V \to V', \quad \phi(v) := \langle v, \cdot \rangle. \tag{2.62}$$

This map is injective, and since V and V' have the same dimension, it is also surjective, and therefore invertible. That is, given $\omega \in V'$, there exists $v = \phi^{-1}(\omega) \in V$ such that $\langle v, \cdot \rangle = \omega$. Note then that $\phi^{-1} : V' \to V$ satisfies,

$$\langle \phi^{-1}(\omega), u \rangle = \langle v, u \rangle = \omega(u).$$

If $A : V \to V$, then $A' : V' \to V'$ can also be considered as an operator between V and V which we will call $A^\star$.

With the help of this map, we define $A^\star : V \to V$ given by: (See figure)

$$\begin{tikzcd} V \arrow[r, "A"] \arrow[d, "(,)^{-1}"'] & V \arrow[d, "(,)"] \arrow[l, dashed, bend left=80, "A^\star"] \\ V' & V' \arrow[l, "A'"] \end{tikzcd}$$

Figure 2.3: Diagram of the star operator.

$$A^\star(v) := \phi^{-1}(A'(\phi(v))). \tag{2.63}$$

In terms of the inner product, this is:

$$\langle A^\star(v), u \rangle = \langle \phi^{-1}(A'(\phi(v))), u \rangle = A'(\phi(v))(u) = \phi(v)(A(u)) = \langle v, A(u) \rangle. \tag{2.64}$$

In its matrix representation, this operator is,

$$A^{\star j}{}_i = t_{il} A^l{}_k (t^{-1})^{kj}, \quad \text{real inner product} \tag{2.65}$$

$$A^{\star j}{}_i = t_{il} \bar{A}^l{}_k (t^{-1})^{kj}, \quad \text{complex inner product} \tag{2.66}$$

where t_{li} is the representation of the inner product and $(t^{-1})^{jk}$ that of its inverse $(t_{il}(t^{-1})^{lk} = \delta^k{}_i)$. If we choose a basis such that $t_{ik} = \delta_{ik}$, the Kronecker delta, the matrix representation of the adjoint is simply the transpose matrix, $A^{\star j}{}_i = A^{\dagger j}{}_i = A^i{}_j$,

A particularly interesting subset of operators is those for which $A = A^\star$. These operators are called **Hermitian** or **Self-adjoint** . [13]

Self-adjoint operators have important properties:

Lemma 2.10 *Let $M = Span\{u_1, u_2, \ldots, u_m\}$, where $\{u_i\}$ are a set of eigenvalues of A, a self-adjoint operator. Then M and $M^\perp$ are invariant spaces of A.*

[13] In the case of infinite dimension, these names do not coincide for some authors.

Proof: The first statement is clear and general, the second depends on the Hermiticity of A. Let $v \in M^{\perp}$ be any vector, let's see that $A(v) \in M^{\perp}$. Let $u \in M$ be arbitrary, then

$$\langle u, A(v) \rangle = \langle A(u), v \rangle = 0, \tag{2.67}$$

since $A(u) \in M$ if $u \in M$ ♠

This property has the following corollary:

Corolary 2.1 *Let $A : H \to H$ be self-adjoint. Then the eigenvectors of A form an orthonormal basis of H.*

Proof: $A : H \to H$ has at least one eigenvector, let's call it u_1. Now consider its restriction to the space perpendicular to u_1 which we also denote by A since by the previous lemma this is an invariant space, $A : \{u_1\}^{\perp} \to \{u_1\}^{\perp}$. This operator also has an eigenvector, say u_2 and $u_1 \perp u_2$. Now consider the restriction of A to $Span\{u_1, u_2\}^{\perp}$, there we also have $A : Span\{u_1, u_2\}^{\perp} \to Span\{u_1, u_2\}^{\perp}$ and therefore an eigenvector of A, u_3 with $u_3 \perp u_1$, $u_3 \perp u_2$. Continuing in this way, we end up with $n = \dim H$ eigenvectors all orthogonal to each other ♠

This theorem has several extensions to the case where the vector space is of infinite dimension. Later in chapter **??** we will see one of them.

Note that in this basis A is diagonal and therefore we have

Corolary 2.2 *Every self-adjoint operator is diagonalizable*

Also note that if u is an eigenvector of A self-adjoint, with eigenvalue λ, then,

$$\bar{\lambda}\langle u, u \rangle = \langle A(u), u \rangle = \langle u, A(u) \rangle = \lambda \langle u, u \rangle \tag{2.68}$$

and therefore $\bar{\lambda} = \lambda$, that is,

Lemma 2.11 *The eigenvalues of a self-adjoint operator are real*

Let's see what the condition of Hermiticity means in terms of the components of the operator in an orthonormal basis. Let A be a self-adjoint operator and let $\{e_i\}$, $i = 1, \ldots, n$ be an orthonormal basis of the space where it acts. We have that $\langle A(e_i), e_j \rangle = \langle e_i, A(e_j) \rangle$ and therefore, noting that $I = \sum_{i=1}^{n} e_i \theta^i$, we obtain,

$$\begin{aligned}
0 &= \langle \sum_{k=1}^{n} e_k \theta^k (A(e_i)), e_j \rangle - \langle e_i, \sum_{l=1}^{n} e_l \theta^l (A(e_j)) \rangle \\
&= \sum_{k=1}^{n} \bar{A}^k{}_i \langle e_k, e_j \rangle - \sum_{l=1}^{n} A^l{}_j \langle e_i, e_l \rangle \\
&= \sum_{k=1}^{n} \bar{A}^k{}_i \delta_{kj} - \sum_{l=1}^{n} A^l{}_j \delta_{li} \\
&= \bar{A}^j{}_i - A^i{}_j
\end{aligned} \tag{2.69}$$

from which we conclude that

$$\bar{A}^j{}_i = A^i{}_j \tag{2.70}$$

that is, the transpose matrix is the complex conjugate of the original. In the case of real matrices, we see that the condition is that in this basis the matrix is equal

2.5 Adjoint Operators

Let A be a linear operator between two vector spaces, $A : V \to W$, that is, $A \in \mathscr{L}(V,W)$. Since V and W have dual spaces, this operator naturally induces a linear operator from W' to V', called its **dual** ,

$$A'(\omega)(v) := \omega(A(v)) \;\; \forall \;\; \omega \in W', \;\;\; v \in V. \tag{2.71}$$

That is, the operator that when applied to an element $\omega \in W'$ gives us the element $A'(\omega)$ of V' which, when acting on $v \in V$, gives the number $\omega(A(v))$. See figure.

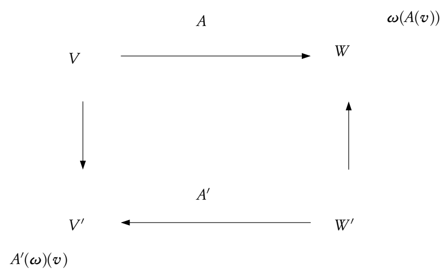

Figure 2.4: Diagram of the dual operator.

Note that this is a linear operator since,

$$\begin{aligned} A'(\alpha\omega + \sigma)(v) &= (\alpha\omega + \sigma)(A(v)) \\ &= \alpha\omega(A(v)) + \sigma(A(v)) \\ &= \alpha A'(\omega)(v) + A'(\sigma)(v). \end{aligned} \tag{2.72}$$

In the matrix representation, this operator is represented merely by the same matrix as the original, but now acting on the left, $A'(w)_i = w_j A^j{}_i$, that is, $A'^j_i = A^j{}_i$.

If there are norms defined in V and W and we define the norm of $A : V \to W$ in the usual way,

$$||A|| := \sup_{||v||_V = 1} \{||A(v)||_W\} \tag{2.73}$$

Then we see that

$$\begin{aligned}
||A'|| &:= \sup_{||\omega||_{W'}=1} \{||A'(\omega)||_{V'}\} \\
&= \sup_{||\omega||_{W'}=1} \{ \sup_{||v||_V=1} \{|A'(\omega)(v)|\}\} \\
&= \sup_{||\omega||_{W'}=1} \{ \sup_{||v||_V=1} \{|\omega(A(v))|\}\} \\
&\leq \sup_{||\omega||_{W'}=1} \{ \sup_{||v||_V=1} \{||\omega||_{W'}||(A(v))||_W\}\} \\
&= \sup_{||v||_V=1} \{||(A(v))||_W\} \\
&= ||A||. \qquad (2.74)
\end{aligned}$$

Thus we see that if an operator is bounded, then its dual is also bounded. In fact, the equality of the norms can be proven, but this would require introducing new tools (in the most general case, the Hahn-Banach theorem) that we do not wish to incorporate in this text.

Let's see what the components of the dual of an operator are in terms of the components of the original operator. Let $\{e_i\}$, $\{\theta^i\}$, $i = 1,..n$ be a basis and respectively a co-basis of V, and let $\{\hat{e}_i\}$, $\{\hat{\theta}^i\}$, $i = 1,..m$ be a pair of basis and co-basis of W. We then have that the components of A with respect to these bases are: $A^i{}_j := \hat{\theta}^i(A(e_j))$, that is, $A(v) = \sum_{i=1}^m \sum_{j=1}^n A^i{}_j \hat{e}_i \theta^j(v)$.

Therefore, if v has components $(v^1, v^2, \dots, v^n)$ in the basis $\{e_i\}$, $A(v)$ has components

$$(\sum_{i=1}^n A^1{}_i v^i, \sum_{i=1}^n A^2{}_i v^i, \dots, \sum_{i=1}^n A^m{}_i v^i)$$

in the basis $\{\hat{e}_i\}$

Now let's see the components of A'. By definition we have,

$$A'^i{}_j := A'(\hat{\theta}^i)(e_j) = \hat{\theta}^i(A(e_j)) = A^i{}_j.$$

That is, the same components, but now the matrix acts on the left on the components $(\omega_1, \omega_2, \dots, \omega_m)$ of an element ω of W' in the co-basis $\{\hat{\theta}^i\}$. The components of $A'(\omega)$ in the co-basis $\{\theta^i\}$ are,

$$(\sum_{i=1}^m A^i{}_1 \omega_i, \sum_{i=1}^m A^i{}_2 \omega_i, \dots, \sum_{i=1}^m A^i{}_n \omega_i).$$

A particularly interesting case of this construction is when $W = V$ and this is a space with an inner product, that is, a Hilbert space. In this case, the inner product gives us a canonical map between V and its dual V':

$$\phi : V \to V', \quad \phi(v) := \langle v, \cdot \rangle. \qquad (2.75)$$

This map is injective, and since V and V' have the same dimension, it is also surjective, and therefore invertible. That is, given $\omega \in V'$, there exists $v = \phi^{-1}(\omega) \in V$ such that $\langle v, \cdot \rangle = \omega$. Note then that $\phi^{-1} : V' \to V$ satisfies,

$$\langle \phi^{-1}(\omega), u \rangle = \langle v, u \rangle = \omega(u).$$

If $A : V \to V$, then $A' : V' \to V'$ can also be considered as an operator between V and V which we will call $A^\star$.

With the help of this map, we define $A^\star : V \to V$ given by: (See figure)

$$\begin{tikzcd}
V \arrow[r, "A"] \arrow[d, "(\,,\,)^{-1}"'] & V \arrow[d, "(\,,\,)"] \arrow[l, dashed, bend left=60, "A^\star"] \\
V' & V' \arrow[l, "A'"]
\end{tikzcd}$$

Figure 2.5: Diagram of the star operator.

$$A^\star(v) := \phi^{-1}(A'(\phi(v))). \tag{2.76}$$

In terms of the inner product, this is:

$$\langle A^\star(v), u \rangle = \langle \phi^{-1}(A'(\phi(v))), u \rangle = A'(\phi(v))(u) = \phi(v)(A(u)) = \langle v, A(u) \rangle. \tag{2.77}$$

In its matrix representation, this operator is,

$$\begin{aligned}
A^{\star j}{}_i &= t_{il} A^l{}_k (t^{-1})^{kj}, \quad \text{real inner product} &(2.78)\\
A^{\star j}{}_i &= t_{il} \bar{A}^l{}_k (t^{-1})^{kj}, \quad \text{complex inner product} &(2.79)
\end{aligned}$$

where $t_{l i}$ is the representation of the inner product and $(t^{-1})^{jk}$ that of its inverse $(t_{il}(t^{-1})^{lk} = \delta^k{}_i)$. If we choose a basis such that $t_{ik} = \delta_{ik}$, the Kronecker delta, the matrix representation of the adjoint is simply the transpose matrix, $A^{\star j}{}_i = A^{\dagger j}{}_i = A^i{}_j$,

A particularly interesting subset of operators is those for which $A = A^\star$. These operators are called **Hermitian** or **Self-adjoint** . [14]

Self-adjoint operators have important properties:

Lemma 2.12 *Let $M = Span\{u_1, u_2, \ldots, u_m\}$, where $\{u_i\}$ are a set of eigenvalues of A, a self-adjoint operator. Then M and $M^\perp$ are invariant spaces of A.*

[14] In the case of infinite dimension, these names do not coincide for some authors.

