Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

A mathematical model of the early visual system and border completion via subriemannian Hamiltonian geodesics

Relatore:
Prof. Rita Fioresi

Presentata da:
Alice Marraffa

Anno Accademico 2021/2022

Sommario

L'obiettivo di questa tesi è discutere un semplice modello matematico che, attraverso il linguaggio della geometria differenziale, descriva due processi fondamentali del sistema visivo umano: l'individuazione di segmenti orientati nel campo visivo e il meccanismo di percezione dei contorni come linee continue. La struttura colonnare della corteccia visiva primaria viene modellizzata tramite il fibrato $\mathbb{R}^2 \times S^1$, il *fibrato degli orientamenti* e, introducendo un campo vettoriale su di esso, si descrive il processo di riconoscimento dei segmenti orientati. Le linee continue vengono percepite elevando i segmenti orientati nello stimolo visivo al fibrato; tali curve nella varietà $\mathbb{R}^2 \times S^1$ sono tangenti a una distribuzione completamente non integrabile. Di conseguenza, è possibile costruire una struttura subriemanniana sulla varietà, tramite cui i contorni percepiti sono identificabili con le geodesiche subriemanniane orizzontali alla distribuzione, soluzioni alle equazioni di Hamilton.

Con questo scopo, nel primo capitolo si illustra il funzionamento delle più fondamentali strutture delle prime vie visive nel cervello umano, dalla retina alla corteccia visiva primaria. Si procede con l'introduzione dei concetti necessari di geometria differenziale e subriemanniana nei capitoli due e tre, con un accenno all'approccio di Cartan alle connessioni. Infine, il modello vero e proprio viene implementato nel capitolo quattro, dove si conclude con un confronto tra i risultati ottenuti e l'esistenza di un *association field*, evidenziata da Heyes, Fields, e Hess in seguito ai loro risultati sperimentali sul problema del completamento dei contorni.

Abstract

In this thesis, we aim to discuss a simple mathematical model for the edge detection mechanism and the boundary completion problem in the human brain in a differential geometry framework. We describe the columnar structure of the primary visual cortex as the fiber bundle $\mathbb{R}^2 \times S^1$, the *orientation bundle* and, by introducing a first vector field on it, explain the edge detection process. Edges are detected through a lift from the domain in $\mathbb{R}^2$ into the manifold $\mathbb{R}^2 \times S^1$ and are horizontal to a completely non-integrable distribution. Therefore, we can construct a subriemannian structure on the manifold $\mathbb{R}^2 \times S^1$, through which we retrieve perceived smooth contours as subriemannian geodesics, solutions to Hamilton's equations.

To do so, in the first chapter, we illustrate the functioning of the most fundamental structures of the early visual system in the brain, from the retina to the primary visual cortex. We proceed with introducing the necessary concepts of differential and subriemannian geometry in chapters two and three. We finally implement our model in chapter four, where we conclude, comparing our results with the experimental findings of Heyes, Fields, and Hess on the existence of an *association field*.

Contents

Introduction

In this dissertation, we aim to present and discuss a mathematical model that, employing a modern differential geometry language, describes those neural mechanisms that perform visual edge detection and boundary completion in the human brain.

Even though we address only the most fundamental and simple structures in the visual system, from the retina to the primary visual cortex, they are enough to provide an insight into how the brain strategically organizes to elaborate complex stimuli through parallel and segregated receptive networks. The modeling of neural networks elaborating sensory inputs, besides being extremely engaging, can lead to many applications and is of practical and intellectual interest. Of course, in the attempt to outline a mathematical model of perceptive phenomena, we must consider and rely on both the effective physiological functioning of the neural structures involved and their phenomenological description. While addressing edge detection, it is fascinating to understand that the primary visual cortex, with its topographical, layered, and columnar functional organization, lands itself to fit in a differential geometry framework through the mathematical abstraction of fiber bundles. Even more fascinating is that this geometrical framework allows for a straightforward solution to the problem of boundary completion, by introducing a subriemannian structure in the primary visual cortex and finding the related Hamiltonian.

The issue of perceptual completion has been widely studied in psychology from a phenomenological point of view. The Gestalt theory identifies the empirical principles under which the human brain integrates local and incomplete visual stimuli into shapes and smooth figures. Among them, continuity is the leading principle for the perception of contours and must be reproduced as a law in any mathematical modeling of boundary completion. By doing so, our modelization and interpretation of these perceptual phenomena mainly draw from the works of Mumford [1], Hoffmann [2], Petitot, Tondut [3], Citti, and Sarti [4].

In the first chapter of this dissertation, we want to give an insight into the neurophysiological aspects of the early visual system. We provide a functional description of the retina, from the retinal receptors to the optic nerve, we rapidly

move to the lateral geniculate nucleus, and finally, we outline the structural organization of the primary visual cortex. Here, we only focus on the aspects concerning oriented edge detection, and a more comprehensive description can be found in books on neural physiology. We mainly draw from [5].

In the second chapter, we lay the foundations for working in a differential geometry environment. We briefly describe differentiable manifolds, tangent vectors and vector fields, fiber bundles, and Lie groups. We try to describe such geometrical abstraction in a generalized, coordinate-free manner, following the approach of L. Tu in [6], [7].

In the third chapter, we discuss Riemannian manifolds and Hamiltonian mechanics on manifolds. The aim is to extend such concepts to the context of subriemannian manifolds and find the related geodesic equations, using the method by Montgomery in [8] as a guide. At the end of the chapter, we define connections, mentioning Cartan's coordinate-free approach ([9], [7]).

Finally, in the fourth chapter, we proceed with the development of the modelization of the primary visual cortex as a fiber bundle. First, we focus on the steps that transform discrete light spots into oriented edges. Then, we develop a subriemannian structure that would let us find the geodesic equations, whose solutions represent reconstructed boundaries.

Chapter 1

The visual pathway

Vision is our prime sensory modality, and the most important instrument we have developed to collect and process information about the world we live in. We tend to trust what we see more than what we hear or smell. Actually, the visual neural system occupies more than a half of our cortical surface in the brain. The visual process is indeed very different from a mere reproduction of the images of the outside world: it is designed to produce an inner description that is useful to the observer, neglecting all the irrelevant information.
In this chapter we only give a quick overview of the visual system's lowest pathway, namely from the *retina* to the *primary visual cortex*. We can say that this segment of the visual pathway, shown in fig.1.1, lays the foundation of our understanding of reality: even though it only accounts for low and intermediate image processing, it generates perceptions like contrasts in brightness, contours, surfaces, and depth, upon which all superior cognitive functionalities, like object recognition and vision guided actions, are based. We invite the reader to consult [5] for more details.

1.1 The retina: an overview

The first neural system involved in the process of transforming images into nerve signals is the *retina*, which is located in the posterior wall of the eye. On the retina we can distinguish the *fovea*, the the retinal region that corresponds to about five degrees around the fixation point in the visual field, and the *optic disc*, from which emerge the *optic nerve*, as it is shown in fig.1.2. Roughly speaking, the eye captures the light input, and projects it onto the retinal photoreceptors that lie at the bottom of the retinal structure. Photoreceptors are able to interact with photons to execute some low level processing. The information is then forwarded to the *retinal ganglion cells*, which represent the output of the retina through the optic nerve.

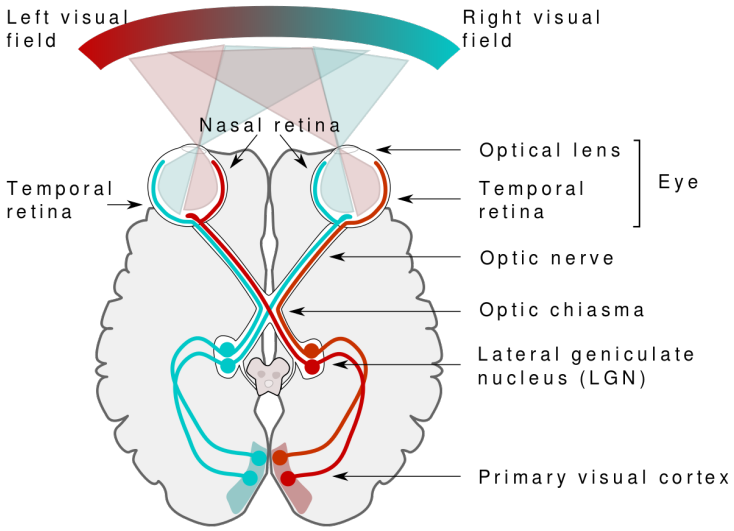


Figure 1.1: *Low-intermediate tract of the human visual pathway, [10]*

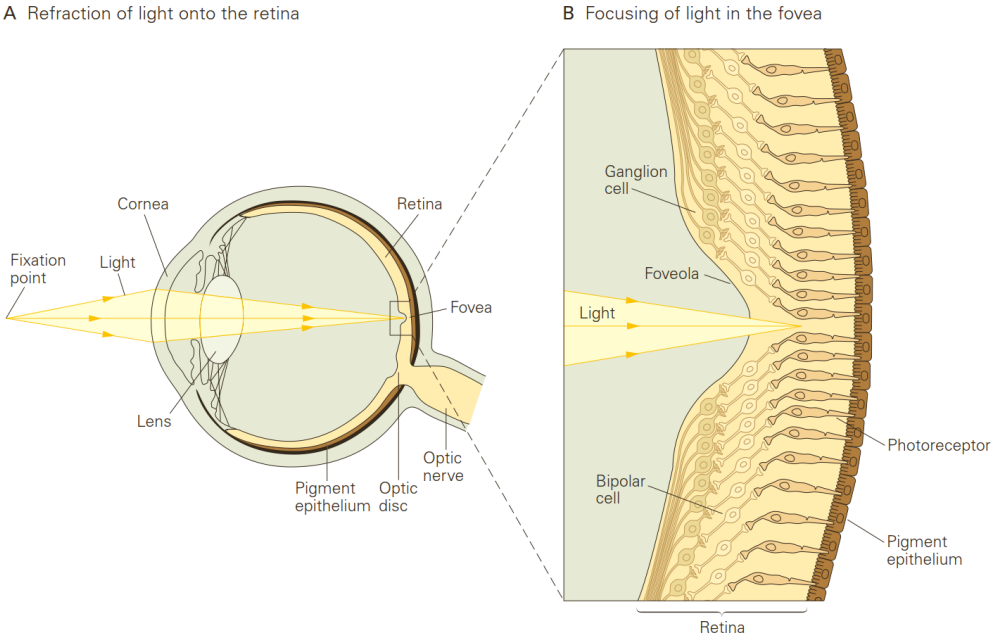


Figure 1.2: *Human eye section view - [5] ch.25 (II)*

1.2 Retinal photoreceptors

Let us now take a closer look at those cells that are first hit by the photons that constitute the visual stimuli: retinal receptor cells. They are found in a depolarized state in the dark, and get hyperpolarized when interacting with light, until they saturate. The extent of the hyperpolarization depends on the intensity of the light input, and on the type of photoreceptor. Two distinct kinds of photoreceptors can be found in the human retina, which differ in shape, role, connectivity, and the way they are distributed on the retina:

- *Rods*: they are about 10 millions and they are distributed throughout the entire receptive surface, except the *fovea*. As they are very reactive to light stimuli, they are responsible for night vision and become of prime importance in low light conditions, while they get rapidly saturated in the daylight, when they are no longer able to convey any information. Further, being endowed with only one kind of molecule that captures light, they do not differentiate wavelengths, and hence do not code for colors. Rods convey the information to *bipolar cells*; each bipolar cell receives synapse from about 15 to 30 rods.

- *Cones*: we can count a number of about 5 million cones, located throughout the whole receptive area, but sharply concentrated in the foveal region only. Accounting for daylight vision, they are much less sensitive to light than rods. They have in return a good spatial resolution, meaning that they have narrower receptive fields that allow them separate nearby points, especially in the foveal region, where they are densely distributed. Indeed, they have one-to-one connections with bipolar cells, to which they convey signals. We can distinguish three types of cones, depending on their color selectivity:

    - L cones (long wavelength, peak sensitivity at around 600 nm);
    - M cones (medium wavelength, peak sensitivity at around 570 nm);
    - S cones (short wavelength, peak sensitivity at around 450 nm).

Let us now take a moment to highlight the fact that the dimension of the receptive field of photoreceptors increases in dimension with eccentricity, i.e distance from the fixation point. As a consequence, we are able to analyse images in detail only in the part of the visual scenario that is projected onto the foveal region, where cones are densely distributed and where our attention leads our gaze. We will meet this concept again later on when studying the *primary visual cortex* and its retinotopical representation, where inputs coming from the center of the visual field account for the majority of the cortical surface.

In essence, we can say that at this level we have a neural representation of the visual scenario that is still pretty simple, with photoreceptors hyperpolarized in lighted areas and depolarized in dark areas, and further complexity needs to be added to the neural representation before letting the visual information leave the retina. This task is performed by *bipolar cells* and *horizontal cells*, as we shall see in the next section.

1.3 Retinal ganglion cells

Retinal ganglion cells (RGC) represent the highest level of retinal processing, casting the information on the *lateral geniculate nucleus* in the thalamus through the *optic nerve*. They differ from photoreceptors and bipolar cells because they respond to light inputs firing *action potentials* as the membrane reaches its threshold potential. These neurons are actually never silent, as they have a basal firing rate that gets modified by the photoreceptors' inputs. Their main computational feature, however, is their circular receptive field, divided into two concentric part in mutual competition: the center and the inhibitory surround. As a consequence, far more complex activation patterns emerge at this level. This complexity arises due to the collaboration of two class of interneurons, bipolar and horizontal cells, that combine signals from different photoreceptors in such a way that the evoked responses in the ganglionic cells critically depend on the specific spatial and temporal light patterns. In particular, the ON-OFF mechanisms are due to the existence of two kinds of bipolar cells, that respond in an opposite way to photoreceptors' neurotransmitter: ON cells, which get hyperpolarized, and OFF cells, which in turn get depolarized. The horizontal cells, with their axoinc trees that spread out laterally in retinal layers, take care of the *horizontal inhibition* mechanism. Through their contact with multiple photoreceptors and shared terminals with bipolar cells, they measure the receptors activation and negative feedback the receptors-bipolar cell synapsis. A representation of the entire structure can be found in fig. 1.3.
We can now make a first distinction between two types of RGCs that are almost equally represented in the retina, on the basis of their receptive field (see fig. 1.4):

- ON-center/OFF-surround: they fire in response to the brightening of the center of the receptive field, while they feature a mechanism of *lateral inhibition* induced by the brightening of the surround.

- OFF-center/ON-surround: these neurons fire in response to a reduction in brightness of the center of the receptive field. This response can be inhibited by a reduction in brightness of the surround.

