Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE
Corso di Laurea Magistrale in Matematica

A Birkhoff's Theorem on closed geodesics

Tesi di Laurea in Analisi Matematica

Relatore:
Chiar.mo Prof.
Vittorio Martino

Presentata da:
Federico Sirocchi

Appello V
Anno Accademico 2023/2024

A mia nonna Ivana

Introduction

This master thesis deals with a geometric problem through the methods of Variational Analysis: the central topic of it is the **Birkhoff's Theorem**, an existence result for geodesics. The statement of the theorem is the following:

"On any compact surface S in $\mathbb{R}^3$ which is C^3-diffeomorphic to the unit sphere $\mathbb{S}^2$ there exists a non-constant closed geodesic."

Geodesics are largely studied in the calculus of variation, since the concept itself of geodesic curve is closely linked to a variational problem. Particularly, Birkhoff's Theorem is an application to geometry of the variational non-direct methods, useful to show the existence of stationary points, given a functional satisfying certain hypotheses. In detail, the method used to show the existence of the critical point in the proof of the theorem is the so called *Minimax principle*, whose easiest version of the statement is the following:

"*Let f be a C^1 real-valued function defined on a Banach space $\mathbb{B}$. Let $\mathcal{A}$ be a family of subsets of $\mathbb{B}$ that is invariant with respect to any semi-flow $\eta : [0, +\infty) \times \mathbb{B} \longrightarrow \mathbb{B}$ such that:*

1. *$\eta(0, x) = x, \forall x \in \mathbb{B}$,*

2. *$f(\eta(t, x))$ is non-increasing in t, $\forall x \in \mathbb{B}$.*

Put

$$c := \inf_{A \in \mathcal{A}} \sup_{x \in A} f(x).$$

If $c \in \mathbb{R}$ and f satisfies the Palais-Smale condition at level c, then there exists $x_0 \in \mathbb{B}$ such that $f(x_0) = c$ and $\mathrm{d}_{x_0} f = 0$ (i.e. c is a critical value for f)".

Thus, some topics we need from analysis are the differentiable calculus on an arbitrary Banach space (in particular the Fréchet and Gâteaux differentials) and the important notion of Palais-Smale condition for a function f: we say that a function $f \in C^1(\mathbb{B}, \mathbb{R})$ satisfies the *Palais-Smale condition* at level c if any sequence $\{x_n\}_{n \in \mathbb{N}} \subset \mathbb{B}$ such that

1. $f(x_n) \overset{n\to\infty}{\longrightarrow} c$,
2. $\|\mathrm{d}_{x_n} f\|_{\mathbb{B}^*} \overset{n\to\infty}{\longrightarrow} 0$,

contains a strongly convergent subsequence. This is a compactness condition that yields us to the *Deformation lemma*, on which the Minimax scheme is based. We will also see a more general version of the Minimax principle for differentiable manifold modeled on Banach spaces and endowed with a *Finsler* structure.

This latter is the variational part, that allows us to prove the existence of closed geodesics, provided these are stationary points of some functional satisfying all the hypotheses of the Minimax principle.

However, before starting with the variational methods, the first chapter of the thesis is more geometric and it is dedicated to the introduction of geodesics on Riemannian manifolds. Indeed, in order to have a proper and general definition of geodesics, we need to work on a differentiable manifold equipped with a Riemannian metric. Informally, when one thinks at geodesics, generally one imagines a curve that minimizes the distance between two nearby points (for example a geodesic in the euclidean plane is a straight line, while a geodesic on a sphere is an arc of great circle). Here geodesics are defined as curves with zero *acceleration* (then one can show that such curves are locally minimum distance paths).

An important notion we need, towards geodesics, is the one of *affine connection*, which is possible to give on a generic differentiable manifold. Then, provided we have an affine connection ∇ on a differentiable manifold M and a curve $\gamma : I \longrightarrow M$, we can define a unique correspondence that associates to each vector field V along γ another vector field $\nabla_{\dot\gamma} V$ along γ satisfying certain properties in some way similar to those of the derivative operator defined on functions of one real variable. The vector field $\nabla_{\dot\gamma} V$ is known as the *covariant derivative* of V. Working on a Riemannian manifold, it turns out that there exists a "special" connection. Indeed, if we have a Riemannian manifold M with metric $\langle\, , \rangle$, we can require that a connection is *compatible* with the metric $\langle\, , \rangle$ and that it is symmetric; the *Levi-Civita Theorem*, which is the fundamental theorem of

Riemannian geometry, tells us: *"If M is a Riemannian manifold, there exists a unique affine connection ∇, called the Levi-Civita connection, which is either compatible with the metric, either symmetric."*

Now, considering a Riemannian manifold M with its Levi-Civita connection ∇, we can give a proper definition of geodesic: we say that a smooth curve $\gamma : I \longrightarrow M$ is a *geodesic* on M if $\nabla_{\dot{\gamma}}\dot{\gamma}(t) = 0$ for all $t \in I$. From this we can see some consequences: it turns out that geodesics have some important minimizing properties. Indeed, given a piecewise smooth curve $\gamma : [a, b] \longrightarrow M$ on a Riemannian manifold, we can consider the *lenght functional* defined as:

$$\ell(\gamma) = \int_a^b \|\dot{\gamma}(t)\|_{\gamma(t)}\, \mathrm{d}t = \int_a^b \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}}\, \mathrm{d}t.$$

Then a geodesic passing through two points $P, Q \in M$ is locally a solution of the minimizing problem:

$$\min\{\ell(\gamma) \,|\, \gamma : [a, b] \longrightarrow M,\ \gamma(a) = P,\ \gamma(b) = Q\}.$$

We can see that, if $t : [c, d] \longrightarrow [a, b]$ is a C^1-diffeomorphism and $\widetilde{\gamma}(\tau) := \gamma(t(\tau))$, then $\ell(\widetilde{\gamma}) = \ell(\gamma)$, meaning that ℓ is invariant under reparametrization. Another functional we will introduce is the *energy functional*: given a curve γ as before, it is defined by the following integral:

$$E(\gamma) = \frac{1}{2} \int_a^b \|\dot{\gamma}(t)\|^2_{\gamma(t)}\, \mathrm{d}t.$$

This second functional is not invariant under diffeomorphism, unlike the previous one; however, it turns out that, if we choose a parametrization for which the velocity of the curve has constant module, then we obtain the relation $\ell(\gamma) = \sqrt{2(b-a)E(\gamma)}$. Hence, among these curves parameterized with constant velocity module, looking for a minimizer of ℓ is equivalent to looking for a minimizer of E. Since the minimizer for E is unique (due to the strict convexity of the functional), then it follows that the minimizer for ℓ is unique (up to reparametrization of the curve). It is more convenient to work with E, instead of ℓ, because E has certain properties of regularity that facilitate the proof of the existence of critical points.

Finally, in the last chapter of this thesis, we will see that for a surface S the space of closed curves on S with finite energy has a structure of Finsler manifold and we will prove the Birkhoff's Theorem; we can see here the most important steps in which the proof is organized:

1. Definition of the energy on the space of closed curves $u : \mathbb{S}^1 \longrightarrow S$:

$$E(u) = \frac{1}{2} \int_0^{2\pi} \|\dot{u}(t)\|^2_{u(t)} \, \mathrm{d}t,$$

   and definition of the space of closed curves with finite energy:

$$H^{1,2}(\mathbb{S}^1; S) := \{u : \mathbb{R} \longrightarrow \mathbb{R}^3 : \; u|_{(0,2\pi)} \in H^{1,2}((0, 2\pi); \mathbb{R}^3), \\ u(t) = u(t + 2\pi), u(t) \in S \; a.e.t \in \mathbb{R}\}.$$

2. Reduction of the problem to searching for critical points of E.

3. Proving that the functional E satisfies the Palais-smale condition.

4. Construction of a flow-invariant family $\mathcal{F}$ as in the statement of the Minimax principle.

5. Proving that the minimax value for E, defined by

$$\beta = \inf_{p \in \mathcal{F}} \sup_{u \in p} E(u),$$

   is strictly positive (hence the closed geodesic corresponding to it is not constant).

This result by Birkhoff dates back to 1917 and, with a later extension to sphere of arbitrary dimensions, mark the beginning of the calculus of variations in the large. Later, in 1929 Lusternik-Schnirelmann will prove the existence of three geometrically distinct closed geodesics free of self-intersections on the 2-dimensional sphere, with an arbitrary Riemannian metric.

Contents

List of Figures

Chapter 1

Geodesics on Riemannian manifolds

In this chapter, we will introduce the notion of *geodesic curves*. In order to do that, we need to work with differentiable manifolds equipped with Riemannian metrics. We will start with some basic definitions and propositions of Riemannian geometry, including the notion of affine connection and the Levi-Civita Theorem, which is known as the *Fundamental Theorem of Riemannian geometry* and immediately precedes the definition of geodesic. Then we will define the *exponential map* (on a subset of the tangent bundle of a differentiable manifold) and see important minimizing properties of geodesics. Finally, besides the length functional, we will introduce the energy functional E and give a characterization of geodesics curves as critical points of this functional.

1.1 Riemannian manifolds and curves

Definition 1.1. Given a differentiable manifold M, a **Riemannian metric** (or *Riemannian structure*) on M is a correspondence which associates to each point $p \in M$ an inner product $\langle\, ,\, \rangle_p$ (symmetric, bilinear, positive-definite form) on the tangent space of M at p T_pM such that $p \longmapsto \langle\, ,\, \rangle_p$ is differentiable in the following sense: if $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$ is a system of coordinates around p, with $\mathrm{x}(x^1, \ldots, x^n) = q \in \mathrm{x}(U)$ and $\frac{\partial}{\partial x^i}(q) := d\mathrm{x}_q(0, \ldots, \underbrace{1}_{i-th}, \ldots, 0)$, then $\langle \frac{\partial}{\partial x^i}(q), \frac{\partial}{\partial x^j}(q) \rangle_q = g_{ij}(x^1, \ldots, x^n)$ is a differentiable function on U. The function $g_{ij}(= g_{ji})$ is called the *local representation of the Riemannian metric in the coordinate system* $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$.

Remark 1.2. The previous definition does not depend on the choice of the coordinate system.

Notation 1.3. When we work with a system of coordinates $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$, we denote by $(x^1, \ldots, x^n)$ the local coordinates for $\mathrm{x}(U)$ (variables on U) and we define $\frac{\partial}{\partial x^i}(q) := \mathrm{dx}_q(0, \ldots, \underbrace{1}_{i-th}, \ldots, 0)$ for $q \in \mathrm{x}(U)$ as in Definition 1.1.

Definition 1.4. A **Riemannian manifold** is a differentiable manifold with a Riemannian metric defined on it.

Example 1.5. The most trivial example of a Riemannian manifold is $M = \mathbb{R}^n$ with $\langle\, , \rangle_p$ standard scalar product for all $p \in \mathbb{R}^n$.

Notation 1.6. If M is a Riemannian manifold, $p \in M$ and $v \in T_pM$, we denote the norm of v by $\|v\|_p := \sqrt{\langle v, v\rangle_p}$.

Definition 1.7. If M and N are two Riemannian manifolds and $f : M \longrightarrow N$ is a diffeomorphism, then we say that f is an *isometry* if the following condition is satisfied:

$$\langle u, v\rangle_p = \langle \mathrm{d}f_p(u), \mathrm{d}f_p(v)\rangle_{f(p)}, \quad \forall p \in M, \forall u, v \in T_pM,$$

where $\mathrm{d}f_p : T_pM \longrightarrow T_{f(p)}N$ is the *differential map* of f at p. If $f : M \longrightarrow N$ is a differentiable map, we say that it is a *local isometry* at $p \in M$ if there exists an open neighborhood $U \subset M$ of p such that $f : U \longrightarrow f(U)$ is an isometry.

Now, we are interested in curves: we will focus on vector fields along curves, particularly the *velocity field* of a curve.

Definition 1.8. A *vector field* X on a differentiable manifold M is a correspondence $X : M \longrightarrow TM$ such that $X(p) \in T_pM$ for any $p \in M$. The field is said to be differentiable if the corresponding mapping is differentiable. If $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$ is a system of coordinates around $p \in M$, then we can write

$$X(p) = \sum_{i=1}^{n} a^i(\mathrm{x}^{-1}(p)) \frac{\partial}{\partial x^i}(p), \tag{1.1}$$

where $a^i : U \longrightarrow \mathbb{R}$.

Remark 1.9. Another way to express the differentiability of the Riemannian metric is to require that $V \ni p \longmapsto \langle X(p), Y(p) \rangle_p$ is a differentiable mapping for any open neighborhood V of M and for any couple of vector fields X and Y that are differentiable on V.

Remark 1.10. If $\mathcal{D}$ is the set of all smooth functions on M and $\mathcal{F}$ is the set of all functions on M, by (1.1), we can think at a vector field X as a mapping $X : \mathcal{D} \longrightarrow \mathcal{F}$ given by

$$Xf(p) := \sum_{i=1}^{n} a^i(\mathrm{x}^{-1}(p)) \frac{\partial (f \circ \mathrm{x})}{\partial x^i}(\mathrm{x}^{-1}(p)), \quad \forall f \in \mathcal{D}.$$

Hence, we can interpret vector fields as differential operators.

Lemma 1.11. *Let X, Y be differentiable vector fields on a differentiable manifold M. Then there exists a unique vector field Z on M such that $Zf = (XY - YX)f$ for any $f \in \mathcal{D}$. Z is usually denoted by $[X, Y]$ and it is called the commutator (or bracket) of X and Y.*

For a proof of the previous lemma see Lemma 5.2 on Chapter 0 of [2].

Definition 1.12. Let M be a differentiable manifold; a *vector field V* along a curve $c : I \longrightarrow M$ is a smooth mapping $V : I \longrightarrow TM$ such that $V(t) \in T_{c(t)}M$ for any $t \in I$. If $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$ is a system of coordinates around $c(t)$, then we can write

$$V(t) = \sum_{i=1}^{n} a^i \left(\mathrm{x}^{-1}(c(t)) \right) \frac{\partial}{\partial x^i}(c(t)),$$

where $a^i : U \longrightarrow \mathbb{R}$ are smooth functions. As before, we can interpret V as a differential operator:

$$Vf(c(t)) := \sum_{i=1}^{n} a^i \left(\mathrm{x}^{-1}(c(t)) \right) \frac{\partial (f \circ \mathrm{x})}{\partial x^i} \left(\mathrm{x}^{-1}(c(t)) \right),$$

where f is a smooth function on M.

Remark 1.13. If V is a vector field along a curve $c : I \longrightarrow M$, then the map $t \longmapsto Vf(c(t))$ is smooth on I for any $f \in \mathcal{D}$.

The following one is the vector field we will work with from now on.

Definition 1.14. Let $c : I \longrightarrow M$ be a smooth curve in a differentiable manifold. For $t_0 \in I$ we can consider the differential map $\mathrm{d}c_{t_0} : T_{t_0}I \longrightarrow T_{c(t_0)}M$. The *velocity field of* c at t_0 is defined by $\dot{c}(t_0) := \mathrm{d}c_{t_0}(\frac{\mathrm{d}}{\mathrm{d}t}(t_0))$ and it acts on any smooth function f on M in the following way:

$$\dot{c}f(c(t)) = \frac{\mathrm{d}(f \circ c)}{\mathrm{d}t}(t).$$

Sometimes we may use the notation $\frac{\mathrm{d}c}{\mathrm{d}t}(t)$ for the velocity field of c (it is a more practical notation especially when the curve c is given by the composition of a curve and a mapping).

Remark 1.15. If X is a vector field on a differentiable manifold M, $f : M \longrightarrow \mathbb{R}$ is a smooth function and $p \in M$, then

$$Xf(p) = \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma(t))\bigg|_{t=0},$$

where $\gamma : (-\varepsilon, \varepsilon) \longrightarrow M$ is an arbitrary smooth curve such that $\gamma(0) = p$ and $\dot{\gamma}(0) = X(p)$.

The advantage of having a Riemannian structure is that we can measure the “length” of a segment of a curve.

Definition 1.16. A *segment* is the restriction of a curve $c : I \longrightarrow M$ to a closed interval $[a, b] \subset I$. If M has a Riemannian structure and c is smooth, then we can define the *length* of the segment by:

$$\ell_a^b(c) := \int_a^b \|\dot{c}(t)\|_{c(t)}\, \mathrm{d}t = \int_a^b \langle \dot{c}(t), \dot{c}(t)\rangle_{c(t)}^{1/2}\, \mathrm{d}t.$$

Sometimes we will omit the two endpoints of the interval of definition of the segment (or curve), writing just $\ell(c)$ in place of $\ell_a^b(c)$.

Remark 1.17. Let M be a Riemannian manifold and $\gamma : [a, b] \longrightarrow M$ a segment of a smooth curve. Consider a reparametrization of γ given by $\tilde{\gamma} : [c, d] \longrightarrow M$ (i.e $\tilde{\gamma} = \gamma \circ \varphi$, where $\varphi : [c, d] \longrightarrow [a, b]$ is a C^1-diffeomorphism). Then $\ell_a^b(\gamma) = \ell_c^d(\tilde{\gamma})$.

Definition 1.18. $c : [a, b] \longrightarrow M$ is an *admissible curve* if it is a continuous mapping and there exists a partition $\{a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b\}$ of $[a, b]$ such that $c|_{[t_{k-1}, t_k]}$ is smooth for any $k = 1, \ldots, n$.

If M has a Riemannian structure, then we can define the length of c:

$$\ell_a^b(c) := \sum_{k=1}^{n} \ell_{t_{k-1}}^{t_k}(c|_{[t_{k-1},t_k]}) = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \langle \dot{c}(t), \dot{c}(t) \rangle_{c(t)}^{1/2} \, \mathrm{d}t.$$

We say that c joins the points $c(a)$ and $c(b)$.

Definition 1.19. We say that $c : [a,b] \longrightarrow M$ is a *unit speed* admissible curve if it is an admissible curve and $\|\dot{c}(t)\|_{c(t)} = 1$ for all t where $\dot{c}(t)$ exists.

Remark 1.20. Any admissible curve admits a unit speed parametrization.

We also have "piecewise smooth" vector fields along curves.

Definition 1.21. Let M be a differentiable manifold, $c : [a,b] \longrightarrow M$ an admissible curve and $V : [a,b] \longrightarrow TM$ a continuous mapping such that $V(t) \in T_{c(t)}M$ for any $t \in [a,b]$. We say that V is a *piecewise smooth vector field* along c if there exists a finite subdivision $\{a = \tilde{a}_0 < \ldots < \tilde{a}_m = b\}$ such that $V|_{[\tilde{a}_{k-1},\tilde{a}_k]}$ is smooth for any $k = 1, \ldots, m$.

Remark 1.22. If M is a connected differentiable manifold, then for any couple of points p and q there exists at least one admissible curve joining p and q.

Definition 1.23. If M is a connected Riemannian manifold and p, q is a couple of points in M, then we can define the *distance between p and q* by

$$d(p,q) := \inf\{\ell(c) : \ c \text{ is an admissible curve joining } p \text{ and } q\}.$$

By the previous remark, $d(p,q) \in \mathbb{R}$ for any couple of points p and q.

Proposition 1.24. *The quantity d introduced in the previous definition is actually a distance and (M,d) is a metric space. Moreover, any Riemannian metric on M defines a distance that induces on M the same topology as the manifold topology.*

For a proof of Proposition 1.24 see Lemma 6.2 on Chapter 6 of [6].

1.2 Affine and Riemannian connections

We denote by $\mathcal{X}(M)$ the set of all smooth vector fields on M and by $\mathcal{D}(M)$ the ring of all smooth real-valued functions defined on M.

Definition 1.25. An **affine connection** ∇ on a differentiable manifold M is a map

$$\begin{aligned}\nabla : \mathcal{X}(M) \times \mathcal{X}(M) &\longrightarrow \mathcal{X}(M)\\ (X, Y) &\longmapsto \nabla_X Y\end{aligned}$$

such that:

1. $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$,
2. $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$,
3. $\nabla_X (fY) = f\nabla_X Y + X(f)Y$,

for all $X, Y, Z \in \mathcal{X}(M)$ and $f, g \in \mathcal{D}(M)$.

Proposition 1.26. *Let M be a differentiable manifold with an affine connection ∇. Given a smooth curve $c : I \longrightarrow M$, there exists a unique correspondence that associates to each vector field V along c another vector field $D_t V$ along c such that:*

1. *$D_t(V + W) = D_t V + D_t W$ for any vector fields V, W along c;*
2. *$D_t(fV) = \frac{\mathrm{d}f}{\mathrm{d}t} V + f D_t V$ for any vector field V along c and for any smooth function f on I;*
3. *if V is induced by a vector field $Y \in \mathcal{X}(M)$ (i.e. $V(t) = Y(c(t))$), then*

$$D_t V = \nabla_{\dot{c}} Y,$$

where $\nabla_{\dot{c}} Y(t) := \nabla_X Y(c(t))$, with $X \in \mathcal{X}(M)$ such that $X(c(t)) = \dot{c}(t)$.

Remark 1.27. We need to check that, if $X, \tilde{X}, Y \in \mathcal{X}(M)$, $p \in M$ and $X(p) = \tilde{X}(p)$, then $\nabla_X Y(p) = \nabla_{\tilde{X}} Y(p)$ (from this the term $\nabla_{\dot{c}} Y$ in *3* is meaningful).

Indeed, let $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$ be a system of coordinates about p and set $X_i := \frac{\partial}{\partial x^i}$. Then we can write

$$X = \sum_i (a^i \circ \mathrm{x}^{-1}) X_i, \quad \tilde{X} = \sum_i (\tilde{a}^i \circ \mathrm{x}^{-1}) X_i, \quad Y = \sum_j (b^j \circ \mathrm{x}^{-1}) X_j,$$

where $a^i, \tilde{a}^i, b^j : U \longrightarrow \mathbb{R}$ are smooth functions. We have

$$\nabla_X Y = \sum_i (a^i \circ \mathrm{x}^{-1}) \nabla_{X_i} \left(\sum_j (b^j \circ \mathrm{x}^{-1}) X_j \right) =$$

$$= \sum_{i,j} (a^i \circ \mathrm{x}^{-1})(b^j \circ \mathrm{x}^{-1}) \nabla_{X_i} X_j + \sum_{i,j} (a^i \circ \mathrm{x}^{-1}) X_i (b^j \circ \mathrm{x}^{-1}) X_j$$

Setting $\nabla_{X_i} X_j = \sum_k \Gamma_{ij}^k X_k$, then Γ_{ij}^k are smooth function on U and

$$\nabla_X Y = \sum_k \left(\sum_{i,j} (a^i \circ \mathrm{x}^{-1})(b^j \circ \mathrm{x}^{-1}) \Gamma_{ij}^k + X(b^k \circ \mathrm{x}^{-1}) \right) X_k.$$

Similarly

$$\nabla_{\tilde{X}} Y = \sum_k \left(\sum_{i,j} (\tilde{a}^i \circ \mathrm{x}^{-1})(b^j \circ \mathrm{x}^{-1}) \Gamma_{ij}^k + \tilde{X}(b^k \circ \mathrm{x}^{-1}) \right) X_k.$$

Since $X(p) = \tilde{X}(p)$, then $(a^i \circ \mathrm{x}^{-1})(p) = (\tilde{a}^i \circ \mathrm{x}^{-1})(p)$ and $X(b^k \circ \mathrm{x}^{-1})(p) = \tilde{X}(b^k \circ \mathrm{x}^{-1})(p)$.[1] Thus, $\nabla_X Y(p) = \nabla_{\tilde{X}} Y(p)$.

Proof. Let us assume there exists a correspondence that associates to each vector field V along a curve $c : I \longrightarrow M$ another vector field $D_t V$ along the same curve satisfying *1*, *2* and *3*. Let $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$ be a parametrization such that $c(I) \cap \mathrm{x}(U) \neq \emptyset$, let $(c^1(t), \ldots, c^n(t))$ be the local expression of $c(t)$, $t \in I$, and set $X_i := \frac{\partial}{\partial x^i}$. Hence, we can express locally

$$V = \sum_j v^j X_j, \quad , j = 1, \ldots, n,$$

where $v^j = v^j(t)$ and $X_j = X_j(c(t))$. By *1* and *2*, we have

$$D_t V(t) = \sum_j \frac{\mathrm{d}v^j(t)}{\mathrm{d}t} X_j(c(t)) + \sum_j v^j(t) D_t(X_j \circ c)(t).$$

Let $X \in \mathcal{X}(M)$, locally $X = \sum_i (a^i \circ \mathrm{x}^{-1}) X_i$, such that $X(c(t)) = \dot{c}(t)$; by *3*, we have

$$D_t(X_j \circ c)(t) = \nabla_{\dot{c}} X_j(t) = \nabla_X X_j(c(t)) = \sum_i \frac{\mathrm{d}c^i(t)}{\mathrm{d}t} \nabla_{X_i} X_j(c(t)), \quad j = 1, \ldots, n.$$

[1] $X(b^k \circ \mathrm{x}^{-1})(p) = \sum_i (a^i \circ \mathrm{x}^{-1})(p) \frac{\partial b^k}{\partial x^i}(\mathrm{x}^{-1}(p)) = \sum_i (\tilde{a}^i \circ \mathrm{x}^{-1})(p) \frac{\partial b^k}{\partial x^i}(\mathrm{x}^{-1}(p)) = \tilde{X}(b^k \circ \mathrm{x}^{-1})(p)$.

All together,

$$D_t V = \sum_j \frac{\mathrm{d}v^j}{\mathrm{d}t} X_j + \sum_{i,j} \frac{\mathrm{d}c^i}{\mathrm{d}t} v^j \nabla_{X_i} X_j, \tag{1.2}$$

therefore we have the uniqueness.

To show the existence, we define $D_t V$ in $\mathrm{x}(U)$ as in (1.2); then it is quite obvious it satisfies the desired properties. If $\mathrm{y}(W)$ is another coordinate neighbourhood such that $\mathrm{x}(U) \cap \mathrm{y}(W) \neq \emptyset$ and we define $D_t V$ in $\mathrm{y}(W)$ as for $\mathrm{x}(U)$, then the two definitions have to agree in $\mathrm{x}(U) \cap \mathrm{y}(W)$ by the uniqueness of $D_t V$ in $\mathrm{x}(U)$; therefore we can extend the definition of $D_t V$ all over M. □

Γ_{ij}^k in the proof of the previous theorem are the so called *Christoffel's symbols*. Since we will use them several times, we fix the following notation.

Notation 1.28. Let M be a differentiable manifold with an affine connection ∇. When we work in a system of coordinates $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$, if $X_i := \frac{\partial}{\partial x^i}$, we denote by Γ_{ij}^k the smooth functions on U such that

$$\nabla_{X_i} X_j = \sum_k \Gamma_{ij}^k X_k.$$

Definition 1.29. If M is a differentiable manifold with an affine connection ∇, c is a smooth curve in M and V is a vector field along c, then the vector field $\nabla_{\dot{c}} V = D_t V$, introduced in the Proposition 1.26, is called the *covariant derivative of V along c.*

We can give a notion of *parallelism*.

Definition 1.30. Let M be a differentiable manifold equipped with an affine connection ∇ and $c : I \longrightarrow M$ a smooth curve in M. If V is a vector field along c, then V is said to be *parallel* if $\nabla_{\dot{c}} V(t) = 0$ for any $t \in I$.

From now on, since we will have several sums, we will use the Einstein summation convention.

Proposition 1.31. *Let M be a differentiable manifold with an affine connection ∇, $c : I \longrightarrow M$ a smooth curve in M, $t_0 \in I$ and $V_0 \in T_{c(t_0)}M$. Then there exists a unique parallel vector field V along c such that $V(t_0) = V_0$; this vector field V is called the parallel transport of $V(t_0)$ along c.*

Proof. We first assume that $c(I)$ is contained in a coordinate neighborhood $\mathrm{x}(U)$, where $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$ is a system of coordinates. Let $(c^1(t), \ldots, c^n(t))$ be the local expression for $c(t)$ and let $V_0 = \sum_j {v_0}^j X_j(c(t_0))$, where $X_j := \frac{\partial}{\partial x^j}$. We assume that there exists a vector field V along c which is parallel and such that $V(t_0) = V_0$. We have that $V = \sum_j v^j X_j$ (where $v^j = v^j(t)$ and $X_j = X_j(c(t))$) satisfies

$$0 = \nabla_{\dot{c}} V = \frac{\mathrm{d}v^j}{\mathrm{d}t} X_j + \frac{\mathrm{d}c^i}{\mathrm{d}t} v^j \nabla_{X_i} X_j = \frac{\mathrm{d}v^j}{\mathrm{d}t} X_j + \frac{\mathrm{d}c^i}{\mathrm{d}t} v^j \Gamma_{ij}^k X_k.$$

Replacing j with k in the first sum, we obtain

$$\nabla_{\dot{c}} V = \sum_k \left(\frac{\mathrm{d}v^k}{\mathrm{d}t} + \frac{\mathrm{d}c^i}{\mathrm{d}t} v^j \Gamma_{ij}^k \right) X_k = 0.$$

If we consider the following system of n differential first-order equations in v^k

$$0 = \frac{\mathrm{d}v^k}{\mathrm{d}t} + \frac{\mathrm{d}c^i}{\mathrm{d}t} v^j \Gamma_{ij}^k, \qquad k = 1, \ldots, n, \tag{1.3}$$

with an initial condition $v^k(t_0) = {v_0}^k$ for any $k = 1, \ldots, n$, then it has a unique solution. Hence, V exists, it is unique and it is defined for all $t \in I$, since the system is linear.

Finally, if $c(I)$ is not contained in a local coordinate neighborhood, by compactness, for any $t_1 \in I$, the segment $c([t_0, t_1]) \subset M$ can be covered by a finite number of coordinate neighborhoods, in each of which V exists and it is unique, as we have seen before. By the uniqueness, the definitions have to coincide when the intersections are not empty and we can define V along all of $[t_0, t_1]$ (therefore we can extend V along all over I). $\square$

Definition 1.32. Let M be a Riemannian manifold with the metric $\langle\, , \rangle$. An affine connection is said to be *compatible* with the metric $\langle\, , \rangle$ if $\langle P(\cdot), P'(\cdot)\rangle_{c(\cdot)}$ is constant for any smooth curve c in M and for any pair of parallel vector fields P and P' along c.

We have the following characterization.

Proposition 1.33. *Let M be a Riemannian manifold. A connection ∇ on M is compatible with the metric if and only if*

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle V(t), W(t)\rangle_{c(t)} = \langle \nabla_{\dot{c}} V(t), W(t)\rangle_{c(t)} + \langle V(t), \nabla_{\dot{c}} W(t)\rangle_{c(t)}, \tag{1.4}$$

for any smooth curve $c : I \longrightarrow M$ and for any vector fields V and W along c.

Proof. It is obvious that if (1.4) holds, then ∇ is compatible with the metric $\langle\, ,\, \rangle$.

