1. The Isoperimetric Problem

We want to prove that the circle is the figure with maximum area among those with given perimeter. More precisely, we are going to prove that, given a piecewise regular[1] C^1 Jordan curve γ in $\mathbb{R}^2$ of lenght L, then the area A of the bounded region enclosed by γ is not greater than the area of the circle of perimeter L. The result is valid in the more general case where the hypothesis *piecewise regular* C^1 is replaced by *continuous*, but the proof is more complicated. First of all, we evaluate the area S of the circle of perimeter L. The radius is given by $\frac{L}{2\pi}$, thus $S = \pi(\frac{L}{2\pi})^2 = \frac{L^2}{4\pi}$. Hence, we have to prove

$$L^2 \geq 4\pi A \tag{1.1}$$

In order to prove (1.1), we recall some properties of Fourier series.

Theorem 1.1. *If f is a Riemann integrable function on $[-\pi, \pi]$, then*

$$\int_{-\pi}^{\pi} f^2(t)\, dt = \pi\Big(\frac{a_0^2}{2} + \sum_{n=1}^{+\infty}(a_n^2 + b_n^2)\Big)$$

where a_n, $n = 0, 1, 2...$ and b_n, $n = 1, 2, 3, ...$ are the Fourier coefficients of f defined by

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(nt)\, dt, \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(nt)\, dt$$

Theorem 1.2. *If f and g are Riemann integrable functions on $[-\pi, \pi]$, then*

$$\int_{-\pi}^{\pi} f(t)g(t)\, dt = \pi\Big(\frac{a_0 c_0}{2} + \sum_{n=1}^{+\infty}(a_n c_n + b_n d_n)\Big)$$

where a_n, $n = 0, 1, 2...$ and b_n, $n = 1, 2, 3, ...$ are the Fourier coefficients of f, and c_n, $n = 0, 1, 2...$ and d_n, $n = 1, 2, 3, ...$ are the Fourier coefficients of g.

Theorem 1.3. *If f is a piecewise C^1 function on $[-\pi, \pi]$ such that $f(-\pi) = f(\pi)$, and a_n, $n = 0, 1, 2...$ and b_n, $n = 1, 2, 3, ...$ are the Fourier coefficients of f, and c_n, $n = 0, 1, 2...$ and d_n, $n = 1, 2, 3, ...$ are the Fourier coefficients of f', we have $c_n = nb_n$, $d_n = -na_n$.*

Theorem 1.1 is a known result in Fourier series. Who is familiar only with the case where f is piecewise C^1, can restrict our considerations to the case where the curve γ is piecewise C^2. Note that under the hypothesis of Theorem 1.1 we are not sure that the Fourier series of f converges to f (pointwise), but nevertheless the Fourier coefficients can be however

[1] regular means $\gamma' \neq 0$

defined. Theorem 1.2 follows from Theorem 1.1, as in the hypothesis of Theorem 1.2 $f+g$ is Riemann integrable so that

$$2\int_{-\pi}^{\pi} f(t)g(t)\,dt = \int_{-\pi}^{\pi} (f(t)+g(t))^2 - f(t)^2 - g(t)^2\,dt =$$

$$\int_{-\pi}^{\pi} (f(t)+g(t))^2\,dt - \int_{-\pi}^{\pi} f(t)^2\,dt - \int_{-\pi}^{\pi} g(t)^2\,dt =$$

$$\pi\Big(\frac{(a_0+c_0)^2}{2} + \sum_{n=1}^{+\infty}((a_n+c_n)^2 + (b_n+d_n)^2)\Big)$$

$$-\pi\Big(\frac{a_0^2}{2} + \sum_{n=1}^{+\infty}(a_n^2+b_n^2)\Big) - \pi\Big(\frac{c_0^2}{2} + \sum_{n=1}^{+\infty}(c_n^2+d_n^2)\Big)$$