As a consequence of this particular behaviour, RGCs give the best possible response to difference in brightness between their center and their surrounds, while

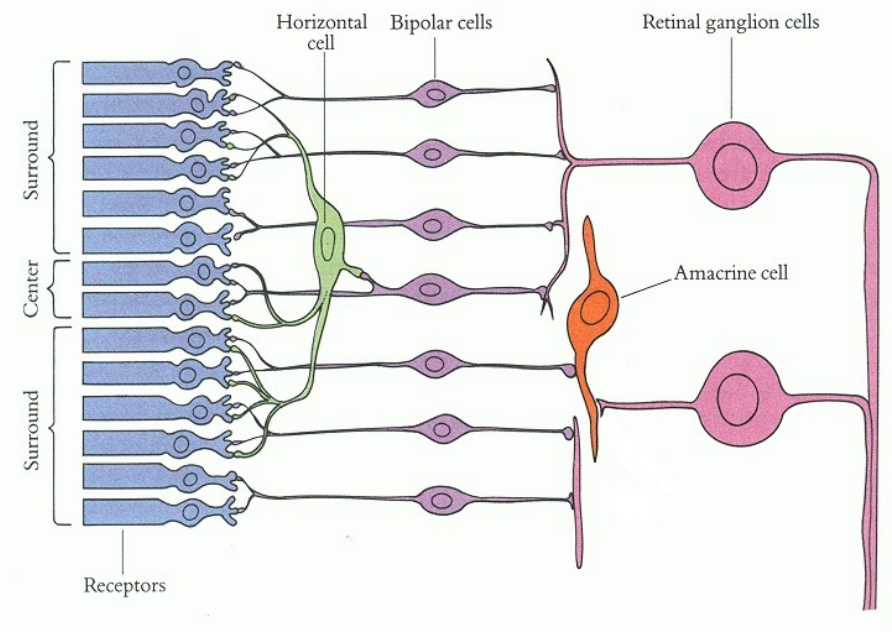


Figure 1.3: *Pictorial representation of neural connections from the retinal receptors to the retinal ganglion cells.*

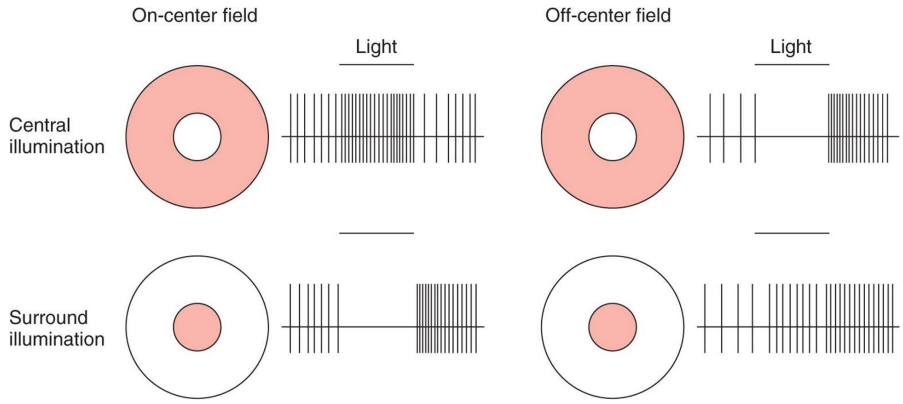


Figure 1.4: *Responses of the two types of retinal ganglion cells to light inputs. - [11]*

they give only weak responses to uniform light stimuli that cover the whole receptive field. This fact tells us that at this level the most useful information is the distribution of contrasts in brightness in the visual scenario, which is indeed the earliest step for the perception of contours.
Another distinction that can be made is between *transient* and *sustained* retinal ganglionic cells: transient cells only give a response with a brief and strong activity shift at the beginning of the light stimulus, immediately followed by a return to basal firing rate; on the other hand, sustained cells have their firing rate mod-

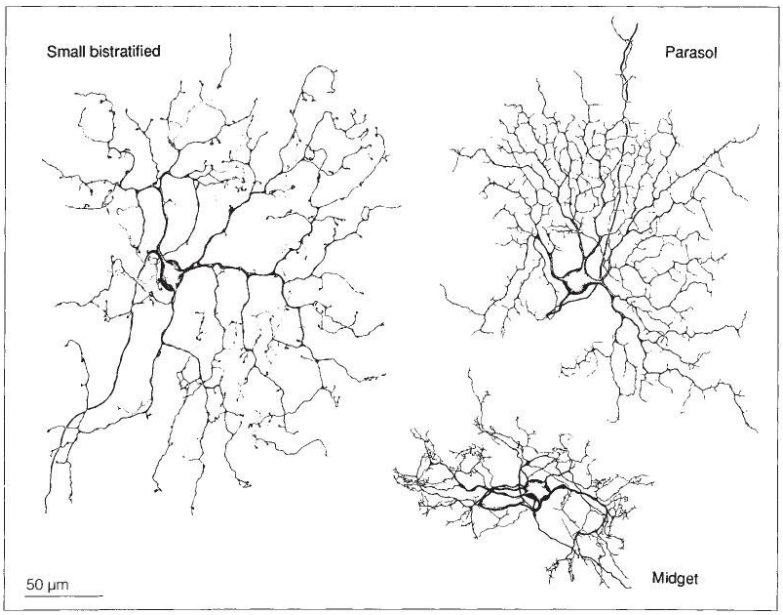


Figure 1.5: *Dendritic morphology of macaque small bistratified, parasol and midget ganglion cells. - [12]*

ified for the whole duration of the stimulus. Retinal ganglion cells can also be distinguished by shape, connectivity, and receptive field size (see fig. 1.5):

- *Parasol* or *M-type* neurons (from *magnus*-big): being big cells with large dendritic trees, thick axons, and wide receptive fields, they quickly respond to small low-contrast light changes in broad regions of the receptive field. They account for about the 10% of the retinal ganglionic population, and are connected to magnocellular layers of the *lateral geniculate nucleus* (LGN).

- *Midget* or *P-type* neurons (from *parvo*-small): they are smaller and more numerous then the M-types, accounting for about the 70% of RGCs. Having tight dendritic trees, thin axons, and narrow receptive fields, they give slower, sustained responses, rather to color changes than contrast changes. They are connected to the parvocellular layers of LGN.

- *Small bistratified* or *K-type* neurons: they have very broad receptive field, and project to the koniocellular layers of LGN.

Relying on all this distinctions, we can divide retinal ganglion cells into many different families that form separate and parallel perceptive networks, each coding for a specific feature of the image hitting the retina. This is a first example of how our brain handles such a broad range of visual information: it strategically

distributes it into many parallel processing pathways, each carrying specialized information that is relayed to superior neural centers which, in the end, are able to form a unique and complete representation of reality.
No new synaptic connections are formed after the retinal ganglion cells, until the information lands in the two *lateral geniculate nuclei*. We must now specify how the neuronal fibers formed by the axons of the RGCs are rearranged over the course of their journey to the thalamus, through a process of *partial decussation* in the *optical chiasm*: the two left visual field representations, *nasal* of the left eye, and *temporal* of the right eye, form the right *optic tract*, and continue on to targets in the right hemisphere, first of all the right *lateral geniculate nucleus*. An analogous process happens to the nasal right and temporal left optic nerve fibers, which join to form the left *optic tract* and get to the left *lateral geniculate nucleus*. In both the lateral geniculate nuclei, the two eye inputs remain separate, each segregated in three of their six layers, and this further segregation continues to the primary visual cortex (to some level of approximation). The information about each visual hemifield is thus split and always kept divided for superior elaboration.

1.4 Lateral geniculate nucleus

We now briefly name the main features of the *lateral geniculate nucleus* (LGN), a small ventral projection of the thalamus connected to the *optic nerve*. LGN is an important relay station in the primary image processing pathway, whose axons are directly sent to the *primary visual cortex*. We have one lateral geniculate nucleus for hemisphere, each composed of six layers: the layer structure help maintaining ocularity and functional segregation of M, P, and K pathways that started in the retina. In fact, the *magnocellular* pathway continues to the two most ventral layers of LGN, while the *parvocellular pathway* leads to the remaining four dorsal layers. The *koniocellular* pathway is not yet well understood, and represented as the fine layers sitting in between the six main layers of LGN. A graphic representation of LGN's layers is sketched in fig. 1.6.

1.5 The primary visual cortex

The *primary visual cortex* (V1) is the earliest cortical visual area; it is located in the occipital lobe, in and around the *calcarine sulcus*. It is defined by its functional structure and the role it plays in the visual system, but it can also be identified with the *striated cortex*, which instead is defined by its anatomical location. It is named after the myelinated *stria of Gennari*, a distinctive stripe also visible by the naked eye, that coincides with the IV^{th} layer of V1, the layer of thalamic

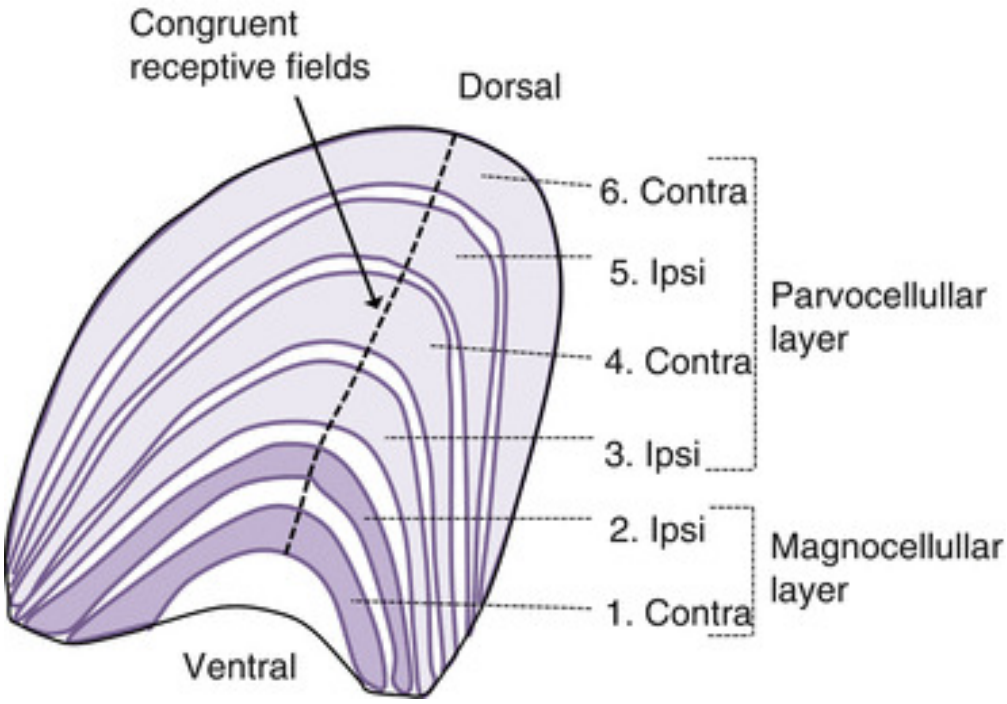


Figure 1.6: *The Lateral Geniculate Nucleus - [13] pp. 201-206*

afferences. Inputs coming from the six layers of the LGN indeed synapse this layer of V1, maintaining functional segregation and eye-specificity: inputs from the magnocellular pathway reach the $IV^{th}C\alpha$ layer of V1, while inputs from the parvocellular pathway reach the $IV^{th}C\beta$ layer, forming concurrently a structure of *ocular dominance columns* in the cortex, namely approximately $400\mu m$-wide stripes of visual neurons that predominantly receive their inputs from one eye or the other.
We are now interested in showing how the cortical representations of the visual space, viewpoint, orientation and color, are organized in specific *maps* that allow a complete coverage of the whole visual field in each of its features. As we want to provide a model for contour perceptions, we will mainly focus only on the orientation cortical mapping.

1.6 Retinotopic maps

From the *retina* to the *primary visual cortex*, the neural representations of the visual scenario keep a tidy organization of spatial information. V1 is indeed arranged to form a 2D map of the visual field where neighborhoods that are close in

the visual image are represented by close neighborhoods in its cortical area. This faithful preservation of spatial information is called *retinotopy*.

The spatial information coming from the retina appears distorted in its mapping to V1, by a process called *cortical magnification*. In fact, the area of the primary visual cortex is not assigned equally to every degree of the visual field, but it is proportional to the density of photoreceptors per retinal area. One can quantify the mm^2 of cortex per degree of visual angle by the *cortical magnification factor*, that was found to fall by several log units from the foveal representation to the monocular crescent representation ([14]). This decreasing of the magnification factor with the eccentricity reflects the fact a larger part of the cortex is allocated for central vision, which must be finer and more detailed than peripheral vision. A precise retinotopic map is shown in fig. 1.7, and more detail about this topic can be found in [14].

1.7 Orientation tuning

From the foregoing, we know that each neuron of V1 contributes with its own receptive field to the retinotopic cortical representation, and exhibits a number of specific properties, such as the above-mentioned ocular dominance, receptive field position in visual space, color selectivity and so on. We now dwell on an emergent property that characterizes primary cortical neurons, described in the Hubel and Wiesel pioneering studies [15] on V1 that got them the Nobel prize in 1981, that is, *orientation tuning*. Neurons found in the retina and in the LGN had circular receptive fields, with a center-surround structure that made them reactive to spot light contrasting stimuli. Cortical neurons were found instead to have elongated receptive fields that would allow responses for linear light stimuli, and an emergent selectivity for light edges presented at specific angles. Distinct neurons can respond at peek proficiency to distinct orientations, and for each of them an *orientation tuning curve* can be computed, quantifying the neuronal response as a function of the presented stimuli orientations. In fig.1.8 we provide an example of a V1 neuron firing action potentials in response to specifically oriented light edges, and the relative orientation tuning curve. A question that might arise is how this orientation tuning property is attained by cortical neurons receiving inputs from the retina through LGN. Hubel and Wiesel explained this fact through a model for the neuronal receptive fields' orientation, suggesting a serial elaboration of lower order receptive field properties. According to the model, a cortical oriented cell should gather together inputs from specific neurons of the LGN, chosen in such a way that their center-surround receptive fields, once aligned, correspond to the cell's preferred orientation. This process of converging of multiple LGN's cells on a single cortical cell to give rise to orientation tuning has been proven by decades

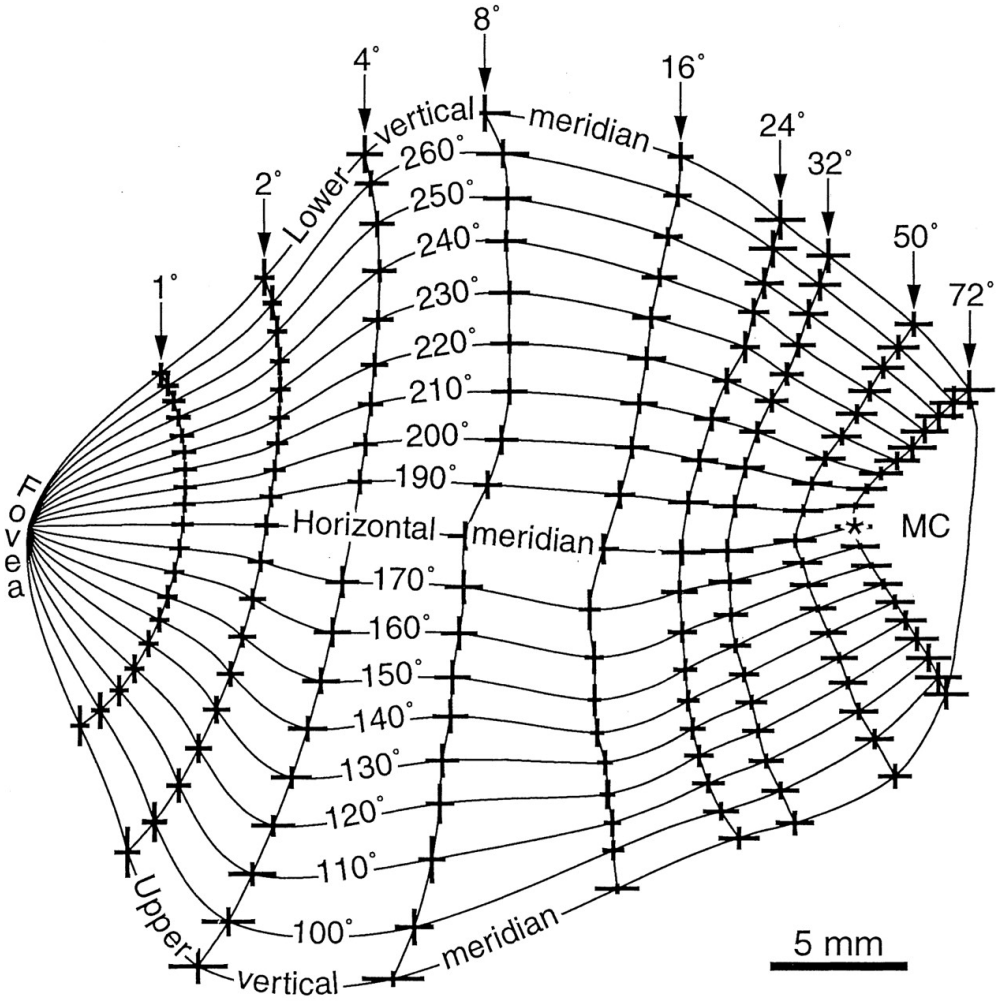


Figure 1.7: *Retinotopic map in the primary visual cortex - Daniel L. Adams and Jonathan C. Horton - [14]*

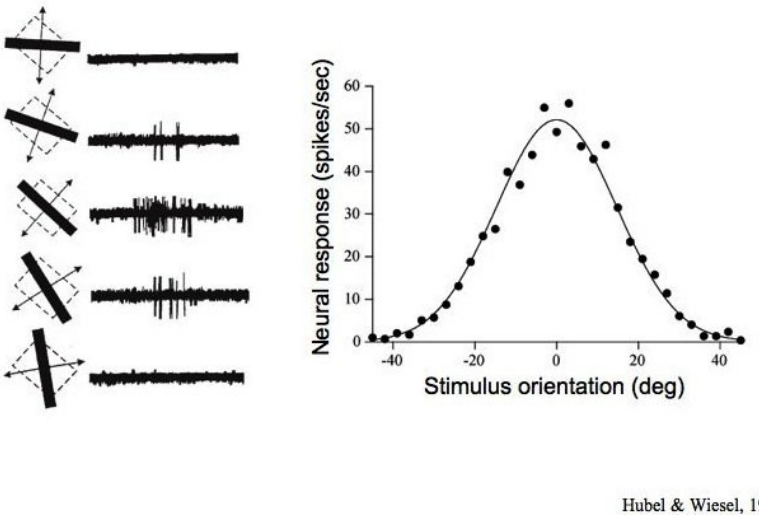


Figure 1.8: *Response and orientation tuning curve of a simple V1 cell - Seminal results of Hubel and Wiesel, figure adopted from Hubel and Wiesel, 1968. On the left, orientation selectivity of a V1 cell. On the right, the orientation tuning.*

of research to be very likely true to some degree of approximation, with the aid of other additional mechanisms, and provides an example of how the visual system works systematically and hierarchically to elaborate sensory inputs.
We can now make a further distinction among oriented cortical cells, based on their receptive field properties and therefore their selectivity.

- *Simple cells*: their elongated receptive fields have separate ON (excitatory) -OFF (inhibitory) areas that can occur in different combinations covering the receptive field, inheriting from receptive fields of neurons in the LGN. This property makes them selective for of course orientation of the light stimuli, but also for the position of the latter in the receptive field (i.e, a stimulus activates a cell only if it falls within the ON-area).