Proof: The first statement is clear and general, the second depends on the Hermiticity of A. Let $v \in M^{\perp}$ be any vector, let's see that $A(v) \in M^{\perp}$. Let $u \in M$ be arbitrary, then

$$\langle u, A(v)\rangle = \langle A(u), v\rangle = 0, \tag{2.80}$$

since $A(u) \in M$ if $u \in M$ ♠

This property has the following corollary:

Corolary 2.3 *Let $A : H \to H$ be self-adjoint. Then the eigenvectors of A form an orthonormal basis of H.*

Proof: $A : H \to H$ has at least one eigenvector, let's call it u_1. Now consider its restriction to the space perpendicular to u_1 which we also denote by A since by the previous lemma this is an invariant space, $A : \{u_1\}^{\perp} \to \{u_1\}^{\perp}$. This operator also has an eigenvector, say u_2 and $u_1 \perp u_2$. Now consider the restriction of A to $Span\{u_1, u_2\}^{\perp}$, there we also have $A : Span\{u_1, u_2\}^{\perp} \to Span\{u_1, u_2\}^{\perp}$ and therefore an eigenvector of A, u_3 with $u_3 \perp u_1$, $u_3 \perp u_2$. Continuing in this way, we end up with $n = \dim H$ eigenvectors all orthogonal to each other ♠

This theorem has several extensions to the case where the vector space is of infinite dimension. Later in chapter **??** we will see one of them.

Note that in this basis A is diagonal and therefore we have

Corolary 2.4 *Every self-adjoint operator is diagonalizable*

Also note that if u is an eigenvector of A self-adjoint, with eigenvalue λ, then,

$$\bar{\lambda}\langle u, u\rangle = \langle A(u), u\rangle = \langle u, A(u)\rangle = \lambda\langle u, u\rangle \tag{2.81}$$

and therefore $\bar{\lambda} = \lambda$, that is,

Lemma 2.13 *The eigenvalues of a self-adjoint operator are real*

Let's see what the condition of Hermiticity means in terms of the components of the operator in an orthonormal basis. Let A be a self-adjoint operator and let $\{e_i\}$, $i = 1, \ldots, n$ be an orthonormal basis of the space where it acts. We have that $\langle A(e_i), e_j\rangle = \langle e_i, A(e_j)\rangle$ and therefore, noting that $I = \sum_{i=1}^{n} e_i \theta^i$, we obtain,

$$\begin{aligned}
0 &= \langle \sum_{k=1}^{n} e_k \theta^k (A(e_i)), e_j\rangle - \langle e_i, \sum_{l=1}^{n} e_l \theta^l (A(e_j))\rangle \\
&= \sum_{k=1}^{n} \bar{A}^k{}_i \langle e_k, e_j\rangle - \sum_{l=1}^{n} A^l{}_j \langle e_i, e_l\rangle \\
&= \sum_{k=1}^{n} \bar{A}^k{}_i \delta_{kj} - \sum_{l=1}^{n} A^l{}_j \delta_{li} \\
&= \bar{A}^j{}_i - A^i{}_j
\end{aligned} \tag{2.82}$$

from which we conclude that

$$\bar{A}^j{}_i = A^i{}_j \tag{2.83}$$

so the transpose matrix is the complex conjugate of the original. In the case of real matrices, we see that the condition is that in that basis the matrix is equal to its transpose, which is usually denoted by saying that the matrix is symmetric.

An interesting property of self-adjoint operators is that their norm is equal to the supremum of the magnitudes of their eigenvalues. Since the calculation demonstrating this will be used later in the course, we provide a demonstration of this fact below.

Lemma 2.14 *If A is self-adjoint then* $||A|| = \sup\{|\lambda_i|\}$.

Proof: Let $F(\boldsymbol{u}) := \langle A(\boldsymbol{u}), A(\boldsymbol{u})\rangle$ defined on the sphere $||\boldsymbol{u}|| = 1$. Since this set is compact (here we are using the fact that the space is finite-dimensional) it has a maximum which we will denote by $\boldsymbol{u}_0$. Note then that $F(\boldsymbol{u}_0) := ||A||^2$. Since $F(\boldsymbol{u})$ is differentiable on the sphere, it must satisfy

$$\frac{d}{d\lambda} F(\boldsymbol{u}_0 + \lambda \delta \boldsymbol{u})|_{\lambda=0} = 0, \tag{2.84}$$

along any curve tangent to the sphere at the point $\boldsymbol{u}_0$, that is, for all $\delta\boldsymbol{u}$ such that $\langle \boldsymbol{u}_0, \delta\boldsymbol{u}\rangle = 0$. See figure.

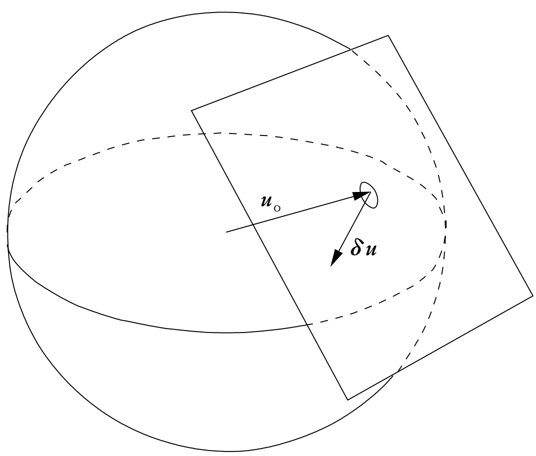

Figure 2.6: Normal and tangent vectors to the sphere.

But

$$\begin{aligned}
\frac{d}{d\lambda}F(u_o+\lambda\delta u)|_{\lambda=0} &= \frac{d}{d\lambda}\langle A(u_o+\lambda\delta u),A(u_o+\lambda\delta u)\rangle|_{\lambda=0} \\
&= \langle A(\delta u),A(u_o)\rangle+\langle A(u_o),A(\delta u)\rangle \\
&= 2\Re\langle A^\star Au_o,\delta u\rangle \\
&= 0 \quad \forall\delta u,\ \ \langle u_o,\delta u\rangle=0 \qquad (2.85)
\end{aligned}$$

Since δu is arbitrary in $\{u_o\}^\perp$ this simply implies

$$A^\star A(u_o)\ \in\ \{\delta u\}^\perp=\{u_o\}^{\perp\perp}=\{u_o\}. \qquad (2.86)$$

and therefore there will exist $\alpha\in C$ such that

$$A^\star A(u_o)=\alpha u_o. \qquad (2.87)$$

Taking the inner product with u_o we get,

$$\begin{aligned}
\langle A^\star A(u_o),u_o\rangle &= \bar\alpha\langle u_o,u_o\rangle \\
&= \langle A(u_o),A(u_o)\rangle \\
&= ||A||^2 \qquad (2.88)
\end{aligned}$$

and therefore we have that $\alpha=\bar\alpha=||A||^2$.

Now let $v:=Au_o-||A||u_o$, then, now using that A is self-adjoint, we have,

$$\begin{aligned}
A(v) &= AA(u_o)-||A||A(u_o) \\
&= A^\star A(u_o)-||A||A(u_o) \\
&= ||A||^2u_o-||A||A(u_o) \\
&= ||A||(||A||u_o-A(u_o)) \\
&= -||A||v, \qquad (2.89)
\end{aligned}$$

therefore either v is an eigenvector of A, with eigenvalue $\lambda=-||A||$, or $v=0$, in which case u_o is an eigenvector of A with eigenvalue $\lambda=||A||$ ♠

2.6 Unitary Operators

Another subclass of linear operators that appears very often in physics when there is a privileged inner product is that of **unitary operators** , that is, those such that their action preserves the inner product,

$$\langle U(u),U(v)\rangle=\langle u,v\rangle,\quad \forall u,v\ \in H. \qquad (2.90)$$

The most typical case of a unitary operator is a transformation that sends one orthonormal basis to another. Usual examples are rotations in $\mathbb{R}^n$.

We also observe that $||U||=\sup_{||v||=1}\{||U(v)||\}=1$.

Note that

$$\langle U(u), U(v)\rangle = \langle U^{\star}U(u), v\rangle = \langle u, v\rangle, \quad \forall u, v \in H, \tag{2.91}$$

that is,

$$U^{\star}U = I \tag{2.92}$$

and therefore,

$$U^{-1} = U^{\star}. \tag{2.93}$$

Let's see what the eigenvalues of a unitary operator U are. Let v_1 be an eigenvector of U (we know it has at least one), then,

$$\begin{aligned} \langle U(v_1), U(v_1)\rangle &= \lambda_1\bar{\lambda}_1\langle v_1, v_1\rangle \\ &= \langle v_1, v_1\rangle \end{aligned} \tag{2.94}$$

and therefore $\lambda_1 = e^{i\theta_1}$ for some angle θ_1.

If the operator U represents a non-trivial rotation in $\mathbb{R}^3$, then, given that we have an odd number of eigenvalues, there will be one that is real, the other two complex conjugates of each other. The eigenvector corresponding to the real eigenvalue defines the axis that remains fixed in that rotation. If we have more than one eigenvalue, then their corresponding eigenvectors are orthogonal, indeed, let v_1 and v_2 be two eigenvectors then

$$\begin{aligned} \langle U(v_1), U(v_2)\rangle &= \bar{\lambda}_1\lambda_2\langle v_1, v_2\rangle \\ &= \langle v_1, v_2\rangle \end{aligned} \tag{2.95}$$

and therefore if $\lambda_1 \neq \lambda_2$ we must have $\langle v_1, v_2\rangle = 0$.

Exercise: Show that if A is a self-adjoint operator, then $U := e^{iA}$ is a unitary operator.

Equivalence Relations.

Definition: An **equivalence relation**, $\approx$, between elements of a set X is a relation that satisfies the following conditions:

1. Reflexive: If $x \in X$, then $x \approx x$.
2. Symmetric: If $x, x' \in X$ and $x \approx x'$, then $x' \approx x$.
3. Transitive: If $x, x', x'' \in X$, $x \approx x'$ and $x' \approx x''$, then $x \approx x''$.

Note that the first property ensures that each element of X satisfies an equivalence relation with some element of X, in this case with itself. Equivalence relations often appear in physics, essentially when we use a mathematical entity to describe a physical process that has superfluous parts with respect to this process and therefore we would like to ignore them. This is achieved by declaring two entities that describe the same phenomenon as equivalent entities.

Example: Let X be the set of real numbers and let $x \approx y$ if and only if there exists an integer n such that $x = n + y$, this is clearly an equivalence relation. This is used when we are interested in describing something using the straight line but that in reality should be described using a circle of unit circumference.
Given an equivalence relation in a set, we can group the elements of this set into **equivalence classes** of elements, that is, into subsets where all their elements are equivalent to each other and there is no element outside this subset that is equivalent to any of the elements of the subset. (If X is the set, $Y \subset X$ is one of its equivalence classes, and if $y \in Y$, then $y \approx y'$ if and only if $y' \in Y$.)

The fundamental property of equivalence relations is the following.