Let us prove the converse. Let us fix $t_0 \in I$ and let $\{P_1(t_0), \ldots, P_n(t_0)\}$ be an orthonormal basis for $T_{c(t_0)}M$. By Proposition 1.31, we can extend each element $P_i(t_0)$ to a parallel vector field P_i along c. Since ∇ is compatible with the metric, then $\{P_1(t), \ldots, P_n(t)\}$ is an orthonormal basis for $T_{c(t)}M$ for any $t \in I$. If V and W are vector fields along c, then we can write

$$V = v^i P_i, \qquad W = w^i P_i,$$

where v^i, w^i are smooth functions on I. Since $\nabla_{\dot c} P_i = 0$, we have

$$\nabla_{\dot c} V = \frac{\mathrm{d} v^i}{\mathrm{d}t} P_i, \qquad \nabla_{\dot c} W = \frac{\mathrm{d} w^i}{\mathrm{d}t} P_i.$$

Therefore

$$\langle \nabla_{\dot c} V(t), W(t)\rangle_{c(t)} + \langle V(t), \nabla_{\dot c} W(t)\rangle_{c(t)} = \\ \sum_{i=1}^{n} \left(\frac{\mathrm{d} v^i(t)}{\mathrm{d}t} w^i(t) + v^i(t) \frac{\mathrm{d} w^i(t)}{\mathrm{d}t} \right) = \frac{\mathrm{d}}{\mathrm{d}t} \left(\sum_{i=1}^{n} v^i(t) w^i(t) \right) = \frac{\mathrm{d}}{\mathrm{d}t} \langle V(t), W(t)\rangle_{c(t)} \,,$$

which is exactly (1.4). □

An immediate consequence is the following corollary.

Corollary 1.34. *A connection ∇ on a Riemannian manifold M is compatible with the metric if and only if*

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle, \qquad X, Y, Z \in \mathcal{X}(M), \tag{1.5}$$

meaning that

$$X \left\langle Y(\cdot), Z(\cdot)\right\rangle_{(\cdot)} (p) = \left\langle \nabla_X Y(p), Z(p)\right\rangle_p + \left\langle Y(p), \nabla_X Z(p)\right\rangle_p$$

for all $X, Y, Z \in \mathcal{X}(M)$ and $p \in M$.

Proof. Suppose that ∇ is compatible with the metric. Let $p \in M$ and c be a smooth curve such that $c(0) = p$ and $\dot c(0) = X(p)$. Then

$$X \left\langle Y(\cdot), Z(\cdot)\right\rangle_{(\cdot)} (p) = \frac{\mathrm{d}}{\mathrm{d}t} \left\langle Y(c(t)), Z(c(t))\right\rangle_{c(t)} \bigg|_{t=0} = \\ = \left\langle \nabla_{\dot c}(Y \circ c)(0), Z(p)\right\rangle_p + \left\langle Y(p), \nabla_{\dot c}(Z \circ c)(0)\right\rangle_p = \\ \left\langle \nabla_X Y(p), Z(p)\right\rangle_p + \left\langle Y(p), \nabla_X Z(p)\right\rangle_p .$$

The converse is quite obvious: if $c : I \longrightarrow M$ is a smooth curve and V, W are vector fields along c, then $\dot{c}$ acts on $\langle V(\cdot), W(\cdot)\rangle_{c(\cdot)}$ by $\dot{c}\,\langle V(\cdot), W(\cdot)\rangle_{c(\cdot)}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\langle V(t), W(t)\rangle_{c(t)}$, which is the left hand side of (1.4) and, using the hypothesis, it coincides with the right hand side of (1.4).

□

Definition 1.35. An affine connection ∇ on a differentiable manifold M is said to be *symmetric* if

$$\nabla_X Y - \nabla_Y X = [X, Y], \qquad \forall X, Y \in \mathcal{X}(M). \tag{1.6}$$

Remark 1.36. In a coordinate system $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$, if $X_i := \frac{\partial}{\partial x^i}$, then the fact that ∇ is symmetric implies that

$$\nabla_{X_i} X_j - \nabla_{X_j} X_i = [X_i, X_j] = 0, \text{ for all } i, j = 1, \dots, n, \tag{1.7}$$

which justifies the terminology. We note that (1.7) is equivalent to the fact that $\Gamma_{ij}^k = \Gamma_{ji}^k$ for all $i, j, k = 1, \dots, n$.

Now, we can state and prove the Fundamental Theorem of Riemannian geometry.

Theorem 1.37 (**Levi-Civita**). *Given a Riemannian manifold M, there exists a unique affine connection ∇ on M such that it is either symmetric either compatible with the Riemannian metric.*

Proof. We assume that such ∇ exists and we show that it is the unique connection satisfying both properties. Since ∇ is compatible with the metric, by Corollary 1.34, we have

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle, \tag{1.8}$$

$$Y\langle Z, X\rangle = \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_Y X\rangle, \tag{1.9}$$

$$Z\langle X, Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle, \tag{1.10}$$

for any $X, Y, Z \in \mathcal{X}(M)$.

If we add (1.8) and (1.9) and subract (1.10), using the fact that ∇ is symmetric, we get

$$X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle = \\ = \langle [X, Z], Y\rangle + \langle [Y, Z], X\rangle + \langle [X, Y], Z\rangle + 2\,\langle Z, \nabla_Y X\rangle\,.$$

From this we obtain

$$\langle Z, \nabla_Y X\rangle = \frac{1}{2}\big\{X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle + \\ - \langle [X,Z],Y\rangle - \langle [Y,Z],X\rangle - \langle [X,Y],Z\rangle\big\}. \quad (1.11)$$

The expression (1.11) shows that ∇ is uniquely determined by the metric. Hence, provided we have the existence, we have the uniqueness.

To show the existence, we define ∇ as in (1.11). Then with an easy computation we can see that

$$X\langle Y,Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle, \quad \forall X,Y,Z \in \mathcal{X}(M),$$

whence the connection is compatible with the metric by Corollary 1.34, and

$$\langle \nabla_X Y, Z\rangle - \langle \nabla_Y X, Z\rangle = \langle [X,Y],Z\rangle\,, \quad \forall X,Y,Z \in \mathcal{X}(M),$$

whence the connection is symmetric. □

Definition 1.38. If M is a Riemannian manifold, the connection ∇ whose existence and uniqueness is guaranteed by the Theorem 1.37 is called the *Levi-Civita connection* (or the *Riemannian connection*) of M.

We can see an important example.

Remark 1.39. If S is a regular orientable surface in $\mathbb{R}^3$ and $N : S \longrightarrow \mathbb{R}^3$ is a differentiable unit-normal vector field on S (i.e $N(p) \perp T_P S$, $\|N(p)\| = 1$ for any $P \in S$), then its Levi-Civita connection is given by

$$\nabla_X Y(p) := Y'_{X(p)}(p) - \left\langle Y'_{X(p)}(p), N(p)\right\rangle N(p), \quad \forall X,Y \in \mathcal{X}(S), \forall p \in S,$$

where $Y'_{X(p)}(p) = \frac{\mathrm{d}}{\mathrm{d}t}\left(Y \circ \gamma(t)\right)\big|_{t=0}$, for a curve $\gamma : (-\varepsilon,\varepsilon) \longrightarrow S$ such that $\gamma(0) = p$ and $\dot\gamma(0) = X(p)$ ($Y'_{X(p)}(p)$ does not depend on the choice of the curve γ satisfying these properties).

1.3 Introduction to geodesics

In this section and in the following ones of this chapter, M will be a n-dimensional Riemannian manifold together with its Levi-Civita connection ∇.

Definition 1.40. Let $\gamma : I \longrightarrow M$ be a smooth curve in M. We say that γ is a **geodesic** if $\nabla_{\dot{\gamma}}\dot{\gamma}\,(t) = 0$ for all $t \in I$. If $[a, b] \subset I$ and $\gamma : I \longrightarrow M$ is a geodesic, then the restriction of γ to $[a, b]$ is called a *geodesic segment joining* $\gamma(a)$ to $\gamma(b)$. Sometimes, with an abuse of language, if a curve $\gamma : I \longrightarrow M$ is a geodesic, we refer also to the imagine $\gamma(I)$ as a geodesic.

Remark 1.41. Let $\gamma : I \longrightarrow M$ be a geodesic. Then

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)} = 2 \langle \nabla_{\dot{\gamma}}\dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)} = 0,$$

meaning that the module of the tangent vector $\dot{\gamma}$ is constant. From now on, we will only consider geodesics γ such that $\|\dot{\gamma}(t)\|_{\gamma(t)} = c \neq 0$, i.e. geodesics which do not reduce to points.

Definition 1.42. If $\gamma : I \longrightarrow M$ is a smooth curve in M, the *arc length* s of γ starting from a fixed $t_0 \in I$ (i.e. $s(t_0) = 0$) is

$$s(t) = \int_{t_0}^{t} \|\dot{\gamma}(\tau)\|_{\gamma(\tau)}\, \mathrm{d}\tau.$$

Remark 1.43. If $\gamma : I \longrightarrow M$ is a geodesic and $\|\dot{\gamma}(t)\|_{\gamma(t)} = c$, then $s(t) = c(t - t_0)$. Therefore the parameter of γ is proportional to arc length and we say that γ is *normalized* if $c = 1$.

Now, we want to determine the local equation satisfied by a geodesic γ.

Proposition 1.44. *Let $\gamma : I \longrightarrow M$ be a smooth curve in M, $t_0 \in I$ and consider* $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$, *a system of coordinates about the point $\gamma(t_0)$. Let $\gamma(t) = (\gamma^1(t), \ldots, \gamma^n(t))$ be the local expression of γ in U. Then γ is a geodesic in U if and only if*

$$0 = \ddot{\gamma}^k + \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j, \quad \forall k = 1, \ldots, n. \tag{1.12}$$

Proof. Let $X_i := \frac{\partial}{\partial x^i}$. We have

$$\dot{\gamma} = \dot{\gamma}^k X_k,$$

where $X_k = X_k(\gamma(t))$, and

$$\nabla_{\dot{\gamma}}\dot{\gamma} = \ddot{\gamma}^k X_k + \dot{\gamma}^j \nabla_{\dot{\gamma}} X_j = \ddot{\gamma}^k X_k + \dot{\gamma}^j\dot{\gamma}^i \nabla_{X_i} X_j = \sum_k (\ddot{\gamma}^k + \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j) X_k.$$

Hence, γ is a geodesic in U if and only if the system (1.12) holds. $\square$

Remark 1.45. If $\gamma : I \longrightarrow M$ is a smooth curve in M, then we can consider the curve $t \longmapsto (\gamma(t), \dot\gamma(t))$ in the tangent bundle TM. Hence, if $\mathrm{X} : U \subset \mathbb{R}^n \longrightarrow M$ is a system of coordinates and $\gamma(t) = (\gamma^1(t), \dots, \gamma^n(t))$ is the local expression of γ in U, we have that γ is a geodesic in U if and only if the curve $t \longmapsto \big(\gamma^1(t), \dots, \gamma^n(t), \dot\gamma^1(t), \dots, \dot\gamma^n(t)\big)$ satisfies the system

$$\begin{cases} \dot\gamma^k = \eta_k \\ \dot\eta^k = -\Gamma^k_{ij}\eta^i\eta^j \qquad k = 1, \dots, n. \end{cases} \tag{1.13}$$

Therefore the second order system (1.12) on U and the first order system (1.13) on TU are equivalent.

The following one is an important result of existence and uniqueness.

Theorem 1.46 (**Existence and Uniqueness of Geodesics**). *Let $p \in M$, $v \in T_pM$ and $t_0 \in \mathbb{R}$. Then there exist an open interval $I \subset \mathbb{R}$ containing t_0 and a unique geodesic $\gamma : I \longrightarrow M$ satisfying $\gamma(t_0) = p$ and $\dot\gamma(t_0) = v$. Furthermore, if γ is a geodesic defined on an open interval I containing t_0 and $\tilde\gamma$ is a geodesic defined on another open interval J containing t_0, both satisfying the initial conditions, then $\gamma = \tilde\gamma$ on $I \cap J$.*

Proof. Let $p \in M$, $v \in T_pM$, $t_0 \in \mathbb{R}$ and let $\mathrm{X} : U \subset \mathbb{R}^n \longrightarrow M$ be a system of coordinates at p. We can consider the Cauchy problem given by

$$\begin{cases} \dot\gamma^k = \eta^k, \\ \dot\eta^k = -\Gamma^k_{ij}\eta^i\eta^j, \qquad k = 1, \dots, n, \\ (\gamma^1(t_0), \dots, \gamma^n(t_0)) = \mathrm{X}^{-1}(p), \\ (\eta^1(t_0), \dots, \eta^n(t_0)) = (\mathrm{dX}_p)^{-1}(v). \end{cases}$$

Applying the existence and uniqueness theorem for first-order ODEs, we conclude that there exist an $\varepsilon > 0$ and a unique solution $\theta : (t_0 - \varepsilon, t_0 + \varepsilon) \longrightarrow U \times \mathbb{R}^n$, let us denote it by $\theta(t) = (\gamma^1(t), \dots, \gamma^n(t), \eta^1(t), \dots, \eta^n(t))$, that is solution of our Cauchy problem. Therefore $(t_0 - \varepsilon, t_0 + \varepsilon) \ni t \longmapsto \mathrm{X}(\gamma^1(t), \dots, \gamma^n(t)) \in \mathrm{X}(U)$ is the unique geodesic defined on this interval and satisfying the initial conditions.

Finally, let $\gamma : I \longrightarrow M$ and $\tilde\gamma : J \longrightarrow M$ be two geodesics defined on two open intervals containing t_0 and satisfying $\gamma(t_0) = \tilde\gamma(t_0)$, $\dot\gamma(t_0) = \dot{\tilde\gamma}(t_0)$. We have that $\gamma = \tilde\gamma$

in an open neighbourhood of t_0. Defining $\beta := \sup\{b \in I \cap J : \gamma = \tilde{\gamma} \text{ in } [t_0, b]\}$, it has to be that $\beta = \sup I \cap J$. Indeed, if $\beta < \sup I \cap J$, then $\beta \in I \cap J$ and, by continuity, $\gamma(\beta) = \tilde{\gamma}(\beta)$, $\dot{\gamma}(\beta) = \dot{\tilde{\gamma}}(\beta)$. Hence, applying the local existence and uniqueness result in a neighbourhood of β, we conclude that $\gamma = \tilde{\gamma}$ in an interval that is larger than $[t_0, \beta]$; this is in contradiction with the definition of β. Similarly, we can show that $\inf\{a \in I \cap J : \gamma = \tilde{\gamma} \text{ in } [a, t_0]\} = \inf I \cap J$. Therefore $\gamma = \tilde{\gamma}$ in $I \cap J$. □

Corollary 1.47. *Let $p \in M$ and $v \in T_pM$. Then there exist a unique open interval I containing 0 and a unique geodesic $\gamma : I \longrightarrow M$ such that $\gamma(0) = p$, $\dot{\gamma}(0) = v$ and that is maximal (it can't be extended to a larger interval containing 0).*

Proof. Let $(I_j)_{j \in J}$ be the set of all open intervals containing 0 where geodesics satisfying the initial conditions are defined. For all j, let $\gamma^{(j)}$ be the geodesic defined on I_j; by the previous theorem, we have that $\gamma^{(j)} = \gamma^{(k)}$ in $I_j \cap I_k$. Let $I := \cup_{j \in J} I_j$ and define $\gamma : I \longrightarrow M$ by $\gamma(t) := \gamma^{(j)}(t)$ if $t \in I_j$. Then γ is well-defined and it is the geodesic we are looking for. □

Definition 1.48. For any $p \in M$ and $v \in T_pM$ we denote by γ_v the geodesic of Corollary 1.47 and we refer to it as the *(maximal) geodesic with initial point p and initial velocity v.*

The following result will be useful in the next section.

Lemma 1.49 (**Rescaling lemma**). *Let $p \in M$ and $v \in T_pM$. Then*

$$\gamma_{cv}(t) = \gamma_v(ct),$$

for all $c, t \in \mathbb{R}$ such that both sides are defined.

Proof. If $c = 0$, then the proof is trivial. Indeed, $\gamma_v(0) = p$ by definition of γ_v and, since $\dot{\gamma}_0(0) = 0$ and $\|\dot{\gamma}_0(t)\|_{\gamma_0(t)}$ is constant (because γ_0 is a geodesic), we have that $\dot{\gamma}_0(t) = 0$ for all t, from which $\gamma_0(t) = p$ for all t.

So we can assume $c \neq 0$. Let $I \subset \mathbb{R}$ be the interval of definition of the geodesic γ_v. We define the curve $\tilde{\gamma}_v(t) := \gamma_v(ct)$ on the interval $I_c := \{t \in \mathbb{R} : ct \in I\}$. We want to show that $\tilde{\gamma}_v$ is a geodesic such that $\tilde{\gamma}_v(0) = p$ and $\dot{\tilde{\gamma}}_v(0) = cv$. It is immediate that $\tilde{\gamma}_v(0) = p$.

Let $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$ be a system of coordinates at p and let $\gamma_v(t) = (\gamma^1(t), \ldots, \gamma^n(t))$ and $\tilde{\gamma}_v(t) = (\tilde{\gamma}^1(t), \ldots, \tilde{\gamma}^n(t)) = (\gamma^1(ct), \ldots, \gamma^n(ct))$ be the local expression of γ_v and $\tilde{\gamma}_v$ respectively. By the chain rule, $\frac{\mathrm{d}}{\mathrm{d}t}(\gamma^i(ct)) = c\frac{\mathrm{d}\gamma^i}{\mathrm{d}t}(ct)$, whence $\dot{\tilde{\gamma}}_v(0) = cv$. Setting $X_k := \frac{\partial}{\partial x^k}$, we have, as in the computation of Proposition 1.44,

$$\begin{aligned}\nabla_{\dot{\tilde{\gamma}}_v}\dot{\tilde{\gamma}}_v(t) &= \sum_k \left(\ddot{\tilde{\gamma}}^k(t) + \Gamma^k_{ij}(\tilde{\gamma}_v(t))\dot{\tilde{\gamma}}^i(t)\dot{\tilde{\gamma}}^j(t)\right) X_k(\tilde{\gamma}_v(t)) = \\ &= \sum_k \left(c^2\ddot{\gamma}^k(ct) + c^2\dot{\gamma}^i(ct)\dot{\gamma}^j(ct)\Gamma^k_{ij}(\gamma_v(ct))\right) X_k(\gamma_v(ct)) = c^2\nabla_{\dot{\gamma}_v}\dot{\gamma}_v(ct) = 0.\end{aligned}$$

Thus, $\tilde{\gamma}_v$ is a geodesic. $\tilde{\gamma}_v$ is a maximal [2] geodesic such that $\tilde{\gamma}_v(0) = p$ and $\dot{\tilde{\gamma}}_v(0) = cv$, whence $\tilde{\gamma} = \gamma_{cv}$. ☐

Before studying minimizing properties of geodesics, we aim to give a geometric interpretation of geodesics on a surface in $\mathbb{R}^3$. We first recall the following fact from differential geometry of surfaces.

Remark 1.50. Let S be a regular orientable surface in $\mathbb{R}^3$ and $\gamma : (a, b) \longrightarrow S$ a smooth curve on S, parameterized by arc length. If $\vec{t}(s)$ is the tangent unit vector to γ at the point $\gamma(s)$ and N is a unit-normal vector field on S, then $\{\vec{t}(s), N(\gamma(s)), \vec{t}(s) \wedge N(\gamma(s))\}$ is an orthonormal basis for $\mathbb{R}^3$ and $\ddot{\gamma}(s) \in \operatorname{span}\{N(\gamma(s)), \vec{t}(s) \wedge N(\gamma(s))\}$ for any s. If k is the *curvature* of γ (i.e. $k(s) = \|\ddot{\gamma}(s)\|$), then there exist two functions $k_n : (a, b) \longrightarrow \mathbb{R}$ and $k_g : (a, b) \longrightarrow \mathbb{R}$ such that $\ddot{\gamma}(s) = k_n(s)N(\gamma(s)) + k_g(s)\big(\vec{t}(s) \wedge N(\gamma(s))\big)$ and $k(s)^2 = k_N^2(s) + k_g^2(s)$ for any s.

We can characterize geodesics on a surface as curves with acceleration orthogonal to the surface itself at any point.

Lemma 1.51. *Let $\sigma : U \longrightarrow \mathbb{R}^3$ be a parametrization of a regular orientable surface S and let $\gamma : (a, b) \longrightarrow \sigma(U)$ be a smooth curve parameterized by arc length. Then γ is a geodesic if and only if $\ddot{\gamma}(t) \perp T_{\gamma(t)}S$ for any t.*

Proof. γ is a geodesic if and only if $0 = \nabla_{\dot{\gamma}}\dot{\gamma}(t) = \ddot{\gamma}(t) - \langle\ddot{\gamma}(t), N(\gamma(t))\rangle\, N(\gamma(t))$ for any t. Therefore γ is a geodesic if and only if $\ddot{\gamma}(t)$ is parallel to $N(\gamma(t))$ for any t, equivalently $\ddot{\gamma}(t) \perp T_{\gamma(t)}S$ for any t (i.e. $k_g(t) = 0$ for any $t \in (a, b)$). ☐

[2] The maximality of $\tilde{\gamma}_v$ follows immediately from the maximality of γ_v

1.4 The exponential map

In this section, we introduce the exponential map on a subset of TM.

Definition 1.52. Let

$$\mathcal{E} := \{(p,v) \in TM : \gamma_v \text{ is defined on an open interval containing } [0,1]\}.$$

We define the *exponential map* $\exp : \mathcal{E} \longrightarrow M$ by

$$\exp(p,v) := \gamma_v(1).$$

For all $p \in M$, we denote by $\exp_p$ the restriction of $\exp$ to $\mathcal{E}_p := \mathcal{E} \cap T_pM$:

$$\exp_p(v) := \exp(p,v), \quad \forall v \in \mathcal{E}_p.$$

We recall some general results about vector fields, before studying the properties of the exponential map.

Definition 1.53. Let X be a differentiable vector field on a differentiable manifold M. We say that $I \ni t \longmapsto \gamma(t) \in M$ is an *integral curve* of X if $\dot\gamma(t) = X(\gamma(t))$ for all t.

Proposition 1.54. *Let X be a C^∞ vector field on a differentiable manifold M and let $p \in M$. Then there exists a unique interval $I(p) = [\alpha(p), \beta(p)]$ containing $t = 0$ and having the following properties:*

1. *1 there exists a C^∞ integral curve of X, $I(p) \ni t \longmapsto \theta_p(t) =: \theta(t,p)$, such that $\theta(0,p) = p$;*
2. *given any other integral curve $G(t)$ with $G(0) = p$, then the domain of G is contained in $I(p)$ and $G(t) = \theta(t,p)$ on the domain of G.*

We say that θ_p is the maximal integral curve of X starting at p.

For a proof of Proposition 1.54 see Theorem IV.4.3 of [1].

Proposition 1.55. *For any C^∞-vector field X the domain of $\theta(t,p)$ is open in $\mathbb{R} \times M$ and θ is a C^∞ map onto M.*

For a proof of Proposition 1.55 see Theorem IV.4.5 of [1].

Proposition 1.56 (Properties of the exponential map). *The map* $\exp : \mathcal{E} \longrightarrow M$ *satisfies the following properties:*

1. *$\mathcal{E}$ is an open subset of TM containing the zero section $M \times \{0\}$ and each set $\mathcal{E}_p$ is star-shaped with respect to* 0[3].

2. *For each $(p, v) \in TM$ the geodesic γ_v is given by*

$$\gamma_v(t) = \exp_p(tv),$$

   for all t such that both sides are defined.

3. *The exponential map is smooth.*

Proof. For each $(p, v) \in TM$, by definition, $\exp_p(tv) = \gamma_{tv}(1)$ and, applying Lemma 1.49, we conclude $\exp_p(tv) = \gamma_{tv}(1) = \gamma_v(t)$, whenever either side is defined. Therefore *2* is proved. Furthermore, if $v \in \mathcal{E}_p$ and $t \in [0, 1]$, again by Lemma 1.49,

$$\exp_p(tv) = \gamma_{tv}(1) = \gamma_v(t)$$

is defined. Thus, $tv \in \mathcal{E}_p$ and this shows that $\mathcal{E}_p$ is star-shaped with respect to 0.

Given a system of coordinates for M, $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$, we also have a system of coordinates for TM, $d\mathrm{x} : TU \cong U \times \mathbb{R}^n \longrightarrow TM$, where $d\mathrm{x}(x, \xi) = (\mathrm{x}(x), d\mathrm{x}_x(\xi))$. Let $(x^1, \dots, x^n, \xi^1, \dots, \xi^n)$ be the local coordinates for $\pi^{-1}(\mathrm{x}(U)) = d\mathrm{x}(TU)$, where $\pi : TM \longrightarrow M$ is the canonical projection. We denote, as usual, $X_i := \frac{\partial}{\partial x^i}$, $\Xi_i := \frac{\partial}{\partial \xi^i}$, Γ_{ij}^k the smooth functions on U such that $\nabla_{X_i} X_j = \Gamma_{ij}^k X_k$; then we consider the vector field

$$G(p, v) = \xi^k(p, v) X_k(p, v) - \xi^i(p, v) \xi^j(p, v) \Gamma_{ij}^k(p)\, \Xi_k(p, v),$$

for $(p, v) \in \pi^{-1}(\mathrm{x}(U))$. Let $t \longmapsto (\gamma^1(t), \dots, \gamma^n(t), \eta^1(t), \dots, \eta^n(t))$ be the local expression of a curve in $\pi^{-1}(\mathrm{x}(U))$; then it is an integral curve of G if and only if it satisfies the

[3] If S is a subset of a vector space, then S is said to be star shaped with respect to $x \in S$ if for any $y \in S$ the whole segment $[x, y] := \{(1 - t)x + ty : t \in [0, 1]\}$ lies in S.

system of ODEs

$$\begin{cases} \dot{\gamma}^k(t) = \eta^k(t), \\ \dot{\eta}^k(t) = -\eta^i(t)\eta^j(t)\Gamma^k_{ij}(\gamma(t)), \quad \forall k = 1, \dots, n. \end{cases} \tag{1.14}$$

(1.14) is exactly the first-order system (1.13) equivalent to the geodesic equation, whence $t \longrightarrow \Gamma(t)$ is an integral curve of G if and only if its projection $\gamma(t) := \pi(\Gamma(t))$ is a geodesic.

We want to show that G extends to a global vector field on TM which is called the *geodesic vector field.* First, we see that G acts on any smooth function $f : TM \longrightarrow \mathbb{R}$ in the following way:

$$Gf(p, v) = \frac{\mathrm{d}}{\mathrm{d}t}\big(f(\gamma_v(t), \dot{\gamma}_v(t))\big)\Big|_{t=0}, \quad \forall (p, v) \in \pi^{-1}(\mathrm{x}(U)).$$

Indeed, writing $\gamma_v(t) = (\gamma^1(t), \dots, \gamma^n(t))$ and $\dot{\gamma}_v(t) = (\eta^1(t), \dots, \eta^n(t))$ in local coordinates, then

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(f(\gamma_v(t), \dot{\gamma}_v(t))\right)\Big|_{t=0} &= \left[\sum_{k=1}^n \left(\frac{\partial f}{\partial x^k}(\gamma_v(t), \dot{\gamma}_v(t))\dot{\gamma}^k(t) + \frac{\partial f}{\partial \xi^k}(\gamma_v(t), \dot{\gamma}_v(t))\dot{\eta}^k(t)\right)\right]_{t=0} = \\ &= \sum_{k=1}^n \left(\frac{\partial f}{\partial x^k}(p, v)\xi^k(p, v) - \frac{\partial f}{\partial \xi^k}(p, v)\xi^i(p, v)\xi^j(p, v)\Gamma^k_{ij}(p)\right) = Gf(p, v), \end{aligned}$$

where we have used that $(\gamma_v, \dot{\gamma}_v)$ satisfies system (1.14) in the second line. Since the formula that expresses the action of G on f is independent of coordinates, we can define G on any coordinate neighbourhood of TM and the definitions have to agree when the intersections are not empty. It follows that G extends to the whole tangent bundle TM.

By Proposition 1.54 and Proposition 1.55, there exist an open neighbourhood $\mathcal{O}$ of $\{0\} \times TM$ in $\mathbb{R} \times TM$ and a smooth map $\theta : \mathcal{O} \longrightarrow TM$ such that each curve $\theta_{(p,v)}(t) := \theta(t, (p, v))$ is the maximal integral curve of G starting at (p, v), defined on an open interval $I((p, v))$ containing 0. We also note that the maximal geodesic starting at p with initial velocity v is the projection of the maximal integral curve of G starting at (p, v), i.e. $\gamma_v = \pi \circ \theta_{(p,v)}$.

Now, let $(p, v) \in \mathcal{E}$: this means that γ_v is defined on an open interval containing $[0, 1]$ and so is $\theta_{(p,v)}$. Then $(1, (p, v)) \in \mathcal{O}$ that is open in $\mathbb{R} \times TM$, whence there exists an open

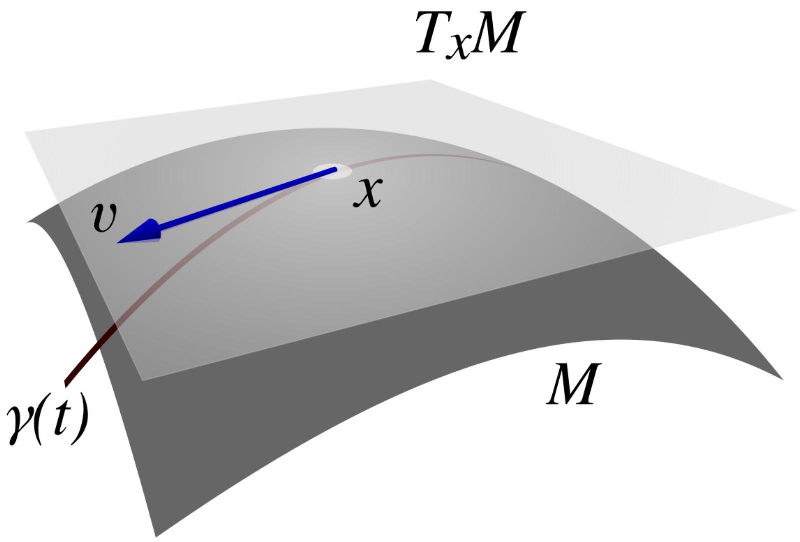

Figure 1.1: The exponential map $\exp_x$ associates to a vector $v \in T_xM$ the second extreme $\gamma(1)$ of the geodesic starting at x with initial velocity v.
See `https://it.wikipedia.org/wiki/Mappa_esponenziale`.

neighbourhood of $(1, (p, v))$ in $\mathbb{R} \times TM$ on which θ is defined. This implies that there exists an open neighbourhood of (p, v) in TM, call it $Z_{(p,v)}$, such that $\theta_{(q,w)}$ is defined on $[0, 1]$ for all $(q, w) \in Z_{(p,v)}$; therefore γ_w is defined on $[0, 1]$ for any $(q, w) \in Z_{(p,v)}$, meaning that $Z_{(p,v)} \subset \mathcal{E}$. This show that $\mathcal{E}$ is open. Moreover, for any $p \in M$ the maximal geodesic with initial point p and initial velocity 0 is nothing less than the point p itself and it is defined for all $t \in [0, 1]$. Thus, $M \times \{0\} \subset \mathcal{E}$ and the proof of *1* is completed.