- *Complex cells*: still selective for orientation, they have rectangular, wider receptive fields, and there is no neat separation between ON and OFF areas. This fact, again according to the Hubel-Wiesel serial model, might arise from a progressive information-convergence process in the primary visual cortex, that is, many V1 *simple cells* conveying information to one single *complex cell* by the overlapping of their receptive field. We then have an emerging *position invariance* property that allows *complex cells* to respond steadily to stimuli with different positions in the receptive field, and that

makes them particularly reactive to moving linearly oriented stimuli that cross their receptive fields.

- *End-stopped cells*: their receptive fields feature a central excitatory region bordered by inhibitory peripheral regions with the same preferential orientation. These patterns might cause, for example, a short linear oriented light segment to be very effective in activating one of these neurons, and by contrast a long oriented light line to be very ineffective. *End-stopped cells* are hence useful to detect angles or curved lines in visual images.

1.8 Color blobs

Before moving forward, we shall mention some (relatively) recently discovered structures that can be found through the second and the third layer of V1 and represent a prominent feature of its architecture: *color blobs*. *Color blobs* are composed of color tuned but unoriented cells, rich of cytochrome-oxidase, that are able to compare the activation of green (M) and red (L) cones. Some of them respond ON to red light and OFF to green light, and some others do the opposite, while in white or yellow light they respond poorly. They form an anatomically separated pathway with respect to orientation tuned cells (*interblobs*), and have a different functional purpose, as they are part of the broader process of color vision and surface property estimation.

1.9 The Ice Cube model

We have seen so far how V1 neurons differentiate and exhibit new coding properties that allow a prior analysis of the visual image. We want now to summarize the main findings about the *primary visual cortex* functional architecture. Hubel and Wiesel's works ([16], [17]) on the topic led them to formulate the so-called *Ice Cube model* of V1, based on the discovery that the cortex's functional organization is columnar. Neurons with similar functional properties, i.e neurons with the same orientation tuning or with the same ocular dominance, and with the same retinotopic positioning, appear indeed close to each other, forming columnar structures. More specifically, a vertical electrode penetration in V1 will encounter neurons that are selective for the same light edges orientation forming *orientation columns*, while a tangential penetration will encounter neurons that progressively change their preferred orientation, covering all the possible orientations. This shifting in the orientation tuning is regular and systematic, approximately continuous, and it is reproduced periodically with period of about $1mm$. As stated above,

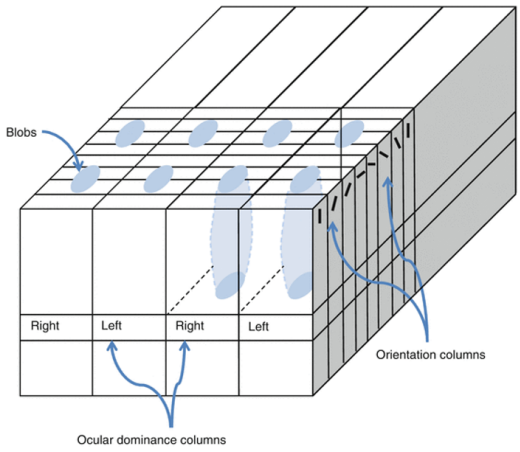


Figure 1.9: *Ocular dominance columns and orientation columns - [13], pp. 207-218*

the LGN afferents eye-segregation creates also contro-lateral and ipsi-lateral *ocular dominance bands*, which alternate along V1 almost every $1mm$.
A scheme of this columnar functional organization is pictured in fig.1.9. To describe the fact that *orientation columns*, *ocular dominance bands* and *blobs* are not independent and work together to obtain cortical representations, Hubel and Wiesel defined the concept of *hypercolumn*, complete set of columns that contains all the cortical structures needed to analyze a specific feature (for example, orientation) of the stimuli in some specific area of the visual field. Orientation hypercolumns include a complete set of orientation columns that span all possible preferred orientations, while a ocular dominance hypercolumn includes one each of a left eye and a right eye column. Orientation hypercolumns and ocular dominance columns have the tendency to run perpendicularly to each other, forming, together with blobs, computational $1mm$ x $1mm$ modules, that somehow resemble to the ice cubes after which they are named. Furthermore, precisely within these modules, cells represent a single precise spatial location. We can therefore say that, containing all anatomo-functional type of V1 neurons, these modules are able to analyse a specific position in the visual field for all possible visual parameters, and are hence repeated thousands of times in order to address the whole visual field. This *Ice Cube model*, sketched in fig.1.10 is a very important and compelling computational model that let us understand how our *primary visual cortex* takes care of all positions in the visual field, by attaching for each point of its retino-

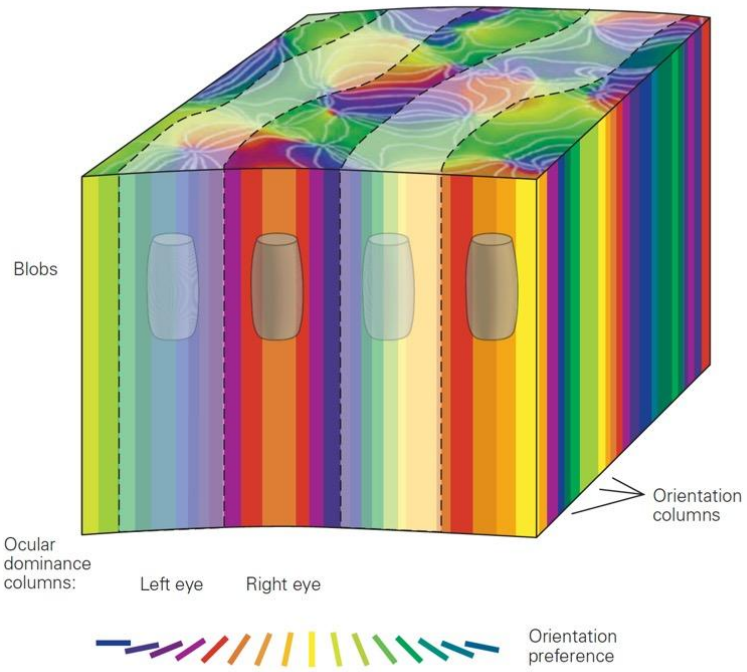


Figure 1.10: *The Hubel and Wiesel's Ice Cube model.*

topic map the correct values of all possible visual parameters, and creating many parallel representations of the visual image.

Chapter 2

Differentiable manifolds

In this chapter we give the basic concepts of the theory of differentiable manifolds, including the notion of tangent space, cotangent space, and differential of a smooth map. We are particularly interested in Lie groups and their action on manifolds, and the notion of a fiber bundle. We assume some basic notion of topology, see [18].

2.1 Differentiable manifolds

In this section we give the notion of differentiable manifold, using as a guide the notion of differentiability on $\mathbb{R}^n$. With the same logic, we get to define smooth map between manifolds. Along the discussion, we examine an important example: the unit sphere as manifold via the stereographic projections. We invite the reader to consult [6] for more details.

Definition 2.1. Let M be a Hausdorff, second countable, topological space. We say M is a *topological manifold of dimension* n if every point p in M has a neighborhood U homeomorphic to an open subset of $\mathbb{R}^n$.
We give the definitions:

- the pair (U, ϕ) is called a *chart*, $\phi \colon U \to \phi(U) \subset \mathbb{R}^n$ open, ϕ homeomorphism
- U is called a *coordinate neighbourhood* of p
- ϕ is called a *coordinate system* on U

Definition 2.2. Let $\phi\colon U \to \mathbb{R}^n$, $\psi\colon V \to \mathbb{R}^n$ be two charts of a topological manifold M. ϕ, ψ are C^∞-*compatible* if the two maps

$$\phi \circ \psi^{-1}\colon \psi(U \cap V) \to \phi(U \cap V),$$
$$\psi \circ \phi^{-1}\colon \phi(U \cap V) \to \psi(U \cap V)$$

are of class C^∞. These two maps are then called *transition functions* between the charts.
A C^∞ *atlas* on a topological manifold M is a collection $\mathcal{U} = \{(U_\alpha, \phi_\alpha)\}$ of pairwise compatible charts such that $M = \bigcup_\alpha U_\alpha$.
An atlas $\mathcal{U}$ is said to be *maximal* if it is not contained in a larger atlas.

We are now ready for the most important definition of this chapter.

Definition 2.3. A *smooth manifold* is a topological manifold M together with a maximal atlas. The maximal atlas is called a *differentiable structure* on M.

Any atlas on a topological manifold is contained in a maximal atlas. In order to check that a topological manifold M is a smooth manifold, then, it is sufficient to exhibit a C^∞ atlas for M. See [6] for more details.

Example 2.1. Let S^2 be the unit sphere in $\mathbb{R}^3$:

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$$

Let us call the two points $N = (0, 0, 1)$ and $S = (0, 0, -1)$ the north and the south pole on the unit sphere, respectively. We can cover S^2 with two open sets $U_1 = S^2 \setminus \{N\}$, $U_2 = S^2 \setminus \{S\}$, and define two charts (U_1, ϕ_1), (U_2, ϕ_2) as follows:

$$\begin{aligned} \phi_1\colon U_1 &\longrightarrow \mathbb{R}^2 & \qquad \phi_2\colon U_2 &\longrightarrow \mathbb{R}^2 \\ (x, y, z) &\longmapsto \left(\frac{x}{1-z}, \frac{y}{1-z}\right) & \qquad (x, y, z) &\longmapsto \left(\frac{x}{1+z}, \frac{y}{1+z}\right) \end{aligned}$$

These two charts are called *stereographic projections*, as they map the sphere onto a plane. Note that ϕ_1 and ϕ_1 map the south pole S and the north pole N to $(0, 0)$, respectively, and so ϕ_1 and ϕ_2 map $(U \cap V)$ onto $\mathbb{R}^2 \setminus \{(0, 0)\}$.
Since ϕ_1 and ϕ_2 are homeomorphisms, we can compute their inverse ϕ_1^{-1} and ϕ_2^{-1}, which are continuous:

$$\phi_1^{-1}((u_1, u_2)) = \left(\frac{2u_1}{u_1^2 + u_2^2 + 1}, \frac{2u_2}{u_1^2 + u_2^2 + 1}, \frac{u_1^2 + u_2^2 - 1}{u_1^2 + u_2^2 + 1}\right)$$
$$\phi_2^{-1}((v_1, v_2)) = \left(\frac{2v_1}{v_1^2 + v_2^2 + 1}, \frac{2v_2}{v_1^2 + v_2^2 + 1}, \frac{1 - v_1^2 - v_2^2}{v_1^2 + v_2^2 + 1}\right)$$

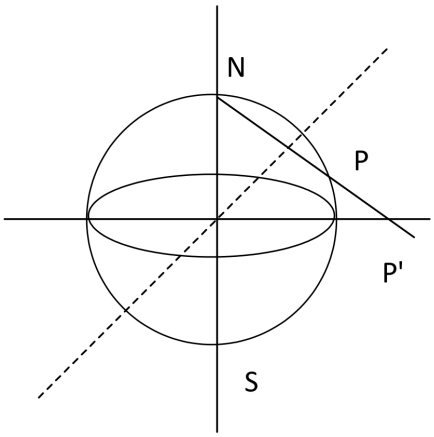


Figure 2.1: *Stereographic projections*

Note that the transition functions

$$\begin{aligned}\phi_1 \circ \phi_2^{-1} \colon \phi_2(U \cap V) &\longrightarrow \phi_1(U \cap V)\\ (u_1, u_2) &\longmapsto (\frac{u_1}{u_1^2 + u_2^2}, \frac{u_2}{u_1^2 + u_2^2})\\ \phi_2 \circ \phi_1^{-1} \colon \phi_1(U \cap V) &\longrightarrow \phi_2(U \cap V)\\ (v_1, v_2) &\longmapsto (\frac{v_1}{v_1^2 + v_2^2}, \frac{v_2}{v_1^2 + v_2^2})\end{aligned}$$

are differentiable functions from $\mathbb{R}^2 \setminus \{(0,0)\}$ to itself. The two stereographic projections (U_1, ϕ_1), (U_2, ϕ_2) therefore form a C^∞-atlas on S^2. Hence S^2 is a differentiable manifold.

We now define the notion of smooth map of manifolds. The notion of smoothness can be referred to a map of manifolds thanks to the introduction of a differentiable structure on the manifold via its charts.

Definition 2.4. Let N and M be smooth manifolds of dimension n and m, respectively. A continuous map $F \colon N \to M$ is *smooth at a point* p in N if there are charts (V, ψ) about $F(p)$ in M, (U, ϕ) about p in N such that

$$\psi \circ F \circ \phi^{-1} \colon \phi(F^{-1}(V) \cap U) \to \mathbb{R}^m$$

is C^∞ at $\phi(p)$. The continuous map $F \colon N \to M$ is said to be C^∞ if it is C^∞ at every point of N.

The notion of smoothness is well defined, as it is independent of the choice of charts. This independence is ensured by the C^∞ compatibility of charts in a differentiable structure stated above.
We now define an important class of bijective smooth functions, that is the diffeomorphism.

Definition 2.5. A *diffeomorphism* of manifolds is a C^∞ bijection $F\colon N \to M$ whose inverse is also C^∞.

Observation 2.1. Any chart on a manifold is a diffeomorphism, and any diffeomorphism of an open subset of a manifold to its image in $\mathbb{R}^n$ can be looked at as a chart. In fact

$$\begin{array}{ccc} U & \xrightarrow{\phi} & \mathbb{R}^n \\ \downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle id} \\ \phi(U) & \xrightarrow{f} & \mathbb{R}^n \end{array}$$

$id \circ \phi \circ \phi^{-1}$ is C^∞, because it is the identity map.

2.2 Tangent vectors and the tangent space

In this section we generalize the notions of tangent vector and tangent space from $\mathbb{R}^n$ to manifolds. The generalization we need would let us define the tangent space intrinsically, without referring to any embeddings of the manifold into $\mathbb{R}^n$. This approach, despite making tangent vectors less intuitive, becomes crucial when studying manifolds that are not naturally embedded in $\mathbb{R}^n$. Given the definition of tangent space to a point in a manifold, we finally define the differential of a smooth map on a manifold, which we will need later on in our discussion.

First, we introduce an equivalence relation on the C^∞ functions defined in some neighborhood of a point in a manifold M.

Definition 2.6. Consider all pairs (f, U) where $U \subset M$ is a neighborhood of p, $f\colon U \to \mathbb{R}$ is a C^∞ function. Two such pairs (f, U), (g, V) are equivalent if and only if there is an open set $W \subset U \cap V$ containing p such that $f\big|_W = g\big|_W$. The equivalence class of (f, U) is called the *germ* of f at p. The set of all germs of C^∞ real-valued functions at p in M is denoted by $C^\infty_p(M)$.

We note that $C_p^\infty(M)$, with ordinary addition scalar multiplication by real numbers, is a vector space over $\mathbb{R}$. In addition, if equipped with the ordinary multiplication and addition of functions (defined pointwise), $C_p^\infty(M)$ is a commutative ring. This is due to the fact that, in general, functions do not have a multiplicative inverse.

Definition 2.7. A *derivation at a point* in M is any linear map $D\colon C_p^\infty(M) \to \mathbb{R}$ that satisfies the Leibniz rule:

$$D(fg) = (Df)g(p) + f(p)Dg.$$

A *tangent vector* at a point p in a manifold M is a derivation at p. The tangent vectors at a point p form a vector space $T_p(M)$, the *tangent space of M at p.*

Let (U, ϕ) be a coordinate neighborhood, $r^1, \ldots, r^n$ the standard coordinates on $\mathbb{R}^n$, $x^i = r^i \circ \phi\colon U \to \mathbb{R}$, f a smooth function in a neighborhood of p. We define the *partial derivative $\frac{\partial f}{\partial x^i}$ of f with respect to x^i at p* to be:

$$\left.\frac{\partial}{\partial x^i}\right|_p f = \left.\frac{\partial}{\partial r^i}\right|_{\phi(p)} (f \circ \phi^{-1}) \in \mathbb{R}.$$

Then, $\frac{\partial}{\partial x^i}|_p$ is a function from $C_p^\infty(M)$ to $\mathbb{R}$ that evidently satisfies the Leibniz rule and is therefore a tangent vector at p.
Now that we have defined smooth maps between manifolds and the tangent space to a manifold, we introduce an important linear map of tangent spaces, the differential of a smooth map.

Definition 2.8. Let $F\colon N \to M$ be a C^∞ map between two manifolds. At each point $p \in N$, F induces a linear map of tangent spaces, its *differential at p*,

$$d_pF\colon T_pN \to T_{F(p)}M$$

so that, if $X_p \in T_pN$, then $d_pF(X_p)$ in $T_{F(p)}M$ is defined by

$$(d_pF(X_p))f = X_p(f \circ F) \in \mathbb{R} \quad \text{for } f \in C_{F(p)}^\infty(M).$$

It can be shown that a simple rule for computing the differential of the composition of functions, the so-called *chain rule*, holds.