$$= \pi\Big(a_0c_0 + \sum_{n=1}^{+\infty}(2a_nc_n + 2b_nd_n)\Big)\,.$$

Theorem 1.3 directly follows from the definition of the Fourier coefficients using a partial integration, taking into account that the by the hypothesis, the limit terms at $-\pi$ and at π are equal. We are now going to prove (1.1). Let $\gamma = (x,y)$. We can assume that γ is parametrized by arc lenght, in particular $\gamma : [0,L] \to \mathbb{R}^2$. In order to use the theory of Fourier series, we reparametrize it in such a way that it is defined on $[-\pi,\pi]$. Namely, let $\alpha : [-\pi,\pi] \to [0,L]$ be defined as $\alpha(t) = \frac{L}{2\pi}(t+\pi)$, and let $\tilde{\gamma} = \gamma\circ\alpha$. Then $\tilde{\gamma}$ has the same image and the same lenght as γ. Moreover, putting $\tilde{\gamma} = (\tilde{x},\tilde{y})$, we have

$$||\tilde{\gamma}'(t)|| = ||\big(x'(\alpha(t))\alpha'(t), y'(\alpha(t))\alpha'(t)\big)|| = ||\frac{L}{2\pi}\big(x'(\alpha(t)), y'(\alpha(t))\big)||$$

$$= \frac{L}{2\pi}||\big(x'(\alpha(t)), y'(\alpha(t))\big)|| = \frac{L}{2\pi}\,,$$

the last equality depending on the fact that γ is parametrized by arc lenght. We have

$$\int_{-\pi}^{\pi} \big(\tilde{x}'(t)^2 + \tilde{y}'(t)^2\big)\,dt = \int_{-\pi}^{\pi} ||\tilde{\gamma}'(t)^2||\,dt = \int_{-\pi}^{\pi} \frac{L^2}{4\pi^2}\,dt = \frac{L^2}{2\pi}\,,$$

hence

$$L^2 = 2\pi\int_{-\pi}^{\pi} \big(\tilde{x}'(t)^2 + \tilde{y}'(t)^2\big)\,dt\,.$$

Let now a_n, b_n be the Fourier coefficients of $\tilde{x}$, c_n, d_n be the Fourier coefficients of $\tilde{y}$. As γ is a Jordan curve we have $\gamma(0) = \gamma(L)$ and by definition of $\tilde{\gamma}$, $\tilde{\gamma}(-\pi) = \tilde{\gamma}(\pi)$, so that $\tilde{x}(-\pi) = \tilde{x}(\pi)$, $\tilde{y}(-\pi) = \tilde{y}(\pi)$. Hence, we can apply Theorem 1.3 and deduce that the Fourier coefficients of $\tilde{x}'$ are nb_n and $-na_n$. Using Theorem 1.1 we thus have

$$\int_{-\pi}^{\pi} \tilde{x}'(t)^2\,dt = \pi\sum_{n=1}^{+\infty} n^2(a_n^2+b_n^2)$$

and by similar considerations

$$\int_{-\pi}^{\pi} \tilde{y}'(t)^2\, dt = \pi \sum_{n=1}^{+\infty} n^2 (c_n^2 + d_n^2),$$

and, in conclusion

$$L^2 = 2\pi^2 \sum_{n=1}^{+\infty} n^2 (a_n^2 + b_n^2 + c_n^2 + d_n^2)\,. \tag{1.2}$$

Next, we evaluate A using the formula $A = \int_{-\pi}^{\pi} \tilde{x}(t)\tilde{y}'(t)\, dt$, consequence of the Green formula. Using Theorems 1.2 and 1.3 we deduce

$$A = \pi \sum_{n=1}^{+\infty} n(a_n d_n - b_n c_n)\,.$$

Using (1.2) we get

$$L^2 - 4\pi A = 2\pi^2 \Big(\sum_{n=1}^{+\infty} n^2(a_n^2 + b_n^2 + c_n^2 + d_n^2) - 2na_n d_n + 2nb_n c_n \Big)$$

$$= 2\pi^2 \Big(\sum_{n=1}^{+\infty} (na_n - d_n)^2 + (nb_n + c_n)^2 + (n^2 - 1)(c_n^2 + d_n^2) \Big) \geq 0$$