Theorem 2.6 *An equivalence relation in a set X allows regrouping its elements into equivalence classes such that each element of X is in one and only one of these.*

Proof: Let $x \in X$ and Y be the subset of all elements of X equivalent to x. Let's see that this subset is an equivalence class. Let y and y' be two elements of Y, that is, $y \approx x$ and $y' \approx x$, but by the transitivity property, $y \approx y'$. If $y \in Y$ and $z \notin Y$, then $y \not\approx z$, because otherwise z would be equivalent to x and therefore would be in Y. Finally, note that by reflexivity, x is also in Y. It only remains to see that if y is in Y and also in another equivalence class, say Z, then $Y = Z$. Since $y \in Y$, then y is equivalent to every element of Y, and since $y \in Z$, then y is equivalent to every element of Z, but by transitivity, every element of Y is equivalent to every element of Z, but since these are equivalence classes and therefore each contains all its equivalent elements, both must coincide.

Exercise: What are the equivalence classes of the previous examples?

2.7 Problems

Problem 2.2 *Let the operator $A : V \to V$ where $\dim V = n$, such that $Ax = \lambda x$. Calculate $\det A$ and trA.*

Problem 2.3 *Let $V = \mathbb{R}^3$ and $\boldsymbol{x}$ be any non-zero vector. Find geometrically and analytically the quotient space $V/W_{\boldsymbol{x}}$, where $W_{\boldsymbol{x}}$ is the space generated by $\boldsymbol{x}$. Take another vector, $\boldsymbol{x}'$, linearly independent of the first and now calculate $V/W_{(\boldsymbol{x},\boldsymbol{x}')}$.*

Problem 2.4 *The norm on operators is defined as:*

$$||A||_{\mathcal{L}} = max_{||\boldsymbol{x}||_V=1}||A(\boldsymbol{x})||_V. \tag{2.96}$$

Find the norms of the following operators, given by their matrix representation with respect to a basis and where the norm in the vector space is the Euclidean norm with respect to that basis.

a)

$$\begin{pmatrix} 3 & 5 \\ 4 & 1 \end{pmatrix} \tag{2.97}$$

b)

$$\begin{pmatrix} 3 & 5 & 2 \\ 4 & 1 & 7 \\ 8 & 3 & 2 \end{pmatrix} \tag{2.98}$$

Problem 2.5 *Let V be any vector space and let $||\cdot||$ be a Euclidean norm in that space. The Hilbert-Schmidt norm of an operator is defined as:*

$$||A||^2_{HS} = \sum_{i,j=1}^{n} |A^j{}_i|^2. \tag{2.99}$$

where the basis used has been orthonormal with respect to the Euclidean norm.

a) Show that this is a norm.

b) Show that $||A||_{\mathcal{L}} \leq ||A||_{HS}$.

c) Show that $\sum_{j=1}^{n} |A^j{}_k|^2 \leq ||A||^2_{\mathcal{L}}$ for each k. Therefore $||A||^2_{HS} \leq n||A||^2_{\mathcal{L}}$, and the two norms are equivalent.

Hint: use that $\theta^j(A(\boldsymbol{u})) = \theta^j(A)(\boldsymbol{u})$ and then that $|\theta(\boldsymbol{u})| \leq ||\theta||||\boldsymbol{u}||$.

Problem 2.6 *Calculate the eigenvalues and eigenvectors of the following matrices:*

a)

$$\begin{pmatrix} 3 & 6 \\ 4 & 1 \end{pmatrix} \tag{2.100}$$

b)

$$\begin{pmatrix} 3 & 6 \\ 0 & 1 \end{pmatrix} \tag{2.101}$$

c)

$$\begin{pmatrix} 2 & 4 & 2 \\ 4 & 1 & 0 \\ 3 & 3 & 1 \end{pmatrix} \tag{2.102}$$

Problem 2.7 *Bring the following matrices to upper triangular form: Note: From the transformation of the bases.*

a)

$$\begin{pmatrix} 3 & 4 \\ 2 & 1 \end{pmatrix} \tag{2.103}$$

b)

$$\begin{pmatrix} 2 & 4 & 2 \\ 4 & 1 & 0 \\ 3 & 3 & 1 \end{pmatrix} \tag{2.104}$$

c)

$$\begin{pmatrix} 1 & 4 & 3 \\ 4 & 4 & 0 \\ 3 & 3 & 1 \end{pmatrix} \tag{2.105}$$

Problem 2.8 *Show again that* $\det e^{A} = e^{trA}$. *Hint: express the matrix representation of A in a basis where it has the Jordan canonical form. Alternatively, use a basis where A is upper triangular and see that the product of upper triangular matrices gives an upper triangular matrix and therefore the exponential of an upper triangular matrix is also upper triangular.*

Bibliography notes: This chapter is based on the following books: [?], [?], [?] and [?]. The following are also of interest, [?] and [?]. Linear algebra is one of the largest and most prolific areas of mathematics, especially when dealing with infinite-dimensional spaces, which is usually called real analysis and operator theory. In my personal experience, most problems end up reducing to an algebraic problem and one feels that progress has been made when that problem can be solved.

Chapter 3

Geometry

3.1 Manifolds

There are several reasons that justify the study of the concept of a manifold, or more generally of differential geometry, by physicists. One is that manifolds naturally appear in physics and therefore we cannot avoid them. Only in elementary courses can they be circumvented through vector calculus in $\mathbb{R}^n$. Thus, for example, to study the movement of a particle restricted to move on a sphere, we imagine the latter embedded in $\mathbb{R}^3$ and use the natural coordinates of $\mathbb{R}^3$ to describe its movements.

The second reason is that the concept of a manifold is of great conceptual utility, since, for example, in the case of a particle moving in $\mathbb{R}^3$, it allows us to clearly distinguish between the position of a particle and its velocity vector as mathematical entities of different nature. This fact is masked in $\mathbb{R}^3$ since this special type of manifold has the structure of a vector space.

Since the time of Galileo, we know that the language of physics is mathematics. Like any language, its utility goes beyond its daily use to understand each other and work on our ideas. Language allows for a synthesis of concepts and knowledge that encapsulates an entire area of knowledge into a smaller number of concepts. This allows future generations to understand an immense amount of knowledge that for previous generations were disparate aspects of reality as particular aspects of the same trunk of knowledge. The clearest example of this is the theories of the standard model of particles, which unify under the same phenomenon what we previously understood as distinct properties of matter. In particular, these theories naturally and deterministically incorporate elements of geometry, such as fiber bundles and connections, symmetries, etc.

A manifold is a generalization of Euclidean spaces $\mathbb{R}^n$ in which one preserves the concept of continuity, that is, its topology in the local sense but discards its character as a vector space. A manifold of dimension n is, in imprecise terms, a set of points that locally is like $\mathbb{R}^n$, but not necessarily in its global form.

An example of a two-dimensional manifold is the sphere, S^2. If we look at a sufficiently small neighborhood, U_p, of any point of S^2, we see that it is similar to a neighborhood of the plane, $\mathbb{R}^2$, in the sense that we can define a continuous and invertible map between both neighborhoods. Globally, the plane and the sphere are topologically distinct, as there is no continuous and invertible map between them. [See figure 3.1.]

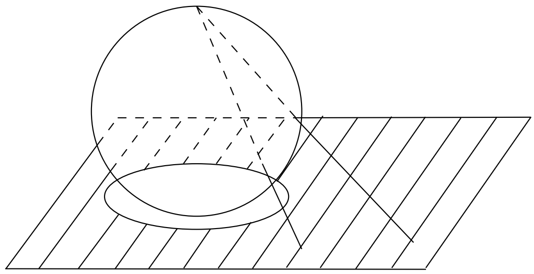

Figure 3.1: An atlas of the sphere.

As we said before, one might object to the need, in the previous example, to introduce the concept of a manifold, since one could consider S^2 as the subset of $\mathbb{R}^3$ such that $x_1^2 + x_2^2 + x_3^2 = 1$. The answer to this objection is that in physics one must follow the rule of economy of concepts and objects and discard everything that is not fundamental to the description of a phenomenon: if we want to describe things happening on the sphere, why do we need a space of more dimensions? This rule of economy forces us to refine concepts and discard everything superfluous. It is in this way that we advance in our maturation as physicists. It is the way we truly penetrate the mysteries of nature.

Note that the first way of working with the sphere is what we normally do when we seek to locate points and paths on the globe. In fact, we use flat maps, formerly called charts, to describe what happens in our cities and countries. When we want to have a collection of maps that cover the entire globe, we acquire an atlas, that is, a set of maps that cover the entire globe and have sectors in common between one and another. Some of these sectors are internal, for example when they describe a city within a country (on the map that covers that country), or on the edges when we go from one sheet to another. Only when we want to see the global structure of the globe, for example if we want to take a plane trip covering a large part of the globe, do we use a small version of the planet as implanted in $\mathbb{R}^3$.

We now give a series of definitions to finally arrive at the definition of an n-dimensional manifold.

Definition: Let M be a set. A **chart of** M is a pair (U, φ) where U is a subset of M and φ an injective map between U and $\mathbb{R}^n$, such that its image, $\varphi[U]$ is open in $\mathbb{R}^n$.

Definition: An **atlas of** M is a collection of charts $\{(U_i, \varphi_i)\}$ satisfying the following conditions: [See figure 3.2.]

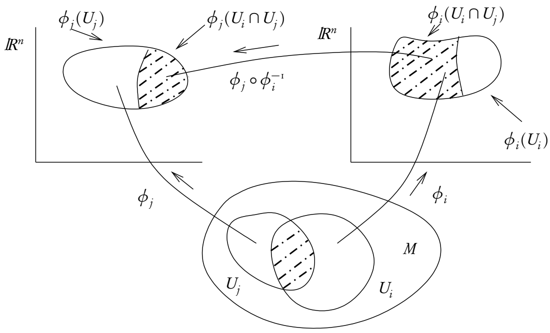

Figure 3.2: Relationship between charts.

1. The U_i cover M, $(M = \bigcup_i U_i)$.
2. If two charts overlap then $\varphi_i(U_i \cap U_j)$ is also an open set in $I\!R^n$.
3. The map $\varphi_j \circ \varphi_i^{-1} : \varphi_i[U_i \cap U_j] \to \varphi_j[U_i \cap U_j]$ is continuous, injective, and surjective.

Condition **??** gives us a notion of *closeness* in M induced from the analogous notion in $I\!R^n$. Indeed, we can say that a sequence of points $\{p_k\}$ in M converges to p in U_i if there exists k_0 such that $\forall\ k > k_0$, $p_k \in U_i$ and the sequence $\{\varphi_i(p_k)\}$ converges to $\varphi_i(p)\}$. Another way to see this is that if after this construction of a manifold we impose that the maps φ_i are continuous, then we induce a unique topology on M. [See figure 3.3.]

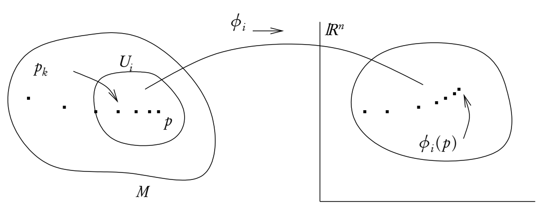

Figure 3.3: Sequences in M.

Condition **??** simply ensures that this notion is consistent. If $p \in U_i \cap U_j$ then the fact that the sequence converges is independent of whether we use the chart (U_i, φ_i) or (U_j, φ_j).

Condition **??** allows us to encode in the maps $\varphi_j \circ \varphi_i^{-1}$ from $I\!R^n$ to $I\!R^n$ the global topological information necessary to distinguish, for example, whether M is a sphere, a plane, or a torus. Therein lies, for example, the information that there is no continuous and invertible map between S^2 and $\mathbb{R}^2$. But also, if we require these maps to be differentiable, it is what will allow us to formulate differential calculus on M. Indeed, note that in condition **??** we speak of the continuity of the map $\varphi_j \circ \varphi_i^{-1}$, which is well-defined because it is a map between $\mathbb{R}^n$ and $\mathbb{R}^n$. Similarly, we can speak of the differentiability of these maps.

We will say that **an atlas** $\{(U_i, \varphi_i)\}$ **is** C^p if the maps $\varphi_j \circ \varphi_i^{-1}$ are p-times differentiable and their p-th derivative is continuous.