Theorem 2.1. *Let $F\colon M \to N$ and $G\colon N \to P$ be two smooth maps of manifolds, and $p \in M$. By definition of differential, we have $d_pF\colon T_pM \to T_{F(p)}N$ and $d_{F(p)}G\colon T_{F(p)}N \to T_{G(F(p))}P$. Then*

$$d_p(G \circ F) = d_{F(p)}G \circ d_pF$$

where $d_p(G \circ F)\colon T_pM \to T_{G(F(p))}$.

This theorem leads to the following very important corollary

Corollary 2.1. If $F\colon M \to N$ is a diffeomorphism of manifolds, then $d_pF\colon T_pM \to T_{F(p)}N$ for any p in M is an isomorphism of vector spaces.

We invite the reader to read [6] for more details. With the aid of these results, we can construct a basis for the tangent space to a point on a smooth manifold M. Let (U, ϕ) be a chart containing p. We know that ϕ is a diffeomorphism, and hence it follows from the previous corollary that $d_p\phi\colon T_pM \to T_{\phi(p)}\mathbb{R}^n = \mathbb{R}^n$ is an isomorphism of vector spaces. In particular, T_pM is an n-dimensional vector space.

Proposition 2.1. *Let $(U, \phi) = (U, x^1, \ldots, x^n)$ be a chart about a point p in M. Then*

$$d_p\phi\left(\frac{\partial}{\partial x^i}\Big|_p\right) = \frac{\partial}{\partial r^i}\Big|_{\phi(p)}$$

By virtue of the well known linear algebra property that an isomorphism of vector spaces carries a basis into a basis, and the fact that $\frac{\partial}{\partial r^1}, \ldots, \frac{\partial}{\partial r^n}$ form a basis for $T_{\phi(p)}\mathbb{R}^n = \mathbb{R}^n$, we can conclude that

$$\frac{\partial}{\partial x^1}\Big|_p, \ldots, \frac{\partial}{\partial x^n}\Big|_p$$

constitute a basis for T_pM.

We now want to compute a local expression for the differential. Let $F\colon N \to M$ be a smooth map of manifolds. Let p be a point in N, $(U, x^1, \ldots, x^n)$ a chart containing p, $(V, y^1, \ldots, y^n)$ a chart for M containing $F(p)$. According to what stated above, $\{\frac{\partial}{\partial x^j}|_p\}$ is a basis for T_pN, and $\{\frac{\partial}{\partial y^j}|_{F(p)}\}$ is a basis for $T_{F(p)}M$. The differential d_pF is a linear map, and hence is completely determined by the numbers a^i_j

$$d_pF\left(\frac{\partial}{\partial x^j}\Big|_p\right) = \sum_{k=1}^{m} a^k_j \frac{\partial}{\partial y^k}\Big|_{F(p)},$$

and applying y^i to both sides

$$\begin{aligned} a^i_j &= d_pF(\frac{\partial}{\partial x^j}\Big|_p)\, y^i \\ &= \frac{\partial}{\partial x^j}\Big|_p (y^i \circ F) = \frac{\partial F^i}{\partial x^j}(p), \end{aligned}$$

where F^i is to be understood as $y^i \circ F$. It follows from these considerations that the differential d_pF at p can be represented locally by the matrix $[\frac{\partial F^i}{\partial x^j}(p)]$

2.3 Curves in a manifold

In the previous section, we defined the notion of smooth map of manifolds. Such notion is useful to define the notion of a smooth curve in an intrinsic way.

Definition 2.9. A *smooth curve* γ in a manifold M is a smooth map $\gamma\colon]a,b[\to M$, $]a,b[$ being an open subset of $\mathbb{R}$. We define the *velocity vector* $\dot\gamma(t_0)$ of γ at some point $t_0 \in]-\epsilon,\epsilon[$ as

$$\dot\gamma(t_0) = d_{t_0}\gamma\,(\frac{d}{dt}\Big|_{t_0}) \in T_{\gamma(t_0)}M.$$

We recall that

$$(d_{t_0}\gamma(\frac{d}{dt}\Big|_{t_0})(f) = (\frac{d}{dt}\Big|_{t_0})(f\circ\gamma)$$

for $f \in C^\infty_{\gamma(t_0)}$

We say that a curve γ is a curve starting at p if $\gamma(0) = p$.
Each smooth curve γ, being a map of manifolds, gives rise to a tangent vector for each t_0 in $T_{\gamma(t_0)}M$ through its differential at t_0. It can be shown that for every tangent vector $X_p \in T_pM$ you can construct a smooth curve γ such that X_p is the velocity of γ at p.

2.4 Regular submanifolds, immersions and submersions

In this section, we provide the definition of regular submanifold, immersion and submersion. We proceed defining the regular points of a smooth map, and state

an important theorem which allows us to recognize a regular level set of a smooth function as a smooth manifold itself. We begin with the definition of a regular submanifold.

Definition 2.10. A subset S of a manifold M of dimension n is a *regular submanifold* of M of dimension k if for every $p \in M$ there exists a coordinate neighborhood $(U, \phi) = (U, x^1, \ldots, x^n)$ in the maximal atlas of M such that $U \cap S$ is defined by the vanishing of $n-k$ coordinate functions, i.e, on $U \cap S$, $\phi = (x^1, \ldots, x^k, 0, \ldots, 0)$.

We call such chart an *adapted chart* for S. Let $\phi_S \colon U \cap S \to \mathbb{R}^k$, $\phi_S = (x^1, \ldots, x^k)$. We note that $(U \cap S, \phi_S)$ is a chart for S in the subspace topology.

Definition 2.11. A map $F \colon N \to M$ is said to be an *immersion at a point* $p \in N$ if its differential at p is injective, and a *submersion at* p if d_pF is surjective.
A point p in N is a *critical point* of F if its differential at p fails to be surjective, and it is a *regular point* of F if d_pF is surjective, i.e, if F is a submersion at p.
Let $F^{-1}(\{c\}) = \{p \in N : F(p) = c\}$ be a level set. $F^{-1}(\{c\})$ is called a *regular level set* if every p in $F^{-1}(\{c\})$ is a regular point of F.

We note that, if $F \colon N \to \mathbb{R}$ is a real-valued function, a point p is a critical point of F if and only if, relative to some chart $(U, x^1, \ldots, x^n)$ containing p, all the partial derivatives satisfy

$$\frac{\partial f}{\partial x^i}(p) = 0.$$

In fact, the differential is represented by the matrix

$$[\frac{\partial f}{\partial x^1}(p), \ldots, \frac{\partial f}{\partial x^1}(p)],$$

and the image of d_pf is a linear subset of $\mathbb{R}$, that can be either 1 or 0-dimensional, i.e, if d_pf fails to be surjective, it is the zero map.
We are now ready to state a criterion useful to show that the regular level set of a map of manifolds is itself a manifold, the *regular level set theorem.* See section 9.3 of [6] for the proof.

Regular Level Set Theorem. *Let $F \colon N \to M$ be a smooth map of manifolds, $dimN = n$ and $dimM = m$. A non-empty regular level set $F^{-1}(\{c\}) = \{p \in N : F(p) = c\}$, $c \in M$, is a regular submanifold of N of dimension $n - m$.*

2.5 Cotangent space

We now define the dual space of the tangent space to a manifold at some point, the cotangent space.

Definition 2.12. Let M be a smooth manifold, $p \in M$. The dual vector space of the tangent space to M at p, T_p^*M, is called the *cotangent space* to M at p, $(T_pM)^*$.

Being the dual of T_pM, and being T_pM a finite-dimensional vector space of dimension n, T_pM and T_p^*M are isomorphic as vector spaces. Hence, T_p^*M is also an n-dimensional vector space, and we now describe a basis.

Definition 2.13. Let $f \in C^\infty(M)$. At every point $p \in M$, we can define the linear map

$$\begin{aligned} d_p \colon C^\infty(M) &\to T_p^*M \\ f &\mapsto d_p f. \end{aligned}$$

We call such a map the *gradient operator* at p, and d_pf the *gradient* of the function f at p. Notice that, according to definition 2.8, $(d_pf)X = X(f)$.

We can use the gradient to construct a basis for the cotangent space to a manifold at some point p in the manifold as follows. Let $p \in M$, $(U, \phi) = (U, x^1, \dots, x^n)$ be a chart containing p. We can construct n covectors $d_px^1, \dots, d_px^n$. Being T_p^*M an n-dimensional vector space, these covectors are good candidate to form a basis for T_p^*M, and indeed they are. In fact, if we let d_px^a act on some basis vector for T_pM, $\frac{\partial}{\partial x^b}\Big|_p$, we have

$$d_px^a\left(\frac{\partial}{\partial x^b}\Big|_p\right) = \frac{\partial}{\partial x^b}\Big|_p x^a = \frac{\partial}{\partial r^b}(x^a \circ \phi^{-1})\Big|_{\phi(p)} = \delta^a_b,$$

and we can let a and b vary from 1 to n. Thus, $d_px^1, \dots, d_px^n$ are the *dual basis* of the dual of T_pM. The d_px^i are commonly denoted dx^i. We shall see in the next sections that the dx^i are *differential 1-forms* on M.

2.6 Vector fields and differential forms

In this section, we first want to give a definition of a vector field. This notion is crucial for our purposes, and we will also see it from a different perspective in

the following sections. Then, we provide some essential knowledge on the topic of differential forms, and we will restrict only to what is needed for the understanding of Hamiltonian mechanics on manifolds. We invite the reader to consult section 3 and 17 of [6] for more details.

Definition 2.14. Let M be a smooth manifold of dimension n. We define a *vector field* X on M as an assignment

$$\begin{aligned} X\colon M &\to T_p(M) \\ p &\mapsto X(p) \,=\, X_p \in T_p(M). \end{aligned}$$

Let $(U, \phi) = (U, x^1, \dots, x^n)$ be a chart on M about some point p. Recall that any tangent vector at p can be written as a linear combinations of tangent vectors $\partial_1, \dots, \partial_n$ defined by

$$\partial_i f = \frac{\partial}{\partial x^i}\Big|_p f = \frac{\partial}{\partial r^i}\Big|_{\phi(p)} (f \circ \phi^{-1}).$$

It follows that, if X is a vector field on M, the value of X at some $x \in U$ is a linear combination

$$X(x) = \sum_i a^i(x)\partial_i.$$

We note that, as x varies in U, the coefficients a^i become functions on U. We can therefore say that the vector field X is smooth if its coefficients are smooth functions. It can be easily checked that the smoothness of a vector field is a notion that is independent from the choice of the chart.
We now turn to differential forms. Differential forms can be thought of as a generalization of real-valued functions on manifolds, which instead of assigning to each point in the manifold a real number, assign some covector acting on its tangent space.

Definition 2.15. A *differential 1-form* on a manifold M is a function ω that assigns to each $p \in M$ a covector ω_p, that is, a *covector field*.

Now let $(U, \phi) = (U, x^1, \dots, x^n)$ be a coordinate chart on M. Then the differentials $dx^1, \dots, dx^n$ are 1-forms on U, that at each point $p \in U$ form a basis for T_p^*M, which moreover is the dual basis of $\{\partial_i\}$. Therefore, every 1-form ω on U can be written as

$$\omega \,=\, \sum_i a_i dx_i,$$

and we have found a local coordinate expression for differential 1-forms.
We now want to define *differential k-forms*. First of all, we recall the definition of a *k-tensor* on a vector space V, that is a k-linear function

$$T\colon V \times V \times \cdots \times V \to \mathbb{R}$$

We say that a k-linear tensor is *alternating* if, for any permutation σ in the set of all permutations of order k, S_k,

$$T(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) \;=\; sgn(\sigma)\, T(v_1, \ldots, v_n)$$

holds. An alternating k-tensor on V is also called a *k-covector*.

Definition 2.16. A *k-differential form* on a manifold M is a function ω that assigns to each $p \in M$ a k-covector ω_p on the vector space T_pM.
If ω is a k-form on M and $X_1, \ldots X_k$ are vector fields on M, then $\omega(X_1, \ldots, X_k)$ is a function on M defined by $(\omega(X_1, \ldots, X_k))(p) \;=\; \omega_p((X_1)_p, \ldots, (X_k)_p)$

To provide a local expression for k-forms, we first need to define the *wedge product* of k-linear tensors, and to find a basis for the space of all k-linear tensors.

Definition 2.17. Let T_1 be a k-tensor and T_2 a l-tensor. We define the *wedge product* $T_1 \wedge T_2$ by

$$(T_1 \wedge T_2)(v_1, \ldots, v_{k+l}) \;=\; \frac{1}{k!l!} \sum_{\sigma \in S_{k+l}} sgn(\sigma)\, T_1(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) T_2(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}),$$

where the factor $\frac{1}{k!l!}$ compensates for repetitions in the sum.

This definition can be easily generalized for the wedge product of n-many covectors. However, we will focus on the wedge product of 1-covectors. Indeed, it can be easily shown that, if $\alpha_1, \ldots, \alpha_k$ are linear functions on a vector space V, then

$$(\alpha_1 \wedge \cdots \wedge \alpha_n)(v_1, \ldots, v_k) \;=\; \sum_{\sigma \in S_k} sgn(\sigma) \alpha_1(v_{\sigma(1)}) \ldots \alpha_k(v_{\sigma(k)}) \;=\; \det(\alpha^i(v_j)).$$

Now, we know that a k-linear function on a vector space V (with a basis $\{e_1, \ldots, e_n\}$) is completely determined by its values on all k-tuples $(e_{i_1}, \ldots, e_{i_k})$. If the k-linear function is alternating, it is sufficient to know its values on all k-tuples $(e_{i_1}, \ldots, e_{i_k})$ with $1 \leq i_1 < \cdots < i_k \leq n$. Let $\{\alpha_1, \ldots, \alpha_n\}$ be the dual basis of V. It can be shown (see [6] for the proof) that the alternating k-linear functions $\alpha_{i_1} \wedge \cdots \wedge \alpha_{i_k}$,

with $1 \leq i_1 < \cdots < i_k \leq n$ form a basis for the space of all alternating k-linear tensors on V.
Turning again to k-vector fields on a manifold M, it follows from the results above that a basis for the alternating k-tensors on T_p^*U is formed by all $(dx_{i_1})_p \wedge \cdots \wedge (dx_{i_k})_p$, $1 \leq i_1 < \cdots < i_k \leq n$. We can therefore locally express at each p in M a k-linear form ω on T_pM as the linear combination

$$\omega_p = \sum a_{i_1,\ldots,i_k}(p)((dx_{i_1})_p \wedge \cdots \wedge (dx_{i_k})_p),$$

and omitting p we write $\omega = \sum a_{i_1,\ldots,i_k}(dx_{i_1} \wedge \cdots \wedge dx_{i_k})$.
As a final step, we define an anti-derivation on the set of all differential forms on M, A, called the *exterior derivative*.

Definition 2.18. An *exterior derivative* on a manifold M is a map $D\colon A \to A$ such that

(i) $D(\omega\tau) = (D\omega) \wedge \tau - \omega \wedge D\tau$,

(ii) $D \circ D = 0$,

(iii) If f is a smooth function on M, and X a smooth vector field, $(Df)(X) = Xf$.

The third condition is equivalent to say that

$$Df \equiv df = \sum_i \frac{\partial f}{\partial x^i} dx_i$$

Now, let $(U, x^i, \ldots, x^n)$ be a coordinate chart on M as usual, and ω a differential k-form, $\omega = \sum a_{i_1,\ldots,i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_1}$. It follows from the definition of exterior derivative that

$$\begin{aligned} D\omega &= D(\sum a_{i_1,\ldots,i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_1}) \\ &= \sum D(a_{i_1,\ldots,i_k}) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_1}) + \sum a_{i_1,\ldots,i_k} D(dx_{i_1} \wedge \cdots \wedge dx_{i_1}) \\ &= \sum D(a_{i_1,\ldots,i_k}) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_1}) \\ &= \sum \sum_j \frac{\partial a_{i_1,\ldots,i_k}}{\partial x^j} dx^j \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_1} \end{aligned}$$

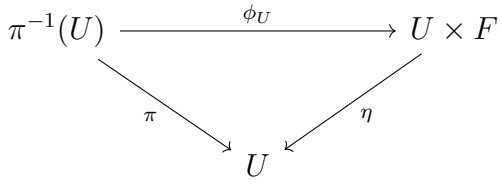
2.7 Fiber bundles

A fiber bundle over a smooth manifold M encodes the data of a family of manifolds, all diffeomorphic to some other manifold F, parametrized by M. Locally, it looks like the product manifold $U \times F$, U open set of M.

Definition 2.19. Let E, M, F be three smooth manifolds, and $\pi\colon E \to M$ a smooth surjection. A *local trivialization* with fiber F is an open cover $\mathcal{U}$ for M together with a collection of diffeomorphisms

$$\{\phi_U\colon \pi^{-1}(U) \to U \times F \ : \ U \in \mathcal{U}\}$$

that agree with the projection to the first factor $\eta\colon U \times F \to U$, i.e, each diagram



should commute.

With those notions clarified, we can provide the definition of a *fiber bundle*.

Definition 2.20. A *fiber bundle* (E, M, π) with fiber F is a smooth surjection $\pi\colon E \to M$ with a local trivialization with fiber F, i.e, it is *locally trivial* with fiber F. The manifold E is called the *total space*, and the manifold M is the *base space* of the fiber bundle. The *fiber* over a point x in M is the set $E_x := \pi^{-1}(x)$.