and (1.1) is proved. We also note that the equality holds in (1.1) if and only if the equality holds in the previous inequality, if and only if we have $c_n = d_n = 0$ for $n \geq 2$, $d_n = na_n$ and $c_n = -nb_n$ for every $n \geq 1$, and this occurs when $a_n = b_n = c_n = d_n = 0$ for every $n \geq 2$, $a_1 = d_1$, $b_1 = -c_1$. Since $\tilde{x}$ and $\tilde{y}$ are piecewise C_1, thus in particular continuous, they amount to the the sum of their Fourier series, hence

$$\tilde{x}(t) = \frac{a_0}{2} + a_1 \cos t + b_1 \sin t \qquad \tilde{y}(t) = \frac{c_0}{2} - b_1 \cos t + a_1 \sin t$$

which represents the equation of a circle[2]. It follows that not only the circle is the plane Jordan curve of given lenght enclosing the maximum area but also that it is the unique curve having such a property. The solution presented here is due to Hurwitz.

2. The Ascoli-Arzelà Theorem

We start by recalling some base facts about compactness. In order to simplify the presentation, we restrict our considerations to Hausdorff topological spaces. We recall that a (Hausdorff) topological space X is compact if for every family of open subsets U_i of X,

[2] Indeed, setting $r = \sqrt{a_1^2 + b_1^2}$, we have $\left(\frac{a_1}{r}\right)^2 + \left(\frac{b_1}{r}\right)^2 = 1$, hence there exists $\bar{t} \in \mathbb{R}$ so that $\frac{a_1}{r} = \cos \bar{t}$, $\frac{b_1}{r} = \sin \bar{t}$. We thus easily see that $\tilde{x}(t) = \frac{a_0}{2} + r\cos(t - \bar{t})$, $\tilde{y}(t) = \frac{c_0}{2} + r\sin(t - \bar{t})$.

$i \in I$ such that $X = \bigcup_{i \in I} U_i$ there exist $i_1, ..., i_m \in I$ such that $X = U_{i_1} \cup U_{i_2} \cup ... \cup U_{i_m}$. We recall that if X is a metric space then X is compact if and only if every sequence in X has a convergent (in X) subsequence. As a consequence, every compact metric space is complete (recall that in any metric space a *Cauchy* sequence having a convergent subsequence is convergent). We recall that if X is a subset of $\mathbb{R}^n$ then X is compact if and only if it is both closed and bounded. It follows that every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence. Instead, when X is subset of an infinite dimensional normed space, if X is compact, then it is both closed and bounded, but the converse does not hold, i.e., a closed-and-bounded set is not necessarily compact.

The Ascoli-Arzelà Theorem is related to the problem of what subsets of the space of the continuous functions from a topological space A to $\mathbb{R}$ with the norm $||f|| = \sup |f(x)|$, are compact. More precisely, under what conditions we can state that a sequence of continuous functions from X to $\mathbb{R}$ has a uniformly convergent subsequence. We can more generally suppose that the functions are valued in a metric space (Y, d). We need some preliminary definitions. We recall that if f_n is a sequence of functions from a topological space X with values in a complete metric space, then f_n is uniformly convergent if and only if it is uniformly Cauchy, that is, for every $\varepsilon > 0$ there exists $\nu \in \mathbb{N}$ such that, if $n \geq \nu$, $m \geq \nu$ then $d(f_n(x), f_m(x)) < \varepsilon$ for every $x \in X$.

Definition 2.1. *We say that a sequence f_n of functions from a Hausdorff topological space X into a metric space Y are equicontinuous if for every $x \in X$ and for every $\varepsilon > 0$ there exists a neighborhood U of x such that for every $y \in U$ and for every n we have $d(f_n(x), f_n(y)) < \varepsilon$.*

Note that if f_n are equicontinuous then every f_n is continuous, but the converse is not true, since in the previous definition we require that the neighborhood U does not depend on n. Note also that, since every neighborhood of x contains an open set containing x, by the definition of a neighborhood, in Def. 2.1 we can suppose that U is open. If X is a metric space with a metric d' then the previous definitions can be expressed in terms of ε and δ, i.e., for every $x \in X$ and for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $y \in X$ such that $d'(x, y) < \delta$ and for every n we have $d(f_n(x), f_n(y)) < \varepsilon$. If X is a metric space with a metric d', we say that f_n are uniformly equicontinuous if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $x, y \in X$ such that $d'(x, y) < \delta$ and for every n we have $d(f_n(x), f_n(y)) < \varepsilon$. In other words, δ is also independent of x. Clearly, if f_n are uniformly equicontinuous then f_n are equicontinuous. It is possible to prove that also the converse holds if X is compact. The argument of the proof is analogous to that used for proving that a continuous function on a compact set is uniformly continuous.