One might be tempted to define the manifold M as the pair consisting of the set M and an atlas $\{(U_i, \varphi_i)\}$, but this would lead us to consider as different manifolds, for example, the plane with an atlas given by the chart $(\mathbb{R}^2, (x,y) \to (x,y))$ and the plane with an atlas given by the chart $(\mathbb{R}^2, (x,y) \to (x,-y))$.

To remedy this inconvenience, we introduce the concept of equivalence between atlases.

Definition: We will say that **two atlases are equivalent** if their union is also an atlas.

Exercise: Prove that this is indeed an equivalence relation $\approx$, that is, it satisfies:

$i) A \approx A$

$ii) A \approx B \implies B \approx A$

$iii) A \approx B, B \approx C \implies A \approx C$.

With this equivalence relation, we can divide the set of atlases of M into different **equivalent classes**. [Remember that each equivalent class is a set where all its elements are equivalent to each other and such that there is no element equivalent to these that is not in it.]

Definition: We will call **manifold** M **of dimension** n **and differentiability** p the pair consisting of the set M and an equivalent class of atlases, $\{\varphi_i : U_i \to \mathbb{R}^n\}$, in C^p.

It can be shown that to uniquely characterize the manifold M it is sufficient to give the set M and an atlas. If we have two atlases of M, then either they are equivalent and thus represent the same manifold, or they are not and then represent different manifolds.

The definition of a manifold that we have introduced is still too general for usual physical applications, in the sense that the allowed topologies can still be pathological from the point of view of physics. Therefore, in this course, we will impose an extra condition on manifolds. We will assume that they are **separable or Hausdorff**. That is, if p and $q \in M$, distinct, then: either they belong to the domain of the same chart U_i (in which case there exist neighborhoods W_p of $\varphi_i(p)$ and W_q of $\varphi_i(q)$ such that $\varphi_i^{-1}(W_p) \cap \varphi_i^{-1}(W_q) = \emptyset$, that is, the points have disjoint neighborhoods) or there exist U_i and U_j with $p \in U_i$, $q \in U_j$ and $U_i \cap U_j = \emptyset$, which also implies that they have disjoint neighborhoods. This is a property in the topology of M that essentially says we can separate points of M. An example of a non-Hausdorff manifold is the following.

Example: M, as a set, consists of three intervals of the line $I_1 = (-\infty, 0]$, $I_2 = (-\infty, 0]$ and $I_3 = (0, +\infty)$.
An atlas of M is $\{(U_1 = I_1 \cup I_3, \varphi_1 = id), (U_2 = I_2 \cup I_3, \varphi_2 = id)\}$. [See figure 3.4.]

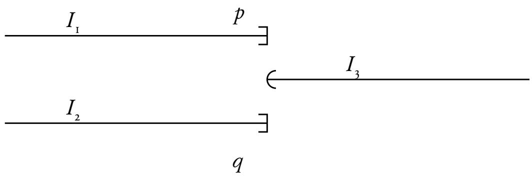

Figure 3.4: Example of a non-Hausdorff manifold.

Exercise: Prove that it is an atlas.
Note that given any neighborhood W_1 of $\varphi_1(0)$ in $\mathbb{R}$ and any neighborhood W_2 of $\varphi_2(0)$ we necessarily have
$\varphi_1^{-1}(W_1) \cap \varphi_2^{-1}(W_2) \neq \emptyset$.

3.2 Differentiable Functions on M

From now on we will assume that M is a C^∞ manifold, that is, all its maps $\varphi_i \circ \varphi_j^{-1}$ are infinitely differentiable. Although mathematically this is a restriction, it is not in physical applications. In these, M is generally the space of possible states of the system and therefore its points cannot be determined with absolute certainty, as every measurement involves some error. This indicates that through measurements we could never know the degree of differentiability of M. For convenience, we will assume it is C^∞.

A function on M is a map $f : M \to \mathbb{R}$, that is, a map that assigns a real number to each point of M. The information encoded in the atlas on M allows us to say how smooth f is.

Definition: We will say that f is **p-times continuously differentiable** at the point $q \in M$, $f \in C^p_q$ if given (U_i, φ_i) with $q \in U_i$, $f \circ \varphi_i^{-1} : \varphi_i(U_i) \subset \mathbb{R}^n \to \mathbb{R}$ is p-times continuously differentiable at $\varphi_i(q)$.
Note that this property is independent of the chart used [as long as we consider only charts from the compatible class of atlases]. We will say that $f \in C^p(M)$ if $f \in C^p_q \ \forall q \in M$. [See figure 3.5.]

In practice, one defines a particular function $f \in C^p(M)$ by introducing functions $f_i : \varphi_i(U_i) \subset \mathbb{R}^n \to \mathbb{R}$ (that is, $f_i(x^j)$ where x^j are the Cartesian coordinates

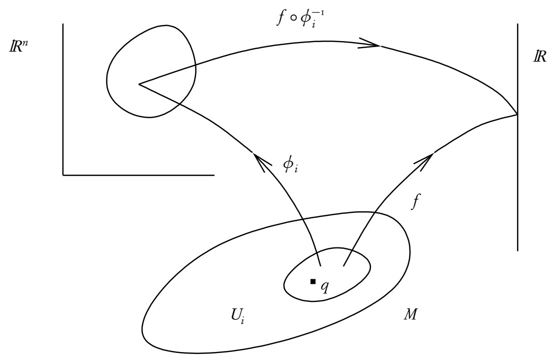

Figure 3.5: Composition of the map of a chart with a function.

in $\varphi_i(U_i) \subset I\!R^n$) that are C^p in $\varphi_i(U_i)$ and such that $f_i = f_j \circ \varphi_j \circ \varphi_i^{-1}$ in $\varphi_i(U_i \cap U_j)$. This guarantees that the set of f_i determines a unique function $f \in C^p(M)$. The set of f_i $(= f \circ \varphi_i^{-1})$ forms a **representation of** f in the atlas $\{(U_i, \varphi_i)\}$. [See figure 3.6.]

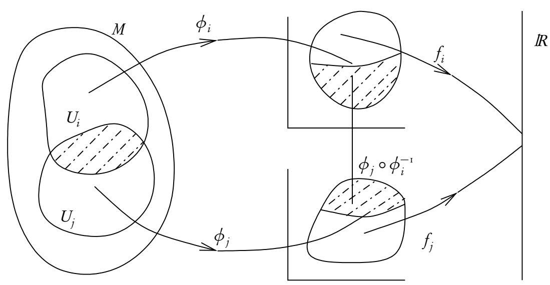

Figure 3.6: The relationship between the f_i.

<u>Exercise</u>: The circle, S^1, can be thought of as the interval $[0, 1]$ with its ends identified. What are the functions in $C^2(S^1)$?

Using the previous construction, one can also define maps from M to $I\!R^m$ that are p-times differentiable. Now we perform the inverse construction, that is, we will define the differentiability of a map from $I\!R^m$ to M. We will do the case $I\!R \to M$, in which case the map thus obtained is called a curve. The general case is obvious.

3.3 Curves in M

Definition: **A curve in** M is a map between an interval $I \subset I\!R$ and M, $\gamma : I \to M$.

Note that the curve is the map and not its graph in M, that is, the set $\gamma[I]$. Thus, it is possible to have two different curves with the same graph. This is not a mathematical whim but a physical necessity: it is not the same for a car to travel the road Córdoba–Carlos Paz at 10 km/h as at 100 km/h, or to travel it in the opposite direction.

Definition: We will say that $\gamma \in C^p_{t_o}$ if given a chart (U_i, φ_i) such that $\gamma(t_o) \in U_i$ the map $\varphi_i \circ \gamma(t) : I_{t_o} \subset I \to I\!R^n$ is p-times continuously differentiable at t_o. [See figure 3.7.]

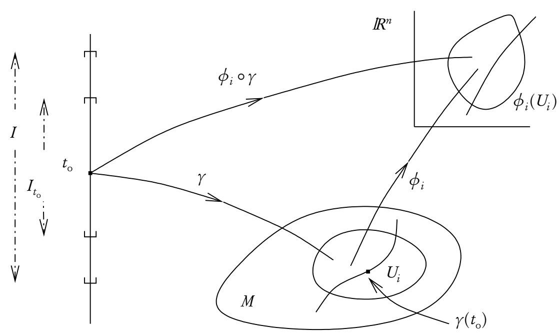

Figure 3.7: Differentiability of curves in M.

Exercise: Prove that the previous definition does not depend on the chart used.

This time we have used the concept of differentiability between maps from $I\!R$ to $I\!R^n$.

Definition: A curve $\gamma(t) \in C^p(I)$ if $\gamma(t) \in C^p_t \ \forall\, t \in I$.

Exercise: How would you define the concept of differentiability of maps between two manifolds?

Of particular importance among these are the maps from M to itself $g : M \to M$ that are continuously differentiable and invertible. They are called **Diffeomorphisms**. From now on we will assume that all manifolds, curves, diffeomorphisms, and functions are smooth, that is, they are C^∞.

3.4 Vectors

To define vectors at points of M we will use the concept of directional derivative at points of $I\!R^n$, that is, we will exploit the fact that in $I\!R^n$ there is a one-to-one correspondence between vectors $(v^1, \dots, v^n)|_{x_0}$ and directional derivatives $v(f)|_{x_0} = v^i \frac{\partial}{\partial x^i} f\Big|_{x_0}$.

As we have defined differentiable functions on M we can define derivations, or directional derivatives, at its points and identify with them the tangent vectors.

Definition: A **tangent vector** v **at** $p \in M$ is a map

$$v : C^\infty(M) \to I\!R$$

satisfying: $\forall f, g \in C^\infty(M),\ a, b \in I\!R$

i) Linearity; $v(af + bg)|_p = a\, v(f)|_p + b\, v(g)|_p$.

ii) Leibniz; $v(fg)|_p = f(p) v(g)|_p + g(p) v(f)|_p$.

Note that if $h \in C^\infty(M)$ is the constant function, $h(q) = c \quad \forall q \in M$, then $v(h) = 0$. [$i) \Longrightarrow v(h^2) = v(ch) = c\, v(h)$ while $ii) \Longrightarrow v(h^2) = 2h(p)v(h) = 2c\, v(h)$]. These properties also show that $v(f)$ depends only on the behavior of f at p.

Exercise: Prove this last statement.

Let T_p be the set of all vectors at p. This set has the structure of a vector space and is called the **tangent space at the point** p. Indeed, we can define the sum of two vectors v_1, v_2 as the vector, that is the map, satisfying *i*) and *ii*), $(v_1 + v_2)(f) = v_1(f) + v_2(f)$ and the product of the vector v by the number a as the map $(av)(f) = a\, v(f)$.

As in $I\!R^n$, the dimension of the vector space T_p, (that is, the maximum number of linearly independent vectors), is n.

Theorem 3.1 *$dim\ T_p = dim M$.*

Proof: This will consist of finding a basis for T_p. Let $dim\ M = n$ and (U, φ) such that $p \in U$ and $f \in C^\infty(M)$ any. For $i = 1, \dots, n$ we define the vectors $x_i : C^\infty(M) \to I\!R$ given by,

$$x_i(f) := \frac{\partial}{\partial x^i}(f \circ \varphi^{-1})\Big|_{\varphi(p)}. \tag{3.1}$$

Note that these maps satisfy *i*) and *ii*) and therefore the x_i are really vectors. Note also that the right-hand side of **??** is well defined since we have the usual partial derivatives of maps between $I\!R^n$ and $I\!R$. These x_i depend on the chart (U, φ) but this does not matter in the proof since T_p does not depend on any chart. These vectors are linearly independent, that is if $x = \sum_{i=1}^{n} c^i \boldsymbol{x}_i = 0$ then $c^i = 0 \ \forall\, i = 1, \dots, n$. This is easily seen by considering the functions (strictly defined only in U), $f^j := x^j \circ \varphi$, since $\boldsymbol{x}_i(f^j) = \delta_i^j$ and therefore $0 = \boldsymbol{x}(f^j) = c^j$. It only remains to show that any vector $\boldsymbol{v}$ can be expressed as a linear combination of the $\boldsymbol{x}_i$. For this we will use the following result whose proof we leave as an exercise.