Note that, as π is a submersion, according to the theorem 2.4, each fiber is a regular submanifold of E. We now provide a first basic example.

Example 2.2. Let M, F be manifolds, then define $E = M \times F$ and

$$\begin{aligned} \pi\colon M \times F &\to M \\ (p, f) &\mapsto p \end{aligned}$$

Then (E, M, π) is a bundle called the trivial bundle.

A fiber bundle $\pi\colon E \to M$ is called a *vector bundle* if each fiber $\pi^{-1}(x)$ is a vector space. The collection of tangent spaces to a manifold has the structure of a vector bundle over the manifold, called the *tangent bundle*.

Definition 2.21. Let M be a smooth manifold. The *tangent bundle* is the set $TM = \bigcup_{p\in M} T_pM$. We define the bundle projection map

$$\begin{aligned} \pi : TM &\to M \\ X &\mapsto p \end{aligned}$$

where p is the point for which $X \in T_pM$.

By analogy with the tangent bundle, the union of cotangent spaces to M in all of its point is called *cotangent bundle* and is denoted by T^*M. A point in the cotangent bundle is therefore a linear functional on the tangent space T_qM at some point q in M, and is referred to as the couple (q, p), where q is the chosen point on M, and p is the functional defined on T_qM. If the manifold M has dimension n, the cotangent bundle has the structure of a smooth manifold with dimension $2n$.

Definition 2.22. Let $\pi\colon E \to M$ be a fiber bundle. A smooth map $\sigma\colon M \to E$ is called a *smooth section* of the bundle if $\pi \circ \sigma = id_M$.

A *frame* for a vector bundle $\pi\colon E \to M$ is a collection of sections $\sigma_1, \ldots, \sigma_n$ of E such that for each p in M $\sigma_1(p), \ldots, \sigma_n(p)$ form a basis for the fiber $E_p = \pi^{-1}(p)$.

With the notion of a smooth section of a vector bundle, it is possible to provide an alternative definition of a vector field (cfr def. 2.14).

Definition 2.23. Let M be a smooth manifold, and TM its tangent bundle, $\pi\colon TM \to M$, π smooth. A *vector field* is a *smooth section* of TM. We denote the set of all the smooth sections of the tangent bundle with $\mathfrak{X}(M)$.

In other words, a vector field can be seen as an assignment to each point of a vector in the fiber over that point.

2.8 Lie groups

We now define a very important class of groups, namely Lie groups, which are groups with a differentiable structure that is compatible with the group operations. Due to this property, they can be studied with the tools of differential geometry. We also provide a general definition of Lie algebras; any Lie group gives rise to a Lie Algebra, that can be identified with its tangent space to the identity, equipped with a bracket operation $[\,,\,]$.

Definition 2.24. A *Lie Group* $(G, \cdot)$ is a group with the group operation $\cdot$, and a smooth manifold with a differentiable structure such that the group multiplication and inverse

$$\begin{aligned} \mu\colon G \times G &\to G & \qquad \iota\colon G &\to G \\ (g, h) &\mapsto gh & \qquad g &\mapsto g^{-1} \end{aligned}$$

are C^∞. A *morphism of Lie groups* is a group homomorphism which is also differentiable.

Lie groups are very common to encounter, and we now provide some examples.

Example 2.3. The unit circle in the complex plane $S^1 = \{z \in \mathbb{C} \mid \|z\| = 1\}$, with the usual complex multiplication, $(S^1, \cdot_{\mathbb{C}})$ is a commutative Lie group.

Example 2.4. $GL(n, \mathbb{R}) = \{f \colon \mathbb{R}^n \to \mathbb{R}^n \mid f \text{ is linear and } detf \neq 0\}$, with the composition of functions, $(GL(n, \mathbb{R}), \circ)$ is a Lie group called the General Linear group. More generally, $GL(V) = \{f \colon V \to V \text{ linear and invertible }\}$ is a Lie group isomorphic to $GL(n, \mathbb{R})$.

For $g \in G$, let $\ell_g \colon G \to G$, $\ell_g(x) = \mu(g, x)$ be the operation of *left translation* by g, and $r_g \colon G \to G$, $r_g(x) = \mu(x, g)$ be the operation of *right translation* by g. For $g \in G$, the left translation ℓ_g is a diffeomorphism of G onto itself, with inverse $\ell_{g^{-1}}$, that maps the identity element to g. The diffeomorphism ℓ_g induces an isomorphism of tangent spaces

$$d(\ell_g)_e \colon T_e(G) \to T_g(G).$$

If we can describe the tangent space $T_e(G)$ at the identity, then $(d\ell_g)_e(T_e(G))$ will give a description of the tangent space $T_g(G)$ at any point $g \in G$.

Definition 2.25. A *Lie algebra* $(L, +, \cdot, [\,,\,])$ is a vector space $(L, +, \cdot)$ equipped with bilinear map $[\,,\,] \colon L \times L \to L$ that is antisymmetric, that is, $[x, y] = -[y, x]$, and satisfies the Jacobi identity

$$[\,x\,,\,[\,y\,,\,z\,]\,] + [\,z\,,\,[\,x\,,\,y\,]\,] + [\,y\,,\,[\,z\,,\,x\,]\,] = 0$$

Example 2.5. We can equip the set $\mathfrak{X}(M)$ defined in 2.23 with two operations

$$\begin{aligned} + \colon \mathfrak{X}(M) \times \mathfrak{X}(M) &\to \mathfrak{X}(M) \\ (\sigma, \tau) &\mapsto \sigma + \tau \end{aligned}$$

defined by $(\sigma + \tau)(p) = \sigma(p) + \tau(p)$, where the $+$ is defined on the vector space T_pM, and

$$\begin{aligned} \cdot \colon C^\infty(M) \times \mathfrak{X}(M) &\to \mathfrak{X}(M) \\ (f, \tau) &\mapsto f \cdot \tau \end{aligned}$$

where $(f \cdot \tau)(p) = f(p) \cdot \tau(p)$ is the scalar multiplication on the vector space T_pM. Therefore $(\mathfrak{X}(M), +, \cdot)$ has the structure of a module over the ring $C^\infty(M)$. Note that $\mathfrak{X}(M)$ is also a vector field over $\mathbb{R}$. Now let X and Y be two smooth

vector fields on a smooth manifold M. Their *Lie Bracket* $[\,X\,,\,Y\,]$ at p is defined by

$$[\,X\,,\,Y\,]_p f = [\,X_p Y - Y_p X\,]f$$

for any germ f of a C^∞ function at p. It can be easily checked that $[\,X\,,\,Y\,]_p$ is a derivation of $C^\infty(M)$ and hence a tangent vector at p. If p varies in M, $[\,X\,,\,Y\,]$ is a vector field on M, that is also smooth. $[\ ,\ ]$ becomes therefore a bilinear map on the vector space of all smooth vector fields on M, $[\ ,\ ]\colon \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$. As a bilinear map, $[\ ,\ ]$ is antisymmetric and satisfies the Jacobi identity, as follows from its pointwise definition. Therefore, $\mathfrak{X}(M)$ is a Lie algebra with Lie brackets $[\ ,\ ]$.

2.9 Lie groups actions on manifolds

The concept of action of a group on a manifold is very important for the applications. Indeed, it encodes the symmetries of the geometrical objects we want to study.

Definition 2.26. Let $(G, \cdot)$ be a Lie group and M a smooth manifold, a smooth map

$$\begin{aligned} \mu\colon G \times X &\to X \\ (\sigma, x) &\mapsto \sigma \cdot x \end{aligned}$$

is a smooth *left action* of G on M if

- (i) $e \cdot x = x$, and
- (ii) $\sigma \cdot (\tau \cdot x) = (\sigma\tau) \cdot x)$

for all $\sigma, \tau \in G$, $x \in X$. We can call M a *right G-manifold*.

Similarly, we define the right action of a group on a manifold

Definition 2.27. A smooth map

$$\begin{aligned} \mu\colon X \times G &\to X \\ (x, \sigma) &\mapsto x \cdot \sigma \end{aligned}$$

is a smooth *right action* of G on M if

- (i) $x \cdot e = x$, and
- (ii) $(x \cdot \sigma) \cdot \tau = x \cdot (\sigma\tau)$

for all $\sigma, \tau \in G$, $x \in X$.

Given a left action on some manifold, it is possible to turn it into a right action, and vice-versa. It is sufficient to define the right action $x \cdot g = g^{-1} \cdot x$, and it can be easily checked that defined this way it satisfies the property needed.
We now provide some definitions for a left action on some smooth manifold M that have their equivalents for right actions.

- For any $p \in M$ we define its *orbit* under the action as the set

$$\mathcal{O}_p = \{q \in M \mid \exists\, g \in G : \mu(g, p) = q\},$$

  $\mathcal{O}_p \subseteq M$.

- We define the equivalence relation

$$p \sim q \iff \exists\, g \in G \mid q = \mu(g, p).$$

  Hence, p and q are equivalent if they lie on the same *orbit*. We denote the *quotient space* $M/\!\sim$ under the equivalence relation $\sim$ as M/G.

- For any $p \in M$ we define the *stabilizer* $\mathcal{S}_p = \{g \in G \mid \mu(g, p) = p\}$, $\mathcal{S}_p \subseteq G$.

- An action is called *free* if for each p in M $\mathcal{S}_p = e$.

We now define right G-equivariant maps of manifolds (and analogously left G-equivariant maps) .

Definition 2.28. Let N and M be right G-manifolds. A smooth map $f\colon N \to M$ is *right G-equivariant* if

$$f(x \cdot g) = f(x) \cdot g,$$

for every $(x, g) \in N \times G$.

Thanks to the possibility of defining a right action from a left G-action on a manifold, we can define G-equivariant maps of manifolds.

Definition 2.29. Let N be a right G-manifold and M a left G-manifold. A smooth map $f\colon N \to M$ is *G-equivariant* if

$$f(x \cdot g) = f(x) \cdot g = g^{-1} \cdot f(x)$$

.

Definition 2.30. A fiber bundle $\pi\colon P \to M$ with fiber G is a *principal G-bundle* if G acts smoothly and freely on P on the right, and every fiber-preserving diffeomorphism

$$\phi_U\colon \pi^{-1}(U) \to U \times G$$

is G-equivariant, where the right G-action on $U \times G$ is defined by

$$(x, h) \cdot g = (x, hg),$$

and hence if $\phi_U(x) = (x, h)$ for some $h \in G$, $\phi_U(xg) = (x, hg)$ must hold.

Chapter 3

Subriemannian geometry

In this chapter, we first define Riemannian manifolds, and recall some concepts of Hamiltonian mechanics on manifolds. Our purpose, here, is to extend such notions to the context of subriemannian manifolds, and find the related geodesic curves. Subriemannian manifolds are basically manifolds endowed with a distribution and an inner product on the distribution (see [8]). In the next chapter we will use the notions introduced in this chapter to address the problem of border completion. In the last section, we briefly introduce connections and Cartan's approach to connections.

3.1 Riemannian manifolds

We begin this section with the definition of a Riemannian structure on a manifold, which makes possible to define the length of a curve and geodesic curves on the manifold.

Definition 3.1. A *Riemannian manifold* (M, g) is a smooth manifold M equipped with a *Riemannian metric*, i.e, a non-degenerate inner product $\langle\,\rangle_g \colon T_xM \times T_xM \to \mathbb{R}$ on each tangent space T_xM of M.

Let $(q, p) \in T^*M$. We can define the *metric tensor* g_q as the function

$$\begin{aligned} g_q \colon T_qM &\to T_q^*M \\ v_q &\mapsto g_q(v_q) \end{aligned}$$

such that $g_q(v_q)(w_q) = \langle v_q, w_q\rangle_g$. We can compute the inverse of the metric tensor, the so-called *cometric tensor*

$$\begin{aligned} g_q^{-1} = \tilde{g}_q \colon T_q^*M &\to T_q^{**}M = T_qM \\ v_q^* &\mapsto \tilde{g}_q(v_q^*) \end{aligned}$$

and its associated scalar product

$$((\ ,\ ))_{\tilde{g}}\colon T_q^*M \times T_q^*M \to \mathbb{R}$$
$$(v_q^*, w_q^*) \mapsto ((v_q^*, w_q^*))_{\tilde{g}} \ = \ (\tilde{g}(v_q^*))\, w_q^*.$$

We now give the definition of *length* of a smooth curve and distance between two points in Riemannian geometry. See [9] for more details.

Definition 3.2. We define the *length* of a smooth curve γ

$$l = l(\gamma) = \int_\gamma \|\dot{\gamma}\| dt$$

where $\|\dot{\gamma}\| = \sqrt{\langle\, \dot{\gamma}(t), \dot{\gamma}(t)\,\rangle_g}$. The *distance between two points* p and q in M is defined by $d(p,q) = \inf\, l(\gamma)$, where the infimum is taken over all smooth curves connecting p and q.

3.2 Subriemannian manifolds

We want to introduce the notion of a subriemannian manifold, and in order to do so we need some fundamental definitions.

Definition 3.3. A *distribution* $\mathcal{D}$ on a smooth manifold M is an assignment $x \mapsto D_x$ for each $x \in M$, where D_x is a linear subspace of T_xM.

It should be noticed that, once a chart is fixed, being D_x a linear subspace of T_x, the assignment above can be performed by selecting for each x in M the basis vectors in T_xM that generate the linear subspace D_x,

$$D_x = \text{span}\{X_1(x), \ldots, X_n(x)\}.$$

Each $X_i\colon M \to TM$ is contained in the space of all vector fields on M, $\mathfrak{X}(M)$, and the collection of sections $X_1, \ldots, X_n$ of TM forms a *local frame* for $\mathcal{D}$. To check whether $\mathcal{D}$ is smooth or not, it is sufficient to check the smoothness of the coefficients of the chosen basis vectors for each D_x as functions of the point x. We invite the reader to consult [7], sec. 12.5 for the proof.

Definition 3.4. Let M be a smooth manifold of dimension $dimM$, $k < dimM$. A smooth distribution $\mathcal{D}$ is said to be *regular of rank* k if all of the D_x have the same dimension as vector subspaces, $dimD_x = k$.

Example 3.1. Let $M = \mathbb{R}^3$, and $\mathcal{D} = \text{span}\{\partial_x, \partial_y\}$. We know that $T_xM = \mathbb{R}^3 = \text{span}\{\partial_x, \partial_y, \partial_z\}$. It is evident that the distribution $\mathcal{D}$ is smooth.

A distribution $\mathcal{D}$, $D_x = \text{span}\{X_1(x), \dots, X_n(x)\}$, is called *completely non-integrable* (or *bracket generating*) if any local frame $X_1, \dots, X_n$, together with all of its iterated Lie brackets $[X_i, X_j]$, $[X_i, [X_j, X_k]]$, $\dots$, spans the whole tangent bundle TM, i.e, if any tangent vector $X \in T_xM$ can be written as a linear combination of vectors of the type $X_i(x), [X_i, X_j](x), [X_i, [X_j, X_k]](x)$, and so on. Let us now introduce some useful terminology:

- $[\mathcal{D}, \mathcal{D}] := \text{span}\{[X, Y] \; \forall \, X, \, Y \in \mathcal{D}\}$;
- $[\mathcal{D}, [\mathcal{D}, \mathcal{D}], \dots] = TM$ if $\mathcal{D}$ is completely non-integrable;
- $[\mathcal{D}, \mathcal{D}] \subset \mathcal{D}$ if $\mathcal{D}$ is *involutive*.
- An immersed submanifold N of M is an *integral manifold* of $\mathcal{D}$ if and only if $D_p = T_pN$ for each p in N.
- $\mathcal{D}$ is a *completely integrable* distribution if and only if there exists an integral manifold of $\mathcal{D}$ for each p in M.

With these notions acknowledged, we quote a fundamental theorem of differential geometry, the *Frobenius Theorem* (see [19] pp.195-209).

Frobenius Theorem. *Let M be a smooth manifold of dimension n. A smooth r-dimensional distribution $\mathcal{D}$ on M is completely integrable if and only if it is involutive.*

The concepts of bracket generating distributions and integrable distributions stand at opposite ends, and the Frobenius theorem can help us understand where the most profound differences lie.
We are now ready to define the notion of subriemannian geometry on a manifold, and extend the concept of distance between two points to this context.

Definition 3.5. A *subriemannian manifold* $(M, \mathcal{D}, g)$ is a smooth manifold M of dimension n with a completely non-integrable smooth regular distribution of rank $k < n$, $\mathcal{D}$, equipped with a non-degenerate inner product $\langle \, , \rangle_g \colon \mathcal{D}_x \times \mathcal{D}_x \to \mathbb{R}$ for any $x \in M$.