Finally,

$$\exp(p, v) = \gamma_v(1) = \pi \circ \theta_{(p,v)}(1) = \pi \circ \theta(1, (p, v)), \quad \forall (p, v) \in \mathcal{E},$$

from which exp is smooth, being the composition of smooth maps. □

Lemma 1.57. *For any $p \in M$ there exist an open neighborhood $\mathcal{V}$ of $0 \in T_pM$ and an open neighbourhood $\mathcal{U}$ of $p \in M$ such that $\exp_p : \mathcal{V} \longrightarrow \mathcal{U}$ is a diffeomorphism.*

Proof. We have $\exp_p : \mathcal{E}_p \subset T_pM \longrightarrow M$ and $\mathrm{d}(\exp_p)_0 : T_0(T_pM) \longrightarrow T_pM$. There is a natural identification $T_0(T_pM) \equiv T_pM$ (because T_pM is a vector space whose dimension is the same as M), then we can consider $\mathrm{d}(\exp_p)_0 : T_pM \longrightarrow T_pM$. For any $v \in T_pM$, we have $\mathrm{d}(\exp_p)_0(v) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\exp_p \circ\gamma(t)\right)\Big|_{t=0}$, where $\gamma : (-\varepsilon, \varepsilon) \longrightarrow T_pM$ is a smooth curve satisfying $\gamma(0) = 0$ and $\dot{\gamma}(0) = v$. Choosing $\gamma(t) = tv$, we get

$$\mathrm{d}(\exp_p)_o(v) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\exp_p(tv)\right)\Big|_{t=0} = \dot{\gamma}_v(t)\big|_{t=0} = v,$$

whence $\mathrm{d}(\exp_p)_0$ is the identity map on T_pM and it is invertible. By the inverse function theorem[4], $\exp_p$ is a local diffeomorphism at 0. □

Definition 1.58. Let $p \in M$. A *normal neighbourhood of* p is an open neighbourhood $\mathcal{U}$ of p that is the diffeomorphic image under $\exp_p$ of an open neighbourhood $\mathcal{V}$ of $0 \in T_pM$. If $\varepsilon > 0$ is small enough such that, given $B_\varepsilon(0) := \{v \in T_pM : ||v||_p < \varepsilon\}$, $\exp_p$ is a diffeomorphism on $B_\varepsilon(0)$, we say that $\exp_p(B_\varepsilon(0))$ is a *geodesic ball* of radius ε in M. If $\overline{B_\varepsilon(0)}$ is contained in an open set $\mathcal{V} \subset T_pM$ such that $\exp_p$ is a diffeomorphism on $\mathcal{V}$, then we say that $\exp_p(\overline{B_\varepsilon(0)})$ is a *closed geodesic ball* of radius ε and $\exp_p(\partial B_\varepsilon(0))$ is a *geodesic sphere*.

Definition 1.59. We define the *injectivity radius* of M at $p \in M$ by

$$\mathrm{inj}_p(M) := \sup\{r > 0 : \exp_p \text{ is a diffeomorphism on } B_r(0) \subset T_pM\},$$

and the *injectivity radius* of M by

$$\mathrm{inj}(M) := \inf\{\mathrm{inj}_p(M) : p \in M\}.$$

Definition 1.60. Let $\{E_i\}_{i=1,\dots,n.}$ be an orthonormal basis for T_pM. Consider the isomorphism $E : \mathbb{R}^n \longrightarrow T_pM$ given by $E(x^1, \dots, x^n) = \sum_{i=1}^n x^i E_i$. Let $\mathcal{U}$ be a normal neighbourhood of p, $U := E^{-1} \circ \exp_p^{-1}(\mathcal{U})$ and consider the system of coordinates about p given by $\mathrm{x} := \exp_p \circ E : U \subset \mathbb{R}^n \longrightarrow M$. We say that $(x^1, \dots, x^n)$ are *normal coordinates* centered at p.

[4] As a consequence of the inverse function theorem in $\mathbb{R}^n$, there is a version for differentiable mapping on differentiable manifolds: if $\varphi : M_1 \longrightarrow M_2$ is a differentiable mapping and $p \in M_1$ such that $\mathrm{d}\varphi_p : T_pM_1 \longrightarrow T_{\varphi(p)}M_2$ is an isomorphism, then φ is a local diffeomorphism at p.

If $\mathrm{x} : U \longrightarrow M$ is a system of normal coordinates centered at p, we can define the *radial distance function* r by

$$r(q) := \left(\sum_{i=1}^{n} x^i(q)^2 \right)^{1/2}, \quad q \in \mathrm{x}(U),$$

and the *unit radial vector field* $\partial/\partial r$ by

$$\frac{\partial}{\partial r} := \sum_{i=1}^{n} \frac{x^i}{r} \frac{\partial}{\partial x^i}.$$

The normal coordinates satisfies the following properties.

Proposition 1.61. *Let* $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow \mathcal{U} \subset M$ *be a system of normal coordinates centered at* p.

1. *For any* $v = \sum_{i=1}^n v^i \frac{\partial}{\partial x^i}(p) \in T_pM$ *the maximal geodesic starting at* p *with initial velocity* v *is represented in normal coordinates by*

$$\gamma_v(t) = (tv^1, \dots, tv^n), \tag{1.15}$$

   for all t *such that* $\gamma_v(t) \in \mathcal{U}$. *Thus, radial paths in normal coordinates are exactly the geodesics through* p.

2. *The coordinates of* p *are* $(0, \dots, 0)$.

3. $g_{ij}(p) := \left\langle \frac{\partial}{\partial x^i}(p), \frac{\partial}{\partial x^j}(p) \right\rangle_p = \delta_{ij}$.

4. *The partial derivatives of the map* $(x^1, \dots, x^n) \longrightarrow g_{ij}(x^1, \dots, x^n)$ *and the Christoffel's symbols vanish at* p.

Proof. There exists an orthonormal basis for T_pM $\{E_i\}_{i=1,\dots,n}$ such that $\mathrm{x} = \exp_p \circ E$, where E is the isomorphism between $\mathbb{R}^n$ and T_pM of Definition 1.60. First of all, we note that

$$\frac{\partial}{\partial x^i}(p) = \mathrm{d}(\exp_p \circ E)_0(e_i) = \big(\underbrace{\mathrm{d}(\exp_p)_0}_{=id_{T_pM}} \circ \underbrace{\mathrm{d}E_0}_{=E}\big)(e_i) = E(e_i) = E_i,$$

from which *3* is immediate.

The maximal geodesic starting at p with initial velocity v is given by $\gamma_v(t) := \exp_p(tv)$, as long as $\gamma_v(t) \in \mathcal{U}$. If $v = \sum_{i=1}^n v^i E_i$, then $E^{-1} \circ \exp_p^{-1}(\gamma_v(t)) = (tv^1, \dots, tv^n)$ for any t such that $\gamma_v(t) \in \mathcal{U}$ and $E^{-1} \circ \exp_p^{-1}(p) = (0, \dots, 0)$. Hence, *1* and *2* are proved.

Let $v \in T_pM$, $v = \sum_{i=1}^m v^i \frac{\partial}{\partial x^i}(p)$ and let $\gamma_v(t) = (tv^1, \dots, tv^n)$ be the local expression of the maximal geodesic starting at p with initial velocity v; by the geodesic equation for γ_v, we have that $0 = v^i v^j \Gamma_{ij}^k(\gamma_v(t))$ and, in particular, for $t = 0$ we get

$$0 = v^i v^j \Gamma_{ij}^k(p), \quad \forall k = 1, \dots, n.$$

This implies that $\Gamma_{ij}^k(p) = 0$ for any i, j, k. Finally, for any i, j, k

$$\begin{aligned} \frac{\partial}{\partial x_k}\left(g_{ij}(x^1, \dots, x^n)\right)\Big|_{(x^1,\dots,x^n)=(0,\dots,0)} &= X_k \left\langle X_i(\cdot), X_j(\cdot)\right\rangle_{(\cdot)}(p) = \\ &= \left\langle \nabla_{X_k} X_i(p), X_j(p)\right\rangle_p + \left\langle X_i(p), \nabla_{X_k} X_j(p)\right\rangle_p = 0, \end{aligned}$$

where we have used Corollary 1.34 and the fact that $\nabla_{X_k} X_i, \nabla_{X_k} X_j$ vanish at p, since the Christoffel's symbols vanish at p. Thus, the proof of *4* is completed. $\square$

Definition 1.62. Because of formula (1.15), if $p \in M$, then geodesics starting at p and lying in a normal neighbourhood of p are called *radial geodesics*.

Remark 1.63. If $p \in M$ and $q \in \exp_p(B_\rho(0))$, where $\rho < \mathrm{inj}_p(M)$, then there exists a unique radial geodesic from p to q lying in $\exp_p(B_\rho(0))$ (up to reparametrization).

Indeed, if $v \in B_\rho(0) \subset T_pM$ such that $q = \exp_p(v)$, then $\gamma_v(t) := \exp_p(tv)$, where $t \in [0, 1]$, is a radial geodesic from p to q lying in $\exp_p(B_\rho(0))$. Now, let $\tilde{\gamma} : [a, b] \longrightarrow \exp_p(B_\rho(0))$ be another radial geodesic from p to q lying in the same geodesic ball; then $\tilde{\tilde{\gamma}} : [0, b-a] \longrightarrow \exp_p(B_\rho(0))$ given by $\tilde{\tilde{\gamma}}(t) := \tilde{\gamma}(t + a)$ is still a radial geodesic from p to q and $\tilde{\tilde{\gamma}}(t) = \exp_p(tz)$, where z is the initial velocity vector of $\tilde{\tilde{\gamma}}$. Furthermore, it has to be $\exp_p((b-a)z) = q = \exp_p(v)$, whence $(b-a)z = v$. It follows that $\tilde{\tilde{\gamma}}(t) = \gamma_v(\frac{1}{b-a}t)$ and $\tilde{\gamma}(t) = \tilde{\tilde{\gamma}}(t-a) = \gamma_v(\frac{1}{b-a}(t-a))$, meaning that $\tilde{\gamma}$ is a reparametrization of γ_v.

Definition 1.64. Let $p \in M$, $\rho < \mathrm{inj}_p(M)$ and $q \in \exp_p(B_\rho(0))$. If $v \in B_\rho(0)$ such that $q = \exp_p(v)$, by the previous remark, we refer to the curve $[0, 1] \ni t \longmapsto \exp_p(tv)$ as the *radial geodesic from p to q.*

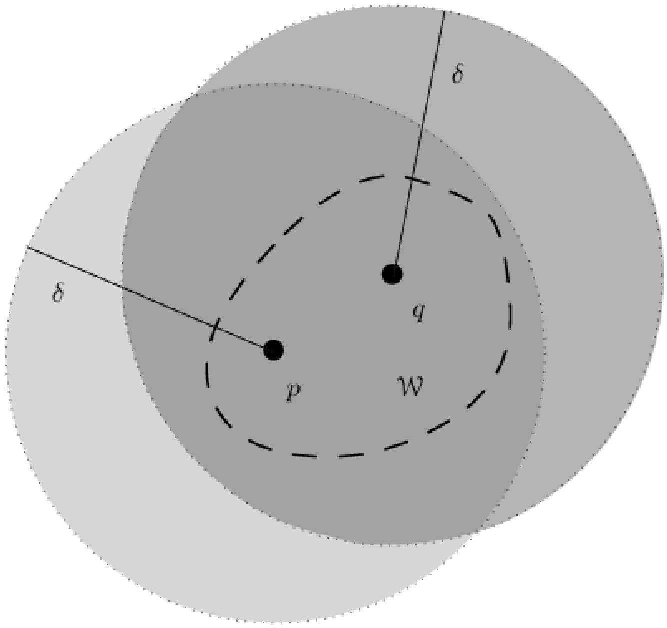

Figure 1.2: Uniformly normal neighbourhood
See Figure 5.4 on [6].

Definition 1.65. An open set $\mathcal{W} \subset M$ is called *uniformly normal* if there exists a value $\delta > 0$ such that for any $p \in \mathcal{W}$ there exists a geodesic ball of radius δ containing $\mathcal{W}$.

Lemma 1.66. *Let $p \in M$ and let $\mathcal{U}$ be a neighbourhood of p. There exists a uniformly normal neighbourhood $\mathcal{W}$ of p such that $\mathcal{W} \subset \mathcal{U}$.*

For a proof of Lemma 1.66 see Lemma 5.12. on chapter 5 of [6].

1.5 Minimizing curves are geodesics

In this section, we aim to show that, if c is an admissible curve minimizing the arc length, then it is a geodesic. We need some preliminary definitions and lemmas.

Definition 1.67. Let $\gamma : [a, b] \longrightarrow M$ be a segment of an admissible curve. We say that γ is *minimizing* if $\ell(\gamma) \leq \ell(c)$ for any arbitrary admissible curve c joining $\gamma(a)$ to $\gamma(b)$.

Definition 1.68. An *admissible family of curves* is a continuous map $\Gamma : (-\varepsilon, \varepsilon) \times [a, b] \longrightarrow M$ such that Γ is smooth on any rectangle of the form $(-\varepsilon, \varepsilon) \times [a_{i-1}, a_i]$, where

$\{a_0 = a < \ldots < a_k = b\}$ is a finite subdivision, and such that $\Gamma_s(t) := \Gamma(s,t)$ is an admissible curve for each $s \in (-\varepsilon, \varepsilon)$.

A *vector field along* Γ is a continuous mapping $V : (-\varepsilon, \varepsilon) \times [a,b] \longrightarrow TM$ such that $V(s,t) \in T_{\Gamma(s,t)}M$ for each (s,t) and such that $V|_{(-\varepsilon,\varepsilon)\times[\tilde{a}_{i-1},\tilde{a}_i]}$ is smooth for a finite subdivision $\{\tilde{a}_0 = a < \ldots < \tilde{a}_m = b\}$.

Given ad admissible family of curves, the *main curves* are $\Gamma_s(t) := \Gamma(s,t)$, defined on $[a,b]$, while the *transverse curves* are $\Gamma^{(t)}(s) := \Gamma(s,t)$, defined on $(-\varepsilon, \varepsilon)$. We note that the transverse curves are smooth on $(-\varepsilon, \varepsilon)$ for each t, while the main curves are, in general, only admissible curves.

Notation 1.69. If Γ is smooth, we denote by

$$\partial_t \Gamma(s,t) := \dot{\Gamma}_s(t),$$

the velocity field of the main curves and by

$$\partial_s \Gamma(s,t) := \dot{\Gamma}^{(t)}(s),$$

the velocity field of the transverse curves.

If V is a vector field along Γ, on any rectangle where Γ is smooth we can consider the covariant derivatives of V either along the main curves or along the transverse curves: we denote them respectively by $D_t V$ and $D_s V$.

We have the following result.

Lemma 1.70 (Symmetry Lemma). *Let $\Gamma : (-\varepsilon, \varepsilon) \times [a,b] \longrightarrow M$ be an admissible family of curves. On any rectangle $(-\varepsilon, \varepsilon) \times (a_{i-1}, a_i)$ where Γ is smooth we have*

$$D_s \partial_t \Gamma = D_t \partial_s \Gamma.$$

Proof. We fix one of such rectangle where Γ is smooth and consider a point $(s_0, t_0) \in (-\varepsilon, \varepsilon) \times (a_{i-1}, a_i)$. Let $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$ be a system of coordinates about $\Gamma(s_0, t_0)$. Let $\Gamma(s,t) = (\Gamma^1(s,t), \ldots, \Gamma^n(s,t))$ be the local expression of Γ, then

$$\partial_t \Gamma = \frac{\partial \Gamma^k}{\partial t} X_k, \qquad \partial_s \Gamma = \frac{\partial \Gamma^k}{\partial s} X_k,$$

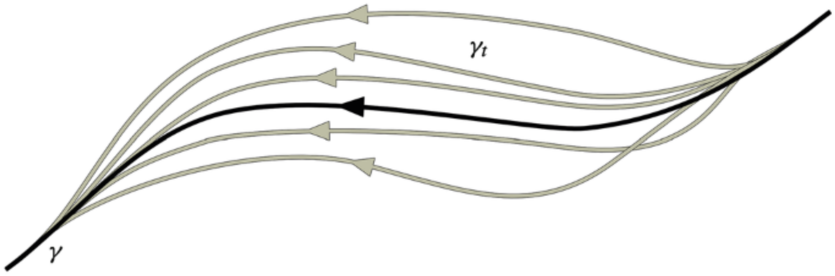

Figure 1.3: Proper variation of a curve
See `https://link.springer.com/chapter/10.1007/978-3-031-39838-4_2`.

where $X_k := \frac{\partial}{\partial x^k}$ and $X_k = X_k(\Gamma(s,t))$. By properties of covariant derivatives and definition of Christoffel's symbols, we have

$$D_s\partial_t\Gamma = \frac{\partial^2\Gamma^k}{\partial s\partial t}X_k + \frac{\partial\Gamma^k}{\partial t}D_sX_k = \sum_{k=1}^n\left(\frac{\partial^2\Gamma^k}{\partial s\partial t} + \frac{\partial\Gamma^i}{\partial t}\frac{\partial\Gamma^j}{\partial s}\Gamma^k_{ij}\right)X_k,$$

and similarly

$$D_t\partial_s\Gamma = \sum_{k=1}^n\left(\frac{\partial^2\Gamma^k}{\partial t\partial s} + \frac{\partial\Gamma^i}{\partial s}\frac{\partial\Gamma^j}{\partial t}\Gamma^k_{ij}\right)X_k.$$

By Remark 1.36, we know that $\Gamma^k_{ij} = \Gamma^k_{ji}$, whence the expression of $D_s\partial_t\Gamma$ and $D_t\partial_s\Gamma$ do coincide. $\square$

Definition 1.71. Let $\gamma : [a,b] \longrightarrow M$ be an admissible curve. A *variation of* γ is an admissible family $\Gamma : (-\varepsilon,\varepsilon)\times[a,b] \longrightarrow M$ such that $\Gamma_0(t) := \Gamma(0,t) = \gamma(t)$ for all $t \in [a,b]$. We say that Γ is a *proper variation* if $\Gamma_s(a) = \gamma(a)$ and $\Gamma_s(b) = \gamma(b)$ for all $s \in (-\varepsilon,\varepsilon)$. Furthermore, if Γ is a variation of γ, the *variation field of* Γ is the vector field along γ given by $V(t) := \partial_s\Gamma(s,t)|_{s=0}$ and we say that V is *proper* if $V(a) = V(b) = 0$.

Remark 1.72. The variation field of a proper variation is proper.

Lemma 1.73. *Let $\gamma : [a,b] \longrightarrow M$ be an admissible curve and let V be a vector field along γ. Then there exists a variation Γ such that V is the variation field of Γ. Furthermore, if V is proper, then we can take Γ to be proper as well.*

Proof. Let us set $\Gamma(s,t) := \exp_{\gamma(t)}(sV(t)) = \exp(\gamma(t), sV(t))$. Since $[a,b]$ is compact (and then $\gamma([a,b])$) is compact), there exists a positive value ε such that Γ is defined on

$(-\varepsilon,\varepsilon)\times[a,b]$. Γ is smooth on any $(-\varepsilon,\varepsilon)\times[a_{i-1},a_i]$ for each subinterval $[a_{i-1},a_i]$ where V is smooth and it is continuous on its whole domain. By the properties of the exponential map, $\partial_s\Gamma(s,t)\big|_{s=0} = \partial_s(\exp_{\gamma(t)}(sV(t)))\big|_{s=0} = V(t)$. Moreover, if $V(a) = V(b) = 0$, then $\Gamma_s(a) := \Gamma(s,a) = \gamma(a)$ and $\Gamma_s(b) := \Gamma(s,b) = \gamma(b)$, so Γ is a proper variation. $\square$

Proposition 1.74 (First variation formula for ℓ). *Let $\gamma : [a,b] \longrightarrow M$ be a unit speed admissible curve. Let $\Gamma : (-\varepsilon,\varepsilon)\times[a,b] \longrightarrow M$ be a proper variation of γ and V its variation field. Let $\{a_0 = a < \ldots < a_k = b\}$ be a subdivision of $[a,b]$ such that Γ is smooth on any rectangle of the form $(-\varepsilon,\varepsilon)\times[a_{i-1},a_i]$ (as in the definition of admissible family of curves). Then*

$$\frac{\mathrm{d}}{\mathrm{d}s}\ell_a^b(\Gamma_s)\bigg|_{s=0} = -\sum_{i=1}^{k}\int_{a_{i-1}}^{a_i}\langle V(t),\nabla_{\dot\gamma}\dot\gamma(t)\rangle_{\gamma(t)}\,\mathrm{d}t - \sum_{i=1}^{k-1}\langle V(a_i),\triangle_i\dot\gamma\rangle_{\gamma(a_i)}\,, \tag{1.16}$$

where $\triangle_i\dot\gamma := \dot\gamma(a_i^+) - \dot\gamma(a_i^-) = \lim_{t\to a_i^+}\dot\gamma(t) - \lim_{t\to a_i^-}\dot\gamma(t)$ is the jump in the tangent vector field $\dot\gamma$ at a_i.

Proof. We denote

$$T(s,t) := \partial_t\Gamma(s,t), \qquad S(s,t) := \partial_s\Gamma(s,t).$$

We have that

$$\ell_{a_{i-1}}^{a_i}(\Gamma_s|_{[a_{i-1},a_i]}) = \int_{a_{i-1}}^{a_i}\|T(s,t)\|_{\Gamma(s,t)}\,\mathrm{d}t = \int_{a_{i-1}}^{a_i}\langle T(s,t),T(s,t)\rangle_{\Gamma(s,t)}^{1/2}\,\mathrm{d}t,$$

for all $i = 1,\ldots,k$. On any subinterval $[a_{i-1},a_i]$ the integrand in $\ell_{a_{i-1}}^{a_i}(\Gamma_s|_{[a_{i-1},a_i]})$ is differentiable and the domain of integration is compact; hence, we can differentiate under the integral sign:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\ell_{a_{i-1}}^{a_i}(\Gamma_s|_[a_{i-1},a_i]) &= \int_{a_{i-1}}^{a_i}\frac{\partial}{\partial s}\langle T(s,t),T(s,t)\rangle_{\Gamma(s,t)}^{1/2}\,\mathrm{d}t = \\ &= \int_{a_{i-1}}^{a_i}\frac{1}{2}\langle T(s,t),T(s,t)\rangle_{\Gamma(s,t)}^{-1/2}\,2\,\langle D_sT(s,t),T(s,t)\rangle_{\Gamma(s,t)}\,\mathrm{d}t = \\ &= \int_{a_{i-1}}^{a_i}\frac{1}{\|T(s,t)\|_{\Gamma(s,t)}}\langle D_tS(s,t),T(s,t)\rangle_{\Gamma(s,t)}\,\mathrm{d}t,\end{aligned}$$

where we have used Proposition 1.33 in the second line to differentiate $\langle T(s,t),T(s,t)\rangle_{\Gamma(s,t)}$ and Lemma 1.70 in the last line to say that $D_s\partial_t\Gamma(s,t) = D_t\partial_s\Gamma(s,t)$.

In particular, $S(0,t) = V(t)$ and $T(0,t) = \dot\gamma(t)$, whence,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\ell_{a_{i-1}}^{a_i}(\Gamma_s|_{[a_{i-1},a_i]})\Big|_{s=0} &= \int_{a_{i-1}}^{a_i} \underbrace{\frac{1}{\|\dot\gamma(t)\|_{\gamma(t)}}}_{=1} \langle \nabla_{\dot\gamma} V(t), \dot\gamma(t)\rangle_{\gamma(t)}\,\mathrm{d}t = \\ &= \int_{a_{i-1}}^{a_i}\left(\frac{\mathrm{d}}{\mathrm{d}t}\langle V(t),\dot\gamma(t)\rangle_{\gamma(t)} - \langle V(t), \nabla_{\dot\gamma}\dot\gamma(t)\rangle_{\gamma(t)}\right)\mathrm{d}t = \\ &= \big\langle V(a_i), \dot\gamma(a_i^-)\big\rangle_{\gamma(a_i)} - \big\langle V(a_{i-1}), \dot\gamma(a_{i-1}^+)\big\rangle_{\gamma(a_{i-1})} - \int_{a_{i-1}}^{a_i}\langle V(t), \nabla_{\dot\gamma}\dot\gamma(t)\rangle_{\gamma(t)}\,\mathrm{d}t.\end{aligned}$$

Summing all over $i = 1, \ldots, k$ and using that $V(a_0) = V(a) = 0$, $V(a_k) = V(b) = 0$, as the variation Γ is proper by hypothesis, we obtain (1.16). $\square$

Remark 1.75. We have seen that any admissible curve has a unit speed parametrization and the length functional is independent of parametrization; therefore the requirement that γ is unit speed in the above proposition is not restrictive.

The following result is crucial.

Theorem 1.76. *Let $\gamma : [a,b] \longrightarrow M$ be a unit speed admissible curve. Suppose that $\ell(\gamma) \le \ell(c)$ for any admissible curve c joining $\gamma(a)$ and $\gamma(b)$, i.e γ is a minimizing curve. Then γ is a geodesic.*

Proof. If $\Gamma : (-\varepsilon,\varepsilon)\times[a,b] \longrightarrow M$ is a proper variation of γ, then $\Gamma_s(t) := \Gamma(s,t)$ is an admissible curve joining $\gamma(a)$ and $\gamma(b)$ for all s, whence $\ell(\Gamma_0) \le \ell(\Gamma_s)$ for all s, because γ minimizes the length functional among those curves joining $\gamma(a)$ and $\gamma(b)$: it follows that $(-\varepsilon,\varepsilon) \ni s \mapsto \ell(\Gamma_s)$ is differentiable[5] and it has a minimum at $s = 0$, whence $\frac{\mathrm{d}}{\mathrm{d}s}\ell(\Gamma_s)\big|_{s=0} = 0$.

If V is a proper vector field along γ, by Lemma 1.73, there exists a proper variation Γ such that V is the variation field of Γ. From what we have seen at the beginning of this proof and Proposition 1.74, it follows that the left hand side in (1.16) vanishes.

Let $\{a_0 = a < \ldots < a_k = b\}$ be a subdivision such that γ is smooth on any subinterval $[a_{i-1}, a_i]$: the first step is to show that $\nabla_{\dot\gamma}\dot\gamma = 0$ on any $[a_{i-1}, a_i]$. We

[5] Indeed, let $\{a_0 = a < \ldots < a_k = b\}$ be a subdivision such that Γ is smooth on any rectangle $(-\varepsilon,\varepsilon)\times[a_{i-1},a_i]$. Then $\ell(\Gamma_s) = \sum_{i=1}^m \int_{a_{i-1}}^{a_i} \|\partial_t\Gamma(s,t)\|_{\Gamma(s,t)}\,\mathrm{d}t$ and the map $s \mapsto \ell(\Gamma_s)$ is given by $s \mapsto \sum_{i=1}^m \int_{a_{i-1}}^{a_i} \|\partial_t\Gamma(s,t)\|_{\Gamma(s,t)}\,\mathrm{d}t$, that is differentiable.

consider a subinterval $[a_{i-1}, a_i]$ and a bump function $\varphi \in C^\infty(\mathbb{R})$ such that $\varphi > 0$ on (a_{i-1}, a_i) and $\varphi = 0$ elsewhere. If we choose $V = \varphi \nabla_{\dot\gamma}\dot\gamma$, it is a proper vector field along γ and (1.16) becomes

$$-\int_{a_{i-1}}^{a_i} \varphi(t)\|\nabla_{\dot\gamma}\dot\gamma(t)\|^2_{\gamma(t)}\,\mathrm{d}t = 0.$$

The integrand is nonnegative, therefore $\nabla_{\dot\gamma}\dot\gamma = 0$ on $[a_{i-1}, a_i]$.

Now, we need to show that $\triangle_i\dot\gamma := \dot\gamma(a_i^+) - \dot\gamma(a_i^-) = 0$ for all $i = 0, \ldots, k$. For $i = 0, \ldots, k$, by using a bump function, it is possible to construct a proper vector field V along γ such that $V(a_i) = \triangle_i\dot\gamma$ and $V(a_j) = 0$ for $j \neq i$[6]. Then (1.16) becomes

$$-\| \triangle_i \dot\gamma\|^2_{\gamma(a_i)} = 0,$$

meaning that $\dot\gamma(a_i^+) = \dot\gamma(a_i^-)$. We have that $\gamma|_{[a_{i-1},a_i]}$ and $\gamma|_{[a_i,a_{i+1}]}$ are two geodesics passing through the point $\gamma(a_i)$ and having the same tangent vector. It follows by Theorem 1.46 that the geodesic $\gamma|_{[a_i,a_{i+1}]}$ is the continuation of the geodesic $\gamma|_{[a_{i-1},a_i]}$, whence γ is smooth all over $[a, b]$ and it is a geodesic. $\square$

From this Theorem we can see a characterization of geodesics as "critical points" of the functional ℓ.

Definition 1.77. Let $\gamma : [a, b] \longrightarrow M$ be an admissible curve. We say that γ is a *critical point* of the length functional ℓ if for any proper variation Γ of γ we have that $\frac{\mathrm{d}}{\mathrm{d}s}\ell(\Gamma_s)\big|_{s=0} = 0$.

Theorem 1.78. *Let $\gamma : [a, b] \longrightarrow M$ be a unit speed admissible curve. Then γ is a geodesic if and only if it is a critical point of the length functional.*

Proof. If γ is a critical point of the length functional, then, by the same proof of Theorem 1.76, we can show that it is a geodesic.

Viceversa, assume that γ is a geodesic. Let Γ be a proper variation of γ, V its variation field and $\{a_0 = a < \ldots a_k = b\}$ a subdivision such that Γ is smooth on any

[6]Let $\mathrm{x} : U \subset \mathbb{R}^n \longrightarrow M$ be a system of coordinates about $\gamma(a_i)$, then $\mathrm{dx}_{\gamma(a_i)} : \mathbb{R}^n \longrightarrow T_{\gamma(a_i)}M$ is a diffeomorphism and we have an identification $\triangle_i\dot\gamma = (\triangle_i\dot\gamma^{(1)}, \ldots, \triangle_i\dot\gamma^{(n)})$. For any $j = 1, \ldots, n$, we define a positive bump function $\varphi_j \in C^\infty(\mathbb{R})$ such that $\varphi_j = \triangle_i\dot\gamma^{(j)}$ in an open neighbourhood of a_i and $\varphi_j(a_l) = 0$ for $l \neq i$. Setting $\varphi = (\varphi_1, \ldots, \varphi_n)$, then we can define $V(t) = \mathrm{dx}_{\gamma(t)} \circ \varphi(t)$.

rectangle $(-\varepsilon,\varepsilon)\times[a_{i-1},a_i]$. By (1.16), we have that

$$\frac{\mathrm{d}}{\mathrm{d}s}\ell_a^b(\Gamma_s)\bigg|_{s=0} = -\sum_{i=1}^{k}\int_{a_{i-1}}^{a_i}\langle V(t),\nabla_{\dot\gamma}\dot\gamma(t)\rangle_{\gamma(t)}\,\mathrm{d}t - \sum_{i=1}^{k-1}\langle V(a_i),\triangle_i\dot\gamma\rangle_{\gamma(a_i)} = 0,$$

because $\nabla_{\dot\gamma}\dot\gamma = 0$, being γ a geodesic, and $\triangle_i\dot\gamma = 0$, being γ smooth (so $\dot\gamma$ has no jumps). □

1.6 Geodesics are locally minimizing

Our aim in this section is to show that geodesics have important minimizing properties.