Theorem 2.2 (Ascoli-Arzelà Theorem). *If f_n are equicontinuous functions from a compact topological space X to a compact metric space Y, then there exists a subsequence of f_n uniformly convergent on X.*

Proof. For each $\varepsilon > 0$ and for each $x \in X$ let $U_{\varepsilon,x}$ be an open set in X containing x

such that for each $y \in U_{\varepsilon,x}$ and for every n, we have $d(f_n(x), f_n(y)) < \varepsilon$. Such a set $U_{\varepsilon,x}$ exists by the assumption that f_n are equicontinuous. As, for a given $s = 1, 2, 3, ...$ the sets $U_{1/s,x}$, $x \in X$ are open sets whose union is X, X being compact there exists a finite subset $A_s = \{x(i,s) : i = 1, ...m(s)\}$ of X such that

$$X = \bigcup_{i=1}^{m(s)} U_{1/s,x(i,s)} \,. \tag{2.1}$$

Let $A = \bigcup_{s=1}^{\infty} A_s$. As every A_s is a finite set, the set A is countable. Put $A = \{a_1, a_2, a_3, ...\}$. Note that for each $x \in X$ the sequence $(f_n(x))$, as it lies in the compact metric space Y, has a convergent subsequence. We are now looking for a subsequence of f_n which converges at all $a_i \in A$, and then we will prove that it uniformly converges on X.

Let $g_n^{(1)}$ be a subsequence of f_n which converge at a_1. For the same reason we can find a subsequence $g_n^{(2)}$ of $g_n^{(1)}$ that converges at a_2, and $g_n^{(2)}$ being a subsequence of $g_n^{(1)}$, it converges at a_1 as well. Next, we can find a subsequence $g_n^{(3)}$ of $g_n^{(2)}$ that converges at a_3, and $g_n^{(3)}$ being a subsequence of $g_n^{(2)}$, it converges at a_1 and a_2 as well. By continuing this process we inductively find sequences $g_n^{(h)}$ for each natural h such that $g_n^{(h+1)}$ is a subsequence of $g_n^{(h)}$ for each h, and $g_n^{(h)}$ converges at $a_1, a_2, ..., a_h$. Clearly, every $g_n^{(h)}$ is a subsequence of f, and the problem is that each of those subsequences only converges (a priori) at a *finite* subset of A.

By what a way we can find a subsequence that converges at *all* points in A? The answer is: we take the *diagonal* subsequence defined by $g_n = g_n^{(n)}$. Indeed, g_n is a subsequence "from h on" of $g_n^{(h)}$ for each h in the sense that there exists a strictly increasing map ψ_h from $\{h, h+1, ...\}$ into itself so that $g_n = g^{(h)}_{\psi_h(n)}$ for every $n \geq h$, and therefore on one hand g_n is a subsequence of f_n, on the other g_n converges at a_h for each h.[3] Prove now that g_n is uniformly convergent on X. As Y is a compact, thus complete, metric space, this amounts to prove that g_n is uniformly Cauchy. *Given* $\varepsilon > 0$, let $s = 1, 2, 3, ...$ be so that $\frac{3}{s} < \varepsilon$. As $g_n(a_i)$ converges for all i, in particular $g_n(x(i,s))$ converges for all $i = 1, ..., m_s$. Hence *there exists* $\nu \in \mathbb{N}$ such that when $h, k \geq \nu$ then

$$d\big(g_h(x(i,s)), g_k(x(i,s))\big) < \frac{1}{s} \tag{2.2}$$

for each $i = 1, ..., m_s$. This as for each $i = 1, ..., m_s$ we find ν_i so that (2.2) holds (for that i) for each $h, k \geq \nu_i$. Then, we take $\nu = \max \nu_i$. *Let now* $x \in X$. In view of (2.1)