We call $\mathcal{D}$ the *horizontal distribution*, and a curve $\gamma \colon]-\epsilon, \epsilon[\to M$ on the manifold M is called *horizontal* if $\dot{\gamma}(t) \in D_{\gamma(t)}$ for every $t \in]-\epsilon, \epsilon[$. We define the *length* $l = l(\gamma)$ of a smooth horizontal curve γ as in Riemannian geometry,

$$l = l(\gamma) = \int_\gamma \|\dot{\gamma}\| dt$$

but in this context $\|\dot{\gamma}\|$ is computed through the inner product $\langle\,,\rangle_g$ on the horizontal spaces $D_{\gamma(t)}$,

$$\|\dot{\gamma}\| = \int_\gamma \sqrt{\langle\dot{\gamma}(t),\dot{\gamma}(t)\rangle_g}dt.$$

We can now define the subriemannian distance between two points X and Y in M

$$d(X,Y) \;=\; \inf\, l(\gamma)$$

where the infimum is taken over all the smooth horizontal curves that connect the two points. If there is no such curve, then $d(X,Y)=\infty$ by definition.
We now quote an important result, crucial to the purposes of our discussion, which states that the subriemannian distance between two points on a manifold M is always finite (see [8] sec. 1.6)

Chow-Rashevskii Theorem. *If $\mathcal{D}$ is a completely non-integrable distribution on a connected smooth manifold M, then any two points in M can be joined by a horizontal path.*

The fundamental hypothesis here is the complete non-integrability of the distribution $\mathcal{D}$: there is no submanifold of M the horizontal paths are restricted to lie on.

Definition 3.6. A smooth horizontal path that realizes the distance between two points is called a *geodesic*.

Because of the Chow-Rashevskii theorem, the existence of a geodesic is always guaranteed in the context of subriemannian geometry.

If we denote with $\mathcal{C}(M)$ the set of all smooth horizontal curves on M, we can define the *curve energy* functional

$$\begin{aligned} E\colon \mathcal{C}(M) &\to \mathbb{R} \\ \gamma &\mapsto E(\gamma) = \int_\gamma \frac{1}{2}||\dot{\gamma}||^2. \end{aligned}$$

If we fix the time T of the path $\gamma\colon [0,T]\to M$, we can state the following proposition.

Proposition 3.1. *The horizontal curve γ minimizes the energy E among all curves joining q_0 to q_1 in time T if and only if it minimizes the length l among all curves joining q_0 and q_1, and is parametrized to have constant speed $c = d(q_0,q_1)/T$*

One can show this proposition making use of the Cauchy-Schwartz inequality. According to this result, the problem of finding the shortest paths connecting two points on a manifold can be addressed minimizing the energy functional, that is nothing more than the constant *Hamiltonian functional*. We shall proceed this way when we compute the geodesic curves in the following sections.

3.3 Hamiltonian mechanics on manifolds and Poisson brackets

Before finding the geodesic equations in subriemannian geometry, we need to introduce some basic notions of Hamiltonian mechanics on Riemannian manifolds. We invite the reader to consult [8] (appendix A) and chapter 7 of [20] for more details.
Let (Q, g) be a Riemannian manifold, the so-called *configuration space*, endowed with local coordinates $(q^1, \dots, q^n)$ near a point q. We call the cotangent bundle T^*Q its *phase space*.

Definition 3.7. A *Hamiltonian* is a function $H : T^*Q \to \mathbb{R}$ on the phase space.

We can associate a cotangent vector p to a tangent vector v through the inner product that defines the Riemannian structure

$$p \mapsto v$$

defined by its action on a tangent vector w, $p(w) = \langle v, w \rangle_q$. In coordinates, $v = \sum v^i \partial q^i$, and $p = \sum p_i dq^i$, and we can write

$$p_i = \sum_j g_{ij}(q) v^j,$$

where g_{ij} is the metric in the chosen coordinates. Hence, the coordinates q^i induce fiber coordinates p_i. We call the 2n functions (q^i, p_i) *canonical coordinates* for our phase space. We can now define the so-called *tautological 1-form* on T*Q, $\Theta = \sum_i p_i dq^i$.

Definition 3.8. The *canonical symplectic form* on T^*Q is the non-degenerate and closed 2-form $-d\Theta$, that has the coordinate expression

$$\omega = \sum_i dq^i \wedge dp_i.$$

Non-degeneracy of ω implies that we can construct a linear isomorphism

$$\begin{aligned} T_{(q,p)}(T^*Q) &\to T^*_{(q,p)}(T^*Q) \\ v &\mapsto i_v \omega = \omega(v, \ ) \end{aligned}$$

Therefore, if we are given a function H, then there is a unique vector field X_H such that $dH = \omega(X_H, \ )$, (note that the gradient of H is a cotangent vector, and

hence a 1-form on T^*Q). The so-defined vector field is called the *Hamiltonian vector field*. In canonical coordinates (q^i, p_i) this vector field can be written as

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i} \tag{3.1}$$

which are a set of ordinary differential equations for the integral curves of X_H, and are called the *Hamilton's equations*.

We can now derive the Hamilton's equations in a different form, through the Poisson bracket's formalism.

Definition 3.9. Let X_g be the Hamiltonian vector field associated to a function $g\colon T^*Q \to \mathbb{R}$, and $f\colon T^*Q \to \mathbb{R}$ a smooth function. We can define the *Poisson brackets*

$$\begin{aligned} \{\,,\,\}\colon C^\infty(T^*Q) \times C^\infty(T^*Q) &\to C^\infty(T^*Q) \\ (f, g) &\mapsto \{f, g\} = X_g(f). \end{aligned}$$

We have thus, by definition of gradient of a function

$$\{f, g\} = df(X_g) = \omega(X_f, X_g) = -\omega(X_g, X_f) = -\{g, f\}$$

It can also be shown that $\{\,,\,\}$ is bilinear and satisfies the Jacobi identity. It follows that $\{\,,\,\}$ forms a Lie algebra structure on the ring $C^\infty(T^*Q)$. In addition, $\{\,,\,\}$ satisfies the Leibniz rule, and hence $\{\,, H\}$ defines a derivation on $C^\infty(T^*Q)$, that is a vector field, that coincides with the Hamiltonian vector field X_H. In canonical coordinates, we can compute

$$\{f, g\} = \sum_i \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i}$$

Finally, we can derive the *Hamilton's equations in bracket forms*. Let $\gamma(t)$ be the integral curve for the vector field X_H, and f a smooth function. Then

$$\frac{d}{dt}(f(\gamma(t)) = (X_{H\gamma}(t))f = \{f, H\}(\gamma(t)).$$

These equations are usually written as $\dot{f} = \{f, H\}$. Letting f run through the canonical coordinates q^i, p_j we obtain our earlier expression for Hamilton's equations.

3.4 Geodesic equations

A Riemannian manifold comes equipped with a non-degenerate inner product defined on the tangent space to the manifold, and a metric tensor. This fact allows to construct a non-degenerate inner product on each cotangent space, that comes together with the *cometric tensor*.

In subriemannian geometry we cannot carry out the same computation, as the inner product on a subriemannian manifold is defined only on the distribution $\mathcal{D}$, and has the same rank as the distribution $k < dimM$. The metric tensor associated to the inner product is clearly not invertible. For this reason, we need to find another way to define the cometric tensor, which will lead us to a definition of the Hamiltonian functional associated with the subriemannian structure, the *subriemannian Hamiltonian*.
Let $(M, \mathcal{D}, g)$ be a subriemannian manifold. We can define the *cometric* β as the symmetric bundle map

$$\begin{aligned} \beta \colon T^*M &\to TM \\ (q,p) &\mapsto \beta_q(p) \in T_qM \end{aligned}$$

that is uniquely determined by the following two conditions

(1) $\mathrm{im}(\beta_q) = D_q$;

(2) $p(v) = \langle \beta_q(p), v \rangle_g$ for $v \in D_q$, $p \in T_q^*M$,

where $\langle , \rangle$ is the inner product defined on the distribution $\mathcal{D}$ associated to the metric g. The cometric β allows us to define a degenerate bilinear form $((,))$ on $T^*M \times T^*M$ as

$$\begin{aligned} ((,))_q \colon T_q^*M \times T_q^*M &\to \mathbb{R} \\ (p,r) &\mapsto ((p,r))_q = p(\beta_q(r)) \end{aligned}$$

and $p(\beta_q(r)) = \langle \beta_q(p), \beta_q(r) \rangle_g$ by condition (2).

Definition 3.10. Let M be a smooth manifold with a subriemannian structure. We define the *Hamiltonian functional* H

$$\begin{aligned} H \colon T^*M &\to \mathbb{R} \\ (q,p) &\mapsto \frac{1}{2}((p,p))_q \end{aligned}$$

The Hamiltonian functional is strictly connected to the module of the velocity of a horizontal curve $\dot{\gamma}(t)$. In fact, by definition of horizontal curve, $\dot{\gamma}(t) \in D_{\gamma(t)}$. Hence $\dot{\gamma}(t) = \beta_{\gamma(t)}(p)$ for some p in $T^*_{\gamma(t)}$, by definition of β. It follows that

$$\begin{aligned}\frac{1}{2}\|\dot{\gamma}\| = \frac{1}{2}\langle\dot{\gamma}(t),\dot{\gamma}(t)\rangle_g = \frac{1}{2}\langle\beta_{\gamma(t)}(p),\beta_{\gamma(t)}(p)\rangle_g = \frac{1}{2}p(\beta_{\gamma(t)}(p)) = \frac{1}{2}((p,p))_{\gamma(t)} \\ = H(p,\gamma(t))\end{aligned}$$

Let $q = \gamma(t)$. We can conclude that $H(q,p) = \frac{1}{2}\|\dot{\gamma}\|$. H uniquely determines β through the *polarization identity*

$$((a,b)) = \frac{1}{4}\left[((a+b,a+b)) - ((a-b,a-b))\right],$$

and β uniquely determines the subriemannian structure.
We now want to compute the Hamiltonian functional associated with a subriemannian structure. In order to do so, we need to define some useful linear functions on the cotangent bundle, that are called the *momentum functions*.

Definition 3.11. Let X be a vector field on a manifold Q. We define the *momentum function* for X on the cotangent bundle

$$\begin{aligned}P_X &: T^*Q \to \mathbb{R} \\ &(q,p) \mapsto P_X(q,p) = p(X(q)).\end{aligned}$$

Let $X_1,\dots,X_n$ be a basis for D_p at each p for a distribution $\mathcal{D}$ on Q. The $P_a = P_{X_a}$ are the momentum functions for our frame. We know the expression of X_a relative to the local coordinates x^i to be $X_a = \sum_i X_a^i(x)\partial_i$, $X_a^i(x)$ the coefficient functions. Hence, it is straightforward to write

$$P_{X_a}(x,p) = \sum_i X_a^i(x)\, p(\partial_i) = \sum_i X_a^i(x)\, P_{\partial_i}(x.p)$$

and express P_{X_a} in terms of the functions $p_i = P_{\partial_i}$,

$$P_{X_a} = \sum_i X_a^i(x)\, p_i.$$

Let $\langle , \rangle$ be the inner product on the distribution $\mathcal{D}$, with respect to which $X_1,\dots,X_n$ are orthonormal. We can represent it as the matrix $g_{ab}(q) = \langle X_a(q), X_b(q)\rangle_q$. The associated degenerate bilinear form $((,))_q$ on T^*Q can be therefore represented

by the matrix $g^{ab}(q)$ that is such that $g^{am}\, g_{mb} = \delta^a_b$. We have now found a way of computing the Hamiltonian functional in terms of the momentum functions

$$H(p,q) \;=\; \frac{1}{2}((p,p))_q \;=\; \frac{1}{2}\sum_{a,b} g^{ab}(q) P_a(q,p) P_b(q,p)$$

In addition, if the X_a are an orthonormal frame for $\mathcal{H}$ with respect to the subriemannian inner product, $\langle X_a(q), X_b(q)\rangle_q = \delta_{ab}$, we can write

$$H \;=\; \frac{1}{2}\sum_{a} P_a^2.$$

The function H is a smooth function on the cotangent bundle T^*Q. We saw in section 3.3 that any smooth function on the cotangent bundle can be associated with a vector field on the cotangent bundle, and generates a system of Hamiltonian differential equations of the form 3.1. These equations are called the *normal geodesic equations*.
We now state an important theorem (for the proof see [8] section 1.9) which ensures that the normal geodesic equations lead to a solution of the problem of finding the geodesics on the horizontal distribution that defines our subriemannian structure.

Theorem 3.1. *Let $\zeta(t) = (\gamma(t), p(t)) \in T^*Q$ be a solution to the normal geodesic equations for the subriemannian Hamiltonian H, $\pi(\zeta(t)) = \gamma(t)$ its projection to Q. Then every sufficiently short arc of γ is a minimizing subriemannian geodesic. Once its endpoints are given, γ is unique.*

We call the projected curves $\gamma(t)$ the *normal* subriemannian geodesics. In Riemannian geometry all geodesics are normal. This fact is no longer true in subriemannian geometry, where we can find minimizing geodesics that do not solve the normal geodesic equations. We call these particular geodesics *singular geodesics*. For our purposes, we are only interested in normal geodesics.

3.5 Connections

In this section we define connections, that are a generalization of directional derivatives to vector fields and tensor fields. We invite the reader to consult chapter 6 of [7] and [9] (appendix B).

Definition 3.12. Let M be an arbitrary smooth manifold, $\mathfrak{X}(M)$ the space of all vector fields on M, and let us set $\mathcal{F} = C^\infty(M)$. We define a *connection* on M as the $\mathbb{R}$-bilinear map

$$\begin{aligned} \nabla\colon \mathfrak{X}(M) \times \mathfrak{X}(M) &\to \mathfrak{X}(M) \\ (X, Y) &\mapsto \nabla_X Y \end{aligned}$$

that for all X, Y in $\mathfrak{X}(M)$, f in $\mathcal{F}$, satisfies

(i) (Leibniz rule) $\nabla_X(fY) = (Xf)\,Y + f\,\nabla_X Y$;

(i) ($\mathcal{F}$-linearity) $\nabla_{fX}Y = f\,\nabla_X Y$.

Example 3.2. An immediate example is the directional derivative in $\mathbb{R}^n$. Let v be a vector in $\mathbb{R}^n$, $v = \{v^i\}$. If $y = \sum_i b^i \partial_i$, the directional derivative D_v is defined by

$$D_v\, y = \sum_{i,j} v^j\, (\partial_j b^i)\partial_i.$$

We can show that the directional derivative satisfies properties (i) and (ii), i.e, it is a connection on $\mathbb{R}^n$.

Given a connection on an n-dimensional smooth manifold M, once local coordinates $x^1, \ldots, x^n$ are fixed, and the basis vectors on each tangent space $\partial_1, \ldots, \partial_n$ are set accordingly (see 2.2), we can construct the *Christoffel symbols*. Let $X = \sum_i a^i \partial_i$, $Y = \sum_j b^j \partial_j$. By definition

$$\begin{aligned}\nabla_X Y &= \nabla_{\sum_i a^i \partial_i} \sum_j b^j \partial_j = \sum_i a^i \nabla_{\partial_i} \sum_j b^j \partial_j \\ &= \sum_{i,j} a^i (\partial_i b^j)\, \partial_j + \sum_{i,j} a^i b^j\, \nabla_{\partial_i} \partial_j\end{aligned}$$

holds. Hence, the connection is fully defined once the $\nabla_{\partial_i}\partial_j \in \mathfrak{X}(M)$ are given. We define the Christoffel symbols Γ_{ij}^k by the equation

$$\nabla_{\partial_i}\partial_j = \Gamma_{ij}^k \partial_k.$$

We can therefore say that, locally, a covariant derivative is identified with its Christoffel symbols, once we fix a local coordinates system.
Before going on with our discussion, we define the *torsion* and the *curvature* of a connection on a manifold M. We set

- the *torsion* vector field $T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$,

- the *curvature* operator $R(X,Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}$.

Example 3.3. We can compute the torsion on the Euclidean flat space $\mathbb{R}^n$. Let $X = \sum_i a^i \partial_i$, and $Y = \sum_j b^j \partial_j$. We know that the directional derivative along, say, X of Y is given by $D_X Y = \sum_{i,j} a^i\, (\partial_i b^j)\, \partial j$. Hence,

$$T(X,Y) = \sum_{i,j} a^i\, (\partial_i b^j)\, \partial_j - b^i\, (\partial_i a^j)\, \partial_j - [a^i \partial_i, b^j \partial_j] = 0.$$

Let us now note that, through the approach we used so far, in order to fully determine a connection it is fundamental to fix coordinate charts. However, this is not the only possible approach. Indeed, when M is a Riemannian manifold, it is possible to develop a chart-independent expression for the connection using orthonormal frames, which is the idea underlying Cartan's formalism. We invite the reader to consult [9], [7] for more details.

First, we define the *solder form*

$$\begin{aligned} \theta \colon T_pM &\to \mathbb{R}^n \\ v &\mapsto \theta(v) \,=\, (v_1, \dots, v_n) \end{aligned}$$

that maps an orthonormal frame with respect to the Riemannian metric in TM (∂_μ in coordinates) to an orthonormal frame $\{e_i\}$ in $\mathbb{R}^n$. Hence, we can write

$$\nabla_{e_i} e_j \,=\, \omega_j^k(e_i)\, e_j,$$

where $\omega \,=\, (\omega_j^k)$ is a matrix of 1-forms. We now state the important proposition that allows us to identify a connection on a Riemannian manifold with the just defined matrix of 1-forms ω (see [9] appendix B).