Lemma 3.1 *Let $F : I\!R^n \to I\!R$, $F \in C^\infty(I\!R^n)$ then for each $x_0 \in I\!R^n$ there exist functions $H_i : I\!R^n \to I\!R \in C^\infty(I\!R^n)$ such that $\forall\, x \in I\!R^n$ it holds*

$$F(x) = F(x_0) + \sum_{i=1}^{n} (x^i - x_0^i) H_i(x) \ \text{and} \tag{3.2}$$

also,

$$\left.\frac{\partial F}{\partial x^i}\right|_{x=x_0} = H_i(x_0). \tag{3.3}$$

We now continue the proof of the previous Theorem. Let $F = f \circ \varphi^{-1}$ and $x_0 = \varphi(p)$, then $\forall\, q \in U$ we have

$$f(q) = f(p) + \sum_{i=1}^{n} (x^i \circ \varphi(q) - x^i \circ \varphi(p)) H_i \circ \varphi(q) \tag{3.4}$$

Using $i)$ and $ii)$ we obtain,

$$\begin{aligned} \boldsymbol{v}(f) \ &= \boldsymbol{v}(f(p)) + \textstyle\sum_{i=1}^{n} (x^i \circ \varphi(q) - x^i \circ \varphi(p))\big|_{q=p} \ \boldsymbol{v}(H_i \circ \varphi) \\ &\quad + \textstyle\sum_{i=1}^{n} (H_i \circ \varphi)\big|_p \ \boldsymbol{v}(x^i \circ \varphi - x^i \circ \varphi(p)) \\ &= \textstyle\sum_{i=1}^{n} (H_i \circ \varphi)\big|_p \ \boldsymbol{v}(x^i \circ \varphi) \\ &= \textstyle\sum_{i=1}^{n} v^i \ \boldsymbol{x}_i(f) \end{aligned} \tag{3.5}$$

where $v^i \equiv \boldsymbol{v}(x^i \circ \varphi)$, and therefore we have expressed $\boldsymbol{v}$ as a linear combination of the $\boldsymbol{x}_i$, thus concluding the proof ♠

The basis $\{\boldsymbol{x}_i\}$ is called a **coordinate basis** and the $\{v^i\}$, the **components of $\boldsymbol{v}$** in that basis.

<u>Exercise</u>: If $(\tilde{U}, \tilde{\varphi})$ is another chart such that $p \in \tilde{U}$, then it will define another coordinate basis $\{\tilde{\boldsymbol{x}}_i\}$. Show that

$$\boldsymbol{x}_j = \sum_{i=1}^{n} \frac{\partial \tilde{x}^i}{\partial x^j} \tilde{\boldsymbol{x}}_i$$

where $\tilde{x}^i$ is the i-th component of the map $\tilde{\varphi} \circ \varphi^{-1}$. Also show that the relationship between the components is $\tilde{v}^i = \displaystyle\sum_{i=1}^{n} \frac{\partial \tilde{x}^i}{\partial x^j} v^j$.

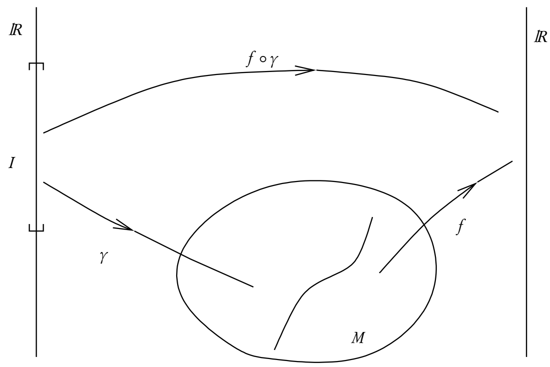

Figure 3.8: Definition of vector.

Example: Let $\gamma : I \to M$ be a curve in M. At each point $\gamma(t_o)$, $t_o \in I$, of M we can define a vector as follows, [See figure 3.8.]

$$\boldsymbol{t}(f) = \frac{d}{dt}\,(f \circ \gamma)|_{t=t_o}. \tag{3.6}$$

Its components in a coordinate basis are obtained through the functions

$$x^i(t) = x^i \circ \varphi \circ \gamma(t) \tag{3.7}$$

$$\begin{aligned}
\frac{d}{dt}(f \circ \gamma) &= \frac{d}{dt}(f \circ \varphi^{-1} \circ \varphi \circ \gamma) \\
&= \frac{d}{dt}(f \circ \varphi^{-1}(x^i(t))) \\
&= \sum_{i=1}^{n} \left(\frac{\partial}{\partial x^i}(f \circ \varphi^{-1})\right)\frac{dx^i}{dt} \\
&= \sum_{i=1}^{n} \frac{dx^i}{dt}\,\boldsymbol{x}_i(f)
\end{aligned} \tag{3.8}$$

3.5 Vector and Tensor Fields

If to each point q of M we assign a vector $\boldsymbol{v}|_q \in T_q$ we will have a **vector field**. This will be in $C^\infty(M)$ if given any $f \in C^\infty(M)$ the function $\boldsymbol{v}(f)$, which at each point p of M assigns the value $\boldsymbol{v}\,|_p\,(f)$, is also in $C^\infty(M)$. We will denote the set of C^∞ vector fields by TM and it is obviously a vector space of infinite dimension.

3.5.1 The Lie Bracket

Now consider the operation in the set TM of vector fields, $[\cdot,\cdot] : TM \times TM \to TM$. This operation is called the **Lie bracket** and given two vector fields (C^∞) it gives us a third:

$$[x,y](f) := x(y(f)) - y(x(f)). \tag{3.9}$$

Exercise:

1) Show that $[x,y]$ is indeed a vector field.

2) See that the **Jacobi identity** is satisfied:

$$[[x,y],z] + [[z,x],y] + [[y,z,x] = 0 \tag{3.10}$$

3) Let x^i and x^j be two vector fields coming from a coordinate system, that is $x^i(f) = \frac{\partial f}{\partial x^i}$, etc. Show that $\left[x^i, x^j\right] = 0$.

4) Given the components of x and y in a coordinate basis, what are those of $[x,y]$?

With this operation TM acquires the character of an algebra, called **Lie Algebra.**

3.5.2 Diffeomorphisms and the Theory of Ordinary Differential Equations

Definition: A **one-parameter group of diffeomorphisms** g^t is a map $I\!R \times M \to M$ such that:

1) For each fixed t it is a diffeomorphism [1] $M \to M$

2) For any pair of real numbers, $t, s \in I\!R$ we have $g^t \circ g^s = g^{t+s}$ (in particular $g^0 = id$).

We can associate with g^t a vector field in the following way: For a fixed p, $g^t(p) : I\!R \to M$ is a curve that at $t = 0$ passes through p and therefore defines a tangent vector at p, $v\,|_p$. Repeating the process for every point in M we have a vector field in M. Note that due to the group property satisfied by g^t, the tangent vector to the curve $g^t(p)$ is also tangent to the curve $g^s(g^t(p))$ at $s = 0$.

We can ask the inverse question: Given a smooth vector field v in M, does there exist a one-parameter group of diffeomorphisms that defines it? The answer to this question, which consists of finding all the integrable curves $g^t(p)$ that pass through each $p \in M$, is the theory of ordinary differential equations, –which will be the subject of our study in the following chapters– since it consists of solving the equations $\frac{dx^i}{dt} = v^i(x^j)$ with initial conditions $x^i(0) = \varphi^i(p)\ \ \forall\ p \in M$. As we will see, the answer is affirmative but only locally, that is, we can only find g^t defined in $I(\subset I\!R) \times U(\subset M) \to M$.

[1] That is, a smooth map with a smooth inverse.

Example: In $I\!R^1$ let the vector have the coordinate component x^2, that is $v(x) = x^2\frac{\partial}{\partial x}$. The ordinary differential equation associated with this vector is $\frac{dx}{dt} = x^2$, whose solution is

$$t - t_0 = \frac{-1}{x} + \frac{1}{x_0} \text{ ó } x(t) = \frac{-1}{t - \frac{1}{x_0}} \tag{3.11}$$

where we have taken $t_0 = 0$. That is, $g^t(x_0) = \frac{-1}{t-\frac{1}{x_0}}$. Note that for any t this map is not defined for all $I\!R$ and therefore is not a diffeomorphism. Also note that for any interval we take for its definition, the time interval of the solution's existence will be finite, either towards the future or the past.

Example: Let g^t be a linear diffeomorphism in $I\!R$, that is $g^t(x+\alpha y) = g^t(x) + \alpha g^t(y)$. Then it has the form $g^t(x) = f(t)x$. The group property implies $f(t)\cdot f(s) = f(t+s)$ or $f(t) = c\,e^{kt} = e^{kt}$, since $g^0 = id$. Therefore $g^t(x) = e^{kt}\,x$. The associated differential equation is: $x(t) = e^{kt}\,x_0 \implies \dot{x} = k\,e^{kt}\,x_0 = \boxed{k\,x = \dot{x}.}$

Exercise: Plot in a neighborhood of the origin in $I\!R^2$ the integral curves and therefore g^t of the following linear systems.

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{3.12}$$

a) $k > 1$ b) $k = 1$ c) $0 < k < 1$, d) $k = 0$, e) $k < 0$

3.5.3 Covector and Tensor Fields

Just as we introduced the notion of a vector field, we can also introduce the notion of a covector field, that is, a smooth map from M to T^*_p. This will act on vector fields giving as a result functions on M. In the following example, we see how the field **differential of** f is defined.

Example: Let $f \in C^\infty_p$. A vector at $p \in M$ is a derivation on functions in C^∞_p, $v(f) \in I\!R$. But given v_1 and $v_2 \in T_p$, $a \in I\!R$ and $f \in C^\infty_p$, $(v_1 + av_2)(f) = v_1(f) + av_2(f)$ and therefore each given f defines a linear functional $\boldsymbol{df}\,|_p$: $T_p \to I\!R$, called the differential of f, that is an element of T^*_p,

$$\boldsymbol{df}(v) := v(f), \quad \forall v \in T_p.$$

In this way, the differential of a function, $\boldsymbol{df}$, is a covector that when acting on a vector v gives us the number *the derivative of f at the point p in the direction of v.*

Let f be a smooth function on M, $a \in I\!R$ and consider the subset S_a of M such that $f(S_a) = a$. It can be seen that if $df \neq 0$ this will be a submanifold of M, that is, a surface embedded in M, of dimension $n-1$. The condition $df|_p(v) = 0$ on vectors of T_p with $p \in S_a$ means that these are actually tangent vectors to S_a, that is, elements of $T_p(S_a)$. On the contrary, if $df(v)|_p \neq 0$ then at that point v *pierces* S_a.

Example: The function $f(x,y,z) = x^2 + y^2 + z^2$ in $I\!R^3$.

$S_a = \{(x,y,z) \in I\!R^3 | f(x,y,z) = a^2, \;\; a > 0\}$ is the sphere of radius a, and as we have already seen, a manifold. Let (v^x, v^y, v^z) be a vector at the point $(x,y,z) \in I\!R^3$, then the condition $\boldsymbol{df}(v) = 2(xv^x + yv^y + zv^z) = 0$ implies that v is tangent to S. Indeed, we see that this is the condition that tells us that v is *perpendicular* to (x,y,z) when we are in the conventional Euclidean structure.

Given a coordinate system (chart) that covers a point $p \in M$, we have seen that we have a canonical basis of T_p associated with it given by the vectors,

$$\boldsymbol{x}_i(f) := \frac{\partial f \circ \phi^{-1}}{\partial x^i}|_{\phi(p)}.$$

What will be the associated cobasis? Note that the coordinate system also gives us a set of n privileged functions, that is, the components of the map ϕ that defines the chart, $\{x^j\}$, $j = 1..n$, $x^i(p)$:=*value of the* i*-th coordinate assigned by* ϕ *to the point* p. Note that $x^i \circ \varphi^{-1}$ is then the identity map for the i-th coordinate. If we apply the basis vectors to these functions, we obtain,

$$\boldsymbol{x}_i(x^j) := \frac{\partial x^j \circ \varphi^{-1}}{\partial x^i}|_{\phi(p)} = \delta^j_i,$$

but then, since $\boldsymbol{dx}^j(\boldsymbol{x}_i) = \boldsymbol{x}_i(x^j) = \delta^j_i$, we see that the differentials $\boldsymbol{dx}^j$ are the cobasis of the coordinate basis. In particular, we have that the coordinate components of a vector v are given by:

$$v^j = \boldsymbol{dx}^j(v),$$

and the components of a covector $\boldsymbol{\omega} \in T^*_p$ by,

$$\omega_i = \boldsymbol{\omega}(\boldsymbol{x}_i).$$

Similarly, we define **tensor fields** as the multilinear maps that when acting on vector and covector fields give functions from the manifold to the reals and that at each point of the manifold only depend on the vectors and covectors defined at that point. This last clarification is necessary because otherwise, we would include among the tensors, for example, line integrals over vector fields.