Lemma 1.79 (The Gauss Lemma). *Let $p\in M$ and $(p,v)\in\mathcal{E}$ (where $\mathcal{E}$ is the domain of* exp*). Then for any $w\in T_pM\equiv T_v(T_pM)$ we have*

$$\left\langle \mathrm{d}(\exp_p)_v(v),\mathrm{d}(\exp_p)_v(w)\right\rangle_{\exp_p(v)} = \langle v,w\rangle_p\,.$$

Proof. The proof is trivial if $v=0$ or $w=0$. Therefore we can assume $v\neq 0$ and $w\neq 0$. We consider the decomposition $w=w_T+w_N$, where $w_T\in\operatorname{span}(v):=\{\lambda v:\lambda\in\mathbb{R}\}$ and $w_N\in v^\perp:=\{z\in T_pM:\langle z,v\rangle_p=0\}$. Hence, by linearity, it is enough to prove the thesis for $w=v$ and $w\perp v$.

Let $w=v$ and set $\gamma_v(t):=\exp_p(tv)$. To compute $\mathrm{d}(\exp_p)_v(v)$, we consider the curve $\delta(t)=tv$ such that $\delta(1)=v$ and $\dot\delta(1)=v$; then we have

$$\mathrm{d}(\exp_p)_v(v) = \frac{\mathrm{d}}{\mathrm{d}t}(\exp_p(tv))\bigg|_{t=1} = \dot\gamma_v(1).$$

Since γ_v is a geodesic, we have $\|\dot\gamma_v(1)\|_{\gamma_v(1)} = \|\dot\gamma_v(0)\|_{\gamma_v(0)}$ and

$$\left\langle \mathrm{d}(\exp_p)_v(v),\mathrm{d}(\exp_p)_v(v)\right\rangle_{\exp_p(v)} = \langle\dot\gamma_v(1),\dot\gamma_v(1)\rangle_{\gamma_v(1)} = \langle\dot\gamma_v(0),\dot\gamma_v(0)\rangle_{\gamma_v(0)} = \langle v,v\rangle_p\,.$$

This complete the proof in the case $w=v$.

Now, let $w\perp v$. In this case, we can find a smooth curve γ_1 in the sphere $\partial B_{||v||_p}(0)\subset T_pM$ such that $\gamma_1(0)=v$ and $\dot\gamma_1(0)=w$. Since $(p,v)=(p,\gamma_1(0))\in\mathcal{E}$, which is open, and $\mathcal{E}_p$ is star-shaped with respect to 0, there exists $\varepsilon>0$ such that $(p,t\gamma_1(s))\in\mathcal{E}$ for all

$0 < t < 1$ and $-\varepsilon < s < \varepsilon$. Let $A := \{(t,s) : 0 < t < 1, -\varepsilon < s < \varepsilon\}$ and consider the smooth map $f : A \longrightarrow M$ given by $f(t,s) = \exp_p(t\gamma_1(s))$. We denote by f_t and f_s the velocity field of the curves $(0,1) \ni t \longmapsto f(t,s)$ and $(-\varepsilon,\varepsilon) \ni s \longmapsto f(t,s)$ respectively. We have

$$f_t(1,0) = \frac{\mathrm{d}}{\mathrm{d}t}(\exp_p(t\gamma_1(0)))\Big|_{t=1} = \frac{\mathrm{d}}{\mathrm{d}t}(\exp_p(tv))\Big|_{t=1} = \mathrm{d}(\exp_p)_v(v),$$

$$f_s(1,0) = \frac{\mathrm{d}}{\mathrm{d}s}(\exp_p(\gamma_1(s)))\Big|_{s=0} = \mathrm{d}(\exp_p)_v(w),$$

thus,

$$\big\langle \mathrm{d}(\exp_p)_v(v), \mathrm{d}(\exp_p)_v(w)\big\rangle_{\exp_p(v)} = \langle f_t(1,0), f_s(1,0)\rangle_{\exp_p(v)}.$$

Furthermore:

- $t \longmapsto f(t,s_0) = \exp_p(t\gamma_1(s_0))$ is the geodesic starting at p with initial velocity $\gamma_1(s_0)$ and $f_t(t,s_0)$ is its velocity field. Therefore $D_t f_t = 0$.
- Being the connection symmetric, $D_s f_t - D_t f_s = [f_s, f_t] = 0$[7], i.e. $D_s f_t = D_t f_s$.
- f_t is the velocity field of the geodesic $t \longmapsto \exp_p(t\gamma_1(s))$, whence $\|f_t(t,s)\|_{f(t,s)}$ is constant.

As a consequence of these facts:

$$\frac{\partial}{\partial t}\langle f_s(t,s), f_t(t,s)\rangle_{f(t,s)} = \langle D_t f_s(t,s), f_t(t,s)\rangle_{f(t,s)} + \Big\langle f_s(t,s), \underbrace{D_t f_t(t,s)}_{=0}\Big\rangle_{f(t,s)} =$$

$$= \langle D_s f_t(t,s), f_t(t,s)\rangle_{f(t,s)} = \frac{1}{2}\frac{\partial}{\partial s}\langle f_t(t,s), f_t(t,s)\rangle_{f(t,s)} = 0,$$

whence $\langle f_s(t,s), f_t(t,s)\rangle_{f(t,s)}$ does not depend on t and

$$\langle f_t(1,0), f_s(1,0)\rangle_{\exp_p(v)} = \langle f_s(t,0), f_t(t,0)\rangle_{\exp_p(tv)}, \quad \forall t \in (0,1). \tag{1.17}$$

Since

$$\lim_{h\to 0} f_s(h,0) = \lim_{h\to 0}\frac{\mathrm{d}}{\mathrm{d}s}(\exp_p(h\gamma_1(s)))\Big|_{s=0} = \lim_{h\to 0}\mathrm{d}(\exp_p)_{hv}(hw) = 0,$$

[7] If $g : M \longrightarrow \mathbb{R}$ is a smooth function, then the action of f_t on g at the point $f(t,s)$ is given by $f_t g(f(t,s)) = \frac{\partial}{\partial t}(g(f(t,s))$ and similarly $f_s g(f(t,s)) = \frac{\partial}{\partial s}(g(f(t,s))$, whence f_t and f_s do commute.

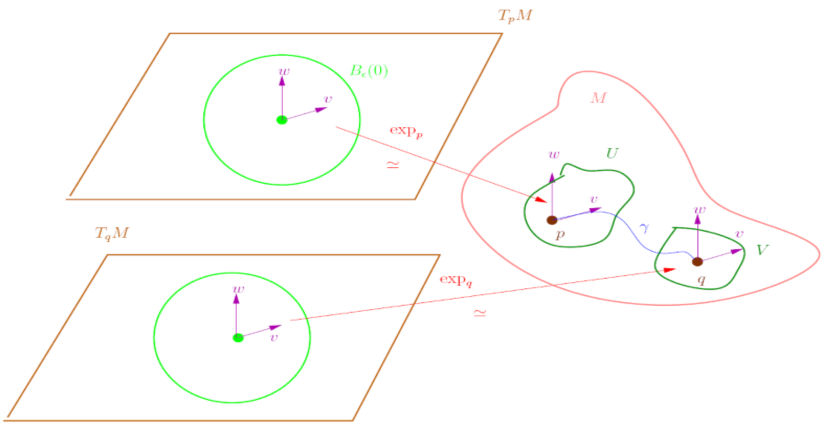

Figure 1.4: The exponential map as a radial isometry
See `https://en.wikipedia.org/wiki/Gauss%27s_lemma_(Riemannian_geometry)`.

taking the limit for $t \to 0$ in (1.17), we get

$$\left\langle \mathrm{d}(\exp_p)_v(v), \mathrm{d}(\exp_p)_v(w) \right\rangle_{\exp_p(v)} = \langle f_t(1,0), f_s(1,0) \rangle_{\exp_p(v)} =$$
$$= \lim_{t \to 0} \langle f_s(t,0), f_t(t,0) \rangle_{\exp_p(tv)} = 0 = \langle v, w \rangle_p .$$

□

The Gauss Lemma has the following geometric interpretation.

Corollary 1.80. *Let $p \in M$ and $\rho < \mathrm{inj}_p(M)$. For any $q \in \exp_p(\partial B_\rho(0))$ the radial geodesic from p to q is orthogonal to $\exp_p(\partial B_\rho(0))$.*

Proof. Let $v \in \partial B_\rho(0) \subset T_pM$ such that $q = \exp_p(v)$. The radial geodesic from p to q is $\gamma(t) := \exp_p(tv)$, where $t \in [0,1]$, and the tangent vector at q is given by $\dot\gamma(1) = \mathrm{d}(\exp_p)_v(v) \in T_{\exp_p(v)}M$.

Let $z \in T_q(\exp_p(\partial B_\rho(0)))$; we aim to show that $\left\langle z, \mathrm{d}(\exp_p)_v(v) \right\rangle_{\exp_p(v)} = 0$. To see this, we consider a smooth curve $\delta : (-\varepsilon, \varepsilon) \longrightarrow \exp_p(\partial B_\rho(0))$ such that $\delta(0) = q$ and $\dot\delta(0) = z$. The curve $\eta := \exp_p^{-1} \circ \delta : (-\varepsilon, \varepsilon) \longrightarrow \partial B_\rho(0)$ satisfies $\eta(0) = v$ and $\dot\eta(0) = \mathrm{d}(\exp_p)_q^{-1}(z)$; this means that $\mathrm{d}(\exp_p)_q^{-1}(z) \in T_v(\partial B_\rho(0))$, from which $\left\langle \mathrm{d}(\exp_p)_q^{-1}(z), v \right\rangle_p = 0$. Finally, applying Gauss Lemma, we get the conclusion:

$$0 = \left\langle \mathrm{d}(\exp_p)_q^{-1}(z), v \right\rangle_p = \left\langle z, \mathrm{d}(\exp_p)_v(v) \right\rangle_{\exp_p(v)} .$$

□

The following result is crucial in order to show later that geodesics minimizes the arc length, at least locally.

Theorem 1.81. *Let $p \in M$ and $\rho < \mathrm{inj}_p(M)$. Then for any $q \in \exp_p(B_\rho(0))$ the radial geodesic from p to q is the only admissible curve connecting p and q with length $d(p,q)$ (up to reparametrization).*

Proof. The radial geodesic from p to q is $[0,1] \ni t \longmapsto \gamma(t) := \exp_p(tv)$, where $v \in B_\rho(0)$ such that $q = \exp_p(v)$. Let $\sigma : [0,1] \longrightarrow M$ be an admissible curve such that $\sigma(0) = p$ and $\sigma(1) = q$. Assume that σ is parameterized with constant speed, whence $\ell_0^1(\sigma) = \|\dot\sigma(t)\|_{\sigma(t)}$ for all t. By definition, we have $\ell_0^1(\sigma) \geq d(p,q)$ and we aim to show that $\ell_0^1(\sigma) = d(p,q)$ if and only if $\sigma = \gamma$. Without loss of generality, we may assume that $p \notin \sigma((0,1])$[8].

First, we assume that $\sigma((0,1]) \subset \exp_p(B_\rho(0))$. In this case, for $t \in (0,1]$ the curve σ can be written uniquely as $\sigma(t) = \exp_p(r(t)w(t))$, where $(0,1] \ni t \longmapsto r(t) \in (0,\rho)$ is a piecewise smooth function and $(0,1] \ni t \longmapsto w(t) \in T_pM$ is a smooth curve such that $\|\dot w(t)\|_{w(t)} = 1$ for all $t \in (0,1]$. For all but a finite quantity of $t \in (0,1]$ we have

$$\dot\sigma(t) = \mathrm{d}(\exp_p)_{r(t)w(t)}(r'(t)w(t) + r(t)\dot w(t)).$$

We note that $\dot w(t) \in T_{w(t)}T_pM \equiv T_pM$ and, since $\|w(t)\|_p = 1$, it follows

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\langle w(t), w(t)\rangle_p = 2\langle \dot w(t), w(t)\rangle_p,$$

i.e. $w(t) \perp \dot w(t)$ for all $t \in (0,1]$. By linearity and Gauss Lemma, we get

$$\begin{aligned}
&\big\langle \mathrm{d}(\exp_p)_{r(t)w(t)}(r'(t)w(t)), \mathrm{d}(\exp_p)_{r(t)w(t)}(r(t)\dot w(t))\big\rangle_{\exp_p(r(t)w(t))} = \\
&\qquad = r'(t)\big\langle \mathrm{d}(\exp_p)_{r(t)w(t)}(r(t)w(t)), \mathrm{d}(\exp_p)_{r(t)w(t)}(\dot w(t))\big\rangle_{\exp_p(r(t)w(t))} = \\
&\qquad\qquad = r'(t)\langle r(t)w(t), \dot w(t)\rangle_p = r'(t)r(t)\langle w(t), \dot w(t)\rangle_p = 0,
\end{aligned}$$

i.e $\mathrm{d}(\exp_p)_{r(t)w(t)}(r'(t)w(t)) \perp \mathrm{d}(\exp_p)_{r(t)w(t)}(r(t)\dot w(t))$. Thus, again by linearity and Gauss

[8] If $p \in \sigma((0,1])$, set $t_0 := \sup\{t \,|\, \sigma(t) = p\}$, then $p \notin \sigma((t_0,1])$ and we can consider $\sigma|_{[t_0,1]}$.

Lemma, we obtain:

$$\begin{aligned}
\|\dot\sigma(t)\|^2_{\sigma(t)} &\geq \left\langle \mathrm{d}(\exp_p)_{r(t)w(t)}(r'(t)w(t)), \mathrm{d}(\exp_p)_{r(t)w(t)}(r'(t)w(t)) \right\rangle_{\exp_p(r(t)w(t))} = \\
&= \frac{|r'(t)|^2}{r(t)} \left\langle \mathrm{d}(\exp_p)_{r(t)w(t)}(r(t)w(t)), \mathrm{d}(\exp_p)_{r(t)w(t)}(w(t)) \right\rangle_{\exp_p(r(t)w(t))} = \\
&= |r'(t)|^2 \underbrace{\left\langle \mathrm{d}(\exp_p)_{r(t)w(t)}(w(t)), \mathrm{d}(\exp_p)_{r(t)w(t)}(w(t)) \right\rangle_{\exp_p(r(t)w(t))}}_{=\langle w(t), w(t)\rangle_p = 1} = |r'(t)|^2.
\end{aligned}$$

Note that, being $\exp_p$ continuous,

$$p = \lim_{t\to 0} \sigma(t) = \lim_{t\to 0} \exp_p(r(t)w(t)) = \exp_p(\lim_{t\to 0} r(t)w(t)),$$

whence, being $\exp_p$ injective around 0,

$$\lim_{t\to 0} r(t) = 0.$$

Finally, if $\{t_0 = 0 < \dots t_k = 1\}$ is a subdivision such that σ and r are smooth on $[t_{i-1}, t_i]$ for all $i = 1, \dots, k$, we have

$$\begin{aligned}
\ell_0^1(\sigma) = \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \|\dot\sigma(t)\|_{\sigma(t)} \,\mathrm{d}t \geq \sum_{i=1}^k \int_{t_{i-1}}^{t_i} |r'(t)| \,\mathrm{d}t \geq \\
\geq \sum_{i=1}^k \int_{t_{i-1}}^{t_i} r'(t) \,\mathrm{d}t = r(1) - \lim_{t\to 0} r(t) = r(1). \quad (1.18)
\end{aligned}$$

We note that $q = \sigma(1) = \exp_p(r(1)w(1))$, from which, being $\exp_p$ a diffeomorphism in $B_\rho(0)$, it has to be $r(1)w(1) = v$ and consequently $|r(1)| = \|v\|_p = \ell_0^1(\gamma)$. By (1.18), it follows that $\ell_0^1(\sigma) \geq \ell_0^1(\gamma)$. Furthermore, if $\ell_0^1(\sigma) = \ell_0^1(\gamma)$, then in (1.18) all inequalities are actually equalities and, in particular, $|r'(t)| = \|\dot\sigma(t)\|_{\sigma(t)} = \ell_0^1(\sigma)$ is constant and $\dot w = 0$. This implies that σ is precisely the geodesic γ.

If $\sigma((0,1]) \not\subset \exp_p(B_\rho(0))$, then let $t_1 \in (0,1]$ be defined by

$$t_1 := \inf\{t \in (0,1] \,|\, \sigma(t) \in \exp_p(\partial B_\rho(0))\}.$$

Let $z \in \partial B_\rho(0)$ such that $\sigma(t_1) = \exp_p(z)$ and consider $\tau(t) = \exp_p(tz)$ for $t \in [0,1]$. Then, by the previous step, we have the conclusion:

$$\ell_0^1(\sigma) \geq \ell_0^{t_1}(\sigma) \geq \ell_0^1(\tau) = \rho > \ell_0^1(\gamma).$$

□

Definition 1.82. A curve $\gamma : I \longrightarrow M$ is *locally minimizing* if for any $t_0 \in I$ there exists $U \subset I$ such that $\gamma|_U$ is minimizing between each pair of its points.

Theorem 1.83. *Every Riemannian geodesic is locally minimizing.*

Proof. Let $\gamma : I \longrightarrow M$ be a geodesic, where we may assume that I is open. Let $\mathcal{W}$ be a uniformly normal neighbourhood of $\gamma(t_0)$ (it exists by Lemma 1.66) and let $U \subset I$ be the connected component of $\gamma^{-1}(\mathcal{W})$ containing t_0. If $t_1, t_2 \in U$ and $q_i := \gamma(t_i)$, there exists a positive value δ such that $q_2 \in \exp_{q_1}(B_\delta(0))$. Let $v \in B_\delta(0)$ such that $q_2 = \exp_{q_1}(v)$; then $\sigma(t) = \exp_{q_1}(tv)$, $t \in [0,1]$, is the unique minimizing curve joining q_1 and q_2 (up to reparametrization) by Theorem 1.81. The restriction $\gamma|_{[t_1,t_2]}$ is a radial geodesic from q_1 to q_2 lying in the same geodesic ball: by remark 1.63, it has to coincide with σ (up to reparametrization). $\square$

Remark 1.84. Geodesics are only locally minimizing, not globally. For example, if we consider a sphere S in $\mathbb{R}^3$, a great circe γ on S and two points p and q belonging to γ such that the segment in $\mathbb{R}^3$ with extremes p and q is not a diameter, then there are two arcs of great circle joining p and q: both of them are geodesics but only one of them has a length equal to $d(p, q)$.

1.7 The energy functional

So far, the only one functional we have seen is the energy functional ℓ, defined on the class of admissible curves. We have seen a characterization of unit speed geodesics as critical points of the functional ℓ. Now, we introduce the energy functional E; in this section, we will see another characterization of geodesics as critical points of the energy functional. The reason why we introduce this second functional is that E has more properties of regularity than ℓ, thanks to which we will be able to apply the variational methods to show the existence of critical points, as we will see in the next chapters.

Definition 1.85. Let $\gamma : [a, b] \longrightarrow M$ be an admissible curve and let $\{a_0 = a < \ldots < a_k = b\}$ be a subdivision such that γ is smooth on $[a_{i-1}, a_i]$ for any $i = 1, \ldots, k$. We define the *energy* of γ as

$$E(\gamma) = \frac{1}{2}\sum_{i=1}^{k}\int_{a_{i-1}}^{a_i} \|\dot\gamma(t)\|^2_{\gamma(t)}\,\mathrm{d}t.$$

Remark 1.86. Unlike the length functional, the energy functional E is not invariant under reparametrization.

Lemma 1.87. *If $\gamma : [a,b] \longrightarrow M$ is an admissible curve, then*

$$2E(\gamma) \geq \frac{\ell(\gamma)^2}{b-a}.$$

Proof. If γ is a smooth curve, applying the Holder inequality,

$$\ell(\gamma) = \int_a^b \|\dot\gamma(t)\|_{\gamma(t)}\,\mathrm{d}t \leq \sqrt{b-a}\left(\int_a^b \|\dot\gamma(t)\|^2_{\gamma(t)}\,\mathrm{d}t\right)^{1/2} = \sqrt{2E(\gamma)(b-a)},$$

from which

$$2E(\gamma) \geq \frac{\ell(\gamma)^2}{b-a}.$$

In the general case, let $\{a_0 = a < \ldots < a_k = b\}$ be a subdivision such that γ is smooth on any $[a_{i-1}, a_i]$; then

$$\begin{aligned}\ell(\gamma) = \sum_{i=1}^{k}\ell(\gamma|_{[t_{i-1},t_i]}) \leq \sum_{i=1}^{k}\sqrt{2E(\gamma|_{[t_{i-1},t_i]})}\sqrt{t_i - t_{i-1}} \leq \\ \leq \left(\sum_{i=1}^{k} 2E(\gamma|_{[t_{i-1},t_i]})\right)^{1/2}\left(\sum_{i=1}^{k}(t_i - t_{i-1})\right)^{1/2} = \sqrt{2E(\gamma)}\sqrt{b-a},\end{aligned}$$

where we have applied the Cauchy-Schwartz inequality in $\mathbb{R}^k$ in the step between the first and the second line. □

Remark 1.88. If $\gamma : [a,b] \longrightarrow M$ is an admissible curve with $\|\dot\gamma(t)\|_{\gamma(t)}$ constant, then

$$2E(\gamma) = \frac{\ell(\gamma)^2}{b-a}.$$

It follows that, among the class of admissible curves parameterized with constant speed and joining two fixed points, a curve γ is a minimizer for the length functional if and only if it is a minimizer for the energy functional.

Corollary 1.89. *If $\gamma : [a,b] \longrightarrow M$ is an admissible curve such that $E(\gamma) \leq E(c)$ for any unit speed admissible curve c joining $\gamma(a)$ and $\gamma(b)$, then γ is a geodesic.*

Proof. By the previous remark, we know that $\ell(\gamma) \leq \ell(c)$ for any unit speed admissible curve c joining $\gamma(a)$ and $\gamma(b)$. If δ is an admissible curve joining $\gamma(a)$ and $\gamma(b)$, there exists a reparametrization c of δ such that $\|\dot{c}\|_{c(t)} = 1$ for all t. Therefore $\ell(\gamma) \leq \ell(c) = \ell(\delta)$ and we can apply Theorem 1.76. $\square$

Proposition 1.90 (First variation formula for E). *Let $\gamma : [a,b] \longrightarrow M$ be an admissible curve, $\Gamma : (-\varepsilon, \varepsilon) \times [a,b] \longrightarrow M$ a proper variation of γ and V its variation field. Let $\{a_0 = a < \ldots < a_k = b\}$ be a subdivision such that Γ is smooth on any rectangle of the form $(-\varepsilon,\varepsilon) \times [a_{i-1}, a_i]$. Then*

$$\frac{\mathrm{d}}{\mathrm{d}s} E(\Gamma_s)\bigg|_{s=0} = -\sum_{i=1}^{k-1} \langle V(a_i), \triangle_i \dot{\gamma}\rangle_{\gamma(a_i)} - \sum_{i=1}^{k} \int_{a_{i-1}}^{a_i} \langle V(t), \nabla_{\dot{\gamma}} \dot{\gamma}(t)\rangle_{\gamma(t)} \,\mathrm{d}t. \tag{1.19}$$

Proof. We have the following computation:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s} E(\Gamma_s) &= \frac{1}{2} \sum_{i=1}^{k} \frac{\mathrm{d}}{\mathrm{d}s} \int_{a_{i-1}}^{a_i} \langle \partial_t \Gamma_s(t), \partial_t \Gamma_s(t)\rangle_{\Gamma_s(t)} \,\mathrm{d}t = \\
= \frac{1}{2} \sum_{i=1}^{k} \int_{a_{i-1}}^{a_i} \frac{\partial}{\partial s} \langle \partial_t \Gamma_s(t), \partial_t \Gamma_s(t)\rangle_{\Gamma_s(t)} \,\mathrm{d}t &= \sum_{i=1}^{k} \int_{a_{i-1}}^{a_i} \langle D_s \partial_t \Gamma_s(t), \partial_t \Gamma_s(t)\rangle_{\Gamma_s(t)} \,\mathrm{d}t = \\
&= \sum_{i=1}^{k} \int_{a_{i-1}}^{a_i} \langle D_t \partial_s \Gamma_s(t), \partial_t \Gamma_s(t)\rangle_{\Gamma_s(t)} \,\mathrm{d}t = \\
= \sum_{i=1}^{k} \left(\int_{a_{i-1}}^{a_i} \frac{\partial}{\partial t} \langle \partial_s \Gamma_s(t), \partial_t \Gamma_s(t)\rangle_{\Gamma_s(t)} \,\mathrm{d}t \right. &- \left. \int_{a_{i-1}}^{a_i} \langle \partial_s \Gamma_s(t), D_t \partial_t \Gamma_s(t)\rangle_{\Gamma_s(t)} \,\mathrm{d}t \right) = \\
= \sum_{i=1}^{k} \Big(\big\langle \partial_s \Gamma_s(a_i^-), \partial_t \Gamma_s(a_i^-)\big\rangle_{\Gamma_s(a_i)} &- \big\langle \partial_s \Gamma_s(a_{i-1}^+), \partial_t \Gamma_s(a_{i-1}^+)\big\rangle_{\Gamma_s(a_{i-1})} \Big) + \\
&- \sum_{i=1}^{k} \int_{a_{i-1}}^{a_i} \langle \partial_s \Gamma_s(t), D_t \partial_t \Gamma_s(t)\rangle_{\Gamma_s(t)} \,\mathrm{d}t,
\end{aligned}$$

where we have applied Lemma 1.70 in the third line. Therefore for $s = 0$, being $V(a_0) = V(a_k) = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}s} E(\Gamma_s)\bigg|_{s=0} = -\sum_{i=1}^{k-1} \big\langle V(a_i), \triangle_i \dot{\gamma}\big\rangle_{\gamma(a_i)} - \sum_{i=1}^{k} \int_{a_{i-1}}^{a_i} \big\langle V(t), \nabla_{\dot{\gamma}} \dot{\gamma}(t)\big\rangle_{\gamma(t)} \,\mathrm{d}t.$$

□

Remark 1.91. The first variation formula for the energy functional E coincides with that one for the length functional.

Therefore we have an analogous version of Theorem 1.76 for the energy functional.

Theorem 1.92. *Let $\gamma : [a, b] \longrightarrow M$ be an admissible curve and assume that $E(\gamma) \leq E(c)$ for any admissible curve c joining $\gamma(a)$ and $\gamma(b)$. Then γ is a geodesic.*

Proof. The proof is exactly the same as Theorem 1.76. □

Definition 1.93. We say that an admissible curve γ is a *critical point* for the functional E if $\frac{\mathrm{d}}{\mathrm{d}s} E(\Gamma_s)\big|_{s=0} = 0$ for any proper variation Γ of γ.

We also have an analogous version of Theorem 1.78.

Theorem 1.94. *Let $\gamma : [a, b] \longrightarrow M$ be an admissible curve. Then γ is a geodesic if and only if it is a critical point for the energy functional E.*

Proof. The proof is exactly the same as Theorem 1.78. □

Remark 1.95. The energy functional is strictly convex: given two points $p, q \in M$, if a minimizer exists in the class of admissible curve joining p and q, then it is unique.

In the next chapters, we will introduce some variational methods and then use them to show the existence of critical points of the functional E (hence geodesics).

Chapter 2

A topological variational method: the Minimax Principle

In this chapter, we will introduce the differential calculus on Banach spaces, including the Fréchet differential and Gâteaux derivatives. We will also see the important notions of Palais-Smale sequence and Palais-Smale condition for a function of class C^1 defined on a Banach space. Finally, we will see how the Palais-Smale condition yields to the Deformation Lemma, whose one of the most important consequence is the Minimax Principle.

2.1 The differential calculus on a Banach space

Remark 2.1. We recall that if X and Y are two Banach spaces, then $L(X,Y) := \{f : X \longrightarrow Y \mid f \text{ is linear and continuous}\}$ is a Banach space too, with the operator norm given by

$$\|f\|_{L(X,Y)} := \sup_{\substack{x\in X\\ x\neq 0}} \frac{\|f(x)\|_Y}{\|x\|_X} = \sup_{\|x\|_X=1} \|f(x)\|_Y, \quad f \in L(X,Y).$$

Definition 2.2. Let X and Y be Banach spaces and $f : X \longrightarrow Y$. We say that f is *Fréchet differentiable* at $x_0 \in X$ if there exists $F \in L(X,Y)$ such that

$$\lim_{h\to 0} \frac{f(x_0+h) - f(x_0) - F(h)}{\|h\|_X} = 0. \tag{2.1}$$

In this case, we set $\mathrm{d}_{x_0}f := F$ and we say that F is the *Fréchet differential* of f at x_0. In particular, we say that f is *Fréchet differentiable* if it is Fréchet differentiable at x_0 for any $x_0 \in X$. Finally, if f is Fréchet differentiable and the map $\mathrm{d}_{(\cdot)}f : X \longrightarrow L(X,Y)$ is continuous, we say that $f \in C^1(X;Y)$.

Remark 2.3. Let $f : X \longrightarrow Y$ and $x_0 \in X$ such that there exists $F \in L(X,Y)$ satisfying (2.1). Then F is uniquely determined (hence $\mathrm{d}_{x_0}f$ is well defined). Indeed, for any $v \in X$, we have

$$0 = \lim_{t\to 0} \frac{f(x_0+tv) - f(x_0) - F(tv)}{t\|v\|} = \lim_{t\to 0} \frac{f(x_0+tv) - f(x_0)}{t\|v\|} - \frac{F(v)}{\|v\|},$$

from which

$$F(v) = \lim_{t\to 0} \frac{f(x_0+tv) - f(x_0)}{t}.$$

Remark 2.4. Let X and Y be Banach spaces and $f : X \longrightarrow Y$ a Fréchet differentiable function. If $\mathrm{d}_{(\cdot)}f : X \longrightarrow L(X,Y)$ is Fréchet differentiable at $x \in X$, we can consider its Fréchet differential at x and denote it by $\mathrm{d}^2_x f \in L(X, L(X,Y))$: this is the *second order Fréchet differential* of f at x. If $\mathrm{d}_{(\cdot)}f : X \longrightarrow L(X,Y)$ is Fréchet differentiable and $\mathrm{d}^2_{(\cdot)}f : X \longrightarrow L(X, L(X,Y))$ is continuous, we say that $f \in C^2(X;Y)$. Inductively, we can define differentiable functions of any orders and function of class C^k for any $k \geq 1$.