Proposition 3.2. *Locally, on a Riemannian manifold M, a connection ∇ is completely determined by both its Christoffel symbols Γ_{ij}^k and the matrix of 1-forms ω_j^k.*

Given a manifold M, there are many possible connections that can be defined on it, as well as many possible metrics. We want now to impose some additional conditions on the connections that can be defined on a manifold, in such a way that the choice of a metric determines a unique connection satisfying such properties.

Definition 3.13. Let $(M, \langle , \rangle)$ be a Riemannian manifold. A connection ∇ is called a *metric connection* or a *Levi-Civita connection* if

1) $Z\langle X, Y\rangle \,=\, \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle$, $\forall X, Y, Z \,\in \mathfrak{X}(M)$,

2) the connection is torsion free, i.e, $T(X, Y) \,=\, 0$ $\forall X, Y \,\in \mathfrak{X}(M)$.

Proposition 3.3. *There exists a metric connection on a Riemannian manifold satisfying properties (1) and (2). Further, such metric connection is unique.*

We invite the reader to consult chapter 6 of [7] or [21] for the proof. We now state an important theorem (by E. Cartan) that allows us to compute both the matrix of 1-forms ω that identifies the connection and the curvature of a metric connection in the orthonormal frame's picture.

Theorem 3.2. *Given a metric connection ∇ on a Riemannian manifold M, the following properties hold*

(i) ω *is skew-symmetric, i.e* $\omega_j^i = -\omega_i^j$

(ii) First structural equation. *Let* $\{e_i\}$ *be the orthonormal frame identified by the solder form, and* $\{\theta^i\}$ *the orthonormal 1-forms of* T_p^*M *found by the identification* $T_pM \cong T_p^*M$ *through the metric* $\langle , \rangle$. *The* ω_k^j *are given by the equation*

$$d\theta^i + \omega_k^i \wedge \theta^k = 0,$$

that is called the first structural equation. *Indeed, it defines the unique metric connection.*

(iii) Second structural equation.

$$\Omega_j^i = dw_j^i + w_k^i \wedge w_j^k$$

where Ω *is defined as*

$$R(X,Y)e_j = \sum_i \Omega_j^i(X,Y)e_i.$$

We now provide an example of the calculation of ω and Ω given a Riemannian manifold.

Example 3.4. (The Poincaré upper-half plane) Let (x, y) where $y > 0$ be the global coordinates for the upper-half plane of $\mathbb{R}^2$. The basis vectors ∂_1, ∂_2 of T_pM are defined accordingly. We define the Poincaré metric

$$ds^2 = \frac{1}{y^2}dx \otimes dx + \frac{1}{y^2}dy \otimes dy.$$

An orthonormal frame with respect to the metric is given by $e_1 = y\partial_1$, $e_2 = y\partial_2$, and its dual is defined by $\theta^i(e_j) = \delta_i^j$. Hence, $\theta^1 = \frac{1}{y}dx$, $\theta^2 = \frac{1}{y}dy$. From (ii), we can calculate an expression for $\omega_j^i = a\,dx + b\,dy$

$$\begin{aligned} d\theta^1 &= -\omega_2^1 \wedge \theta^2 = -\frac{1}{y^2}dy \wedge dx \\ d\theta^2 &= -\omega_1^2 \wedge \theta^1 = 0, \end{aligned}$$

that give as a result $\omega_2^1 = \omega_1^2 = -\frac{1}{y}dx$. We can compute he curvature through (iii)

$$\Omega_2^1 = -\Omega_1^2 = d\omega_2^1 = \frac{1}{y^2}dy \wedge dx.$$

Chapter 4

Border perception and border completion

In chapter 1, we gave a general overview of the visual pathway, from the eye to the primary visual cortex, with a focus on the retina, the lateral geniculate nucleus, and the primary visual cortex itself. Though not exhaustive, our description of such neural structures highlights the role these structures play in elaborating the visual input, and specifically what are the neural mechanisms that underlie edge detection. The primary visual cortex, with its specialized simple, complex, and end-stopped cells, and through its columnar functional organization, is the most important site where this aspect of the input is processed. Nevertheless, we saw that the neural process that, from the simple light inputs hitting the retinal photoreceptors, gives rise to edge detection actually begins in the retina, where the retinal ganglion cells play an extremely important role. In this chapter we first provide a simple mathematical modeling for the low visual pathway, according to that given in [22], with regard to those mechanisms responsible for edge detection and border perception. Then, through the introduction of a contact structure in the primary visual cortex, [2], [3], [4], [1], we give an interpretation of the border completion mechanism.

4.1 Edge detection

We now want to define a function that describes how the photoreceptors that lie at the bottom of the retina are activated by the light inputs. If we consider only the hemiretinal receptoral layer, we can think of it as a compact, simply connected domain $E \subset \mathbb{R}^2$, where each point in this domain corresponds to a receptor. As we saw in section 1.2, photoreceptors interact with photons in such a way that the extent of their activation is proportional to the light intensity, until

they saturate. We can therefore imagine to associate to each point $p = (x, y)$ in E a number representing the activation rate of the photoreceptor that corresponds to that point. It is not important, here, to distinguish between rods and cones, and on and off receptors: in such a model, we consider only those features that have a different logical impact on the downstream neurons. Hence, we define a field $\mathcal{R}$

$$\begin{aligned} \mathcal{R}\colon E &\to \mathbb{R} \\ (x, y) &\mapsto \mathcal{R}(x, y). \end{aligned}$$

Of course, we cannot expect this function to be continuous, but we can certainly deduce, from the previous considerations, that it is bounded.
As stated above, the layer of retinal ganglion cells, whose axons constitute the output of the retina, plays a crucial role in elaborating the visual information to give rise to border perception, and more complex activation patterns arise. Due to the competing center and surround in their concentric receptive fields, which we described in 1.3, the retinal ganglion cells best respond to contrasts in brightness, rather than to uniform light stimuli. As we did for the receptoral layer, we can think of the retinal ganglion layer as a compact, simply connected subset of $\mathbb{R}^2$, which we call $\tilde{E}$. In this model, we identify through a map

$$\begin{aligned} G\colon E &\to \tilde{E} \\ (x, y) &\mapsto (x', y') \end{aligned}$$

receptors and ganglions, in the sense that for the receptor (x, y) there exists a ganglion whose receptive field (bigger than one of a receptor) is centered at (x', y'). Such a map G must preserve distances, i.e G is an isometry according to the metric d_2 of $\mathbb{R}^2$. As to the activation pattern, we should take into account that a retinal ganglion receives inputs from many receptors in a neighborhood of $G^{-1}(x', y')$, and that it has a center-surround receptive field. We can take care of these features describing the ganglionic activation pattern through a function

$$\begin{aligned} \tilde{\mathcal{R}}\colon \tilde{E} &\to \mathbb{R} \\ (x', y') &\mapsto \tilde{\mathcal{R}}(x', y') \end{aligned}$$

defined by

$$\tilde{\mathcal{R}}(x', y') = \int_{U_\rho(G^{-1}(x', y'))} K(u, v)\mathcal{R}(u, v)dudv$$

where

$$\begin{aligned} U_\rho(x, y) &= \{(u, v) \in \mathbb{R}^2 : (u - x)^2 + (v - y)^2 \leq \rho^2\} \\ K(u, v) &= \begin{cases} \pm 1 & \text{if } (u - x)^2 + (v - y)^2 \leq (\rho - \epsilon)^2 \\ \mp 1 & \text{if } (\rho - \epsilon)^2 \leq (u - x)^2 + (v - y)^2 \leq \rho^2 \end{cases} \end{aligned}$$

The subset of E, $U_\rho(x,y)$, represents the receptive field of the corresponding ganglion, while the the integral kernel $K(u,v)$ encodes the fact that the center and the circular anulus (see fig.1.4) are in mutual competition. The two cases defining $K(u,v)$ stand for the ON-center/OFF-surround cells and the OFF-center/ON-surround cells, respectively.
An interesting property of the retinal ganglion activation pattern is that the function $\tilde{\mathcal{R}}$ is *Lipschitz continuous* in both variables on $\tilde{E}$. Even though the light image is non-continuous, its perception is. This fact provides a first example of how the human brain can reconstruct border perceptions from a collection of light inputs. We may now state the latter property as a more general proposition.

Proposition 4.1. *Let D be a compact domain in $\mathbb{R}^2$, and consider*

$$\begin{aligned} &\mathcal{S}\colon D \to \mathbb{R} \\ &(x,y) \mapsto \mathcal{S}(x,y) \;=\; \int_{U_\rho(x,y)} K(u,v)S(u,v)dudv \end{aligned}$$

where

$$\begin{aligned} U_\rho(x,y) &= \{(u,v) \in \mathbb{R}^2 : (u-x)^2 + (v-y)^2 \leq \rho^2\} \\ K(u,v) &= \begin{cases} \pm 1 & \text{if } (u-x)^2 + (v-y)^2 \leq (\rho-\epsilon)^2 \\ \mp 1 & \text{if } (\rho-\epsilon)^2 \leq (u-x)^2 + (v-y)^2 \leq \rho^2 \end{cases} , \end{aligned}$$

and S is an arbitrary function with $S(D)$ bounded. The function $\mathcal{S}$ is Lipschitz continuous.

An interesting application of such property can be found in [22]. A further smoothing reconstruction is carried out in the lateral geniculate nucleus; hence we are allowed to consider, from now on, $\tilde{R}$ a smooth function.
We finally turn to study the primary visual cortex, whose anatomical and functional properties are discussed in sections 1.6, 1.7, 1.9, and we now recall those involved in the process of border perception. First of all, we saw that in V1 there is a faithful preservation of spatial information, i.e, there is a homeomorphism between the hemiretinal receptoral layer and V1, the *retinotopic map*. This is basically how the primary visual cortex encodes for the position of the point in the retinal layer that receives the light stimulus. The existence of such retinotopic map allows us to identify V1 with a compact domain in $\mathbb{R}^2$, which we may call D. Thanks to the center-surround elongated receptive fields of its simple and complex cells, the primary visual cortex can encode for the orientation θ of some edge through the point at the center of the receptive fields of such cells (see fig.1.8). Moreover, through the so called end-stopped cells, the perception of the curvature

of those perceived edge is created. According to the fundamental work [15], V1 is organised in functional units called *hypercolumns*, that allow to analyze different aspects of the visual information separately and in parallel. In this model, we assume that at each point of the domain D there is a full set of simplex, complex and hypercomplex orientation columns, i.e, a full orientation hypercolumn. For this reason, at each point of the domain D we have available every possible orientation for any light edge through the point. For this purpose, we make use of the concept of *fiber bundle*, which we described in section 2.7,and we define the *orientation bundle* $\mathcal{E} := D \times S^1 \to D$. We now define the vector field Z on $D \times S^1$

$$Z(x, y, \theta) \;=\; -sin\theta\, \partial_x + cos\theta\, \partial_y,$$

where ∂_x and ∂_y form a basis for the tangent space at each point (x, y) in the domain D, and θ is the coordinate for the manifold S^1. Thanks to the action of the lateral geniculate nucleus, we can identify each image with a smooth function $F\colon D \to \mathbb{R}$. We can therefore define the *orientation of a function.*

Definition 4.1. Let $F\colon D \to \mathbb{R}$ be a smooth function, and $reg(D) \in D$ the subset of the regular points of F (see definition 2.11). We define the *orientation of* F as

$$\begin{aligned} \Theta\colon reg(D) &\to S^1 \\ (x, y) &\mapsto \Theta(x, y) \;=\; \text{argmax}_{\theta \in S^1}\{Z(\theta)F(x, y)\} \end{aligned}$$

The map Θ here is in charge of reproducing the behavior of simple and complex hypercolumns, so that an oriented edge is assigned at each point. We now show that the map Θ is well defined.

Proposition 4.2. *Let $F\colon D \to \mathbb{R}$ be a smooth function, and $(x_0, y_0) \in D$ a regular point for F. Then:*

1) *There exists a unique $\theta_{x_0,y_0} \in S^1$ for which the function $\zeta_{x_0,y_0}\colon S^1 \to \mathbb{R}$, $\zeta_{x_0,y_0}(\theta) = Z(\theta)F(x_0, y_0)$ reaches its maximum.*

2) *The map $\Theta\colon reg(D) \to S^1$, $\theta(x, y) = \theta_{x,y}$ is well defined and differentiable.*

3) *The set*

$$\phi \;=\; \{(x, y, \Theta(x, y)) \in D \times S^1 \;:\; \Theta(x, y) = \theta_{x,y}\}$$

is a regular submanifold of $D \times S^1$.

Proof. 1. For each (x, y) regular point of F, $\zeta_{x,y}$ is a differentiable function (and hence of course continuous), on a compact domain. Its image set must also be compact, and hence bounded and closed, i.e the function admits a maximum. We must now show the maximum is unique. Let us write explicitly

$$\zeta_{x,y}(\theta) = -sin\theta\,\partial_x F + cos\theta\,\partial_y F,$$

where we know that $(\partial_x F, \partial_y F) \neq (0, 0)$, as (x, y) is a regular point. The uniqueness of the maximum can be easily seen carrying out the first and second derivative with respect to θ.
2. The map Θ is well defined by (1), and it is clearly differentiable.
3. It is a consequence of the implicit function theorem, and theorem 2.4. □

We may now imagine a smooth contour to be represented by the level set of some smooth function $F\colon D \to \mathbb{R}$,

$$\Gamma = \{(x, y, \theta) \in \mathcal{E} \,:\, F(x, y) = c \in \mathbb{R}\},$$

and we construct the lift of Γ via the map Θ

$$\Sigma = \{(x, y, \theta) \in \mathcal{E} \,:\, (x, y) \in \Gamma,\, \theta = \Theta(x, y)\}.$$

Let $\sigma(t) = (x(t), y(t), \theta(x, y))$ be a regular parametrisation for Σ. We can show that Σ is tangent by construction to

$$\begin{cases} X = cos\theta\,\partial_x + sin\theta\,\partial_y \\ Y = \partial_\theta \end{cases},$$

as $\Theta(x, y)$ corresponds to the angle between the tangent line to the perceived orientation in (x, y) and the x axis:

$$\begin{aligned} \frac{d\sigma(t)}{dt} &= \dot{x}(t)\,\partial_x + \dot{y}(t)\,\partial_y + \dot{\theta}(t)\,\partial_\theta \\ &= \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}(cos\theta\,\partial_x + sin\theta\,\partial_y) + \dot{\theta}(t)\,\partial_\theta \\ &= \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}X + \dot{\theta}(t)Y \end{aligned}$$

We now define the distribution $\Delta = \text{span}\{X, Y\}$. By computing the Lie bracket

$$[X, Y] = -(-sin\theta\,\partial_x + cos\theta\,\partial_y) = -Z,$$

it is clear that $\text{span}\{X(q), Y(q), [X, Y](q)\} = T_q\mathcal{E}$, for each $q \in \mathcal{E}$. Our distribution Δ is therefore a *bracket generating* distribution (see section 3.2), and the lift of the smooth edge Γ is horizontal to it. In the following section, we will make use

of this horizontal distribution to construct a subriemannian structure on $\mathcal{E}$, and address the problem of the border completion mechanism in the primary visual cortex. We need now to highlight the fact that, from the neurophysiological point of view, this kind of perceptual completion performed in the human cortex must be closely related to the perception of continuity, given a set of oriented edges (see [3]).

4.2 Border completion via subriemannian geodesics

In the previous section, we dealt with the modeling of the feed-forward filtering performed by the simple and complex cells of the primary visual cortex, that allows to detect oriented edges in the visual input.
We now describe the border completion mechanism, which is the basis of the perception of objects and shapes as separated entities from the background. A mathematical model of this perceptual completion process must consider, as a phenomenological basis, that given an initial and a final segment, i.e, two boundary inducers, there is always one special path that is preferentially reconstructed as a perceived border in the human brain ([23]). In this section, we want to model such particular path as the projection onto D (see section 4.1) of a normal subriemannian geodesic in $\mathcal{E}$. To compute this geodesic curve, we need to introduce a subriemannian structure to find the related Hamiltonian, and we will follow the method in [8]. Furthermore, this grouping of isolated oriented edges into a smooth contour happens only when the relative orientations of the elements conform to curves without points of inflection ([24]). This statement finds its equivalence in the Gestalt psychology *law of good continuation.*
In the previous section we provided a model for the functioning of the hypercolumnar structure of V1 as the manifold $\mathcal{E} = D \times S^1$, which we called the orientation bundle, and for which we fixed the coordinates (x, y, θ). We saw also that, if Γ is a smooth edge on D in a given visual image, its lift Σ via the map Θ of definition 4.1 is tangent to $X = cos\theta\, \partial_x + sin\theta\, \partial_y$ and $Y = \partial_\theta$, and that the distribution spanned by these two vector fields is *bracket generating.* We now need to define an inner product on this distribution, and we choose the standard euclidean product inherited from $\mathbb{R}^3$, restricted to Δ:

$$\begin{aligned} \langle , \rangle \colon \Delta_q \times \Delta_q &\to \mathbb{R} \\ (v, w) &\mapsto \langle v, w \rangle. \end{aligned}$$

Hence, we can define the subriemannian manifold $(\mathcal{E}, \Delta, \langle , \rangle)$. We note that X and Y are orthonormal, with respect to the subriemannian inner product $\langle , \rangle$.