[3] To see this, note that if $m \geq h$ $g_n^{(m)}$ is a subsequence of $g_n^{(h)}$ thus there exists $\phi_m : \{1, 2, ...\}$ to itself so that $g_n^{(m)} = g^{(h)}_{\phi_m(n)}$. As $g_n^{(m+1)}$ is a subsequence of $g_n^{(m)}$ we find inductively ϕ_m, in particular $g_n^{(m+1)} = g^{(m)}_{\sigma(n)} = g^{(h)}_{\phi_m(\sigma(n))}$ for some strictly increasing σ, hence $\phi_{m+1} = \phi_m \circ \sigma$, and $\phi_{m+1} \geq \phi_m$. Thus, for $n \geq h$, $g_n = g_n^{(n)} = g^{(h)}_{\psi_h(n)}$ with $\psi_h(n) = \phi_n(n)$ and as $\phi_{n+1}(n+1) \geq \phi_n(n+1) > \phi_n(n)$, ψ_h is strictly increasing.

there exists $i = 1, ..., m_s$ such that $x \in U_{1/s,x(i,s)}$, so that, by the definition of $U_{\varepsilon,x}$ we have $d\big(g_m(x), g_m(x(i,s))\big) < \frac{1}{s}$ for all m as every g_m is of the form f_n for some n. Thanks to (2.2), it follows that for $h, k \geq \nu$,

$$d(g_k(x), g_h(x)) \leq d\big(g_k(x), g_k(x(i,s)\big) + d\big(g_k(x(i,s)), g_h(x(i,s))\big) + d\big(g_h(x(i,s)), g_h(x)\big)$$
$$< \frac{3}{s} < \varepsilon\,.$$

As ε is an arbitrary positive number, g_n is uniformly Cauchy, thus it uniformly converges. ■

We cannot apply the previous theorem when $Y = \mathbb{R}^M$ as in such a case Y is not compact. However, if there exists $K > 0$ such that

$$||f_n(x)|| \leq K \quad \forall\, x \in X \quad \forall n, \tag{2.3}$$

we can consider $f_n : X \to \overline{B(0,K)}$ and as $\overline{B(0,K)}$ is compact we can apply Theorem 2.2 again. When (2.3) holds the function f_n are said to be equibounded as they are bounded by a constant which is independent of n. We thus have the following corollary, which is one of the most usual forms of the Ascoli-Arzelà Theorem.

Corollary 2.3. *If f_n are equicontinuous and equibounded functions from a compact topological space X to $\mathbb{R}^M$, then there exists a subsequence of f_n uniformly convergent on X.* ■

When X is a metric space with distance d', a typical case in which the functions f_n are equicontinuous is that in which they are *equilipshitzian*, i.e., there exists $K > 0$ so that $d(f_n(x), f_n(y)) \leq Kd'(x,y)$ for each $x, y \in X$ and for each n. In other words, they satisfy a Lipschitz condition with a constant independent of n. In fact, in this case it suffices to take $\delta = \frac{\varepsilon}{K}$ in the definition of (uniform) equicontinuity. As a particular case, if f_n are functions defined on an interval in $\mathbb{R}$ with values in $\mathbb{R}$, they are equilipshitzian when they have equibounded derivatives. Indeed, by the mean value Theorem, $|f_n(x) - f_n(y)| \leq \big(\sup|f_n'|\big)|x - y|$.

Exercise 2.1. Prove that Theorem 2.2 (or Corollary 2.3) is no longer valid if $X = \mathbb{R}$.

Exercise 2.2. Prove that if f_n are equibounded and equicontinuous functions from $\mathbb{R}^N$ to $\mathbb{R}^M$ (more generally if they are equibounded on every compact subset of $\mathbb{R}^N$ and equicontinuous), then there exists a subsequence of f_n uniformly convergent on the compact subsets of $\mathbb{R}^N$.