3.5.4 The Metric

Let M be an n-dimensional manifold. We have previously defined on M the notions of curves, vector fields, and covector fields, etc., but not a notion of distance between its points, that is, a function $d: M \times M \to I\!R$ that takes any two points, p and q of M and gives us a number $d(p,q)$ satisfying,

1. $d(p,q) \geq 0$.
2. $d(p,q) = 0 \longleftrightarrow p = q$.
3. $d(p,q) = d(q,p)$.
4. $d(p,q) \leq d(p,r) + d(r,q)$.

This, and in some cases a notion of pseudo-distance [where 1) and 2) are not satisfied], is fundamental if we want to have a mathematical structure that is useful for the description of physical phenomena. For example, Hooke's law, which tells us that the force applied to a spring is proportional to its elongation (a distance), clearly needs this entity. Next, we will introduce a notion of infinitesimal distance, that is, between two infinitesimally separated points, which corresponds to the Euclidean notion of distance and allows us to develop a notion of global distance, that is, between any two points of M.

The idea is then to have a concept of distance (or pseudo-distance) between two *infinitesimally close* points, that is, two points connected by an *infinitesimal displacement*, that is, connected by a vector. The notion we need is then that of the norm of a vector. Since a manifold is locally like $\mathbb{R}^n$, in the sense that the space of tangent vectors at a point p, T_pM is $\mathbb{R}^n$, it is reasonable to consider there the notion of Euclidean distance, that is, the distance between two points x_0 and $x_1 \in \mathbb{R}^n$ is the square root of the sum of the squares of the components (in some coordinate system) of the vector connecting these two points. The problem with this is that such a notion depends on the coordinate system being used and therefore there will be as many distances as coordinate systems covering the point p. This is just an indication that the structure we have so far does not contain a privileged or natural notion of distance. This must be introduced as an additional structure. One way to obtain infinitesimal distances independent of the coordinate system (that is, geometric) is by introducing at each point $p \in M$ a tensor of type $\binom{0}{2}$, symmetric $[\mathbf{g}(\mathbf{u},\mathbf{v}) = \mathbf{g}(\mathbf{v},\mathbf{u})\ \forall\, \mathbf{u},\ \mathbf{v}\ \in T_pM]$ and non-degenerate $[\mathbf{g}(\mathbf{u},\mathbf{v}) = \mathbf{0}\ \forall \mathbf{v}\ \in T_pM\ \ \Rightarrow \mathbf{u} = \mathbf{0}]$. If we also require that this tensor be positive definite $[\mathbf{g}(\mathbf{u},\mathbf{u}) \geq \mathbf{0}\ (=\ \Leftrightarrow\ \mathbf{u} = \mathbf{0})]$ it can be easily seen that this defines an inner product in T_pM (or pseudo-inner product if $\mathbf{g}(\mathbf{u},\mathbf{u}) = \mathbf{0}$ for some $u \neq 0 \in T_pM$). [2] If we make a smooth choice for this tensor at each point of M we will obtain a smooth tensor field called the **metric** of M. This extra structure, a tensor field with certain properties, is what allows us to build the mathematical foundations to then construct much of physics on it.

Let $\mathbf{g}$ be a metric on M, given any point p of M there exists a coordinate system in which its components are

$$g_{ij} = \delta_{ij}$$

and therefore gives rise to the Euclidean inner product, however, in general, this result cannot be extended to a neighborhood of the point and in general, its components will depend there on the coordinates. Note that this is what we wanted to do

[2] Later we will see that an inner product gives rise to a distance, correspondingly a pseudo-inner product gives rise to a pseudo-distance.

initially, but now by defining this norm via a vector we have given it an invariant character.

Restricting ourselves now to positive definite metrics, we define **the norm** of a vector $v \in T_p$ as $|v| = \sqrt{|g(v,v)|}$, that is, as the infinitesimal distance divided by ϵ between p and the point $\gamma(\epsilon)$ where $\gamma(t)$ is a curve such that $\gamma(0) = p$, $\frac{d\gamma(t)}{dt}|_{t=0} = v$. Similarly, we can define the length of a smooth curve $\gamma(t) : [0,1] \to M$ by the formula,

$$L(\lambda) = \int_0^1 \sqrt{\mathbf{g}(\mathbf{v},\mathbf{v})}\, dt, \tag{3.13}$$

where $\mathbf{v}(t) = \frac{d\gamma(t)}{dt}$. We see then that we define the length of a curve by measuring the infinitesimal lengths between nearby points on it and then integrating with respect to t.

Exercise: Prove that the length $L(\gamma)$ is independent of the chosen parameter.

We define the distance between two points $p, q \in M$ as,

$$d_g(p,q) = \inf_{\{\gamma(t)\,:\,\gamma(0)=p,\gamma(1)=q\}} |L(\gamma)| \tag{3.14}$$

That is, as the infimum of the length of all curves connecting p with q.

Exercise: Find an example of a manifold with two points such that the infimum in the previous definition is not a minimum. That is, where there is no curve connecting the two points with the minimum distance between them.

Exercise: a) The Euclidean metric in $\mathbb{R}^2$ is $(dx)^2 + (dy)^2$, where $\{dx, dy\}$ is the cobasis associated with $\{\partial x, \partial y\}$. What is the distance between two points in this case?

Exercise: b) What is the form of the Euclidean metric in $\mathbb{R}^3$ in spherical coordinates? And in cylindrical coordinates?

Exercise: c) The metric of the sphere is $(d\theta)^2 + \sin^2\theta\,(d\varphi)^2$. What is the distance in this case? For which points p, q are there multiple curves γ_i with $L(\gamma_i) = d(p,q)$?

Exercise: d) The metric $(dx)^2+(dy)^2+(dz)^2-(dt)^2$ in $\mathbb{R}^4$ is the Minkowski metric of special relativity. What is the *distance* between the point with coordinates $(0,0,0,0)$ and $(1,0,0,1)$?

A metric gives us a privileged map between the space of tangent vectors at p, T_p, and its dual T_p^* for each p in M, that is, the map that assigns to each vector $v \in T_p$ the covector $g(v, \ ) \in T_p$. Since this is valid for each p, we thus obtain a map between vector and covector fields.

3.5.5 Diffeomorphisms and the Theory of Ordinary Differential Equations

Definition: A **one-parameter group of diffeomorphisms** g^t is a map $I\!R \times M \to M$ such that:

1) For each fixed t it is a diffeomorphism [3] $M \to M$

2) For any pair of real numbers, $t, s \in I\!R$ we have $g^t \circ g^s = g^{t+s}$ (in particular $g^0 = id$).

We can associate with g^t a vector field in the following way: For a fixed p, $g^t(p) : I\!R \to M$ is a curve that at $t = 0$ passes through p and therefore defines a tangent vector at p, $v\,|_p$. Repeating the process for every point in M we have a vector field in M. Note that due to the group property satisfied by g^t, the tangent vector to the curve $g^t(p)$ is also tangent to the curve $g^s(g^t(p))$ at $s = 0$.

We can ask the inverse question: Given a smooth vector field v in M, does there exist a one-parameter group of diffeomorphisms that defines it? The answer to this question, which consists of finding all the integrable curves $g^t(p)$ that pass through each $p \in M$, is the theory of ordinary differential equations, –which will be the subject of our study in the following chapters– since it consists of solving the equations $\frac{dx^i}{dt} = v^i(x^j)$ with initial conditions $x^i(0) = \varphi^i(p) \ \ \forall \ p \in M$. As we will see, the answer is affirmative but only locally, that is, we can only find g^t defined in $I(\subset I\!R) \times U(\subset M) \to M$.

Example: In $I\!R^1$ let the vector have the coordinate component x^2, that is $v(x) = x^2 \frac{\partial}{\partial x}$. The ordinary differential equation associated with this vector is $\frac{dx}{dt} = x^2$, whose solution is

$$t - t_0 = \frac{-1}{x} + \frac{1}{x_0} \text{ ó } x(t) = \frac{-1}{t - \frac{1}{x_0}} \tag{3.15}$$

where we have taken $t_0 = 0$. That is, $g^t(x_0) = \frac{-1}{t - \frac{1}{x_0}}$. Note that for any t this map is not defined for all $I\!R$ and therefore is not a diffeomorphism. Also note that for any interval we take for its definition, the time interval of the solution's existence will be finite, either towards the future or the past.

Example: Let g^t be a linear diffeomorphism in $I\!R$, that is $g^t(x + \alpha y) = g^t(x) + \alpha\, g^t(y)$. Then it has the form $g^t(x) = f(t)x$. The group property implies $f(t) \cdot$

[3] That is, a smooth map with a smooth inverse.

$f(s) = f(t+s)$ or $f(t) = c\,e^{kt} = e^{kt}$, since $g^0 = id$. Therefore $g^t(x) = e^{kt}\,x$. The associated differential equation is: $x(t) = e^{kt}\,x_0 \implies \dot{x} = k\,e^{kt}\,x_0 = \boxed{k\,x = \dot{x}.}$

Exercise: Plot in a neighborhood of the origin in $\mathbb{R}^2$ the integral curves and therefore g^t of the following linear systems.

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{3.16}$$

a) $k > 1$ b) $k = 1$ c) $0 < k < 1$, d) $k = 0$, e) $k < 0$

3.5.6 Covector and Tensor Fields

Just as we introduced the notion of a vector field, we can also introduce the notion of a covector field, that is, a smooth map from M to T_p^*. This will act on vector fields giving as a result functions on M. In the following example, we see how the field **differential of** f is defined.

Example: Let $f \in C_p^\infty$. A vector at $p \in M$ is a derivation on functions in C_p^∞, $v(f) \in \mathbb{R}$. But given v_1 and $v_2 \in T_p$, $a \in \mathbb{R}$ and $f \in C_p^\infty$, $(v_1 + a v_2)(f) = v_1(f) + a v_2(f)$ and therefore each given f defines a linear functional $\boldsymbol{df}\,|_p$: $T_p \to \mathbb{R}$, called the differential of f, that is an element of T_p^*,

$$\boldsymbol{df}(v) := v(f), \quad \forall v \in T_p.$$

In this way, the differential of a function, $\boldsymbol{df}$, is a covector that when acting on a vector v gives us the number *the derivative of f at the point p in the direction of v.*

Let f be a smooth function on M, $a \in \mathbb{R}$ and consider the subset S_a of M such that $f(S_a) = a$. It can be seen that if $df \neq 0$ this will be a submanifold of M, that is, a surface embedded in M, of dimension $n-1$. The condition $df|_p(v) = 0$ on vectors of T_p with $p \in S_a$ means that these are actually tangent vectors to S_a, that is, elements of $T_p(S_a)$. On the contrary, if $df(v)|_p \neq 0$ then at that point v *pierces* S_a.

Example: The function $f(x,y,z) = x^2 + y^2 + z^2$ in $\mathbb{R}^3$.

$S_a = \{(x,y,z) \in \mathbb{R}^3 | f(x,y,z) = a^2,\ a > 0\}$ is the sphere of radius a, and as we have already seen, a manifold. Let (v^x, v^y, v^z) be a vector at the point $(x,y,z) \in \mathbb{R}^3$, then the condition $\boldsymbol{df}(v) = 2(xv^x + yv^y + zv^z) = 0$ implies that v is tangent to S. Indeed, we see that this is the condition that tells us that v is *perpendicular* to (x,y,z) when we are in the conventional Euclidean structure.