Definition 2.5. Let X and Y be Banach spaces. A function $f : X \longrightarrow Y$ is said to be *Gâteaux differentiable* at $x_0 \in X$ along $h \in X$ if there exists the following limit:

$$\lim_{t\to 0} \frac{f(x_0+th) - f(x_0)}{t} \in Y.$$

In this case, the value of the previous limit is called the *Gâteaux derivative* of f at x_0 along h and we denote it by $\partial_h f(x)$.

Remark 2.6. Let X and Y be Banach spaces and $f : X \longrightarrow Y$ a Fréchet differentiable function at $x_0 \in X$. Then there exists the Gâteaux derivative of f at x_0 along h for any $h \in X$ and $\partial_h f(x_0) = \mathrm{d}_{x_0}f(h)$.

The following proposition gives a sufficient condition for a function to be Fréchet differentiable.

Proposition 2.7. *Let $f : X \longrightarrow Y$ be a Gâteaux differentiable function at x along h for any $x, h \in X$. If $\partial_{(\cdot)} f(x) \in L(X, Y)$ for any $x \in X$ and $X \ni x \longmapsto \partial_{(\cdot)} f(x) \in L(X, Y)$ is continuous, then f is Fréchet differentiable and $\mathrm{d}_x f(h) = \partial_h f(x)$ for any $x, h \in X$. In particular, $f \in C^1(X; Y)$.*

For a proof of Proposition 2.7 see for instance Proposition 4.1.7 on [3].

Lemma 2.8. *Let X, Y, Z be Banach spaces, $U \subset X$ open and $V \subset Y$ open. Consider two functions $f : U \longrightarrow Y$ and $g : V \longrightarrow Z$ such that $f(U) \subset V$. If f is Fréchet differentiable at $x \in U$ and g is Fréchet differentiable at $f(x) \in V$, then $g \circ f$ is Fréchet differentiable at x and $\mathrm{d}_x(g \circ f) = (\mathrm{d}_{f(x)} g) \circ (\mathrm{d}_x f)$.*

For a proof of Lemma 2.8 see for instance Proposition 4.1.12 on [3].

Definition 2.9. Let X be a Banach space, $I \subset \mathbb{R}$ an open interval and $\varphi : I \longrightarrow X$. If there exists

$$\lim_{h \to 0} \frac{\varphi(t_0 + h) - \varphi(t_0)}{h} \in X,$$

for $t_0 \in I$, we say that the value of this limit is the *first derivative* of φ at t_0 and we denote it by $\varphi'(t_0)$. Similarly, if $I \subset \mathbb{R}$ is an open interval, $\Omega \subset X$ is open, $\eta : I \times \Omega \longrightarrow X$ and there exists

$$\lim_{h \to 0} \frac{\eta(t_0 + h, x_0) - \eta(t_0, x_0)}{h} \in X,$$

for $(t_0, x_0) \in I \times \Omega$, we say that the value of this limit is the *first partial derivative* of η with respect to the variable t at the point (t_0, x_0) and we denote it by $\frac{\partial \eta}{\partial t}(t_0, x_0)$.

The following two theorems are classical results from Analysis.

Theorem 2.10 (Local existence and uniqueness). *Let X be a Banach space, $\Omega \subset X$ open, $I \subset \mathbb{R}$ an open interval, $t_0 \in I$, $x_0 \in \Omega$ and $f : I \times \Omega \longrightarrow X$ a function that is continuous and locally Lipschitz in $x \in \Omega$, uniformly with respect to $t \in I$. Consider the Cauchy problem*

$$\begin{cases} x'(t) = f(t, x(t)), \\ x(t_0) = x_0. \end{cases}$$

Then there exists $\delta > 0$ such that the Cauchy problem has a unique solution in the open interval $(t_0 - \delta, t_0 + \delta)$.

Notation 2.11. In the same hypotheses of Theorem 2.10, we denote by $x(\cdot;x_0)$ the maximal solution of the Cauchy problem

$$\begin{cases} x'(t) = f(t,x(t)), \\ x(t_0) = x_0. \end{cases}$$

Theorem 2.12 (**Continuous dependence on initial data**). *Let X be a Banach space, $\Omega \subset X$ open, $I \subset \mathbb{R}$ an open interval and $f : I \times \Omega \longrightarrow X$ a function that is continuous and locally Lipschitz in $x \in \Omega$, uniformly with respect to $t \in I$. For any $K \subset I \times \Omega$ compact there exist $L, \delta > 0$ (depending only on K and f) such that*

$$\|x(t;x_0) - x(t;\tilde{x}_0)\| \leq e^{L|t-t_0|}\|x_0 - \tilde{x}_0\|, \quad \forall (t_0,x_0),(t_0,\tilde{x}_0) \in K, \, \forall t \in (t_0 - \delta, t_0 + \delta).$$

Notation 2.13. We fix the following notation for this chapter:

- $\mathbb{H}$ will be a separable Hilbert space with inner product $\langle \, , \, \rangle$;
- $\mathbb{B}$ will be a separable Banach space with norm $\|.\|$.

Remark 2.14. $L(\mathbb{B},\mathbb{R})$ is nothing less than the topological dual space of $\mathbb{B}$, which is usually denoted by $\mathbb{B}^*$. It is not true, in general, that $L(\mathbb{B},\mathbb{R})$ is separable: for example $L^1(\mathbb{R}^n)$ is separable, but its dual space $L^\infty(\mathbb{R}^n)$ is not separable. However, if $\mathbb{B}$ is a reflexive space (i.e $(\mathbb{B}^*)^* = \mathbb{B}$), then $L(\mathbb{B},\mathbb{R})$ is a separable space.

Definition 2.15. Let $f : \mathbb{B} \longrightarrow \mathbb{R}$ be Fréchet differentiable. We say that $x_0 \in \mathbb{B}$ is a *critical point* for f if $\mathrm{d}_{x_0}f = 0$, i.e $\mathrm{d}_{x_0}f(h) = 0$ for any $h \in \mathbb{B}$. In this case, we say that $f(x_0) \in \mathbb{R}$ is a *critical value* for f. If $c \in \mathbb{R}$ is not a critical value for f, we say that it is a *regular value.*

Notation 2.16. Let $f : \mathbb{B} \longrightarrow \mathbb{R}$ be a Fréchet differentiable function. We fix the following notation:

- $\mathrm{Crit}(f) = \{x \in \mathbb{B} \,|\, \mathrm{d}_x f = 0\}$, that is the set of critical points for f;
- $\mathrm{Crit}^c(f) = \{x \in \mathbb{B} \,|\, \mathrm{d}_x f = 0, \, f(x) = c\}$, that is the set of critical points for f at level c;

- $\mathrm{Vcrit}(f) = f\big(\mathrm{Crit}(f)\big)$, that is the set of critical values for f.

Definition 2.17. Let $f : \mathbb{H} \longrightarrow \mathbb{R}$ be a Fréchet differentiable function. Applying Riesz Representation theorem, for any $x \in \mathbb{H}$ there exists a unique $v \in \mathbb{H}$ such that $\mathrm{d}_x f(h) = \langle v, h \rangle$ for any $h \in \mathbb{H}$. We say that v is the *gradient* of f at x and we denote it by $\nabla f(x)$.

Remark 2.18. In the situation of the previous proposition, we have $\|\nabla f(x)\| = \|\mathrm{d}_x f\|_*$.

2.2 The Palais-Smale condition

This section is focused on the notion of Palais-Smale sequence and Palais-Smale condition for a function $f \in C^1(\mathbb{B};\mathbb{R})$. We will see that this last one is a compactness condition which will allow us to perform some kind of "deformation".

Definition 2.19. Let $f \in C^1(\mathbb{B};\mathbb{R})$. We say that a sequence $\{x_k\}_{k\in\mathbb{N}}$ is a **Palais-Smale** sequence for f at level $c \in \mathbb{R}$ if

$$\begin{cases} f(x_k) \overset{k\to\infty}{\longrightarrow} c, & \text{in } \mathbb{R}, \\ \mathrm{d}_{x_k} f \overset{k\to\infty}{\longrightarrow} 0, & \text{in } \mathbb{B}^*. \end{cases} \tag{2.2}$$

Moreover, we say that a function f satisfies the *Palais-Smale condition* at level c if any Palais-Smale sequence for f at level c has a converging subsequence.

Notation 2.20. If a function $f \in C^1(\mathbb{B};\mathbb{R})$ satisfies the Palais-Smale condition at level c, we will write that f satisfies $(PS)_c$. If a function $f \in C^1(\mathbb{B};\mathbb{R})$ satisfies the Palais-Smale condition at level c for any $c \in \mathbb{R}$, we will write that f satisfies (PS).

Remark 2.21. If $f \in C^1(\mathbb{H};\mathbb{R})$, the system (2.2) is equivalent to

$$\begin{cases} f(x_k) \overset{k\to\infty}{\longrightarrow} c, & \text{in } \mathbb{R}, \\ \nabla f(x_k) \overset{k\to\infty}{\longrightarrow} 0, & \text{in } \mathbb{H}. \end{cases}$$

Remark 2.22. The Palais-Smale condition is somehow a compactness condition. Indeed, if $f \in C^1(\mathbb{B};\mathbb{R})$ satisfies the Palais-Smale condition at level $c \in \mathbb{R}$, then $\mathrm{Crit}^c(f)$ is a compact set.

Let us see some examples:

Example 2.23. Consider the function $f(x) = \cos(x)$ defined on the real line.

We have that f satisfies the Palais-Smale condition at level c for any $c \in \mathbb{R} \setminus \{-1, 1\}$. Indeed, for any $c \in \mathbb{R} \setminus \{-1, 1\}$ there are no Palais-Smale sequences for f at level c, whence $(PS)_c$ is trivially satisfied.

Let $c = 1$. In this case, the sequence $\{2k\pi\}_{k\in\mathbb{N}}$ is a Palais-Smale sequence for f at level 1, but it is a diverging sequence and then it has no converging subsequences. Thus, the condition $(PS)_1$ is not satisfied. Similarly, we can see that $(PS)_{-1}$ is not satisfied too.

Example 2.24. Consider the function $f(x) = e^{-x}$ defined on the real line.

Let $c = 0$. In this case, the sequence $\{k\}_{k\in\mathbb{N}}$ is a Palais-Smale sequence for f, but it has no converging subsequences. Thus, $(PS)_0$ is not satisfied.

As in the previous example, for any $c \in \mathbb{R} \setminus 0$ there are no Palais-Smale sequences for f at level c, therefore $(PS)_c$ is satisfied.

Example 2.25. Let $f \in C^1(\mathbb{R}^n; \mathbb{R})$ be a coercive function (i.e. $\lim_{\|x\|\to+\infty} f(x) = +\infty$). Then f satisfies (PS). Indeed, if $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ is such that

$$\begin{cases} f(x_k) \overset{k\to\infty}{\longrightarrow} c, \\ \nabla f(x_k) \overset{k\to\infty}{\longrightarrow} 0, \end{cases}$$

then $\{x_k\}_{k\in\mathbb{N}}$ has to be a bounded sequence (from which it is possibile to extract a converging subsequence by Bolzano-Weierstrass theorem). To see that $\{x_k\}_{k\in\mathbb{N}}$ is bounded: if it was not, then there would be a subsequence $\{x_{k_j}\}_{j\in\mathbb{N}}$ such that $\|x_{k_j}\| \overset{j\to\infty}{\longrightarrow} +\infty$ and, by coercivity, $f(x_{k_j}) \overset{j\to\infty}{\longrightarrow} +\infty$, which is not possible.

However, it is not true, in general, that a coercive function defined on an infinite-dimensional Banach space satisfies the Palais-Smale condition.

Example 2.26. Let $g : \mathbb{R} \longrightarrow \mathbb{R}$:

$$g(t) := \begin{cases} 0, & t \in [-2, 2], \\ (|t| - 2)^2, & t \notin [-2, 2], \end{cases}$$

and define $f : \mathbb{B} \longrightarrow \mathbb{R}$ by $f(x) := g(\|x\|)$, where $\dim(\mathbb{B}) = +\infty$. We have that $\lim_{\|x\|\to+\infty} f(x) = +\infty$, i.e. f is a coercive function. If we consider the unit sphere in $\mathbb{B}$

$\{x \in \mathbb{B}\,;\, \|x\| = 1\}$, then it is a subset of $\mathrm{Crit}^0(f)$, but it is not compact[1]. Thus, f does not satisfy $(PS)_0$.

Lemma 2.27. *Let $f \in C^1(\mathbb{B};\mathbb{R})$. If f satisfies $(PS)_c$ and $c \notin \mathrm{Vcrit}(f)$, then there exists $r > 0$ such that $\|\mathrm{d}_x f\|_* \geq r$ for any $x \in f^{-1}([c-r, c+r])$.*

Proof. By contradiction, we assume that, fixed an arbitrary $r > 0$, there exists a value $x \in f^{-1}([c-r, c+r])$ such that $\|\mathrm{d}_x f\|_* \leq r$. In particular, for any $n \in \mathbb{N}$ there exists $x_n \in f^{-1}([c - \frac{1}{n}, c + \frac{1}{n}])$ such that $\|\mathrm{d}_{x_n} f\|_* \leq \frac{1}{n}$. We have

$$\begin{cases} f(x_n) \overset{n\to\infty}{\longrightarrow} c, \\ \mathrm{d}_{x_n} f \overset{n\to\infty}{\longrightarrow} 0, \end{cases}$$

meaning that $\{x_n\}_{n\in\mathbb{N}}$ is a Palais-Smale sequence at level c. Thus, by $(PS)_c$, there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ converging to some $x_0 \in \mathbb{B}$. It follows that

$$\begin{cases} f(x_{n_k}) \overset{n\to\infty}{\longrightarrow} f(x_0) = c, \\ \mathrm{d}_{x_{n_k}} f \overset{n\to\infty}{\longrightarrow} 0, \end{cases}$$

and, in particular, $c \in \mathrm{Vcrit}(f)$, against the hypothesis. $\square$

Example 2.28. Consider the function $f(x) = e^{-x}$ defined on the real line. In this case, $0 \notin \mathrm{Vcrit}(f)$, but, since $f'(x) \overset{x\to+\infty}{\longrightarrow} 0$, the thesis of Lemma 2.27 does not hold: in fact f does not satisfy $(PS)_0$.

We recall the definition of compact map between Banach spaces.

Definition 2.29. Let $K : \mathbb{B}_1 \longrightarrow \mathbb{B}_2$ be a continuous map between Banach spaces. We say that K is a *compact map* if one of the following equivalent conditions is satisfied:

- for any bounded subset V of $\mathbb{B}_1$, then $K(V)$ is relatively compact in $\mathbb{B}_2$ (i.e. the closure of $K(V)$ in $\mathbb{B}_2$ is compact),
- for any bounded sequence $\{x_k\}_{k\in\mathbb{N}}$ in $\mathbb{B}_1$, then $\{K(x_k)\}_{k\in\mathbb{N}}$ admits a converging subsequence.

[1] As a consequence of Riesz Lemma, if $\mathbb{B}$ is a Banach space, then the unit sphere in $\mathbb{B}$ is compact if and only if $\dim\mathbb{B} < \infty$.

Remark 2.30. The composition of a continuous and a compact map is a compact map.

Example 2.31. A continuous map $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is always compact.

Lemma 2.32. *Let $f \in C^1(\mathbb{B};\mathbb{R})$ and suppose that $\mathrm{d}_x f = L(x) + K(x)$ for any $x \in \mathbb{B}$, where $L : \mathbb{B} \longrightarrow \mathbb{B}^*$ is a linear, continuous and invertible operator, and $K : \mathbb{B} \longrightarrow \mathbb{B}^*$ is a compact map. Then any bounded Palais-Smale sequence for f admits a converging subsequence.*

Proof. Let $\{x_k\}_{k\in\mathbb{N}}$ be a bounded Palais-Smale sequence for f at level c; this means that

$$\begin{cases} f(x_k) \overset{k\to\infty}{\longrightarrow} c, \\ \mathrm{d}_{x_k} f \overset{k\to\infty}{\longrightarrow} 0. \end{cases}$$

We have that $\mathrm{d}_{x_k} f := L(x_k) + K(x_k)$. Then $x_k = -L^{-1}(K(x_k)) + L^{-1}(\mathrm{d}_{x_k} f)$. Since $L^{-1} \circ K$ is a compact map, there exists a converging subsequence $\{L^{-1}(K(x_{k_j}))\}_{j\in\mathbb{N}}$ to some $x_0 \in \mathbb{B}$. In particular, $x_{k_j} = -L^{-1}(K(x_{k_j})) + L^{-1}(\mathrm{d}_{x_{k_j}} f)$ converges to $-x_0 \in \mathbb{B}$ (note that L^{-1} is both linear and continuous, being the inverse of a linear and continuous map, hence $L^{-1}\left(\mathrm{d}_{x_{k_j}} f\right) \overset{j\to\infty}{\longrightarrow} 0$ in $\mathbb{B}$). □

Corollary 2.33. *Let $f \in C^1(\mathbb{H};\mathbb{R})$ and suppose that $\nabla f(x) = x + K(x)$ for any $x \in \mathbb{H}$, where $K : \mathbb{H} \longrightarrow \mathbb{H}$ is a compact map. Then any bounded Palais-Smale sequence for f admits a converging subsequence.*

Proof. For any $x, h \in \mathbb{H}$ we have that $d_x f(h) = \langle \nabla f(x), h\rangle = \big\langle x, h\big\rangle + \big\langle K(x), h\big\rangle$. Let $L : \mathbb{H} \longrightarrow \mathbb{H}^*$ be defined by $L(x) := \langle x, \cdot\rangle$ and $\tilde{K} : \mathbb{H} \longrightarrow \mathbb{H}^*$ be defined by $\tilde{K}(x) := \langle K(x), \cdot\rangle$; then $\mathrm{d}_x f = L(x) + \tilde{K}(x)$ for any $x \in \mathbb{H}$.

L is linear, it is continuous ($\|L(x)\|_* \le \|x\|$ for any $x \in \mathbb{H}$) and it is a bijective map by Riesz representation theorem.

$\tilde{K}$ is a compact map. Indeed, if $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{H}$ is a bounded sequence, then $\{K(x_k)\}_{k\in\mathbb{N}}$ admits a converging subsequence $\{K(x_{k_j})\}_{j\in\mathbb{N}}$ to some $v \in \mathbb{H}$. Let $\varphi \in \mathbb{H}^*$ be defined by $\varphi(\cdot) = \big\langle v, \cdot\big\rangle$; we have

$$\|\tilde{K}(x_{k_j}) - \varphi\|_* = \sup_{\|h\|_{\mathbb{H}}=1} \big\langle K(x_{k_j}) - v, h\big\rangle \le \|K(x_{k_j}) - v\|_{\mathbb{H}} \overset{j\to\infty}{\longrightarrow} 0,$$

hence $\tilde{K}(x_{k_j})$ converges to φ in $\mathbb{H}^*$. Applying Lemma 2.32, we get the conclusion. □

2.3 The Deformation Lemma

The main object of this section is the "Deformation Lemma". First, we will see an easier version and then we will generalize it.

Definition 2.34. Let X be a topological space. We say that φ is a *deformation* in X if it is homotopic to the identity map in X, i.e. there exists $\eta \in C([0,1] \times X; X)$ such that:

1. $\eta(0,x) = x$, $\forall x \in X$;

2. $\eta(1,x) = \varphi(x)$.

Definition 2.35. If X is a topological space and $f : X \longrightarrow \mathbb{R}$, for any $c \in \mathbb{R}$ we define the *sublevel set* for f relating to c:

$$f^c := \{x \in X \,|\, f(x) \leq c\}.$$

When c varies, the number of the connected components and the topology of the sublevel sets can vary too. Let's see some examples.

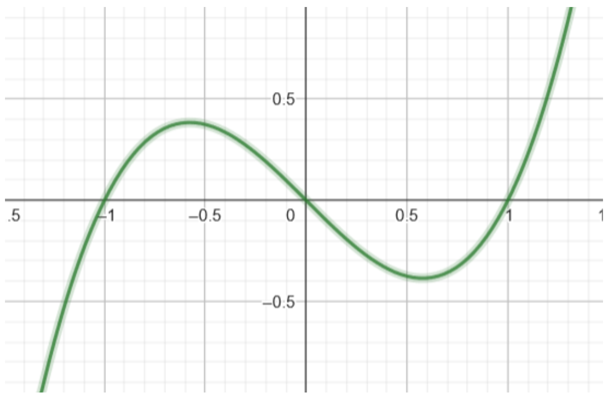

Figure 2.1: Graph of the function $f(x) = x^3 - x$

Example 2.36. Consider the function $f(x) = x^3 - x$: it has a minimum at $x_1 = \frac{\sqrt{3}}{3}$, with $f(x_1) = -\frac{2\sqrt{3}}{9}$, and a maximum at $x_2 = -\frac{\sqrt{3}}{3}$, with $f(x_2) = \frac{2\sqrt{3}}{9}$. We have that:

- $f^c = (-\infty, z_0]$, where z_0 satisfies $f(z_0) = c$, for $c < \frac{-2\sqrt{3}}{9}$;

- $f^c = (-\infty, z_0] \cup \{-\frac{\sqrt{3}}{3}\}$, where z_0 satisfies $f(z_0) = c$ and $z_0 \neq x_2$, for $c = \frac{-2\sqrt{3}}{9}$;

- $f^c = (-\infty, z_0] \cup [z_1, z_2]$, where z_i satisfies $f(z_i) = c$ and $z_0 < z_1 < z_2$, for $\frac{-2\sqrt{3}}{9} < c < \frac{2\sqrt{3}}{9}$;

- $f^c = (-\infty, z_0]$, where z_0 satisfies $f(z_0) = c$ and $z_0 \neq x_1$, for $c = \frac{2\sqrt{3}}{9}$;

- $f^c = (-\infty, z_0]$, where z_0 satisfies $f(z_0) = c$, for $c > \frac{2\sqrt{3}}{9}$.

In this case, if a variation in the number of the connected components of the sublevel sets of f occurs, this means that we have crossed a critical value for f. We also note that f satisfies (PS) (If $\{x_n\}_{n\in\mathbb{N}}$ is a Palais-Smale sequence for f at level $c \in \mathbb{R}$, then $f'(x_n) \overset{n\to\infty}{\longrightarrow} 0$, i.e. $|x_n| \overset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{3}}$, from which $\{x_n\}_{n\in\mathbb{N}}$ is bounded and it contains a converging subsequence by Bolzano-Weierstrass theorem).

Example 2.37. Consider the function $f(x, y) = x^2 - y^2$ defined on $\mathbb{R}^2$: its unique critical point is $(x_0, y_0) = (0, 0)$, with $f(x_0, y_0) = 0$.

- If $c > 0$, then the sublevel set $f^c = \{(x, y) \in \mathbb{R}^2 : x^2 - y^2 \leq c\}$ has a unique connected component: it is the region of $\mathbb{R}^2$ in between the two branches of the hyperbola $x^2 - y^2 = c$.

- If $c = 0$, then the sublevel set $f^c = \{(x, y) \in \mathbb{R}^2 : x^2 - y^2 \leq 0\}$ has a unique connected component: it is the region of $\mathbb{R}^2$ in between the two straight lines $y = x$ and $y = -x$, containing the y-axes.

- If $c < 0$, then the sublevel set $f^c = \{(x, y) \in \mathbb{R}^2 : x^2 - y^2 \leq c\}$ has two connected components: they are the region of $\mathbb{R}^2$ delimited from below by the upper branch of the hyperbola $x^2 - y^2 = c$ and the region delimited from above by the lower branch of the same hyperbola.

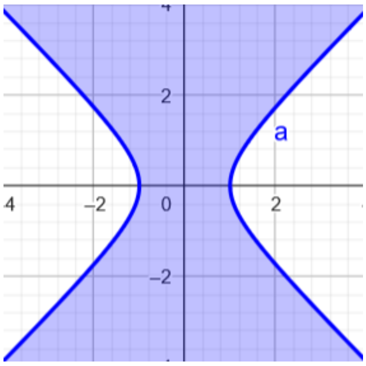

Figure 2.2: Sublevel set for $f(x, y) = x^2 - y^2$ with $c > 0$

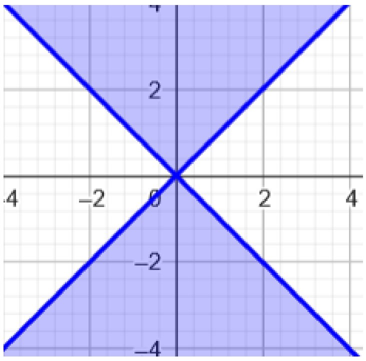

Figure 2.3: Sublevel set for $f(x, y) = x^2 - y^2$ with $c = 0$

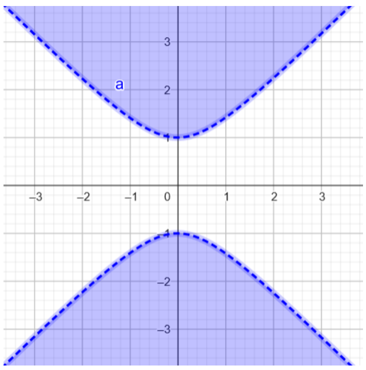

Figure 2.4: Sublevel set for $f(x, y) = x^2 - y^2$ with $c < 0$

As in the previous example, if a variation in the number of the connected components of the sublevel sets of f occurs, this means that we have crossed a critical value for f.

Also in this case, f satisfies (PS) (if $\{(x_n, y_n)\}_{n\in\mathbb{N}}$ is a Palais-Smale sequence for f at level 0, then $x_n \overset{n\to\infty}{\longrightarrow} 0$ and $y_n \overset{n\to\infty}{\longrightarrow} 0$, from which $\{(x_n, y_n)\}_{n\in\mathbb{N}}$ is bounded and it has a converging subsequence by Bolzano-Weierstrass theorem: this means $(PS)_0$ is satisfied. On the other hand, $(PS)_c$ is trivially satisfied for any $c \neq 0$).

Example 2.38. Consider the function $f(x) = -e^{-x} - \frac{2x^2}{x^2+1} + 2$. In this case:

- $f^c = (-\infty, x_0]$, where x_0 is the unique value satisfying $f(x_0) = c$, for $c \leq 0$;
- $f^c = (-\infty, x_0] \cup [x_1, +\infty)$, where x_0 and x_1 satisfy $f(x_0) = f(x_1) = c$ and $x_0 < x_1$, for $0 < c < \max_{x\in\mathbb{R}} f$;
- $f^c = \mathbb{R}$ for $c \geq \max_{x\in\mathbb{R}} f$.

We have a variation in the number of the connected components of the sublevel sets of f when we cross the value $c = 0$; however, $0 \notin \mathrm{Vcrit}(f)$. Note that f does not satisfy $(PS)_0$.

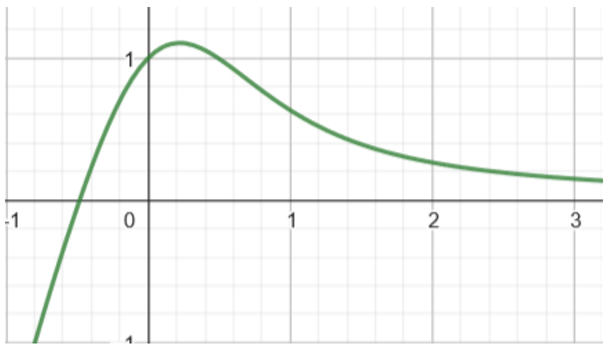

Figure 2.5: Graph of the function $f(x) = e^{-x} - \frac{2x^2}{x^2+1} + 2$

2.3.1 The Deformation Lemma for Banach spaces

The easiest version of the Deformation Lemma is for Hilbert spaces and functions of class C^2.

Theorem 2.39 (Deformation Lemma for Hilbert spaces). *Let* $f \in C^2(\mathbb{H}; \mathbb{R})$ *and* $c \in \mathbb{R}$*. If* $c \notin \mathrm{Vcrit}(f)$ *and* f *satisfies* $(PS)_c$*, then there exist* $\varepsilon > 0$ *and a deformation* $\varphi \in C(\mathbb{H}; \mathbb{H})$ *such that:*

1. $\varphi(x) = x$, $\forall x \notin f^{-1}([c-2\varepsilon, c+2\varepsilon])$;

2. $\varphi(f^{c+\varepsilon}) \subset f^{c-\varepsilon}$.

Proof. Since $f \in C^2(\mathbb{H};\mathbb{H})$, then $\nabla f \in C^1(\mathbb{H};\mathbb{H})$ and, in particular, $\nabla f \in Lip_{loc}(\mathbb{H};\mathbb{H})$. By hypothesis, $c \notin \mathrm{Vcrit}(f)$ and f satisfies $(PS)_c$, therefore, by Lemma 2.27, there exists $\varepsilon > 0$ such that

$$\|\nabla f(x)\| \geq 2\varepsilon, \quad \forall x \in f^{-1}([c-2\varepsilon, c+2\varepsilon]) \quad (\star).$$

We define $A := f^{-1}([c-2\varepsilon, c+2\varepsilon])$, $B := f^{-1}([c-\varepsilon, c+\varepsilon])$ and $\psi : \mathbb{H} \longrightarrow \mathbb{R}$ by

$$\psi(x) := \frac{\mathrm{dist}(x, \mathbb{H} \setminus A)}{\mathrm{dist}(x, \mathbb{H} \setminus A) + \mathrm{dist}(x, B)}.$$

ψ is well-defined (at least one between $\mathrm{dist}(x, B)$ and $\mathrm{dist}(x, \mathbb{H} \setminus A)$ is not 0 for any $x \in \mathbb{H}$) and $\psi \in Lip_{loc}(\mathbb{H};\mathbb{R})$. Moreover:

- $0 \leq \psi(x) \leq 1, \quad \forall x \in \mathbb{H}$;
- $\psi(x) = 0, \quad \forall x \in \mathbb{H} \setminus A$;
- $\psi(x) = 1, \quad \forall x \in B$.

We define another function $\Psi : \mathbb{H} \longrightarrow \mathbb{H}$ by

$$\Psi(x) := \begin{cases} \psi(x)\frac{\nabla f(x)}{\|\nabla f(x)\|}, \quad x \in A, \\ 0, \quad x \in \mathbb{H} \setminus A. \end{cases}$$

Ψ is well-defined by $(\star)$ and $\Psi \in Lip_{loc}(\mathbb{H};\mathbb{H})$. We consider the following Cauchy problem

$$\begin{cases} \dot{\eta}(t,x) = -\Psi(\eta(t,x)), \\ \eta(0,x) = x, \end{cases} \tag{2.3}$$

where we have denote $\dot{\eta}(t,x) := \frac{\partial \eta}{\partial t}(t,x)$. By the local existence and uniqueness theorem, the Cauchy problem (2.3) admits a unique solution $\eta(\cdot, x) \in C^1(\mathbb{R};\mathbb{H})$ for any $x \in \mathbb{H}$ (the local solution extends to the whole real line, being the system autonomous). Moreover, $\eta \in C([0,1] \times \mathbb{H};\mathbb{H})$ by the continuous dependence on initial data theorem. We set $\varphi := \eta(1, \cdot)$.