Exercise 2.3. Prove that the conclusion of the previous exercise is still valid if $\mathbb{R}^N$ is replaced by any open subset of $\mathbb{R}^N$.

Exercise 2.4. Find a sequence of equibounded functions from $[0,1]$ to $\mathbb{R}$ which has no subsequence *pointwise* convergent.

3. Curves of Minimum Lenght

The purpose of this section is to prove the following

Theorem 3.1. *Given a closed subset A of $\mathbb{R}^N$ and two points $P, Q \in A$ such that*
a) there exists a continuous curve in A connecting them having finite lenght,
then there exists a continuous curve in A connecting them having minimum lenght.

Note that a) in Theorem 3.1 for any $P, Q \in A$, is a condition stronger than arcwise connectedness, in the sense that arcwise connectedness requires that any two points $P, Q \in A$ can be connected by a continuous curve but not necessarily having finite lenght. In order to clarify the statement in Theorem 3.1, first of all, we recall the definitions concerning the lenght of a curve. Given a closed interval $[a, b]$ (with $a, b \in \mathbb{R}$, $a < b$), a *partition* of $[a, b]$ is an object of the form $(t_0, t_1, ..., t_n)$ such that $a = t_0 < t_1 < ... < t_n = b$. We denote by $\mathcal{P}_{a,b}$ the set of the partitions of $[a, b]$. A continuous curve in a subset A of $\mathbb{R}^N$ is a continuous function from a closed interval $[a, b]$, $(a < b)$, to A. Given $\Pi = (t_0, t_1, ..., t_n) \in \mathcal{P}_{a,b}$, and a continuous curve from $[a, b]$ to $\mathbb{R}^N$, we denote by $\Lambda_\gamma(\Pi)$ the real number

$$\sum_{i=1}^{n} ||\gamma(t_i) - \gamma(t_{i-1})||$$

and we define the lenght of γ to be the nonnegative, possibly infinite, value

$$L(\gamma) := \sup_{\Pi \in \mathcal{P}_{a,b}} \Lambda_\gamma(\Pi) \,.$$

We recall that if γ is piecewise C^1, then we have the formula

$$L(\gamma) = \int_a^b ||\gamma'(t)|| \, dt \,.$$

We also recall that the lenght of a curve is invariant up to a reparametrization. In order to clarify this, we recall that given a continuous curve $\gamma : [a, b] \to \mathbb{R}^N$, a reparametrization of it is a curve $\tilde{\gamma} : [c, d] \to \mathbb{R}^N$ defined by $\tilde{\gamma} = \gamma \circ \phi^{-1}$ where ϕ is a continuous bijection from $[a, b]$ onto $[c, d]$ (Note that, in such a case, by a well-known theorem, the inverse ϕ^{-1} is continuous as well). Then the lenght of a curve amounts to the lenght of any reparametrization.

We now equip the set $\mathcal{C}_{a,b;N}$ of the continuous curves from a fixed interval $[a, b]$ to $\mathbb{R}^N$ with the norm

$$||\gamma||_\infty = \sup_{t \in [a,b]} ||\gamma(t)||.$$

We recall that the convergence induced by such a norm is the uniform convergence, in other words, $\gamma_n \underset{n\to\infty}{\longrightarrow} \gamma$ in $|| \; ||_\infty$ if and only if $\gamma_n \underset{n\to\infty}{\longrightarrow} \gamma$ uniformly. If we consider $L(\gamma)$

as a function of γ, we realize that L is not continuous, as we can approximate a curve of finite lenght by a sequence of curves having lenght tending to infinity. For example, it is easy to see that the curve $(t, 0)$ on $[0, 1]$ in $\mathbb{R}^2$ has lenght 1 and the approximating curves $\big(t, \frac{1}{\sqrt{n}} \sin(2\pi nt)\big)$ have lenght tending to infinity. Nevertheless, the intuition suggests that, if we approximate a curve γ in $|| \, ||_\infty$, the lenght could greatly increase but not greatly decrease, in other words the function L from the set of the continuous curves into $\mathbb{R} \cup \{+\infty\}$ is not continuous, but *lower semicontinuous*. We recall the following definition.