Given a coordinate system (chart) that covers a point $p \in M$, we have seen that we have a canonical basis of T_p associated with it given by the vectors,

$$\boldsymbol{x}_i(f) := \frac{\partial f \circ \phi^{-1}}{\partial x^i}|_{\phi(p)}.$$

What will be the associated cobasis? Note that the coordinate system also gives us a set of n privileged functions, that is, the components of the map ϕ that defines the chart, $\{x^j\}$, $j = 1..n$, $x^i(p) :=$ *value of the i-th coordinate assigned by ϕ to the point p*. Note that $x^i \circ \varphi^{-1}$ is then the identity map for the i-th coordinate. If we apply the basis vectors to these functions, we obtain,

$$\boldsymbol{x}_i(x^j) := \frac{\partial x^j \circ \varphi^{-1}}{\partial x^i}|_{\phi(p)} = \delta^j_i,$$

but then, since $\boldsymbol{dx}^j(\boldsymbol{x}_i) = \boldsymbol{x}_i(x^j) = \delta^j_i$, we see that the differentials $\boldsymbol{dx}^j$ are the cobasis of the coordinate basis. In particular, we have that the coordinate components of a vector $\boldsymbol{v}$ are given by:

$$v^j = \boldsymbol{dx}^j(v),$$

and the components of a covector $\boldsymbol{\omega} \in T^*_p$ by,

$$\omega_i = \boldsymbol{\omega}(\boldsymbol{x}_i).$$

Similarly, we define **tensor fields** as the multilinear maps that when acting on vector and covector fields give functions from the manifold to the reals and that at each point of the manifold only depend on the vectors and covectors defined at that point. This last clarification is necessary because otherwise, we would include among the tensors, for example, line integrals over vector fields.

3.5.7 The Metric

Let M be an n-dimensional manifold. We have previously defined on M the notions of curves, vector fields, and covector fields, etc., but not a notion of distance between its points, that is, a function $d : M \times M \to I\!R$ that takes any two points, p and q of M and gives us a number $d(p,q)$ satisfying,

1. $d(p,q) \geq 0$.
2. $d(p,q) = 0 \longleftrightarrow p = q$.
3. $d(p,q) = d(q,p)$.
4. $d(p,q) \leq d(p,r) + d(r,q)$.

This, and in some cases a notion of pseudo-distance [where 1) and 2) are not satisfied], is fundamental if we want to have a mathematical structure that is useful for the description of physical phenomena. For example, Hooke's law, which tells us that the force applied to a spring is proportional to its elongation (a distance), clearly needs this entity. Next, we will introduce a notion of infinitesimal distance, that is, between two infinitesimally separated points, which corresponds to the Euclidean notion of distance and allows us to develop a notion of global distance, that is, between any two points of M.

The idea is then to have a concept of distance (or pseudo-distance) between two *infinitesimally close* points, that is, two points connected by an *infinitesimal displacement*, that is, connected by a vector. The notion we need is then that of the norm of a vector. Since a manifold is locally like $I\!R^n$, in the sense that the space of tangent vectors at a point p, T_pM is $I\!R^n$, it is reasonable to consider there the notion of Euclidean distance, that is, the distance between two points x_0 and $x_1 \in I\!R^n$ is the square root of the sum of the squares of the components (in some coordinate system) of the vector connecting these two points. The problem with this is that such a notion depends on the coordinate system being used and therefore there will be as many distances as coordinate systems covering the point p. This is just an indication that the structure we have so far does not contain a privileged or natural notion of distance. This must be introduced as an additional structure. One way to obtain infinitesimal distances independent of the coordinate system (that is, geometric) is by introducing at each point $p \in M$ a tensor of type $\binom{0}{2}$, symmetric $[\mathbf{g}(\mathbf{u},\mathbf{v}) = \mathbf{g}(\mathbf{v},\mathbf{u})\ \forall\ \mathbf{u},\ \mathbf{v}\ \in T_pM]$ and non-degenerate $[\mathbf{g}(\mathbf{u},\mathbf{v}) = \mathbf{0}\ \forall \mathbf{v}\ \in T_pM\ \Rightarrow \mathbf{u} = \mathbf{0}]$. If we also require that this tensor be positive definite $[\mathbf{g}(\mathbf{u},\mathbf{u}) \geq \mathbf{0}\ (=\ \Leftrightarrow\ \mathbf{u} = \mathbf{0})]$ it can be easily seen that this defines an inner product in T_pM (or pseudo-inner product if $\mathbf{g}(\mathbf{u},\mathbf{u}) = \mathbf{0}$ for some $u \neq 0 \in T_pM$). [4] If we make a smooth choice for this tensor at each point of M we will obtain a smooth tensor field called the **metric** of M. This extra structure, a tensor field with certain properties, is what allows us to build the mathematical foundations to then construct much of physics on it.

Let g be a metric on M, given any point p of M there exists a coordinate system in which its components are

$$g_{ij} = \delta_{ij}$$

and therefore gives rise to the Euclidean inner product, however, in general, this result cannot be extended to a neighborhood of the point and in general, its components will depend there on the coordinates. Note that this is what we wanted to do initially, but now by defining this norm via a vector we have given it an invariant character.

Restricting ourselves now to positive definite metrics, we define **the norm** of a vector $v \in T_p$ as $|v| = \sqrt{|g(v,v)|}$, that is, as the infinitesimal distance divided by ϵ between p and the point $\gamma(\epsilon)$ where $\gamma(t)$ is a curve such that $\gamma(0) = p$, $\frac{d\gamma(t)}{dt}|_{t=0} = v$. Similarly, we can define the length of a smooth curve $\gamma(t) : [0,1] \to M$ by the formula,

$$L(\lambda) = \int_0^1 \sqrt{\mathbf{g}(\mathbf{v},\mathbf{v})}\, dt, \tag{3.17}$$

where $\mathbf{v}(t) = \frac{d\gamma(t)}{dt}$. We see then that we define the length of a curve by measuring the infinitesimal lengths between nearby points on it and then integrating with respect to t.

[4]Later we will see that an inner product gives rise to a distance, correspondingly a pseudo-inner product gives rise to a pseudo-distance.

Exercise: Prove that the length $L(\gamma)$ is independent of the chosen parameter.

We define the distance between two points $p,q \in M$ as,

$$d_g(p,q) = \inf_{\{\gamma(t)\,:\,\gamma(0)=p,\gamma(1)=q\}} |L(\gamma)| \tag{3.18}$$

That is, as the infimum of the length of all curves connecting p with q.

Exercise: Find an example of a manifold with two points such that the infimum in the previous definition is not a minimum. That is, where there is no curve connecting the two points with the minimum distance between them.

Exercise: a) The Euclidean metric in $I\!R^2$ is $(dx)^2+(dy)^2$, where $\{dx,dy\}$ is the cobasis associated with $\{\partial x,\partial y\}$. What is the distance between two points in this case?

Exercise: b) What is the form of the Euclidean metric in $I\!R^3$ in spherical coordinates? And in cylindrical coordinates?

Exercise: c) The metric of the sphere is $(d\theta)^2 + \sin^2\theta\,(d\varphi)^2$. What is the distance in this case? For which points p,q are there multiple curves γ_i with $L(\gamma_i) = d(p,q)$?

Exercise: d) The metric $(dx)^2+(dy)^2+(dz)^2-(dt)^2$ in $I\!R^4$ is the Minkowski metric of special relativity. What is the *distance* between the point with coordinates $(0,0,0,0)$ and $(1,0,0,1)$?

A metric gives us a privileged map between the space of tangent vectors at p, T_p, and its dual T_p^* for each p in M, that is, the map that assigns to each vector $v \in T_p$ the covector $g(v,\ ) \in T_p$. Since this is valid for each p, we thus obtain a map between vector and covector fields.

Since g is non-degenerate, this map is invertible, that is, there exists a symmetric tensor of type $\binom{2}{0}$, g^{-1}, such that

$$g(g^{-1}(\theta,\ ),\ ) = \theta \tag{3.19}$$

for any co-vector field θ. This indicates that when we have a manifold with a metric, it becomes irrelevant to distinguish between vectors and co-vectors or, for example, between tensors of type $\binom{0}{2}$, $\binom{2}{0}$, or $\binom{1}{1}$.

3.5.8 Abstract Index Notation

When working with tensorial objects, the notation used so far is not the most convenient because it is difficult to remember the type of each tensor, in which slot it "eats" other objects, etc. One solution is to introduce a coordinate system and work with the components of the tensors, where having indices makes it easy to know what objects they are or to introduce general bases. In this way, for example, we represent the vector $\boldsymbol{l} = l^i \frac{\partial}{\partial x^i}$ by its components $\{l^i\}$. A convenience of this notation is that "eating" becomes "contracting," since, for example, we represent the vector $\boldsymbol{l}$ "eating" a function f by the contraction of the coordinate components of the vector and the differential of f:

$$l(f) = \sum_{i=1}^{n} l^i \frac{\partial f}{\partial x^i}.$$

But a serious drawback of this representation is that it initially depends on the coordinate system and therefore all the expressions we construct with it have the potential danger of depending on such a system.

We will remedy this by introducing **abstract indices** (which will be Latin letters) that indicate where the coordinate indices would go but nothing more, that is, they do not depend on the coordinate system and do not even take numerical values, that is, l^a does not mean the n-tuple $(l^1, l^2, \dots, l^n)$ as if they were indices. In this way, l^a will denote the vector l, θ_a the co-vector θ, and g_{ab} the metric g. A contraction such as $g(v, \ )$ will be denoted $g_{ab}v^a$ and we will denote this co-vector by v_b, that is, the action of g_{ab} is to lower the index of v^a and give the co-vector $v_b \equiv v^a g_{ab}$. Similarly, we will denote $\boldsymbol{g}^{-1}$ (the inverse of $\boldsymbol{g}$) as g^{ab}, that is, $\boldsymbol{g}$ with the indices raised.

The symmetry of $\boldsymbol{g}$ is then equivalent to $g_{ab} = g_{ba}$.

Exercise: How would you denote an antisymmetric tensor of type $\binom{0}{2}$?

Using repeated indices for contraction, we see that $l(f)$ can be denoted by $l^a \nabla_a f$ where $\nabla_a f$ denotes the differential co-vector of f, while the vector $\nabla^a f := g^{ab} \nabla_b f$ is called the **gradient** of f and we see that it depends not only on f but also on $\boldsymbol{g}$.

3.6 Covariant Derivative

We have seen that in M there is the notion of the derivative of a scalar field f, which is the differential co-vector of f that we denote $\nabla_a f$. Is there the notion of the derivative of a tensor field? For example, is there an extension of the operator ∇_a to vectors such that if l^a is a differentiable vector then $\nabla_a l^b$ is a tensor of type $\binom{1}{1}$? To fix ideas, let us define this extension of the differential ∇_a, called the covariant derivative, by requiring it to satisfy the following properties: [5]

i) Linearity: If $A^{a_1 \cdots a_k}_{b_1 \dots b_l}, B^{a_1 \cdots a_k}_{b_1 \dots b_l}$ are tensors of type $\binom{k}{l}$ and $\alpha \in I\!R$ then

$$\nabla_c \left(\alpha A^{a_1 \cdots a_k}_{b_1 \dots b_l} + B^{a_1 \cdots a_k}_{b_1 \dots b_l} \right) = \alpha \nabla_c A^{a_1 \cdots a_k}_{b_1 \dots b_l} + \nabla_c B^{a_1 \cdots a_k}_{b_1 \dots b_l} \qquad (3.20)$$

[5] Note that they are an extension of those required to define derivations.