We have that $\varphi \in C(\mathbb{H};\mathbb{H})$ and $\varphi(x) = x$ for any $x \in \mathbb{H} \setminus A$.[2] So *1* is proved.

We have the following computation:

$$\frac{\partial}{\partial t}\big(f(\eta(t,x))\big) = \langle \nabla f(\eta(t,x)), \dot{\eta}(t,x)\rangle = \begin{cases} -\psi(\eta(t,x))\|\nabla f(\eta(t,x))\|, & \eta(t,x) \in A, \\ 0, & \eta(t,x) \in \mathbb{H} \setminus A. \end{cases}$$

Thus,

$$\frac{\partial}{\partial t}\big(f(\eta(t,x))\big) \leq 0, \ \forall x \in \mathbb{H}, \ \forall t \in [0,1],$$

i.e. $[0,1] \ni t \longrightarrow f(\eta(t,x))$ is non-increasing for any $x \in \mathbb{H}$.

Now, let $x \in f^{c+\varepsilon}$. We claim there exists $t \in [0,1]$ such that $f(\eta(t,x)) \leq c - \varepsilon$. Indeed, if $f(\eta(t,x)) \in (c-\varepsilon, c+\varepsilon]$ for any $t \in [0,1]$ (implying that $\eta(t,x) \in B$ for any $t \in [0,1]$), then

$$f(\varphi(x)) = f(\eta(1,x)) = f(\eta(0,x)) + \int_0^1 \frac{\partial}{\partial t}\big(f(\eta(t,x))\big)\, \mathrm{d}t = f(x) - \int_0^1 \underbrace{\|\nabla f(\eta(t,x))\|}_{\geq 2\varepsilon} \, \mathrm{d}t \leq c + \varepsilon - 2\varepsilon = c - \varepsilon,$$

which is in contradiction with our assumption.

So let $t \in [0,1]$ such that $f(\eta(t,x)) \leq c - \varepsilon$; then

$$f(\varphi(x)) = f(\eta(1,x)) \leq f(\eta(t,x)) \leq c - \varepsilon,$$

i.e $\varphi(x) \in f^{c-\varepsilon}$. This completes the proof of *2*. □

The next step is a generalization of the Deformation Lemma to Banach spaces and functions of class C^1. We need the notion of "pseudo-gradient", which replaces the role of the gradient in a Banach space.

Notation 2.40. If $\mathbb{B}$ is a Banach space and $f \in C^1(\mathbb{B};\mathbb{R})$, we denote the set of regular points of f by $\mathbb{B}^r := \{x \in \mathbb{B} \,|\, \mathrm{d}_x f \neq 0\}$.

[2] If $x \in \mathbb{H} \setminus A$, then, being $\Psi(x) = 0$, the function given by $[0,1] \ni t \mapsto \mu(t,x) := x$ is a solution of (2.3). The solution of (2.3) is unique, whence $\eta(t,x) = \mu(t,x) = x$ for any $t \in [0,1]$ and, in particular, $\varphi(x) = \eta(1,x) = x$.

Definition 2.41. Let $f \in C^1(\mathbb{B};\mathbb{R})$. A function $V \in Lip_{loc}(\mathbb{B}^r;\mathbb{B})$ is said to be a *pseudo-gradient* (PG) for f if

1. $\|V(x)\| \leq 2\min\{\|\mathrm{d}_x f\|_*, 1\}$,

2. $\mathrm{d}_x f(V(x)) \geq \min\{\|\mathrm{d}_x f\|_*, 1\}\|\mathrm{d}_x f\|_*$,

for any $x \in \mathbb{B}^r$.

We recall the following definition from topology.

Definition 2.42. A topological space X is said to be *paracompact* if for any open covering $\mathcal{A} := \{A_j,\, j \in J\}$ (i.e A_j is open for any $j \in J$ and $X \subset \cup_{j\in J} A_j$) there exists $\mathcal{B} := \{B_i,\, i \in I\}$ such that:

- for any $i \in I$ there exists $j \in J$ such that $B_i \subset A_j$, B_i is open and $X \subset \cup_{i\in I} B_i$;

- for any $x \in X$ there exists an open neighbourhood U_x of x such that
$$\mathrm{card}(\{i \in I \,:\, B_i \cap U_x \neq \emptyset\}) < \infty.$$

The following theorem is a well-known topological result. For more details see [8].

Theorem 2.43 (Stone). *Any metric space is paracompact.*

Going back to the pseudo-gradient, we have the following result.

Lemma 2.44. *Let $f \in C^1(\mathbb{B};\mathbb{R})$. Then there exists a pseudo-gradient for f.*

Proof. Let $x_0 \in \mathbb{B}^r$; there exists $w = w(x_0) \in \mathbb{B}$ such that

$$\begin{cases} \|w\| < 2\min\{\|\mathrm{d}_{x_0} f\|_*, 1\}, \\ \mathrm{d}_{x_0} f(w) > \min\{\|\mathrm{d}_{x_0} f\|_*, 1\}\|\mathrm{d}_{x_0} f\|_*. \end{cases} \tag{2.4}$$

To see that such w exists:

- If $\|\mathrm{d}_{x_0} f\|_* \leq 1$, then let $v \in \mathbb{B}$, $\|v\| = 1$, such that $\mathrm{d}_{x_0} f(v) \geq \frac{2}{3}\|\mathrm{d}_{x_0} f\|_*$ and set $w = \frac{3}{2}\|\mathrm{d}_{x_0} f\|_* v$;

- If $\|\mathrm{d}_{x_0} f\|_* > 1$, there must exist $w \in \mathbb{B}, \|w\| < 2$, such that $|\mathrm{d}_{x_0} f(w)| > \|\mathrm{d}_{x_0} f\|_*$. To see this: if $|\mathrm{d}_{x_0} f(w)| \leq \|\mathrm{d}_{x_0} f\|_*$ for any $\|w\| < 2$, then

$$\|\mathrm{d}_{x_0} f\|_* = \sup_{\|w\|=1} |\mathrm{d}_{x_0} f(w)| = \sup_{\|w\|=\frac{3}{2}} \frac{|\mathrm{d}_{x_0} f(w)|}{\|w\|} \leq \frac{2}{3} \|\mathrm{d}_{x_0} f\|_*,$$

which is not possible.

By continuity of $\mathrm{d}_{(\cdot)} f$, for any $x_0 \in \mathbb{B}^r$ there exists an open neighbourhood $W(x_0)$ of x_0 such that (2.4) holds with $w = w(x_0)$ for any $x \in W(x_0)$. Hence, $\{W(x_0)\}_{x_0 \in \mathbb{B}^r}$ is an open covering of $\mathbb{B}^r$ and by Stone's theorem there exists $\mathcal{B} = \{B_i,\, i \in I\}$ as in Definition 2.42. Hence, for any $i \in I$ there exists $v_i \in \mathbb{B}$ such that

$$\begin{cases} \|v_i\| < 2 \min\{\|\mathrm{d}_x f\|_*, 1\}, \\ \mathrm{d}_{x_0} f(v_i) > \min\{\|\mathrm{d}_x f\|_*, 1\} \|\mathrm{d}_x f\|_*, \end{cases} \tag{2.5}$$

for any $x \in B_i$. Now let

$$r_i : \mathbb{B}^r \longrightarrow \mathbb{R}, \quad r_i(x) := \mathrm{dist}(x, \mathbb{B}^r \setminus B_i),$$

and set

$$\eta_i : \mathbb{B}^r \longrightarrow \mathbb{R}, \quad \eta_i(x) := \frac{r_i(x)}{\sum_{j \in I} r_j(x)}.$$

η_i is well-defined because for any $x \in \mathbb{B}^r$ we have that $x \in B_j$ for only a positive finite number of $j's$, whence $r_j(x) \neq 0$ for only a positive finite number of $j's$. For any $i \in I$ we have that $\eta_i \in Lip_{loc}(\mathbb{B}^r; \mathbb{R})$ and

$$\begin{cases} 0 \leq \eta_i(x) \leq 1, \quad \forall x \in \mathbb{B}; \\ \eta_i(x) = 0, \quad \forall x \in \mathbb{B} \setminus B_i; \\ \sum_{i \in I} \eta_i(x) = 1, \quad \forall x \in \mathbb{B}. \end{cases}$$

Finally, let

$$V : \mathbb{B}^r \longrightarrow \mathbb{R}, \quad V(x) := \sum_{i \in I} v_i \eta_i(x).$$

V is well-defined: indeed, for any $x \in \mathbb{B}^r$, if $I(x) := \{i \in I \,|\, \eta_i(x) \neq 0\}$, then $I(x)$ is finite and $V(x)$ is a convex combination of $\{v_i\}_{i \in I(x)}$. Let $x \in \mathbb{B}^r$; then, by triangle inequality and linearity of $\mathrm{d}_x f$, we have:

- $$\|V(x)\| = \| \sum_{i\in I(x)} v_i\eta_i(x)\| \leq \sum_{i\in I(x)} \|v_i\|\eta_i(x) \overset{(2.5)}{<} 2\min\{\|\mathrm{d}_x f\|_*, 1\},$$

  which is *1* in the definition of pseudo-gradient for f;

- $$\mathrm{d}_x f(V(x)) = \sum_{i\in I(x)} \eta_i(x)\mathrm{d}_x f(v_i) \overset{(2.5)}{>} \min\{\|\mathrm{d}_x f\|_*, 1\}\|\mathrm{d}_x f\|_*,$$

which is *2* in the definition of pseudo-gradient for f. □

Remark 2.45. We have actually proved that, given a function $f \in C^1(\mathbb{B};\mathbb{R})$, there exists a pseudo-gradient for f satisfying the strict version of the inequalities of Definition 2.41.

Before seeing a more general version of the Deformation Lemma, with Banach spaces and functions of class C^1, we fix some notations and see a preliminary result.

Notation 2.46. If $f : \mathbb{B} \longrightarrow \mathbb{R}$ is a function of class C^1, $\beta \in \mathbb{R}$, $\delta > 0$ and $\rho > 0$, we denote:

- $N_{\beta,\delta} := \{x \in \mathbb{B} \,:\, |f(x) - \beta| < \delta,\ \|\mathrm{d}_x f\|_* < \delta\}$,
- $U_{\beta,\rho} := \cup_{x\in \mathrm{Crit}^\beta(f)}\{y \in \mathbb{B} \,:\, \|y - x\| < \rho\}$.

Lemma 2.47. *Let $f \in C^1(\mathbb{B},\mathbb{R})$ and suppose that f satisfies (PS). For any $\beta \in \mathbb{R}$ we have:*

1. $\mathrm{Crit}^\beta(f)$ *is compact;*
2. $\{U_{\beta,\rho}\}_{\rho>0}$ *is a fundamental system of neighbourhoods of* $\mathrm{Crit}^\beta(f)$*;*
3. $\{N_{\beta,\delta}\}_{\delta>0}$ *is a fundamental system of neighbourhoods of* $\mathrm{Crit}^\beta(f)$*.*

Proof. If $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in $\mathrm{Crit}^\beta(f)$, then it is a Palais-Smale sequence for f at level β and, by $(PS)_\beta$, it has a converging subsequence to some $x_0 \in \mathbb{B}$. Since f and $\mathrm{d}_{(\cdot)}f$ are continuous maps, then $f(x_0) = \beta$ and $\mathrm{d}_{x_0}f = 0$, meaning that $x_0 \in \mathrm{Crit}^\beta(f)$. Hence, $\mathrm{Crit}^\beta(f)$ is compact and *1* is proved.

If $\rho > 0$, then $U_{\beta,\rho}$ is a neighbourhood of $\mathrm{Crit}^\beta(f)$. Let N be a neighbourhood of $\mathrm{Crit}^\beta(f)$. By contradiction, we assume that $U_{\beta,\rho} \not\subset N$ for any $\rho > 0$. Thus, we can find a sequence $\rho_m \to 0$ and a sequence of points $\{x_m\}_{m\in\mathbb{N}}$ such that $x_m \in U_{\beta,\rho_m} \setminus N$ for any m. For any $m \in \mathbb{N}$ there exists $y_m \in \mathrm{Crit}^\beta(f)$ such that $\|x_m - y_m\| < \rho_m$. By the compactness of $\mathrm{Crit}^\beta(f)$, up to subsequence, we may assume that $y_m \to y_0 \in \mathrm{Crit}^\beta(f)$. It follows that $x_m \to y_0 \in \mathrm{Crit}^\beta(f) \subset N$, whence for $m >> 1$ we have $x_m \in N$, which is not possible. This means that, given a neighbourhood N of $\mathrm{Crit}^\beta(f)$, there exists $\rho > 0$ such that $U_{\beta,\rho} \subset N$. Therefore *2* is proved.

If $\delta > 0$, then $N_{\beta,\delta}$ is a neighbourhood of $\mathrm{Crit}^\beta(f)$. Let N be a neighbourhood of $\mathrm{Crit}^\beta(f)$. By contradiction, we assume that $N_{\beta,\delta} \not\subset N$ for any $\delta > 0$. As before, there exist a sequence $\delta_m \to 0$ and a sequence of points $\{x_m\}_{m\in\mathbb{N}}$ such that $x_m \in N_{\beta,\delta_m} \setminus N$ for any m. It follows that

$$\begin{cases} f(x_m) \overset{m\to\infty}{\longrightarrow} \beta, \\ \|\mathrm{d}_{x_m} f\|_* \overset{m\to\infty}{\longrightarrow} 0, \end{cases}$$

i.e $\{x_m\}_{m\in\mathbb{N}}$ is a Palais-Smale sequence for f at level β. By $(PS)_\beta$, up to subsequence, $\{x_m\}_{m\in\mathbb{N}}$ converges to some $x_0 \in \mathrm{Crit}^\beta(f) \subset N$, whence for $m >> 1$ we have $x_m \in N$, which is not possible. Therefore for any neighbourhood N of $\mathrm{Crit}^\beta(f)$ there exists $\delta > 0$ such that $N_{\beta,\delta} \subset N$ and the proof of *3* is completed. □

Remark 2.48. In the previous lemma, if $\beta \in \mathbb{R}$ is fixed, we need only that f satisfies $(PS)_\beta$ to show that *1*, *2* and *3* hold.

Theorem 2.49 (**Deformation lemma for Banach spaces**). *Let $f \in C^1(\mathbb{B};\mathbb{R})$ such that f satisfies (PS). Let $\beta \in \mathbb{R}$, $\overline{\varepsilon} > 0$ and let N be a neighbourhood of $\mathrm{Crit}^\beta(f)$. Then there exists $\varepsilon \in]0,\overline{\varepsilon}[$ and a continuous 1-parameter family of homeomorphisms $\Phi(t,\cdot)$ of $\mathbb{B}$, $0 \le t < \infty$, with the properties:*

1. *$\Phi(t,x) = x$, if $t = 0$, or $\mathrm{d}_x f = 0$, or $|f(x) - \beta| \ge \overline{\varepsilon}$;*

2. *$f(\Phi(t,x))$ is non-increasing in t for any $x \in \mathbb{B}$;*

3. *$\Phi(1, f^{\beta+\varepsilon} \setminus N) \subset f^{\beta-\varepsilon}$ and $\Phi(1, f^{\beta+\varepsilon}) \subset f^{\beta-\varepsilon} \cup N$.*

Moreover, $\Phi : [0,\infty[\times\mathbb{B} \longrightarrow \mathbb{B}$ has the semi-group property, i.e $\Phi\big(t,\Phi(s,\cdot)\big) = \Phi(t+s,\cdot)$ for any $t, s \ge 0$.

Proof. By Lemma 2.47, there exist $\delta, \rho > 0$ such that:

$$N \supset U_{\beta,2\rho} \supset U_{\beta,\rho} \supset N_{\beta,\delta},$$

where we may suppose $\delta, \rho \leq 1$. Let $\eta \in Lip_{loc}(\mathbb{B};\mathbb{R})$ such that $0 \leq \eta \leq 1$, $\eta \equiv 1$ in $\mathbb{B} \setminus N_{\beta,\delta}$ and $\eta \equiv 0$ in $N_{\beta,\delta/2}$. Also let $\varphi \in Lip_{loc}(\mathbb{R};\mathbb{R})$ such that $0 \leq \varphi \leq 1$, $\varphi(s) = 0$ if $|\beta - s| \geq \min\{\bar{\varepsilon}, \delta/4\}$, $\varphi(s) = 1$ if $|\beta - s| \leq \min\{\bar{\varepsilon}/2, \delta/8\}$. Finally, let $V : \mathbb{B}^r \longrightarrow \mathbb{B}$ be a pseudo-gradient for f (it exists by Lemma 2.44). We define

$$\Psi(x) := \begin{cases} -\eta(x)\varphi(f(x))V(x), & \text{if } x \in \mathbb{B}^r, \\ 0, & \text{elsewhere.} \end{cases}$$

By definition of φ and η, it follows that Ψ vanishes (and therefore is Lipschitz) near critical points of f[3], hence $\Psi \in Lip_{loc}(\mathbb{B};\mathbb{B})$. Moreover, since $\|V\|_{L(\mathbb{B}^r,\mathbb{B})} \leq 2$, then $\|\Psi\|_{L(\mathbb{B},\mathbb{B})} \leq 2$. We consider the Cauchy problem:

$$\begin{cases} \frac{\partial}{\partial t}\Phi(t,x) = \Psi(\Phi(t,x)), \\ \Phi(0,x) = x. \end{cases} \tag{2.6}$$

By the local existence and uniqueness theorem, the Cauchy problem (2.6) has a unique solution $\Phi(\cdot, x) \in C^1(\mathbb{R};\mathbb{B})$ for any $x \in \mathbb{B}$ (the system is autonomous, whence the solution extends to $\mathbb{R}$) and Φ is continuous in the variable $x \in \mathbb{B}$ by the continuous dependence on initial data theorem. Moreover, $\Phi(t, \Phi(s, \cdot)) = \Phi(t + s, \cdot)$ for any $t, s \in \mathbb{R}$. From this, $\Phi(t, \cdot) \circ \Phi(-t, \cdot) = \mathrm{Id}_{\mathbb{B}} = \Phi(-t, \cdot) \circ \Phi(t, \cdot)$, whence $\Phi(t, \cdot)$ is a homeomorphism of $\mathbb{B}$ for any $t \in \mathbb{R}$.

If $t = 0$, then $\Phi(t, x) = x$ for any $x \in \mathbb{B}$ by definition. If $\mathrm{d}_x f = 0$ or $x \in \mathbb{B}^r$ but $|f(x) - \beta| \geq \bar{\varepsilon}$ (in this second case $\varphi(f(x)) = 0$), then $\Psi(x) = 0$ and the function

[3] $\Psi \equiv 0$ in $N_{\beta,\delta/2}$ because $\eta \equiv 0$ in $N_{\beta,\delta/2}$. If $x_0 \in \mathrm{Crit}(f)$, by continuity of $\mathrm{d}_{(\cdot)}f$, we have that $||\mathrm{d}_x f||_* < \frac{\delta}{2}$ for any x in an open neighbourhood of x_0 V_{x_0}. Let $x \in V_{x_0}$; if $x \in \mathrm{Crit}(f)$, then it is immediate that $\Psi(x) = 0$, otherwise:

- if $\eta(x) = 0$, then $\Psi(x) = 0$;
- if $\eta(x) \neq 0$, then $x \notin N_{\beta,\delta/2}$. This implies that $|f(x) - \beta| \geq \frac{\delta}{2} \geq \min\{\bar{\varepsilon}, \delta/4\}$, whence $\varphi(f(x)) = 0$ and $\Psi(x) = 0$.

$t \longmapsto \tilde{\Phi}(t,x) := x$ is a solution of (2.6), whence, by uniqueness, $\Phi(t,x) = \tilde{\Phi}(t,x) = x$ for all $t \in \mathbb{R}$. Thus *1* is proved.

We see *2*:

$$\frac{\partial}{\partial t}\big(f(\Phi(t,x))\big) = \mathrm{d}_{\Phi(t,x)} f\left(\frac{\partial}{\partial t}(\Phi(t,x))\right) = \mathrm{d}_{\Phi(t,x)} f\big(\Psi(\Phi(t,x))\big) =$$
$$= \begin{cases} -\eta(\Phi(t,x))\varphi(f(\Phi(t,x)))\mathrm{d}_{\Phi(t,x)} f\big(V(\Phi(t,x))\big), & \text{if}\, \Phi(t,x) \in \mathbb{B}^r, \\ 0, & \text{else}, \end{cases}$$
$$\leq \begin{cases} -\eta(\Phi(t,x))\varphi(f(\Phi(t,x))) \min\{\|\mathrm{d}_{\Phi(t,x)} f\|_*, 1\}\|\mathrm{d}_{\Phi(t,x)} f\|_*, & \text{if}\, \Phi(t,x) \in \mathbb{B}^r, \\ 0, & \text{else}, \end{cases}$$

whence

$$\frac{\partial}{\partial t}\big(f(\Phi(t,x))\big) \leq 0, \quad \forall x \in \mathbb{B}.$$

Let us prove *3*. Let $\varepsilon \leq \min\{\overline{\varepsilon}/2, \delta/8\}$ and $x \in f^{\beta+\varepsilon}$. If $f(\Phi(1,x)) > \beta - \varepsilon$, then from *2* it follows that

$$\beta + \varepsilon \geq f(x) = f(\Phi(0,x)) \geq f(\Phi(t,x)) \geq f(\Phi(1,x)) > \beta - \varepsilon, \quad \forall t \in [0,1],$$

whence $|f(\Phi(t,x)) - \beta| \leq \varepsilon$ and consequently $\varphi(f(\Phi(t,x))) = 1$ for all $t \in [0,1]$. We note that, if $\Phi(t,x) \notin N_{\beta,\delta}$, then $\Phi(t,x) \in \mathbb{B}^r$.[4]

We have the following computation:

$$f(\Phi(1,x)) = f(\Phi(0,x)) + \int_0^1 \underbrace{\frac{\partial}{\partial t}\big(f(\Phi(t,x))\big)}_{\leq 0} \mathrm{d}t \leq$$
$$\leq \beta + \varepsilon - \int_{\{t \in [0,1] \,:\, \Phi(t,x) \notin N_{\beta,\delta}\}} \underbrace{\eta(\Phi(t,x))}_{=1} \underbrace{\mathrm{d}_{\Phi(t,x)} f(V(\Phi(t,x)))}_{\geq \min\{\|\mathrm{d}_{\Phi(t,x)} f\|_*, 1\}\|\mathrm{d}_{\Phi(t,x)} f\|_* \geq \delta^2} \mathrm{d}t \leq$$
$$\leq \beta + \varepsilon - \mathcal{L}^1(\{t \in [0,1] \,:\, \Phi(t,x) \notin N_{\beta,\delta}\})\delta^2. \quad (2.7)$$

We note that:

[4] If $\Phi(t,x)$ is a critical point for f, then $\|\mathrm{d}_{\Phi(t,x)} f\|_* = 0 < \delta$ and $|f(\Phi(t,x)) - \beta| \leq \varepsilon < \delta$, whence $\Phi(t,x) \in N_{\beta,\delta}$.

- If $x \notin N$, then $\Phi(t,x) \notin N_\delta$ for any $0 \le t < \frac{\rho}{2}$. Indeed, $\mathbb{B} \setminus N$ and N_δ are separated by the "annulus" $U_{\beta,2\rho} \setminus U_{\beta,\rho}$ of width ρ, from which $\mathrm{dist}(\mathbb{B} \setminus N, N_\delta) \ge \rho$ and

$$\|\Phi(t,x) - \underbrace{\Phi(0,x)}_{=x}\| = \left\| \int_0^t \Psi\big(\Phi(s,x)\big)\,\mathrm{d}s \right\| \le \int_0^t \underbrace{\|\Psi\big(\Phi(s,x)\big)\|}_{\le 2}\,\mathrm{d}s < \rho, \quad \forall\, 0 \le t < \frac{\rho}{2}.$$

  This means that

$$[0, \rho/2[\subset \{t \in [0,1] \,:\, \Phi(t,x) \notin N_{\beta,\delta}\};$$

- similarly, if $\Phi(1,x) \notin N$, then $\Phi(t,x) \notin N_\delta$ for any $t \in](2-\rho)/2, 1]$, because

$$\|\Phi(1,x) - \Phi(t,x)\| \le \int_t^1 \|\Psi(\Phi(s,x))\| \mathrm{d}s \le 2(1-t) < \rho, \quad \forall\, \frac{2-\rho}{2} < t \le 1.$$

  Thus, we have

$$](2-\rho)/2, 1] \subset \{t \in [0,1] \,:\, \Phi(t,x) \notin N_{\beta,\delta}\}.$$

Therefore, if either $x \notin N$ or $\Phi(1,x) \notin N$, we have that

$$\mathcal{L}^1(\{t \in [0,1] \,:\, \Phi(t,x) \notin N_{\beta,\delta}\}) \ge \frac{\rho}{2}. \tag{2.8}$$

Summarizing, let $\varepsilon \le \min\{\bar{\varepsilon}/2, \delta/8, (\delta^2\rho)/4\}$. We have that, if $x \in f^{\beta+\varepsilon}$ and $x \notin N$ or $\Phi(1,x) \notin N$, then $f(\Phi(1,x)) \le \beta - \varepsilon$. Indeed, if it was that $f(\Phi(1,x)) > \beta - \varepsilon$, then, combining (2.7) and (2.8), we would have that $f(\Phi(1,x)) \le \beta - \varepsilon$, which is not possible. This means that $\Phi(1, f^{\beta+\varepsilon} \setminus N) \subset f^{\beta-\varepsilon}$ and $\Phi(1, f^{\beta+\varepsilon}) \subset f^{\beta-\varepsilon} \cup N$, completing the proof of *3*. $\square$

Remark 2.50. As in Lemma 2.47, if $\beta \in \mathbb{R}$ is fixed, we need only that f satisfies $(PS)_\beta$ to show that the conclusion holds for that value of β.

2.3.2 The Deformation Lemma for Finsler manifold

We aim to generalize the Deformation Lemma to manifolds modeled on Banach spaces.

Definition 2.51. Let M be a Banach manifold of class C^p, $p \geq 0$[5], called the base space; let E be a topological space, called the total space; let $\pi : E \longrightarrow M$ be a surjective continuous map. Suppose that for any $x \in M$, the fiber $E_x := \pi^{-1}(x)$ has a structure of Banach space. Let $\{U_i \,|\, i \in I\}$ be an open cover of M and suppose that for each $i \in I$ there exist a Banach space X_i and a map τ_i

$$\tau_i : \pi^{-1}(U_i) \longrightarrow U_i \times X_i$$

satisfying the following properties:

- τ_i is a homeomorphism such that $p_1 \circ \tau_i = \pi$, where $p_1 : U_i \times X_i \longrightarrow U_i$ is the projection on U_i, and such that

  $$\tau_{i_x} : E_x \longrightarrow X_i, \quad \tau_{i_x} := \tau_i\big|_{E_x}$$

  is an isomorphism for any $x \in U_i$.

- if U_i and U_j are two members of the open cover, then the map

  $$\begin{aligned} U_i \cap U_j &\longrightarrow \mathrm{L}(X_i, X_j) \\ x &\longmapsto \tau_{j_x} \circ \tau_{i_x}^{-1} \end{aligned}$$

  is a differentiable map of class C^p, where $L(X_i, X_j)$ is the space of all continuous linear maps from X_i to X_j.

The collection $\{(U_i, \tau_i) \,|\, i \in I\}$ is said to be a *trivializing covering* for $\pi : E \longrightarrow M$ and the maps τ_i are called *trivializing maps*. Two trivializing coverings are said to be equivalent if their union satisfies again the two conditions above. An equivalence class of such trivializing coverings is said to determine the structure of a *Banach bundle* on $\pi : E \longrightarrow M$. If all the spaces X_i are isomorphic as topological spaces, then they can be assumed all to be equal to the same space X. In this case, we say that $\pi : E \longrightarrow M$ is a *Banach bundle with fibre X*.

[5]A Banach manifold M is a manifold modeled on Banach spaces; this means that M is a topological space in which each point has a neighbourhood that is homeomorphic to an open set in a Banach space. The manifold is of class C^p if the transition maps between open sets of an atlas of M are of class C^p. As for the Euclidean case, if M has a differential structure, we have an analogous construction of the tangent bundle TM and, given a differentiable map between Banach Manifolds $f : M \longrightarrow N$, we can consider the differential map $\mathrm{d}f_p : T_pM \longrightarrow T_{f(p)}N$, $p \in M$.

Definition 2.52. Let $\pi : F \longrightarrow M$ be a Banach bundle over M. Let

$$\|.\| : F \longrightarrow \mathbb{R}$$

be a continuous map such that the restriction $\|.\|_u$ of $\|.\|$ to each fiber $F_u := \pi^{-1}(u)$ is an admissible norm for F_u. Let us also assume that we have a trivializing covering $\mathcal{A} := \{(U_i, \tau_i)\}_{i \in I}$ satisfying the following properties:

1. for any fixed $u_0 \in M$ there exists a trivialization in $\mathcal{A}$

$$\tau : \pi^{-1}(U) \longrightarrow U \times F_{u_0},$$

   such that $\|.\|_u$ is a norm on F_{u_0} for any $u \in U$;

2. for any $k > 1$

$$\frac{1}{k}\|v\|_u < \|v\|_{u_0} < k\|v\|_u, \quad \forall v \in F_{u_0} \setminus \{0\},$$

   if u is in a small neighbourhood of u_0 contained in U and depending on k.

Then we say that $\|.\|$ is a *Finsler structure* for the bundle $\pi : F \longrightarrow M$.

Remark 2.53. Let M be a differentiable (Banach or Euclidan) manifold equipped with a Riemannian metric $\langle \, , \, \rangle$. Then

$$\|\xi\|_{\pi(\xi)} := \sqrt{\langle \xi, \xi \rangle_{\pi(\xi)}}, \quad \forall \xi \in TM,$$

defines a natural Finsler structure for the bundle $\pi : TM \longrightarrow M$.

Definition 2.54. A *Finsler manifold* of class C^r, $r \geq 1$, is a regular[6] C^r-Banach manifold M, together with a Finsler structure $\|.\|$ on the tangent bundle TM. Also the cotangent bundle T^*M carries a natural Finsler structure by defining

$$\|v^*\| = \sup\{|v^*(v)| \,;\, v \in T_uM, \, \|v\|_u \leq 1\}, \quad \forall u \in M, \forall v^* \in T_u^*M.$$

[6]We recall that a topological space X is said to be *regular* if for any point $x \in X$ and for any neighbourhood U of x there exists a closed neighbourhood F of x such that $F \subset U$.

Remark 2.55. Let M be a connected Finsler manifold M of class C^r, $r \geq 1$, and denote by $\|.\|$ the Finsler structure on TM. Then $\|.\|$ induces a metric on M:

$$d(u,v) = \inf_p \int_0^1 \|\dot{p}(t)\|_{p(t)}\,\mathrm{d}t, \tag{2.9}$$

where the infimum is taken all over C^1 paths $p : [0,1] \longrightarrow M$ such that $p(0) = u$ and $p(1) = v$. We say that M is *complete* if (M,d) is a complete metric space. Moreover, the topology on M induced by the distance (2.9) do coincide with the manifold topology.