Definition 3.2. *Let F be a function from a topological space X to $\mathbb{R} \cup \{+\infty\}$. We say that F is lower semicontinuous (abbreviated as l.s.c.) at a point $x \in X$ if for every $M \in \mathbb{R}$ such that $M < F(x)$ there exists U neighborhood of x in X such that for every $y \in U$ we have $F(y) > M$. We say that F is lower semicontinuous (on X) if F is lower semicontinuous at each point in X.*

We could give the definition of l.s.c. in the more natural setting of functions with values in $\mathbb{R}$, but we prefer to do this in the setting of functions with values in $\mathbb{R} \cup \{+\infty\}$, as we will study it in the case of the lenght of a curve that can well assume the value $+\infty$. Moreover, we will study the sup of a family of l.s.c. functions, which can assume the value $+\infty$ even if all the functions take finite values. Note that if the function F only assumes finite values, then the definition of lower semicontinuity can be expressed as: F is l.s.c. at x if for every $\varepsilon > 0$ the exists U neighborhood of x in X such that for every $y \in U$ we have $f(y) > f(x) - \varepsilon$. So, we see the difference with respect to the definition of continuity at x, where we require that in a suitable neighborhood of x we have $f(x) + \varepsilon > f(y) > f(x) - \varepsilon$, in other words, in the definition of semicontinuity we require that in a neighborhood of x the function is not too smaller than at x, but not necessarily not too greater than at x.

If (X, d) is a metric space and F only takes finite values, of course the semicontinuity can be also expressed using ε and δ, i.e., F is l.s.c. at x if for every $\varepsilon > 0$ the exists $\delta > 0$ such that for every $y \in X$ such that $d(x, y) < \delta$ we have $f(y) > f(x) - \varepsilon$. Of course, every continuous function at x is l.s.c. at x, but the converse is not true, for example the function $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) = 2$ if $x \neq 0$, $F(0) = 1$, is l.s.c. but not continuous at 0. Note that in the definition of lower semicontinuity we use the order stucture of $\mathbb{R}$ or of $\mathbb{R} \cup \{+\infty\}$, so that such a definition, unlike the definition of continuity, does not make sense for functions with values in an arbitrary topological (or even metric) space.

We recall that if X is a metric space the continuity at x can be expressed in terms of convergences of sequences, namely F is continuous at x if and only if, for every sequence (x_n) in X tending to x, we have $F(x_n) \underset{n \to +\infty}{\longrightarrow} F(x)$. A similar characterization holds for lower semicontinuity, namely (under the hypothesis of Def. 3.2) F is l.s.c. at $x \in X$ if and only if for every sequence (x_n) in X tending to x, we have $\liminf\limits_{n \to +\infty} F(x_n) \geq F(x)$. We omit the proof, which resembles that for the continuity. We only note that the part $\Rightarrow$ is rather simple and does not use the fact that X is a metric space, and the part $\Leftarrow$ is proved by contradiction and would not be valid in the general case of X topological space.

We now are going to prove that the sup of l.s.c. functions (at a point) is l.s.c. (at that point). Note that in general the continuity does not have the same property, e.g., the functions $1 - x^n$ from $[0,1]$ to $\mathbb{R}$ are continuous, but the sup of them, when n varies on $1, 2, 3, ...$, is the function f defined by $f(x) = 1$ if $x < 1$, $f(1) = 0$, which is discontinuous at 1.

Lemma 3.3. *Suppose f_i, $i \in I$ are functions from a topological space X with values in $\mathbb{R} \cup \{+\infty\}$, and $f = \sup_{i \in I} f_i$. If f_i are l.s.c. at $x \in X$ then f is l.s.c. at x.*

Proof. Let $M \in \mathbb{R}$ be such that $M < f(x)$. As $f(x) = \sup_{i \in I} f_i(x)$, there exists $i \in I$ such that $f_i(x) > M$, and as f_i is l.s.c. at x there exists U neighborhood of x such that for every $y \in U$ we have $f_i(y) > M$. Hence, for every $y \in U$ we have $f(y) \geq f_i(y) > M$, and as M is an arbitrary number less than $f(x)$, f is l.s.c. at x. ∎