ii) Leibnitz:

$$\nabla_e \left(A^{a_1 \cdots a_k}_{b_1 \dots b_l} B^{c_1 \dots c_m}_{d_1 \dots d_n} \right) = A^{a_1 \cdots a_k}_{b_1 \dots b_l} \left(\nabla_e B^{c_1 \dots c_m}_{d_1 \dots d_n} \right) + \left(\nabla_e A^{a_1 \cdots a_k}_{b_1 \dots b_l} \right) B^{c_1 \dots c_m}_{d_1 \dots d_n} \tag{3.21}$$

iii) Commutativity with contractions:

$$\nabla_e \left(\delta^c{}_d A^{a_1, \dots, d, \dots, a_k}_{b_1, \dots, c, \dots, b_k} \right) = \delta^c{}_d \nabla_e A^{a_1, \dots, d, \dots, a_k}_{b_1, \dots, c, \dots, b_k}, \tag{3.22}$$

where $\delta^c{}_d$ is the identity tensor. That is, if we first contract some indices of a tensor and then take its derivative, we obtain the same tensor as if we first take the derivative and then contract.
iv) Consistency with the differential: If l^a is a vector field and f a scalar field, then

$$l^a \nabla_a f = l(f) \tag{3.23}$$

v) Zero torsion: If f is a scalar field then

$$\nabla_a \nabla_b f = \nabla_b \nabla_a f \tag{3.24}$$

Example: Let $\{x^i\}$ be a global coordinate system in $I\!R^n$ and let ∇_c be the operator that, when acting on $A^{a_1 \cdots a_k}_{b_1 \dots b_l}$, generates the tensor field that in these coordinates has components

$$\partial_j A^{i_1 \dots i_k}_{j_1 \dots j_l} \tag{3.25}$$

By definition, it is a tensor and clearly satisfies all the conditions of the definition, since it satisfies them in this coordinate system, so it is a covariant derivative. If we take another coordinate system, we will obtain another covariant derivative, generally different from the previous one. For example, let us act ∇_c on the vector l^a, then

$$(\nabla_c l^a)^i_j = \partial_j l^i. \tag{3.26}$$

In another coordinate system $\{\bar{x}^i\}$ this tensor has components

$$(\nabla_c l^a)^l_k = \sum_{i,j=1}^{n} \frac{\partial \bar{x}^l}{\partial x^i} \frac{\partial x^j}{\partial \bar{x}^k} \frac{\partial l^i}{\partial x^j} \tag{3.27}$$

which are not, in general, the components of the covariant derivative $\bar{\nabla}_c$ that these new coordinates define, indeed

$$\begin{aligned}
\left(\bar{\nabla}_c l^a\right)^l_k &\equiv \frac{\partial \bar{l}^l}{\partial \bar{x}^k} = \textstyle\sum_{j=1}^n \left(\frac{\partial x^j}{\partial \bar{x}^k}\right) \frac{\partial}{\partial x^j} \sum_{i=1}^n \left(\frac{\partial \bar{x}^l}{\partial x^i} l^i\right) = \\
&= \textstyle\sum_{i,j=1}^n \left(\frac{\partial \bar{x}^l}{\partial x^i}\right)\left(\frac{\partial x^j}{\partial \bar{x}^k}\right) \frac{\partial l^i}{\partial x^j} + \sum_{i,j=1}^n \frac{\partial x^j}{\partial \bar{x}^k} \left(\frac{\partial^2 \bar{x}^l}{\partial x^j \partial x^i}\right) l^i \\
&= (\nabla_c l^a)^i_j + \textstyle\sum_{i,j=1}^n \frac{\partial x^j}{\partial \bar{x}^k} \left(\frac{\partial^2 \bar{x}^l}{\partial x^j \partial x^i}\right) l^i,
\end{aligned} \tag{3.28}$$

which clearly shows that they are two different tensors and that their difference is a **tensor** that depends linearly and **not differentially** on l^a. Is this true in general? That is, given two connections, ∇_c and $\bar{\nabla}_c$, is their difference a tensor (and not a differential operator)? We will see that this is true.

Theorem 3.2 *The difference between two connections is a tensor.*

Proof: Note that by properties iii) and iv) of the definition, if we know how ∇_c acts on co-vectors, we know how it acts on vectors and thus by i) and ii) on any tensor. Indeed, if we know $\nabla_c w_a$ for any w_a, then $\nabla_c l^a$ is the tensor of type $\binom{1}{1}$ such that when contracted with an arbitrary w_a gives us the co-vector

$$(\nabla_c l^a) w_a = \nabla_c (w_a l^a) - l^a (\nabla_c w_a), \tag{3.29}$$

which we know since by iv) we also know how ∇_c acts on scalars. Therefore, it is sufficient to see that

$$\left(\bar{\nabla}_c - \nabla_c\right) w_a = C^b{}_{ca} w_b \tag{3.30}$$

for some tensor $C^b{}_{ca}$. First, let us prove that given any $p \in M$, $\left(\bar{\nabla}_c - \nabla_c\right) w_a|_p$ depends only on $w_a|_p$ and not on its derivative. Let w'_a be any other co-vector such that at p they coincide, that is, $(w_a - w'_a)|_p = 0$. Then given a smooth co-basis $\{\mu^i_a\}$ in a neighborhood of p, we will have that $w_a - w'_a = \sum_i f_i \mu^i_a$ with f_i smooth functions that vanish at p. At p we then have

$$\begin{array}{rcl} \bar{\nabla}_c (w_a - w'_a) - \nabla_c (w_a - w'_a) & = & \sum_i \bar{\nabla}_c \left(f_i \mu^i_a\right) - \sum_i \nabla_c \left(f_i \mu^i_a\right) \\ & = & \sum_i \mu^i_a \left(\bar{\nabla}_c f_i - \nabla_c f_i\right) = 0 \end{array} \tag{3.31}$$

since by *iv)* $\bar{\nabla}_c$ and ∇_c must act in the same way on scalars –and in particular on the f_i–. This shows that

$$\left(\bar{\nabla}_c - \nabla_c\right) w'_a = \left(\bar{\nabla}_c - \nabla_c\right) w_a \tag{3.32}$$

and therefore that $\left(\bar{\nabla}_c - \nabla_c\right) w_a$ depends only on $w_a|_p$ and obviously in a linear way. But then $\left(\bar{\nabla}_c - \nabla_c\right)$ must be a tensor of type $\binom{1}{2}$ that is waiting to "eat" a co-vector to give us the tensor of type $\binom{0}{2}$, $\left(\bar{\nabla}_c - \nabla_c\right) w_a$. That is, $\left(\bar{\nabla}_c - \nabla_c\right) w_a = C^b{}_{ca} w_b$, which proves the theorem.

Note that condition v) tells us that $\nabla_a \nabla_b f = \nabla_b \nabla_a f$, taking $w_a = \nabla_a f$ we get

$$\begin{array}{rcl} \bar{\nabla}_a \bar{\nabla}_b f & = & \bar{\nabla}_a \nabla_b f \\ & = & \bar{\nabla}_a w_b \\ & = & \nabla_a w_b + C^c_{ab} \nabla_c f \\ & = & \nabla_a \nabla_b f + C^c_{ab} \nabla_c f. \end{array} \tag{3.33}$$

Since $\nabla_c f|_p$ can be any co-vector, we see that the condition of no torsion implies that $C^c{}_{ab}$ is symmetric in the lower indices, $C^c{}_{ab} = C^c{}_{ba}$.

Exercise: How does $\left(\bar{\nabla}_c - \nabla_c\right)$ act on vectors?

Exercise: Express the Lie bracket in terms of any connection and then explicitly prove that it does not depend on the connection used.

Exercise: Let $A_{b,\ldots,z}$ be a totally antisymmetric tensor. Show that $\nabla_{[a}A_{b,\ldots,z]}$, that is, the total antisymmetrization of $\nabla_a A_{b,\ldots,z}$, does not depend on the covariant derivative used.

The difference between any connection ∇_c and one coming from a coordinate system $\{x^i\}$ is a tensor called the **Christoffel symbol** of ∇_c with respect to the coordinates $\{x^i\}$, Γ^b_{ca},

$$\nabla_c w_a = \partial_c w_a + \Gamma^b_{ca} w_b. \tag{3.34}$$

The knowledge of this tensor is very useful in practice, as it allows us to express ∇_c in terms of the corresponding coordinate connection, ∂_c.

As we have seen, in a manifold M there are infinite ways to *take the derivative of a tensor*. Is there any natural or privileged one? The answer is no, unless we add more structure to M. Intuitively, the reason for this is that in M we do not know how to compare $l^a|_p$ with $l^a|_q$ if p and q are two different points. [6]

Is the presence of a metric in M sufficient to make this comparison? The answer is yes!

Theorem 3.3 *Let g_{ab} be a (smooth) metric on M, then there exists a unique covariant derivative ∇_c such that $\nabla_c g_{ab} = 0$.*

Proof: Let $\bar{\nabla}_c$ be any connection and let ∇_c be such that $\nabla_c g_{ab} = 0$, it is sufficient to show that this condition uniquely determines the difference tensor, $C^d{}_{ca}$. But,

$$0 = \nabla_a g_{bc} = \bar{\nabla}_a g_{bc} - C^d{}_{ab} g_{dc} - C^d{}_{ac} g_{bd} \tag{3.35}$$

that is

$$C_{cab} + C_{bac} = \bar{\nabla}_a g_{bc} \tag{3.36}$$

but also

$$\begin{array}{llcl} a \leftrightarrow b & C_{cba} + C_{abc} & = & \bar{\nabla}_b g_{ac} \\ c \leftrightarrow b & C_{bca} + C_{acb} & = & \bar{\nabla}_c g_{ab}. \end{array} \tag{3.37}$$

Adding the last two, subtracting the first and using the symmetry of the last two indices we get

$$2C_{abc} = \bar{\nabla}_b g_{ac} + \bar{\nabla}_c g_{ab} - \bar{\nabla}_a g_{bc} \tag{3.38}$$

[6] Note that one way to compare infinitesimally close vectors, given a vector field m^a, is with the Lie bracket of m^a with l^a, $[m,l]^a$. This is not appropriate since $[m,l]^a|_p$ depends on the derivative of m^a at p.

or

$$C^a{}_{bc} = \frac{1}{2} g^{ad}\{\bar{\nabla}_b g_{dc} + \bar{\nabla}_c g_{db} - \bar{\nabla}_d g_{bc}\} \tag{3.39}$$

Note that the existence of a metric is not equivalent to the existence of a connection. There are connections ∇_c for which there is no metric g_{ab} such that $\nabla_a g_{bc} = 0$, that is, there are tensors $C^a{}_{bc}$ for which there is no g_{ab} that satisfies (??).

Exercise: If $\bar{\nabla}_c$ is a derivative corresponding to a coordinate system $\bar{\nabla}_c = \partial_c$ the corresponding Christoffel symbol is

$$\Gamma^a{}_{bc} = \frac{1}{2} g^{ad}\{\partial_b g_{dc} + \partial_c g_{db} - \partial_d g_{bc}\}. \tag{3.40}$$

Show that its components in that coordinate system are given by,

$$\Gamma^i{}_{jk} = \frac{1}{2} g^{il}\{\partial_j g_{lk} + \partial_k g_{lj} - \partial_l g_{jk}\}, \tag{3.41}$$

where g_{ij} are the components of the metric in the same coordinate system.

Exercise: The Euclidean metric of $I\!R^3$ in spherical coordinates is

$$ds^2 = (dr)^2 + r^2((d\theta)^2 + \sin^2\theta(d\phi)^2).$$

Calculate the Laplacian $\Delta = g^{ab}\nabla_a\nabla_b$ in these coordinates.

***The Riemann Tensor**

Given a covariant derivative on a manifold, is it possible to define tensor fields that depend only on it and therefore give us information about it?

The answer is yes! The following tensor is called the **Riemann or Curvature Tensor** and depends only on the connection:

$$2\nabla_{[a}\nabla_{b]}l^c := (\nabla_a\nabla_b - \nabla_b\nabla_a)l^c := R^c{}_{dab}l^d \quad \forall l^c \in TM.$$

Exercise: Show that the definition above makes sense, that is, that the left-hand side, evaluated at any point p in M, depends only on $l^c|_p$ and therefore we can write the right-hand side for some tensor $R^c{}_{dab}$.

Exercise: Let $\bar{\nabla}_a$ be another covariant derivative. Calculate the difference between the respective Riemann tensors in terms of the tensor that appears as the difference of the two connections.

Bibliography notes: I recommend the book by Wald [**?**], especially for its modern notation, see also [**?**]. The language of intuition is geometry, it is the tool that allows us to visualize problems, touch them, turn them around to our liking and then reduce them to algebra. Understanding geometry is the most efficient way to understand physics since it is only fully understood when translated into a geometric language. Do not abuse it, it is a very vast area and it is easy to get lost, learn the basics well and then only what is relevant to you.