The proof of Remark 2.55 is not trivial: for more details see p.201 − 202 on [7].

We can generalize the notion of pseudo-gradient to functions defined on Finsler manifolds.

Definition 2.56. Let M be a C^2-Finsler manifold and $f \in C^1(M)$. Let us denote $\tilde{M} := \{u \in M \mid \mathrm{d}_u f \neq 0\}$, the set of regular points of f. A *pseudo-gradient* vector field for f is a Lipschitz continuous vector field $v : \tilde{M} \longrightarrow TM$ such that $v(u) \in T_uM$ and

1. $\|v(u)\|_u \leq 2\|\mathrm{d}_u f\|_u^2$,
2. $\mathrm{d}_u f(v(u)) \geq \|\mathrm{d}_u f\|_u^2$,

for all $u \in \tilde{M}$, where $||.||_u$ is the norm in the tangent space T_uM.

Lemma 2.57. *If M is a C^2-Finsler manifold and $f \in C^1(M)$, then there exists a pseudo-gradient vector field $v : \tilde{M} \longrightarrow TM$ for f.*

For a proof of Lemma 2.57 see Lemma 3.2 and Lemma 3.3 on [7].

Remark 2.58. As for Banach spaces, by using the Finsler structure and the distance (2.9) on M, we can define:

$$\begin{aligned}
f^\beta &:= \{u \in M \mid f(u) \leq \beta\},\\
\mathrm{Crit}(f) &:= \{u \in M \mid \mathrm{d}_u f = 0\},\\
\mathrm{Crit}^\beta(f) &:= \{u \in M \mid f(u) = \beta,\, \mathrm{d}_u f = 0\},\\
N_{\beta,\delta} &:= \{u \in M \,:\, |f(u) - \beta| < \delta,\, ||\mathrm{d}_u f||_u < \delta\},\\
U_{\beta,\rho} &:= \cup_{u \in \mathrm{Crit}^\beta(f)} \{v \in M \mid d(u,v) < \rho\},
\end{aligned}$$

where $\beta \in \mathbb{R}$ and $\delta, \rho > 0$.

We can extend the notion of Palais-Smale sequence and Palais-Smale condition to functions defined on Finsler manifolds.

Definition 2.59. If M is a connected C^1-Finsler manifold, $f \in C^1(M)$ and $c \in \mathbb{R}$, we say that $\{x_n\}_{n\in\mathbb{N}} \subset M$ is a *Palais-Smale sequence* for f at level c if

$$\begin{cases} f(x_n) \overset{n\to\infty}{\longrightarrow} c, \\ \|\mathrm{d}_{x_n} f\|_{x_n} \overset{n\to\infty}{\longrightarrow} 0. \end{cases}$$

We say that f satisfies the *Palais-Smale condition* at level c and write $(PS)_c$ if any Palais-Smale sequence for f at level c admits a converging subsequence (with respect to the metric (2.9)). We write that f satisfies (PS) if it satisfies $(PS)_c$ for any $c \in \mathbb{R}$.

Now, we give the statement of the Deformation Lemma for function on Finsler manifolds. We will omit the proof of this theorem, since it is a very technical generalization to that one seen for Banach spaces. However, it turns out we need to strengthen the hypotheses on the function in the statement of the theorem.

Theorem 2.60 (Deformation Lemma for Finsler manifold). *Let M be a connected C^2-Finsler manifold and $f \in C^1(M)$ such that f satisfies (PS). Let us also assume that f satisfies the following hypotheses:*

- *f is bounded from below, i.e $\inf_{u\in M} f(u) \in \mathbb{R}$;*
- *the sublevel sets of f are all complete in M, with respect to the metric given by (2.9).*

Let $\beta \notin \mathrm{Vcrit}(f)$. Then there exist an $\varepsilon > 0$ and a flow $\Phi : \mathbb{R} \times M \longrightarrow M$ (i.e. Φ is continuous and $\Phi(t, \cdot) \circ \Phi(s, \cdot) = \Phi(t + s, \cdot)$ for any $t, s \in \mathbb{R}$) such that:

1. *$\Phi(t, \cdot)$ is a homeomorphism of M for any t;*
2. *$\Phi(0, \cdot) = id_M$;*
3. *$t \longrightarrow f(\Phi(t, u))$ is non-increasing for any $u \in M$;*
4. *$\Phi(1, f^{\beta+\varepsilon}) \subset f^{\beta-\varepsilon}$.*

See Theorem 4.6 on [7] for the proof of Theorem 2.60.

Remark 2.61. For a fixed $\beta \in \mathbb{R}$ it suffices that f satisfies $(PS)_\beta$ so that the conclusion of Theorem 2.60 holds.

2.4 The Minimax Principle

In this section, we will state and prove the variational method we will use in the next chapter to show the existence of closed geodesics on the sphere. As for the Deformation Lemma, we will see two versions: one for Banach spaces and one for Finsler Manifolds.

Definition 2.62. Consider a Finsler manifold M (or equivalently a Banach space $\mathbb{B}$) and let $\Phi : [0, +\infty[\times M \longrightarrow M$ be a semi-flow (i.e Φ is continuous and $\Phi(t, \cdot) \circ \Phi(s, \cdot) = \Phi(t + s, \cdot)$ for all $t, s \geq 0$). A family $\mathcal{F}$ of subsets of M is called *(positively)* Φ*-invariant* if $\Phi(t, F) \in \mathcal{F}$ for any $F \in \mathcal{F}$ and $t \geq 0$.

Theorem 2.63 (**Minimax Principle for Banach spaces**). *Let $f \in C^1(\mathbb{B}; \mathbb{R})$ such that f satisfies (PS). Let $\mathcal{F}$ be a family of subsets of $\mathbb{B}$ which is invariant with respect to any semi-flow $\Phi : [0, +\infty[\times\mathbb{B} \longrightarrow \mathbb{B}$ such that $\Phi(0, \cdot) = id_{\mathbb{B}}$, $\Phi(t, \cdot)$ is a homeomorphism of $\mathbb{B}$ for any $t \geq 0$ and $t \longmapsto f(\Phi(t, u))$ is non-increasing for any $u \in \mathbb{B}$. If*

$$\beta := \inf_{F\in\mathcal{F}} \sup_{u\in F} f(u) \in \mathbb{R},$$

then β is a critical value for f.

Proof. By contradiction, we assume that β is not a critical value for f. We fix $\bar{\varepsilon} = 1$ and $N = \emptyset$, which is a neighbourhood of $\mathrm{Crit}^\beta(f) = \emptyset$. Then let $\varepsilon > 0$ and $\Phi : [0, +\infty[\times\mathbb{B} \longrightarrow \mathbb{B}$ as in Theorem 2.49. By definition of β, there exists $\tilde{F} \in \mathcal{F}$ such that

$$\sup_{u\in\tilde{F}} f(u) < \beta + \varepsilon,$$

i.e $\tilde{F} \subset f^{\beta+\varepsilon}$. $\mathcal{F}$ is Φ-invariant (by *2* in Theorem 2.49, $t \longmapsto f(\Phi(t, x))$ is non-increasing for any $x \in \mathbb{B}$), hence, $F_1 := \Phi(1, \tilde{F}) \in \mathcal{F}$ and, by *3* in Theorem 2.49, being $N = \emptyset$, it follows that $F_1 \subset f^{\beta-\varepsilon}$. Therefore we have

$$\beta := \inf_{F\in\mathcal{F}} \sup_{u\in F} f(u) \leq \sup_{u\in F_1} f(u) \leq \beta - \varepsilon,$$

which is not possible. □

Remark 2.64. As in Theorem 2.49, we only need that f satisfies $(PS)_\beta$.

We can see two easy applications of the the previous theorem.

Example 2.65. Let $f \in C^1(\mathbb{B};\mathbb{R})$ and let $\mathcal{F} = \{\mathbb{B}\}$, which is positively Φ-invariant with respect to any semi-flow Φ satisfying the hypotheses of Theorem 2.63. If

$$\beta := \inf_{F\in\mathcal{F}} \sup_{u\in F} f(u) = \sup_{u\in\mathbb{B}} f(u)$$

is finite and f satisfies $(PS)_\beta$, then β is a critical value for f. Thus, β is actually the maximum of f on $\mathbb{B}$: there exists $u_0 \in \mathbb{B}$ such that $f(u_0) = \beta$ and $\mathrm{d}_{u_0} f = 0$.

Example 2.66. Let $f \in C^1(\mathbb{B};\mathbb{R})$ and let $\mathcal{F} = \{\{u\} \,|\, u \in \mathbb{B}\}$, which is positively Φ-invariant with respect to any semi-flow Φ satisfying the hypotheses of Theorem 2.63. If

$$\beta := \inf_{F\in\mathcal{F}} \sup_{u\in F} f(u) = \inf_{u\in\mathbb{B}} f(u)$$

is finite and f satisfies $(PS)_\beta$, then β is a critical value for f. Thus, β is actually the minimum of f on $\mathbb{B}$: there exists $u_0 \in \mathbb{B}$ such that $f(u_0) = \beta$ and $\mathrm{d}_{u_0} f = 0$.

Theorem 2.67 (**Minimax Principle for Finsler manifold**). *Let M be a connected C^2-Finsler manifold and let $f \in C^1(M)$ such that f satisfies (PS). Let us also assume that f is bounded from below and that the sublevel sets of f are all complete in M (with respect to the Finsler distance on M). Let $\mathcal{F}$ be a family of subsets of M that is invariant with respect to any semi-flow $\Phi : [0, +\infty[\times M \longrightarrow M$ such that $\Phi(0, \cdot) = id_M$, $\Phi(t, \cdot)$ is a homeomorphism of M for any $t \geq 0$ and $t \longrightarrow f(\Phi(t, u))$ is non-increasing for any $u \in M$. If*

$$\beta := \inf_{F\in\mathcal{F}} \sup_{u\in F} f(u) \in \mathbb{R},$$

then β is a critical value for f.

Proof. By contradiction, we assume that β is not a critical value for f. Then there exist an $\varepsilon > 0$ and flow $\psi : \mathbb{R} \times M \longrightarrow M$ as in Theorem 2.60. By definition of β, there exists $\tilde{F} \in \mathcal{F}$ such that

$$\sup_{u\in\tilde{F}} f(u) < \beta + \varepsilon.$$

Then $F_1 := \psi(1, \tilde{F}) \in \mathcal{F}$ and $F_1 \subset f^{\beta-\varepsilon}$, whence

$$\beta := \inf_{F \in \mathcal{F}} \sup_{u \in F} f(u) \leq \sup_{u \in F_1} f(u) \leq \beta - \varepsilon,$$

which is not possible. $\square$

In the next chapter, we will see the Birkhoff's theorem which is a very interesting application of the Minimax Principle.

Chapter 3

Closed geodesics on the 2-dimensional sphere in $\mathbb{R}^3$

In this chapter, we will apply the Minimax Principle to solve a geometric problem: the existence of non costant closed geodesics on a surface S diffeomorphic to the 2-dimensional sphere. In the first section, we will recall some preliminary notions about Sobolev spaces. Then in the second section, we will introduce the space of closed curves with finite energy on a compact Riemannian manifold M and we will see a sketch of the proof that it has a structure of Hilbert manifold. Finally, we will state and prove the central theorem of this thesis: the Birkhoff's theorem. The scheme of its proof will be the following:

1. Introduction to the energy functional E and the manifold of closed curves on S with finite energy;
2. Proof of the fact that the closed geodesics on S correspond exactly to the critical points of E in the manifold previously introduced;
3. Proof of the fact that the functional E satisfies the Palais-Smale condition;
4. Construction of a flow invariant family $\mathcal{F}$ as in the statement of the Minimax Principle;
5. Proof of the fact that the Minimax value is not 0.

3.1 Some preliminary notions about Sobolev spaces

Definition 3.1. Let $n \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^n$ open. For $k \in \mathbb{N}$ and $p \in \mathbb{R}$, $1 \leq p < +\infty$, we define

$$W^{k,p}(\Omega;\mathbb{R}) := \{u \in L^p(\Omega;\mathbb{R}) \,:\, \partial^\alpha u \in L^p(\Omega;\mathbb{R}),\, |\alpha| \leq k\},$$

where α is a multi-index and $\partial^\alpha u$ is the weak derivative of u, of order α. Defining

$$\|u\|_{W^{k,p}} := \left(\sum_{|\alpha|\leq k} \|\partial^\alpha u\|_{L^p}^p\right)^{1/p}, \quad u \in W^{k,p}(\Omega;\mathbb{R}),$$

then $W^{k,p}(\Omega;\mathbb{R})$ becomes a separable Banach space and, in particular, $W^{k,2}(\Omega;\mathbb{R})$ is endowed with a structure of Hilbert space.

Definition 3.2. Let $n, k \in \mathbb{N}$, $\Omega \subset \mathbb{R}^n$ open and $p \in \mathbb{R}$, $1 \leq p < +\infty$. We define

$$H^{k,p}(\Omega;\mathbb{R}) = \overline{C^\infty(\Omega;\mathbb{R}) \cap W^{k,p}(\Omega;\mathbb{R})}^{\|\cdot\|_{W^{k,p}}}.$$

When p is omitted, it is implied that $p = 2$.

The following results hold.

Theorem 3.3 (**Meyers-Serrin**). *Let $n, k \in \mathbb{N}$, $\Omega \subset \mathbb{R}^n$ open, $p \in \mathbb{R}$, with $1 \leq p < +\infty$. Then we have that*

$$W^{k,p}(\Omega;\mathbb{R}) = H^{k,p}(\Omega;\mathbb{R}).$$

For a proof of Theorem 3.3 see Theorem 7.9 on [4].

Theorem 3.4 (**Sobolev-Rellich-Kondrakov embedding**). *Let $n \in \mathbb{N}$, $\Omega \subset \mathbb{R}^n$ open, bounded and such that $\partial\Omega$ is Lipschitz. Let $k \in \mathbb{N}$, $p \in \mathbb{R}$, $1 \leq p < +\infty$. We have three cases:*

1. *If $kp < n$, we set $p^* = \frac{np}{n-kp}$; then*

$$i : H^{k,p}(\Omega;\mathbb{R}) \hookrightarrow L^q(\Omega;\mathbb{R})$$

is continuous for any $1 \leq q \leq p^*$. Moreover, i is compact if $1 \leq q < p^*$;

2. *If $kp = n$, then*
$$i : H^{k,p}(\Omega;\mathbb{R}) \hookrightarrow L^q(\Omega;\mathbb{R})$$
is compact for any $1 \leq q < +\infty$;

3. *If $kp > n$, then*
$$i : H^{k,p}(\Omega;\mathbb{R}) \hookrightarrow C^{r,\alpha}(\overline{\Omega};\mathbb{R})$$
is continuous, where $r = k - \lfloor \frac{n}{p} \rfloor - 1$ and
$$\alpha = \begin{cases} \lfloor \frac{n}{p} \rfloor + 1 - \frac{n}{p}, & \text{if } \frac{n}{p} \notin \mathbb{N}, \\ \text{any value in } (0,1), & \text{if } \frac{n}{p} \in \mathbb{N}. \end{cases}$$
Moreover, if $\alpha < \lfloor \frac{n}{p} \rfloor + 1 - \frac{n}{p}$, then i is compact.

For a proof of Theorem 3.4 see Theorem 7.26 on [4].

Definition 3.5. Let $n, k, m \in \mathbb{N}$, $p \in \mathbb{R}$, $1 \leq p < +\infty$ and $\Omega \subset \mathbb{R}^n$ open. We define
$$W^{k,p}(\Omega;\mathbb{R}^m) := \{u \in L^p(\Omega;\mathbb{R}^m) \,:\, \partial^\alpha u = (\partial^\alpha u_1, \ldots, \partial^\alpha u_m) \in L^p(\Omega;\mathbb{R}^m),\ |\alpha| \leq\ k\},$$
and
$$\|u\|_{W^{k,p}} := \left(\sum_{|\alpha| \leq k} \|\partial^\alpha u\|_{L^p}^p \right)^{1/p}, \qquad u \in W^{k,p}(\Omega;\mathbb{R}^m).$$
Also in this case, $W^{k,p}(\Omega;\mathbb{R}^m)$ is a Banach space and, in particular, $W^{k,2}(\Omega;\mathbb{R}^m)$ is a Hilbert space.

Moreover, we define
$$H^{k,p}(\Omega;\mathbb{R}^m) := \overline{W^{k,p}(\Omega;\mathbb{R}^m) \cap C^\infty(\Omega;\mathbb{R}^m)}^{\|\cdot\|_{W^{k,p}}}.$$

Remark 3.6. The thesis of Theorem 3.3 still holds under the assumption of Definition 3.5.

We also have Sobolev spaces of periodic one-variable functions.

Definition 3.7. Let $m \in \mathbb{N}$. We define the following spaces:

- for any $1 \leq q < +\infty$,

$$L^q(\mathbb{S}^1;\mathbb{R}^m) := \{u : \mathbb{R} \longrightarrow \mathbb{R}^m \,:\, u|_{(0,2\pi)} \in L^q((0,2\pi);\mathbb{R}^m),\ u(t) = u(t+2\pi)\ \forall t \in \mathbb{R}\};$$

- for any $k \in \mathbb{N}$ and $1 \leq p < +\infty$,

$$H^{k,p}(\mathbb{S}^1;\mathbb{R}^m) := \{u : \mathbb{R} \longrightarrow \mathbb{R}^m : u|_{(0,2\pi)} \in H^{k,p}((0,2\pi);\mathbb{R}^m),\ u(t) = u(t+2\pi)\ \forall t \in \mathbb{R}\};$$

- for any $r \geq 0$ and $\alpha \in [0,1]$,

$$C^{r,\alpha}(\mathbb{S}^1;\mathbb{R}^m) := \{u : \mathbb{R} \longrightarrow \mathbb{R} : u|_{(0,2\pi)} \in C^{r,\alpha}((0,2\pi);\mathbb{R}^m),\ u(t) = u(t+2\pi)\ \forall t \in \mathbb{R}\}.$$

Remark 3.8. If $k, m \in \mathbb{N}$, $p \in \mathbb{R}$, $1 \leq p < +\infty$, then $H^{k,p}(\mathbb{S}^1;\mathbb{R}^m)$ is a Banach space with norm given by

$$\|u\|_{H^{k,p}} := \|u|_{(0,2\pi)}\|_{H^{k,p}((0,2\pi);\mathbb{R}^m)},$$

and, in particular, it is a Hilbert space if $p = 2$.

Remark 3.9. There is an analogous version of Theorem 3.4 with the spaces introduced in Definition 3.7.

3.2 The Manifold of closed curves on a compact Riemannian manifold

Definition 3.10. A *Hilbert Manifold* M is a differentiable manifold with a countable atlas such that each chart has its image in a Hilbert space; this means that M is a separable Hausdorff space in which each point has a neighbourhood that is homeomorphic to an open set in a Hilbert space. As for Euclidean manifold, there is associated to M the tangent bundle of M and a projection $\tau = \tau_M : TM \longrightarrow M$. Given an atlas $(\phi_\alpha, U_\alpha)_{\alpha \in I}$ of M, we obtain an atlas $(T\phi_\alpha, TU_\alpha)_{\alpha \in I}$ of TM, proceeding exactly as in the Euclidean case.

Remark 3.11. If $\mathbb{H}$ is a Hilbert space, then it is a Hilbert manifold with a single global chart given by $(\mathbb{H}, id_{\mathbb{H}})$. Moreover, if $\langle\,,\,\rangle_{\mathbb{H}}$ is the inner product on $\mathbb{H}$, then $\mathbb{H}$ can be endowed with a natural Riemannian metric $\langle\,,\,\rangle$ given by

$$\langle u, v\rangle_p := \langle u, v\rangle_{\mathbb{H}}, \quad \forall p \in \mathbb{H}, \forall u, v \in T_p\mathbb{H} \cong \mathbb{H}.$$

Remark 3.12. If E is a topological space and M is a Hilbert manifold, we can consider Banach bundles $\pi : E \longrightarrow M$ over M as in Definition 2.51. Then there is an atlas such that for any chart (U, ϕ) we have a commutative diagram of morphisms

$$\begin{array}{ccc} \pi^{-1}(U) & \xrightarrow{\Phi} & \phi(U) \times \mathbb{E} \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\mathrm{pr}_1} \\ U & \xrightarrow{\phi} & \phi(U) \end{array}$$

with the following properties:

1. $\mathbb{E}$ is a Banach space, Φ is a homeomorphism and $\Phi_p := \Phi|_{E_p}$, where $E_p := \pi^{-1}(p)$ for any $p \in U$, is a topological linear isomorphism;
2. if

$$\begin{array}{ccc} \pi^{-1}(U') & \xrightarrow{\Phi'} & \phi(U') \times \mathbb{E}' \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\mathrm{pr}_1} \\ U' & \xrightarrow{\phi'} & \phi'(U') \end{array}$$

   is a commutative diagram satisfying 1 for another chart (U', ϕ'), then the map

$$\begin{aligned} U \cap U' &\longrightarrow L(\mathbb{E}, \mathbb{E}') \\ p &\longmapsto \Phi'_p \circ \Phi_p^{-1} \end{aligned}$$

   is differentiable.

We say that (Φ, ϕ, U) is a local representation for the bundle $\pi : E \longrightarrow M$.

Remark 3.13. If $\pi : E \longrightarrow M$ is a bundle as in Remark 3.12, where E has a differentiable structure, and (Φ, ϕ, U) is a local representation of $\pi : E \longrightarrow M$, then we have a local representation $(T\Phi, \Phi, \pi^{-1}(U))$ for the tangent bundle $\tau_E : TE \longrightarrow E$:

$$\begin{array}{ccc} T\pi^{-1}(U) & \xrightarrow{T\Phi} & \phi(U) \times \mathbb{E} \times \mathbb{H} \times \mathbb{E} \\ {\scriptstyle\tau_E}\downarrow & & \downarrow{\scriptstyle\mathrm{pr}_{1,2}} \\ \pi^{-1}(U) & \xrightarrow{\Phi} & \phi(U) \times \mathbb{E}. \end{array}$$

We consider the following special induced bundle.

Remark 3.14. Let $\pi : E \longrightarrow M$ be a bundle. We denote $\mathbb{S}^1 := [0, 2\pi]/\{0, 2\pi\}$ (that is the interval $[0, 2\pi]$ where the two extremes are identified, i.e the $1-$dimensional unit sphere) and let $c : \mathbb{S}^1 \longrightarrow M$ be a differentiable map. We have an induced bundle $c^*\pi : c^*E \longrightarrow \mathbb{S}^1$, with a commutative diagram

$$\begin{array}{ccc} c^*E & \xrightarrow{\pi^*c} & E \\ \downarrow{\scriptstyle c^*\pi} & & \downarrow{\scriptstyle \pi} \\ \mathbb{S}^1 & \xrightarrow{c} & M \end{array} \tag{3.1}$$

where $c^*E := \{(t, e) \in \mathbb{S}^1 \times E \,:\, c(t) = \pi(e)\}$, $c^*\pi$ is the projection on $\mathbb{S}^1$ and π^*c is the projection on E.

From now on, when not specified, M will be a compact Euclidean Riemannian manifold with a metric $\langle\,,\,\rangle$ and ∇ will be its Levi-Civita connection.

Definition 3.15. We define the space of closed curve on M with finite energy:

$$H^1(\mathbb{S}^1; M) := \{c : \mathbb{S}^1 \longrightarrow M |\, c \,\text{is absolute continuous and}\, \int_0^{2\pi} \langle \dot{c}(t), \dot{c}(t)\rangle_{c(t)}\, \mathrm{d}t < \infty\}.$$

Remark 3.16. We have the following inclusion

$$C^\infty(\mathbb{S}^1; M) \subset H^1(\mathbb{S}^1; M) \subset C^0(\mathbb{S}^1; M),$$

where we can endow $C^0(\mathbb{S}^1; M)$ with the compact-open topology.

Definition 3.17. As in (3.1), if $c \in C^\infty(\mathbb{S}^1; M)$, we can consider the following commutative diagram:

$$\begin{array}{ccc} c^*TM & \xrightarrow{\pi_M^*c} & TM \\ \downarrow{\scriptstyle c^*\pi_M} & & \downarrow{\scriptstyle \pi_M} \\ \mathbb{S}^1 & \xrightarrow{c} & M \end{array}$$

We define $H^0(c^*TM)$ as the set of square-integrable sections $\xi : \mathbb{S}^1 \longrightarrow c^*TM$ in the bundle $c^*\pi_M$; setting

$$\langle \xi, \xi\rangle_0 := \int_0^{2\pi} \langle (\pi_M^* c \circ \xi)(t), (\pi_M^* c \circ \xi)(t)\rangle_{c(t)}\, \mathrm{d}t < \infty,$$

we endow $H^0(c^*TM)$ with a structure of Hilbert space.

Then we define $H^1(c^*TM)$ as the set of those sections $\xi \in H^0(c^*TM)$ such that $\nabla_{\dot{c}}(\pi_M^* c \circ \xi)(t)$ exists for almost every $t \in \mathbb{S}^1$ and

$$\langle \xi, \xi \rangle_1 := \int_0^{2\pi} \langle (\pi_M^* c \circ \xi)(t), (\pi_M^* c \circ \xi)(t) \rangle_{c(t)} \, \mathrm{d}t + \\ + \int_0^{2\pi} \langle \nabla_{\dot{c}}(\pi_M^* c \circ \xi)(t), \nabla_{\dot{c}}(\pi_M^* c \circ \xi)(t) \rangle_{c(t)} \, \mathrm{d}t < \infty.$$

Also $H^1(c^*TM)$ has a structure of Hilbert space with inner product given by $\langle \, , \, \rangle_1$.

We recall the following lemma from Riemannian geometry.

Lemma 3.18. *Let M be a compact Euclidean Riemannian manifold and $\tau_M : TM \longrightarrow M$ the canonical projection. For $\varepsilon > 0$ let*

$$\mathcal{O}_\varepsilon := \{\xi \in TM \, ; \, \|\xi\|_{\tau_M(\xi)} < \varepsilon\}.$$

Then there exists $\varepsilon > 0$ such that

$$\begin{aligned} (\tau_M, \exp) : \mathcal{O}_\varepsilon &\longrightarrow M \times M \\ \xi &\longmapsto (\tau_M(\xi), \exp(\xi)) \end{aligned}$$

is a diffeomorphism onto an open neighbourhood of $M \times M$.

For a proof of Lemma 3.18 see the first part of the proof of Lemma 5.12. on [6].

Lemma 3.19. *Let M be a compact Euclidean Riemannian manifold, $\pi_M : TM \longrightarrow M$ and $\mathcal{O} \equiv \mathcal{O}_\varepsilon$ as in Lemma 3.18. Let $c \in C^\infty(\mathbb{S}^1; M)$ and $\mathcal{O}_c := \pi_M^* c^{-1}(\mathcal{O}) \subset c^*TM$, that is an open neighbourhood of the $0-$section of $c^*\pi_M$. We define*

$$H^1(\mathcal{O}_c) := \{\xi \in H^1(c^*TM) \, ; \, \xi(t) \in \mathcal{O}_c \text{ for all } t \in \mathbb{S}^1\}.$$

Then $H^1(\mathcal{O}_c)$ is open in $H^1(c^*TM)$.

The previous lemma is a consequence of Proposition 1.2.3 on [5].

Definition 3.20. Under the same assumptions of Lemma 3.19, we can consider the following map:

$$\begin{aligned} \exp_c : H^1(\mathcal{O}_c) &\longrightarrow H^1(\mathbb{S}^1; M) \\ (t \mapsto \xi(t)) &\longmapsto (t \mapsto \exp((\pi_M^* c \circ \xi)(t))) \, . \end{aligned}$$

This is an injective map and

$$\exp_c(H^1(\mathcal{O}_c)) = \{e \in H^1(\mathbb{S}^1; M)\,;\, e(t) \in \exp(\mathcal{O} \cap T_{c(t)}M)\}.$$

We put $\mathcal{U}(c) = \exp_c(H^1(\mathcal{O}_c))$.

Lemma 3.21. *Let M be a compact Euclidean Riemannian manifold, $\pi : TM \longrightarrow M$ and $\mathcal{O} \equiv \mathcal{O}_\varepsilon$ as in Lemma 3.18. If $c, d \in C^\infty(\mathbb{S}^1, M)$, then*

$$\exp_d^{-1} \circ \exp_c \,:\, \exp_c^{-1}(\mathcal{U}(c) \cap \mathcal{U}(d)) \longrightarrow \exp_d^{-1}(\mathcal{U}(c) \cap \mathcal{U}(d))$$

is a diffeomorphism.

For a proof of Lemma 3.21 see Lemma 1.2.8 on [5].

Now, assuming the previous results, we can see that $H^1(\mathbb{S}^1; M)$ can be endowed with a structure of Hilbert manifold.

Theorem 3.22. *If M is a compact connected Euclidean Riemannian manifold, then $H^1(\mathbb{S}^1; M)$ is a Hilbert manifold. The differentiable structure is given by the natural atlas*

$$(\exp_c^{-1}, \mathcal{U}(c))_{c \in C^\infty(\mathbb{S}^1; M)}.$$

Proof.

1. For any $c \in \mathbb{C}^\infty(\mathbb{S}^1; M)$ the corresponding chart has its image in the Hilbert space $H^1(c^*TM)$.

2. If $d \in H^1(\mathbb{S}^1; M)$, then we can approximate it by $c \in C^\infty(\mathbb{S}^1; M)$ in the metric given by
$$\mathrm{d}_\infty(c, c') := \sup_{t \in \mathbb{S}^1} \mathrm{d}_M(c(t), c'(t)),$$
(where d_M is the Riemannian distance on M of Definition 1.23) so that $d \in \mathcal{U}(c)$. This shows that $(\mathcal{U}(c))_{c \in C^\infty(\mathbb{S}^1; M)}$ is an open covering of $H^1(\mathbb{S}^1; M)$.

3. It follows from Lemma 3.21 that the natural atlas is of class C^∞. We want to show that the natural atlas has a countable subatlas.