Therefore, according to what discussed in section 3.4, we can define a cometric for every $q \in \mathcal{E}$,

$$\begin{aligned} \beta_q \colon T_q^*\mathcal{E} &\to T_q\mathcal{E} \\ p &\mapsto \beta_q(p). \end{aligned}$$

such that $\mathrm{Im}\beta_q = \mathrm{span}\{X(q), Y(q)\}$. Let us introduce the local coordinate system $(x, y, \theta, p_x, p_y, p_\theta)$ on $T^*\mathcal{E}$, where (p_x, p_y, p_θ) are the local coordinates on the cotangent bundle corresponding to (x, y, θ), defined by writing any covector p as $p = p_x\,dx + p_y\,dy + p_\theta\,d\theta$. We obtain the expression for the cometric in local coordinates

$$\beta_q = \begin{pmatrix} cos\theta & 0 \\ sin\theta & 0 \\ 0 & 1 \end{pmatrix}$$

The subriemannian Hamiltonian associated with β, is the functional

$$\begin{aligned} H \colon T^*\mathcal{E} &\to \mathbb{R} \\ (q, p) &\mapsto \frac{1}{2}\langle \beta_q(p), \beta_q(p) \rangle, \end{aligned}$$

and from the section 3.4 we know that we can express H as

$$H = \frac{1}{2}(P_X^2, P_Y^2).$$

Here, P_X and P_Y are the momentum functions (3.11) of X and Y:

$$\begin{aligned} P_X &= cos\theta\, p_x + sen\theta\, p_y \\ P_Y &= p_\theta, \end{aligned}$$

where $p_x = P_{\partial_x}$, $p_y = P_{\partial_y}$, $p_\theta = P_{\partial_\theta}$, coincide with the local coordinates for $T_q^*\mathcal{E}$, defined above. Hence, in local coordinates, we can write

$$H = \frac{1}{2}[(cos\theta\, p_x + sin\theta\, p_y)^2 + p_\theta^2].$$

From section 3.3, we know that to each Hamiltonian functional we can associate an Hamiltonian vector field, defined by the Hamilton's equations 3.1. Furthermore, we know that $\dot{f} = \{f, H\}$, for any smooth function f on the cotangent bundle. If we define the auxiliary functions

$$\begin{aligned} p_1 &= P_X = cos\theta\, p_x + sen\theta\, p_y \\ p_2 &= P_Y = p_\theta \\ p_3 &= P_Z = -sin\theta\, p_x + cos\theta\, p_y. \end{aligned}$$

We can obtain equations 3.1 letting f vary over the coordinate functions (x, y, θ), and the auxiliary functions on the cotangent bundle (p_1, p_2, p_3). Before going on with our derivation of the Hamilton's equations, let us note that

$$\{P_X, P_Y\} \;=\; P_Z \;=\; -P_{[X,Y]}.$$

We can now easily calculate the Hamilton's equations for the position coordinates (x, y, θ)

$$\dot{q}^i = \{q^i, H\} \longrightarrow \begin{cases} \dot{x} = cos\theta\, p_1 \\ \dot{y} = sin\theta\, p_1 \\ \dot{\theta} = p_2 \end{cases}$$

and for $(p_1 p_2, p_3)$

$$\dot{p}_i = \{p_i, H\} \longrightarrow \begin{cases} \dot{p_1} = p_3 p_2 \\ \dot{p_2} = -p_3 p_1 \\ \dot{p_3} = -p_1 p_2 \end{cases}$$

As we know that the Hamiltonian functional assumes a constant value along the Hamiltonian flow (i.e, the solutions to Hamilton's equations), we can write

$$E \;=\; p_1^2 + p_2^2,$$

where $E/2$ is a constant Hamiltonian value. Then, we can introduce an auxiliary variable $\gamma(t)$ such that

$$\begin{cases} p_1 &= \sqrt{E}\, sin(\frac{\gamma}{2}) \\ p_2 &= \sqrt{E}\, cos(\frac{\gamma}{2}) \\ p_3 &= \frac{1}{2}\,\dot{\gamma}, \end{cases}$$

and therefore we can express the other variables as

$$\begin{cases} \dot{x} &= \sqrt{E}\, sin(\frac{\gamma}{2})\, cos\theta \\ \dot{y} &= \sqrt{E}\, sin(\frac{\gamma}{2})\, sin\theta \\ \dot{\theta} &= \sqrt{E}\, cos(\frac{\gamma}{2}). \end{cases}$$

The variable γ satisfies a pendulum-like differential equation

$$\ddot{\gamma} \;+\; E\, sin\gamma \;=\; 0,$$

which is not analytically solvable. Hence, in order to give a graphical representation of the geodesic curves, we need to resort to numeric integration. The solutions

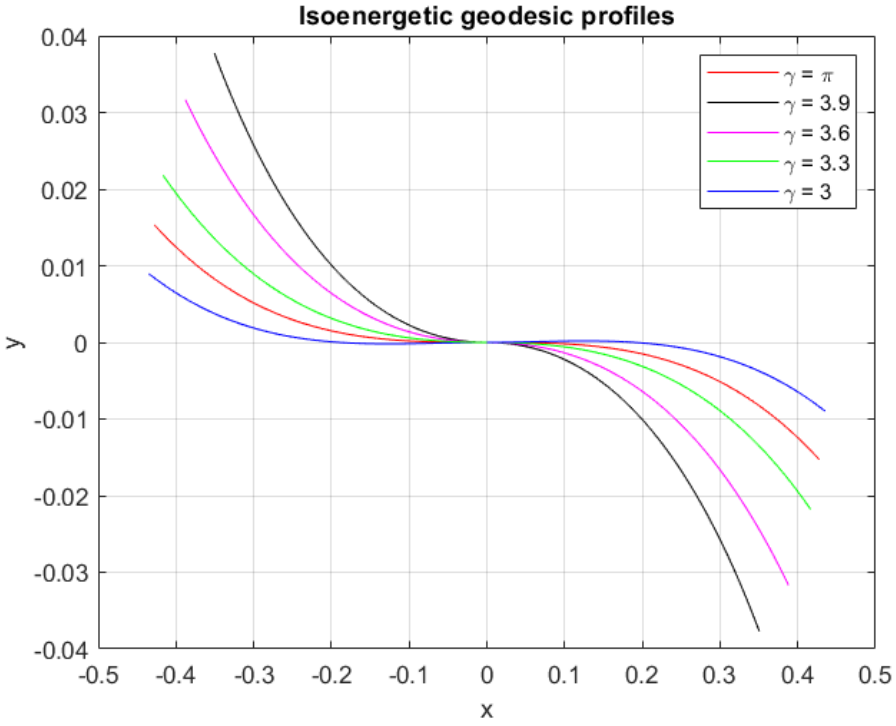


Figure 4.1: *Solutions of the geodesic equations for some values of γ, projected onto the domain D. The energy is fixed to $E = 0.2$.*

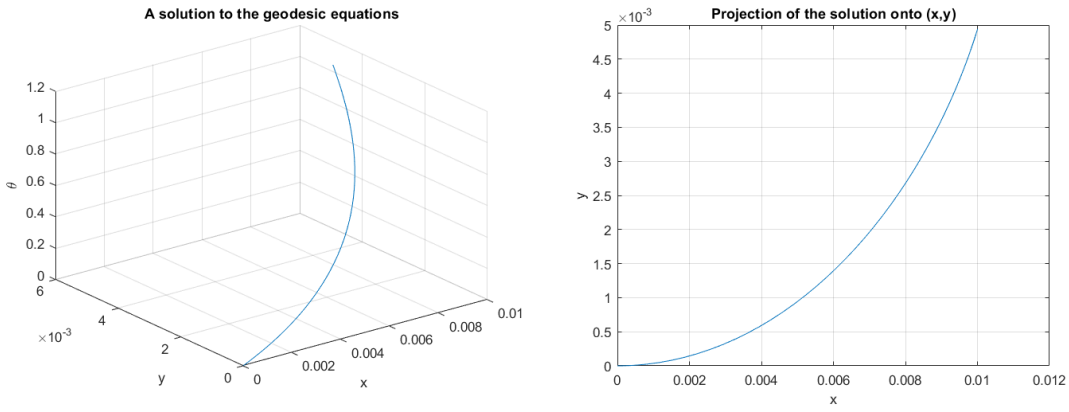


Figure 4.2: *A solution to the geodesic equations that joins $(0, 0, 0)$ and $(0.01, 0.005, \frac{\pi}{3})$ (on the right), and its projection onto the domain D (on the left).*

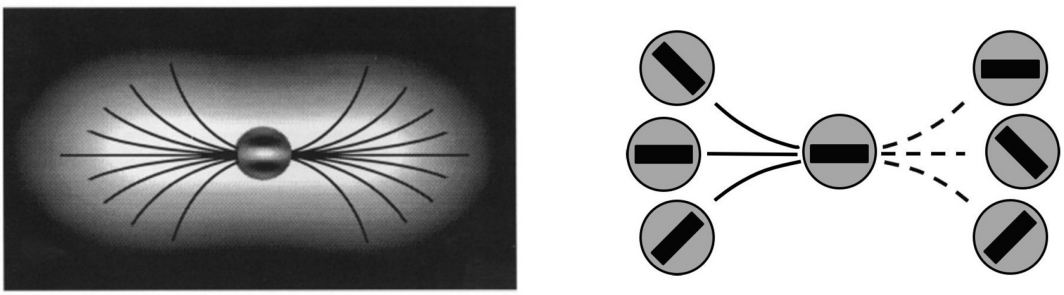


Figure 4.3: *The* association field *(on the left): the rays extending from the ends of the central oriented element represent the optimal orientations at different positions. On the right, the specific rules of alignment are represented.*
- [24]

of this differential equations are the the lifts in $\mathcal{E}$ of perceived borders. In fig. 4.1, we show the projections on D of some solutions with the energy fixed to a value $E = 0.2$, obtained varying the initial value of the parameter γ. A solution to the geodesic equations that joins the points $(0, 0, 0)$ and $(0.01, 0.005, \frac{\pi}{3})$ is shown in fig.4.2, mimicking the border completion mechanism in the brain, given two boundary inducers, i.e, given initial and final coordinates (x, y, θ).

To conclude, we note the similarity of the geodesic curves depicted in fig. 4.1 with the local *association fields* from [24] shown in fig. 4.3. Fields, Heyes and Hess investigate, through a set of experiments, how the relative alignment of neighboring oriented elements is related to the perception of continuity in the human brain. The information detected by single orientation selective cells is supposed to propagate locally through some *long-range connections* in an orientation and position-specific modality ([25], [26]). According to our model, the natural interpretation of the local association field is the representation of the projection onto D of a family of integral curves of the Hamiltonian vector field, i.e, the solutions to the geodesic equations corresponding to the joint constraints of position and orientation.

Ringraziamenti

Voglio ringraziare qui la prof. Fioresi, che è stata importantissima nel mio percorso sia come insegnante, dai primi mesi di università, sia come relatrice e persona, nel coinvolgermi e guidarmi in questo lavoro. Con l'algebra lineare ho avuto esperienza per la prima volta di quel modo nuovo, elegante compatto e potente, di pensare la fisica e gli spazi a cui ho dedicato il tempo più importante del mio studio di questi anni, non ultimo quello per scrivere questa tesi. Nella geometria differenziale ho ritrovato, anzi compreso, cosa della fisica mi aveva attratto e continua a dare un senso alle mie scelte. Studiare per scrivere queste pagine ha soddisfatto da un lato il mio bisogno di approfondire questo tipo di matematica, dall'altro quello di indagare i problemi che un tentativo di modellizzare aspetti del reale necessariamente pone. Ringrazio ancora la prof. per avermi aiutato a vederne la bellezza.

Sono stati, questi, tre anni caratterizzati da una certa forse fisiologica solitudine, e allo stesso tempo da una fondamentale riscoperta dell'essere in relazione. Voglio quindi dire grazie a tutte le persone, le cose, i luoghi, le occasioni che sono parte di questa riscoperta e che non riesco in queste righe a nominare.
Un grazie speciale ai miei genitori, per il loro diverso e complementare sostenermi, il loro amore e le loro contraddizioni, che ho fatto mie. Ai miei nonni, per la loro originale, eccentrica ma preziosa visione del mondo; alla mia famiglia tutta per l'ironia e l'insegnarmi ogni volta lo stare insieme.
A Lorenzo, al suo comunicarsi e comunicarmi per immagini, alla sua creatività e dolcezza, al suo amore per la scienza. Al suo essere inafferrabile che mi ha insegnato il valore di ciò che è incomunicabile.
A Maria Chiara, la sua intensità e il suo coinvolgente amore per le cose belle, al nostro sentire; a Elettra, con cui ho riscoperto l'ascolto e la condivisione e il volersi bene.

Ringrazio infine quelli che in questo percorso hanno condiviso con me momenti intensi o anche piccoli, conflittuali, più e meno divertenti, sulle sedie più e meno comode di Irnerio 46 o nei suoi sotterranei, o dove è capitato. Vi ammiro molto e vi porto con me. (E anche Teresa, che mi ha convinta a iscrivermi a fisica, purtroppo o per fortuna).

Bibliography

[1] D. Mumford. «Elastica and computer vision». In: *Algebraic geometry and its applications* (1994), pp. 491–506.

[2] W.C.Hoffmann. «The visual cortex is a contact bundle». In: *Applied mathematics and computation* 32 (1989), pp. 137–167.

[3] Y. Tondut J. Petitot. «Vers une neurogeometrie. Fibrations corticales structures de contact et countours subjectifs modaux». In: *Mathmatiques, Informatique et Sciences humaines* 145 (1999). DOI: 10.1007/s10851-005-3630-2.

[4] A. Sarti G. Citti. «A cortical based model of perceptual completion in the roto-translation space». In: *J Math Imaging* 24 (2006), pp. 307–326. DOI: 10.1007/s10851-005-3630-2.

[5] Kandel et al. *Principles of neural science*. McGraw Hill, 2012.

[6] Loring Tu. «Introduction to Manifolds». In: (2011).

[7] Loring Tu. *Differential Geometry: Connections, Curvature, and Characteristic Classes*. Springer Nature, 2017.

[8] Montgomery R. *A tour of subriemannian geometries, their geodesics, and applications*. AMS Chelsea publishing. American Mathematical Society, 2002.

[9] Peter Petersen. *Riemannian Geometry*. Springer Nature, 2018.

[10] Miquel Perello Nieto. «A simplified schema of the human visual pathway». In: *https://en.wikipedia.org/wiki/Visual_system* ().

[11] S W Kuffler. «Discharge patterns and functional organization of mammalian retina.» In: *J. Neurophysiol* 16(1) (1953), pp. 37–68.

[12] D. M. Dacey and B. B. Lee. «The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type.» In: *Nature* 367(6465) (1994), pp. 731–735.

[13] Simon E. Skalicky. *Ocular and Visual Physiology*. SpringerLink, 2016. ISBN: 978-981-287-846-5.

[14] Daniel L. Adams and Jonathan C. Horton. «A Precise Retinotopic Map of Primate Striate Cortex Generated from the Representation of Angioscotomas». In: *Journal of Neuroscience* (2003).

[15] D. H. Hubel and T. N. Wiesel. «Receptive Fields of Single Neurones in the Cat's Striate Cortex». In: *J. Physiol* 148 (1959), pp. 574–591.

[16] D. H. Hubel and T. N. Wiesel. «Receptive fields, binocular interaction and functional architecture in the cat's visual cortex». In: *J. Physiol* 160 (1962), pp. 106–154.

[17] D. H. Hubel and T. N. Wiesel. «Receptive fields and functional architecture of monkey striate cortex». In: *J. Physiol* 195 (1968), pp. 215–243.

[18] Czes Kosniowski. *A First Course in Algebraic Topology*. Cambridge University Press, 1980. DOI: 10.1017/CBO9780511569296.

[19] Jocelyn Quaintance Jean Gallier. *Differential Geometry and Lie Groups*. Springer Nature, 2020.

[20] V.I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, 1978.

[21] Serge Carrol Sean M. *Lecture Notes on General Relativity*. arXiv:gr-qc/9712019, 1997.

[22] J. Petkovic R. Fioresi. «A precortical module for robust CNNs to light variations». In: *https://arxiv.org/abs/2202.07432* ().

[23] Kanisza G. *Organization in vision*. Praeger Pub Text, 1979.

[24] David J. Field, Anthony Hayes, and Robert F. Hess. «Contour integration by the human visual system: Evidence for a local "association field"». In: *Vision Research* 33.2 (1993), pp. 173–193. ISSN: 0042-6989. DOI: https://doi.org/10.1016/0042-6989(93)90156-Q. URL: https://www.sciencedirect.com/science/article/pii/004269899390156Q.

[25] C D Gilbert et al. «Spatial integration and cortical dynamics». In: *Proceedings of the National Academy of Sciences USA* 93 (1996), pp. 615–622.

[26] Mitesh K.Kapadia et al. «Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys». In: *Neuron* 15 (1995), pp. 843–856.