   To do this, for each integer $l > 0$ we define
$$H^1(\mathbb{S}^1; M)^l := \left\{ c \in H^1(\mathbb{S}^1; M)\,;\, \int_0^{2\pi} \left\langle \dot{c}(t), \dot{c}(t) \right\rangle_{c(t)} \mathrm{d}t < 2l \right\}.$$

Let $\varepsilon > 0$ as in Lemma 3.18 and let $m = m(l, \varepsilon)$ be an integer for which $4\pi l < \frac{m\varepsilon^2}{9}$. Let $e \in H^1(\mathbb{S}^1; M)^l$ and set $e_j := e(2\pi\frac{j}{m})$ for any $j = 1, \ldots, m$; then we have

$$\mathrm{d}_M(e_{j-1}, e_j)^2 \leq \left(\int_{2\pi\frac{j-1}{m}}^{2\pi\frac{j}{m}} \|\dot{e}(t)\|_{e(t)}\,\mathrm{d}t\right)^2 \leq \frac{2\pi}{m}\int_0^{2\pi} \|\dot{e}(t)\|^2_{e(t)}\,\mathrm{d}t < \frac{4\pi l}{m} < \frac{\varepsilon^2}{9}.$$

It follows that $e|_{[2\pi\frac{j-1}{m}, 2\pi\frac{j}{m}]}$ lies entirely in a ball or radius $\frac{\varepsilon}{3}$ contained in M. If $B(p, \frac{\varepsilon}{3}) := \{q \in M \,:\, \mathrm{d}_M(p,q) < \frac{\varepsilon}{3}\}$, then $\{B(p, \frac{\varepsilon}{3})\}_{p\in M}$ is an open covering of M and it admits a finite sub-covering: there exists a finite set P of points in M such that $\{B(p, \frac{\varepsilon}{3})\}_{p\in P}$ covers M. In particular, there exists a finite sequence $\{p_1, \ldots, p_m\} \subset P$ such that $e_j \in B(p_j, \frac{\varepsilon}{3})$ for any $j = 1, \ldots, m$. For each such sequence of m elements (there exists a finite number of these sequences, since P is finite) we can find $c \in C^\infty(\mathbb{S}^1; M) \cap H^1(\mathbb{S}^1; M)^l$ such that $c(2\pi\frac{j}{m}) = p_j$; then $e \in \mathcal{U}(c)$ for one of these $c's$.

Summarizing, for any integer $l > 0$ we can find a finite covering of $H^1(\mathbb{S}^1; M)^l$: the union of all these coverings is a countable covering for $H^1(\mathbb{S}^1; M)$.

□

3.3 The Birkhoff's Theorem

Theorem 3.23 (**Birkhoff**). *On any compact surface $S \subset \mathbb{R}^3$ which is C^3-diffeomorphic to the 2-dimensional unit sphere there exists a non-constant closed geodesic.*

Proof. Step 1*: Introduction to the energy functional and the space of closed curve on S with finite energy.*

Let

$$H^{1,2}(\mathbb{S}^1; S) := \{u : \mathbb{R} \longrightarrow \mathbb{R}^3 : \, u|_{(0,2\pi)} \in H^{1,2}((0,2\pi); \mathbb{R}^3),$$
$$u(t) = u(t + 2\pi),\, u(t) \in S \text{ for almost every } t \in \mathbb{R}\}$$

be the space of closed curves on S with finite energy

$$E(u) := \frac{1}{2}\int_0^{2\pi} |\dot{u}|^2\,\mathrm{d}t.$$

For any $t, s \in [0, 2\pi]$ we have

$$|u(s) - u(t)| \leq \int_{\min\{s,t\}}^{\max\{s,t\}} |\dot{u}| \, \mathrm{d}\tau \leq \left(|t-s| \int_{\min\{s,t\}}^{\max\{s,t\}} |\dot{u}|^2 \, \mathrm{d}\tau \right)^{1/2} \leq (2|t-s|E(u))^{1/2}, \quad (3.2)$$

where we have applied the Holder's inequality, whence

$$\sup_{\substack{s,t \in [0,2\pi], \\ s \neq t}} \frac{|u(s) - u(t)|}{|t-s|^{1/2}} \leq \sqrt{2E(u)}. \quad (3.3)$$

By (3.3), it follows that, if $u \in H^{1,2}(\mathbb{S}^1; S)$ with $E(u) \leq \gamma$, then $u \in C^{0,1/2}(\mathbb{S}^1; S)$ with $\|u\|_{C^{0,1/2}} \leq \sqrt{2\gamma}$. By Theorem 3.22, $H^{1,2}(\mathbb{S}^1; S)$ is a connected[1] C^2−complete Hilbert submanifold of the Hilbert space $H^{1,2}(\mathbb{S}^1; \mathbb{R}^3)$ and the tangent space is

$$T_u H^{1,2}(\mathbb{S}^1; S) = \{\varphi \in H^{1,2}(\mathbb{S}^1; \mathbb{R}^3) \, ; \, \varphi(t) \in T_{u(t)} S \cong \mathbb{R}^2\},$$

for any $u \in H^{1,2}(\mathbb{S}^1; S)$.[2] Moreover, $H^{1,2}(\mathbb{S}^1; S)$ has a structure of Finsler manifold.[3] We also have, by (3.2), that, if $E(u)$ is small enough, then the image of u is covered by a single coordinate chart of S and

$$T_u H^{1,2}(\mathbb{S}^1; S) \cong H^{1,2}(\mathbb{S}^1; \mathbb{R}^2).$$

Step 2: *E is a C^1 functional on $H^{1,2}(\mathbb{S}^1; S)$ and its critical points are exactly the closed geodesics on S.*

E is a C^1 functional on the Hilbert space $H^{1,2}(\mathbb{S}^1; \mathbb{R}^3)$: indeed, we can see that for any $u, \varphi \in H^{1,2}(\mathbb{S}^1; \mathbb{R}^3)$ there exists the Gâteaux derivative

$$\partial_\varphi E(u) = \int_0^{2\pi} \langle \dot{u}, \dot{\varphi} \rangle \, \mathrm{d}t,$$

[1] S is simply connected; in particular, if $p \in S$ is fixed, we can contract each closed curve $u \in H^{1,2}(\mathbb{S}^1; S)$ to the constant curve $u_c(t) = p$ via a map $H_u : \mathbb{R} \times [0,1] \longrightarrow S$ such that $H(\cdot, 0) = u$, $H(\cdot, 1) = u_c$ and $H(\cdot, s) \in H^{1,2}(\mathbb{S}^1; S)$ for any $s \in [0,1]$.

[2] The elements of $T_u H^{1,2}(\mathbb{S}^1; S)$ are given by $\varphi = \frac{\mathrm{d}}{\mathrm{d}s} \Gamma\big|_{s=0}$, where $\Gamma : (-\varepsilon, \varepsilon) \longrightarrow H^{1,2}(\mathbb{S}^1; S)$ is a differentiable map such that $\Gamma(0) = u$. We can consider $\tilde{\Gamma} : (-\varepsilon, \varepsilon) \times \mathbb{R} \longrightarrow S$ given by $\tilde{\Gamma}(s,t) = \Gamma(s)(t)$: then we have that $\tilde{\Gamma}(0,t) = u(t)$ and $\frac{\partial}{\partial s}(\tilde{\Gamma}(s,t))\big|_{s=0} = \varphi(t)$, meaning that $\varphi(t) \in T_{u(t)} S$ for any t.

[3] $H^{1,2}(\mathbb{S}^1; S)$ inherits a Riemannian structure from $H^{1,2}(\mathbb{S}^1; \mathbb{R}^3)$ (see Remark 3.11). Then the Riemannian structure on $H^{1,2}(\mathbb{S}^1; S)$ defines a natural Finsler structure on $TH^{1,2}(\mathbb{S}^1; S)$ by Remark 2.53.

the map $\partial_{(\cdot)}E(u) \in L(H^{1,2}(\mathbb{S}^1;\mathbb{R}^3),\mathbb{R})$ and the map $u \longmapsto \partial_{(\cdot)}E(u)$ is continuous on $H^{1,2}(\mathbb{S}^1;\mathbb{R}^3)$, therefore we can apply Proposition 2.7. It follows that the restriction of E to the submanifold $H^{1,2}(\mathbb{S}^1;S)$ is still of class C^1.

We want to show that, if $u \in H^{1,2}(\mathbb{S}^1;S)$ is a critical point for E, then it is a geodesic. If $u \in C^2(\mathbb{S}^1;S)$ is a critical point for E, upon integrating by parts, we have

$$0 = \int_0^{2\pi} \langle \dot{u}, \dot{\varphi} \rangle \,\mathrm{d}t = -\int_0^{2\pi} \langle \ddot{u}, \varphi \rangle \,\mathrm{d}t, \quad \forall \varphi \in T_u H^{1,2}(\mathbb{S}^1;S). \tag{3.4}$$

Since

$$L^2(\mathbb{S}^1;\mathbb{R}^3) = T_u H^{1,2}(\mathbb{S}^1;S) \oplus (T_u H^{1,2}(\mathbb{S}^1;S))^{\perp},$$

by (3.4), we have that $\ddot{u} \in (T_u H^{1,2}(\mathbb{S}^1;S))^{\perp}$, whence $\ddot{u}(t) \perp T_{u(t)}S$ for all t, meaning that u is a geodesic on S, parameterized by arc length. Now, let $u \in H^{1,2}(\mathbb{S}^1;S)$ be a critical point for E and let $n : S \longrightarrow \mathbb{R}^3$ be one of the two (C^2-)unit normal vector fields on S. For any $\varphi \in H^{1,2}(\mathbb{S}^1;\mathbb{R}^3)$ we have

$$\varphi(t) - n(u(t)) \, \langle n(u(t)), \varphi(t) \rangle \in T_{u(t)}S, \quad \forall t \in \mathbb{R},$$

meaning that $\psi_\varphi := \varphi - n(u) \, \langle n(u), \varphi \rangle \in T_u H^{1,2}(\mathbb{S}^1;S)$, and

$$\begin{aligned}\int_0^{2\pi} \langle \dot{u}, \dot{\varphi} \rangle \,\mathrm{d}t &= \underbrace{\int_0^{2\pi} \left\langle \dot{u}, \dot{\psi}_\varphi \right\rangle \mathrm{d}t}_{=0} + \int_0^{2\pi} \langle \dot{u}, \langle n(u), \varphi \rangle \, Dn(u)\, \dot{u} \rangle \, dt + \\ &+ \underbrace{\int_0^{2\pi} \langle \dot{u}, (\langle Dn(u)\, \dot{u}, \varphi \rangle + \langle n(u), \dot{\varphi} \rangle) \; n(u) \rangle \,\mathrm{d}t}_{=0} = \int_0^{2\pi} \langle \langle \dot{u}, Dn(u)\, \dot{u} \rangle \; n(u), \varphi \rangle \,\mathrm{d}t.\end{aligned} \tag{3.5}$$

By (3.5), we obtain

$$\ddot{u} = -\langle \dot{u}, Dn(u)\, \dot{u} \rangle \; n(u). \tag{3.6}$$

By (3.6) and (3.4), we get that $u \in H^{2,2}(\mathbb{S}^1;S) \hookrightarrow C^1(\mathbb{S}^1;S)$. Thus, by (3.6), we conclude that $u \in C^2(\mathbb{S}^1;S)$ and it is a geodesic for the previous discussion.

Step 3*: The functional E is bounded from below, its sublevel sets are all complete and it satisfies the Palais-Smale condition on $H^{1,2}(\mathbb{S}^1;S)$.*

Clearly $E \geq 0$ and $E^c := E^{-1}((-\infty, c]) = E^{-1}([0, c])$ is complete for any $c \geq 0$, since $H^{1,2}(\mathbb{S}^1;S)$ is complete.

Let $\{u_m\}_{m\in\mathbb{N}} \subset H^{1,2}(\mathbb{S}^1; S)$ be a Palais-Smale sequence for E at level c:

$$\begin{cases} E(u_m) \overset{m\to\infty}{\longrightarrow} c, \\ \|\mathrm{d}_{u_m} E\| := \sup_{\substack{\varphi\in T_{u_m}H^{1,2}(\mathbb{S}^1;S) \\ \|\varphi\|_{H^{1,2}}\leq 1}} \left|\int_0^{2\pi} \dot{u_m}\, \dot{\varphi}\, \mathrm{d}t\right| \longrightarrow 0; \end{cases} \tag{3.7}$$

we aim to show that it contains a strongly convergent subsequence.

Clearly there exists $\gamma > 0$ such that $E(u_m) \leq \gamma$ for any m, therefore by (3.2) we have that $\{u_m\}_{m\in\mathbb{N}}$ is equi-continuous. Moreover, for any $m \in \mathbb{N}$ we have that $\sup_{t\in\mathbb{S}^1} \|u_m(t)\| \leq \sup_{x\in S} \|x\| \in \mathbb{R}$, being S compact, whence $\{u_m\}_{m\in N}$ is equi-bounded. By Ascoli-Arzelà theorem, we know there exists $u \in C(\mathbb{S}^1; S)$ such that $u_m \longrightarrow u$ uniformly (up to subsequence). In addition, $\{u_m\}_{m\in\mathbb{N}}$ is bounded in $H^{1,2}(\mathbb{S}^1;\mathbb{R}^3)$[4], therefore, up to subsequence, $u_m \rightharpoonup v$ weakly in $H^{1,2}(\mathbb{S}^1;\mathbb{R}^3)$ to some $v \in H^{1,2}(\mathbb{S}^1;\mathbb{R}^3)$. It follows that $v = u$ and $u \in H^{1,2}(\mathbb{S}^1; S)$. Via the unit normal vector field n we define the projection

$$\begin{aligned} \pi_v : H^{1,2}(\mathbb{S}^1;\mathbb{R}^3) &\longrightarrow T_v H^{1,2}(\mathbb{S}^1; S) \\ \varphi &\longmapsto \varphi - n(v)\,\langle n(v), \varphi\rangle\,. \end{aligned}$$

Let $\varphi_m := \pi_{u_m}(u_m - u) \in T_{u_m}H^{1,2}(\mathbb{S}^1; S)$: since $\{u_m\}_{m\in\mathbb{N}} \subset H^{1,2}(\mathbb{S}^1;\mathbb{R}^3)$ is bounded and n is of class C^2, then $\{\varphi_m\}_{m\in\mathbb{N}}$ is bounded in $H^{1,2}$. Thus, by the second equation of (3.7), we have that $\mathrm{d}_{u_m}E(\varphi_m) \overset{m\to\infty}{\longrightarrow} 0$. Moreover, since $u_m \overset{m\to\infty}{\longrightarrow} u$ uniformly and $u_m \rightharpoonup u$ weakly in $H^{1,2}$, then $\varphi_m \overset{m\to\infty}{\longrightarrow} 0$ uniformly and $\varphi_m \rightharpoonup 0$ weakly in $H^{1,2}$. Denoting by $o(1)$ a quantity that goes to 0 when $m \to \infty$, we have the following computation:

$$\begin{aligned} o(1) = \mathrm{d}_{u_m}E(\varphi) = \int_0^{2\pi} \langle \dot{u}_m, \dot{\varphi}_m\rangle\ \mathrm{d}t = \int_0^{2\pi} \langle \dot{u}_m - \dot{u}, \dot{\varphi}_m\rangle\ \mathrm{d}t + \underbrace{\int_0^{2\pi} \langle \dot{u}, \dot{\varphi}_m\rangle\ \mathrm{d}t}_{=o(1)} = \\ = \int_0^{2\pi} \left(\|\dot{u}_m - \dot{u}\|^2 - \left\langle \dot{u}_m - \dot{u}, \frac{\mathrm{d}}{\mathrm{d}t}[n(u_m)\,\langle n(u_m), u_m - u\rangle]\right\rangle\right) \mathrm{d}t + o(1) = \\ = \int_0^{2\pi} \|\dot{u}_m - \dot{u}\|^2 \mathrm{d}t - \underbrace{\int_0^{2\pi} \langle \dot{u}_m - \dot{u}, Dn(u_m)\dot{u}_m\rangle\,\langle n(u_m), u_m - u\rangle\ \mathrm{d}t}_{=o(1)} + \end{aligned}$$

[4] If it is not bounded, there exists $\{u_{m_k}\}_{k\in\mathbb{N}}$ such that $\|u_{m_k}\|_{H^{1,2}} = \sqrt{\|u_{m_k}\|^2_{L^2} + \|\dot{u}_{m_k}\|^2_{L^2}} \overset{k\to\infty}{\longrightarrow} +\infty$, from which $\|\dot{u}_{m_k}\|^2_{L^2} \overset{k\to\infty}{\longrightarrow} +\infty$, meaning that $E(u_{m_k}) \overset{k\to\infty}{\longrightarrow} +\infty$, which is not possible.

$$-\underbrace{\int_0^{2\pi} \langle \dot{u}_m - \dot{u}, n(u_m)\rangle \ \langle Dn(u_m)\dot{u}_m, u_m - u\rangle \ \mathrm{d}t}_{=o(1)} - \int_0^{2\pi} \langle \dot{u}_m - \dot{u}, n(u_m)\rangle \ \langle n(u_m), \dot{u}_m - \dot{u}\rangle \ \mathrm{d}t +$$

$$+ o(1) = \int_0^{2\pi} \left(\|\dot{u}_m - \dot{u}\|^2 - |\langle \dot{u}_m - \dot{u}, n(u_m)\rangle|^2 \ \mathrm{d}t \right) + o(1). \quad (3.8)$$

Since

$$\langle n(u_m), \dot{u}_m\rangle = 0 = \langle n(u), \dot{u}\rangle ,$$

it follows that

$$\int_0^{2\pi} |\langle \dot{u}_m - \dot{u}, n(u_m)\rangle|^2 \ \mathrm{d}t = \int_0^{2\pi} |\langle \dot{u}, n(u_m)\rangle|^2 \ \mathrm{d}t = \int_0^{2\pi} |\langle \dot{u}, n(u_m) - n(u)\rangle|^2 \ \mathrm{d}t \leq$$

$$\leq \int_0^{2\pi} \|\dot{u}\|^2 \ \|n(u_m) - n(u)\|^2 \, \mathrm{d}t \leq 2 \sup_{t\in[0,2\pi]} \|n(u_m(t)) - n(u(t))\|^2 E(u) \overset{m\to\infty}{\longrightarrow} 0.$$

Thus, (3.8) becomes

$$\int_0^{2\pi} \|\dot{u}_m - \dot{u}\|^2 \, \mathrm{d}t = o(1),$$

whence $u_m \longrightarrow u$ strongly in $H^{1,2}(\mathbb{S}^1; S)$.

Step 4: Construction of a family $\mathcal{F}$ of subsets of $H^{1,2}(\mathbb{S}^1; S)$ that is $\Phi-$invariant with respect to any semi-flow Φ satisfying the hypotheses of Theorem 2.67.

By hypothesis, there exists a C^3-diffeomorphism $\Psi : S \longrightarrow \mathbb{S}^2$. Let us consider $p : [-\frac{\pi}{2}, \frac{\pi}{2}] \longrightarrow H^{1,2}(\mathbb{S}^1; S)$, a 1-parameter family of closed curves $u = p(\theta) \in H^{1,2}(\mathbb{S}^1; S)$ such that $p(\pm\frac{\pi}{2})$ are constant curves. We have an identification between $\mathbb{S}^2$ and $\big((-\frac{\pi}{2}, \frac{\pi}{2}) \times [0, 2\pi[\big) \cup \{(-\frac{\pi}{2}; 0), (\frac{\pi}{2}; 0)\}$, via the continuous map

$$\Gamma : \left(\left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \times [0, 2\pi[\right) \cup \left\{\left(-\frac{\pi}{2}; 0\right), \left(\frac{\pi}{2}; 0\right)\right\} \longrightarrow \mathbb{S}^2$$

$$(\theta, \phi) \longmapsto \begin{pmatrix} \cos(\theta)\cos(\phi) \\ \cos(\theta)\sin(\phi) \\ \sin(\theta) \end{pmatrix}$$

We set $\tilde{p}(\theta, \phi) := \Psi(p(\theta)(\phi))$ and define

$$P := \left\{ p \in C^0\left(\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]; H^{1,2}(\mathbb{S}^1; S)\right) \ ; \ p\left(\pm\frac{\pi}{2}\right) \equiv \text{const} \in S \right\},$$

$$\mathcal{F} := \left\{ \{p(\theta)\}_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]},\ p \in P,\ \tilde{p} \circ \Gamma^{-1} \text{ is homotopic to } id|_{\mathbb{S}^2} \right\}.$$

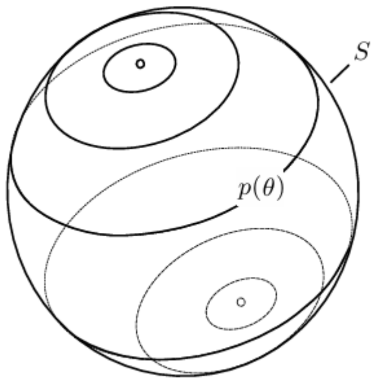

Figure 3.1: An element of $\mathcal{F}$ is as a collection of closed curves on S $\{p(\theta)\}_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]}$; in particular, the two curves corresponding to the values $\theta = \pm\frac{\pi}{2}$ reduces to points. See Fig. 4.1 on [9].

$\mathcal{F}$ is not empty[5]. We also note that $P \ni p \longmapsto \tilde{p} \in C^0([-\frac{\pi}{2},\frac{\pi}{2}] \times [0,2\pi[;\mathbb{S}^2)$ is a continuous map.

Moreover, if $\{p(\theta)\}_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]} \in \mathcal{F}$ and Φ is a homeomorphism of $H^{1,2}(\mathbb{S}^1;S)$ homotopic to the identity and mapping constant maps to constant maps, then $\{\Phi(p(\theta))\}_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]} \in \mathcal{F}$[6]

Now, if $\Phi(s,\cdot)$, $s \geq 0$, is a family of homeomorphisms of $H^{1,2}(\mathbb{S}^1;S)$ such that $\Phi(0,\cdot) = id|_{H^{1,2}(\mathbb{S}^1;S)}$ and $s \longmapsto E(\Phi(s,u))$ is non-increasing for any $u \in H^{1,2}(\mathbb{S}^1;S)$, then each $\Phi(s,\cdot)$ maps constant maps to constant maps. Indeed, if $u \in H^{1,2}(\mathbb{S}^1;S)$ is a

[5]For any $\theta \in [-\frac{\pi}{2},\frac{\pi}{2}]$ let $\gamma_\theta : \mathbb{R} \longrightarrow \mathbb{S}^2$ given by $\gamma_\theta(\phi) := \begin{pmatrix}\cos(\theta)\cos(\phi)\\ \cos(\theta)\sin(\phi)\\ \sin(\theta)\end{pmatrix}$ and $p(\theta)(\phi) := \Psi^{-1}(\gamma_\theta(\phi))$; then $p \in C^0\big([-\frac{\pi}{2},\frac{\pi}{2}];H^{1,2}(\mathbb{S}^1;S)\big)$, $p(\pm\frac{\pi}{2})$ are constant curves and $\tilde{p}\circ\Gamma^{-1} = id|_{\mathbb{S}^2}$, whence $\{p(\theta)\}_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]} \in \mathcal{F}$.

[6]Consider a continuous map $\delta : [0,1] \times H^{1,2}(\mathbb{S}^1;S) \longrightarrow H^{1,2}(\mathbb{S}^1;S)$ such that $\delta(0,u) = u$ and $\delta(1,u) = \Phi(u)$ for any $u \in H^{1,2}(\mathbb{S}^1;S)$. Clearly $\Phi\circ p \in C^0([-\frac{\pi}{2},\frac{\pi}{2}];H^{1,2}(\mathbb{S}^1;S))$ and $(\Phi\circ p)(\pm\frac{\pi}{2})$ is a constant map; moreover, if $\Gamma^{-1}(v) = (\Gamma_1^{-1}(v),\Gamma_2^{-1}(v))$ for $v \in \mathbb{S}^2$, then

$$\tilde{\eta} : [0,1] \times \mathbb{S}^2 \longrightarrow \mathbb{S}^2,$$

given by

$$\tilde{\eta}(t,v) = \Psi\big(\delta(t,p(\Gamma_1^{-1}(v)))(\Gamma_2^{-1}(v))\big),$$

is a continuous map such that $\tilde{\eta}(0,\cdot) = \tilde{p}\circ\Gamma^{-1}$ and $\tilde{\eta}(1,\cdot) = \widetilde{(\Phi\circ p)}\circ\Gamma^{-1}$, i.e. $\widetilde{(\Phi\circ p)}\circ\Gamma^{-1}$ is homotopic to $\tilde{p}\circ\Gamma^{-1}$ that is homotopic to $id|_{\mathbb{S}^2}$.

constant map, then for any $s \geq 0$

$$0 \leq E(\Phi(s,u)) \leq E(\Phi(0,u)) = E(u) = 0,$$

from which $\Phi(s,u)$ is a constant curve on S, being its energy 0. It follows that $\mathcal{F}$ is $\Phi-$invariant with respect to any semi-flow Φ as in Theorem 2.67.

Step 5*: Application to the Minimax Principle and proof of the fact that the Minimax value is not* 0.

Let

$$\beta := \inf_{\{p(\theta)\}_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]}\in\mathcal{F}} \sup_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]} E(p(\theta)),$$

then $\beta \in \mathbb{R}$ (since $\mathcal{F}$ is not empty and $\sup_\theta E(p(\theta))$ is actually $\max_\theta E(p(\theta))$ for any $\{p(\theta)\}_\theta \in \mathcal{F}$). By Theorem 2.67, we know that β is a critical value for E, i.e. there exists $u \in \mathbb{H}^{1,2}(\mathbb{S}^1;S)$ such that $E(u) = \beta$ and $\mathrm{d}_u E = 0$, meaning that u is a closed geodesic. We need to check that $\beta > 0$ to rule out the possibility that u is a constant curve.

There exists $\delta > 0$ such that for any $x \in F_\delta := \{x \in \mathbb{R}^3; \operatorname{dist}(x,S) \leq \delta\}$ there is a unique $\pi(x) \in S$ satisfying

$$\|\pi(x) - x\| = \inf_{y\in S} \|x-y\|,$$

and $\pi : F_\delta \longrightarrow S$ is of class C^2 (because the unit-normal vector field to S is of class C^2). We have

$$diam(u) := \sup_{0\leq\phi,\phi'\leq 2\pi} \|u(\phi) - u(\phi')\| \overset{(3.2)}{\leq} \sup_{0\leq\phi,\phi'\leq 2\pi} (2|\phi-\phi'|E(u))^{1/2} \leq \delta, \tag{3.9}$$

for any $u \in H^{1,2}(\mathbb{S}^1;S)$ such that $E(u) \leq \gamma := \frac{\delta^2}{4\pi}$. We aim to show that $\beta \geq \gamma$.

By contradiction, we assume $\beta < \gamma$. We can choose an element $\{p(\theta)\}_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]} \in \mathcal{F}$ such that $E(p(\theta)) \leq \gamma$ for any $\theta \in [-\frac{\pi}{2},\frac{\pi}{2}]$.[7] Let $\phi_0 \in [0,2\pi]$ be fixed and define $\rho_\theta : [0,1]\times[0,2\pi] \longrightarrow F_\delta$ by

$$\rho_\theta(s,\phi) := (1-s)p(\theta)(\phi) + sp(\theta)(\phi_0).$$

[7] If $\{p(\theta)\}_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]} \in \mathcal{F}$ such that $E(p(\tilde{\theta})) > \gamma$ for some $\tilde{\theta} \in [-\frac{\pi}{2},\frac{\pi}{2}]$, let $M := \sup_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]} E(p(\theta))$. Then $\{\sqrt{\frac{\gamma}{M}}p(\theta)\}_{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]} \in \mathcal{F}$ (because the multiplication for a positive constant is a homeomorphism of $H^{1,2}(\mathbb{S}^1;S)$ homotopic to the identity and which maps constant maps to constant maps) and $E\left(\sqrt{\frac{\gamma}{M}}p(\theta)\right) \leq \gamma$ for any $\theta \in [-\frac{\pi}{2},\frac{\pi}{2}]$.

We note that the image of ρ_θ is actually contained in F_δ: for any $(s,\phi) \in [0,1] \times [0,2\pi]$ we have

$$\operatorname{dist}(\rho_\theta(s,\phi), S) \leq \|\rho_\theta(s,\phi) - \underbrace{p(\theta)(\phi)}_{\in S}\| = |s|\, \|p(\theta)(\phi) - p(\theta)(\phi_0)\| \overset{(3.9)}{\leq} \delta.$$

Since $\rho_\theta(0,\cdot) = p(\theta)(\cdot)$ and $\rho_\theta(1,\cdot) = p(\theta)(\phi_0)$, we have an homotopy between the curve $p(\theta)(\cdot)$ and the constant curve $p(\theta)(\phi_0)$. If we compose it with π and consider

$$\mu(s,\theta,\phi) := \pi(\rho_\theta(s,\phi)), \qquad s \in [0,1],\ \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}],\ \phi \in [0,2\pi],$$

we have an homotopy between the map $[-\frac{\pi}{2}, \frac{\pi}{2}] \ni \theta \longmapsto \mu(0,\theta,\cdot) = p(\theta)$ and the map $[-\frac{\pi}{2}, \frac{\pi}{2}] \ni \theta \longrightarrow \mu(1,\theta,\cdot) =: p_1(\theta)$, where $p_1(\theta)(\phi) := p(\theta)(\phi_0)$ for any $\phi \in [0,2\pi]$. Hence, $p, p_1 \in P$ are homotopic. Furthermore, considering the composition

$$\sigma(s,\theta,\phi) := \Psi(\mu(s,\theta,\phi)) = \Psi\big(\pi(\rho_\theta(s,\phi))\big), \qquad s \in [0,1],\ \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}],\ \phi \in [0,2\pi],$$

we have an homotopy between the map $\tilde{p} = \sigma(0,\cdot,\cdot)$ and $\tilde{p}_1 = \sigma(1,\cdot,\cdot)$. Therefore $\tilde{p}_1 \circ \Gamma^{-1}$ is homotopic to $\tilde{p} \circ \Gamma^{-1}$ that is homotopic to $id|_{\mathbb{S}^2}$. Finally, defining

$$\tau(r,\theta,\phi) := \tilde{p}_1(r\theta,\phi) = \Psi(p_1(r\theta)(\phi)) = \Psi(p(r\theta)(\phi_0)), \quad r \in [0,1],\ \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}],\ \phi \in [0,2\pi],$$

we have an homotopy between $\tilde{p}_1 = \tau(1,\cdot,\cdot)$ and the constant map $\tau(0,\cdot,\cdot)$. It follows that $\tilde{p}_1 \circ \Gamma^{-1}$ is homotopic to a constant map, whence $id|_{\mathbb{S}^2}$ is homotopic to a constant map, which is not possible. $\square$

Bibliography

[1] W.M. Boothby; *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, Orlando, second edition, 1986

[2] M.P. do Carmo; *Riemannian Geometry*, Birkhäuser, Boston, 1992

[3] L. Gasinski, N.S. Papageorgiou; *Nonlinear Analysis*, Chapman & Hall/CRC, 2005

[4] D. Gilbarg, N.S.Trudinger; *Elliptic Partial Differential Equations of Second Order* Springer-Verlag, Berlin Heidelberg, 2001

[5] W. Klingenberg; *Lectures on Closed Geodesics*, Springer-Verlag, Berlin Heidelberg, 1978

[6] J.M. Lee; *Riemannian Manifolds: An Introduction to Curvature*, Springer-Verlag, New York, 1997

[7] R.S. Palais; *Critical Point Theory and the Minimax Principle*. Proc. Symp. Pure Math., Vol. XV, 185-212, 1970

[8] A.H.Stone; *Paracompactness and product spaces*. Bull. Amer. Math. Soc. Volume 54, Number 10, 977-982, 1948

[9] M. Struwe; *Variational Methods: Applications to Nonlinear Partial Differential Equations and Hamiltonian Systems*, Fourth Edition, Springer-Verlag, Berlin Heidelberg, 2008