Corollary 3.4. *Let $a, b \in \mathbb{R}$, $a < b$. Then the function L defined on $\left(\mathcal{C}_{a,b;N}, || \; ||_\infty\right)$ by $\gamma \mapsto L_\gamma$ is l.s.c.*

Proof. In view of Lemma 3.3, it suffices to prove that, for each $\Pi \in \mathcal{P}_{a,b}$, the map $\gamma \mapsto \Lambda_\gamma(\Pi)$ from $\mathcal{C}_{a,b;N}$ to $\mathbb{R}$, is continuous. Let $\Pi = (t_0, t_1, ...t_n)$. Since, clearly, the map $\gamma \mapsto \gamma(t)$ from $\mathcal{C}_{a,b;N}$ to $\mathbb{R}^N$ is continuous for every $t \in [a,b]$, hence so is the map $\gamma \mapsto \gamma(t_i) - \gamma(t_{i-1})$ for $i = 1, ..., n$, as the difference of continuous functions. Hence, the map $\gamma \mapsto \Lambda_\gamma(\Pi)$ is continuous as the sum of the composition of the norm function, which is continuous from $\mathbb{R}^N$ to $\mathbb{R}$, with continuous functions. ∎

We now sketch the plan of the proof of Theorem 3.1. It is possible to prove that a l.s.c. function from a (nonempty) compact topological space to $\mathbb{R}$ has a minimum. Now, the space $\mathcal{C}_{a,b;N}$ is not compact, but we can restrict the l.s.c. function L to a suitable subset X of $\mathcal{C}_{a,b;N}$. As X, being a subset of the metric space $\mathcal{C}_{a,b;N}$ is a metric space as well, in order to see whether X is a compact, we have to check whether every sequence of functions in X has a subsequence convergent to an element of X with respect to the norm $|| \; ||_\infty$, that is, uniformly. Thus, the idea constists in finding a suitable X composed by equibounded and equicontinuous functions, so that we can apply the Ascoli-Arzelà Theorem. First, we can consider the space of the curves in $\mathcal{C}_{a,b;N}$ having lenght less than or equal to a fixed real number k. These curves are not necessarily equicontinuous, but we could reparametrize them by arcleght, so that, as easily verfied, they are Lipshitzian with a Lipshitz constant equal to k. The problem is that not all curves can be reparametrized by arclenght, for example it suffices to consider a curve which is constant on some interval. So, in the following we will perform a slight modification of the above idea.

Given $\gamma \in \mathcal{C}_{a,b;N}$ with $L(\gamma) < +\infty$, and $c, d \in [a,b]$, $c \leq d$, we put $L_{c,d}(\gamma) = L(\gamma|_{[c,d]})$, where of course, $\gamma|_{[c,d]}$ denote the restriction of γ to the interval $[c,d]$, with the convention $L_{c,d}(\gamma) = 0$ if $c = d$. It is well known that, if $a \leq c \leq d \leq u \leq b$, then $L_{c,u}(\gamma) = L_{c,d}(\gamma) + L_{d,u}(\gamma)$. We now consider the arclenght function $\tilde{\phi} : [a,b] \to \mathbb{R}$ defined by

$\tilde{\phi}(t) = L_{a,t}(\gamma)$. We easily see that L is increasing, but not necessarily strictly increasing. We now prove:

Lemma 3.5. *$\tilde{\phi}$ is continuous.*

Proof. We first prove that if $\bar{t} > a$ then $\tilde{\phi}$ is continuous at $\bar{t}$ on the left. Let $\varepsilon > 0$. By the definition of $L_{a,t}(\gamma)$, there exists $\Pi \in \mathcal{P}(a, \bar{t})$ such that $\Lambda_\gamma(\Pi) > \tilde{\phi}(\bar{t}) - \frac{\varepsilon}{2}$. We write $\Pi = (a = t_0, t_1, ..., t_n = \bar{t})$. Let

$$t \in]t_{n-1}, t_n[.$$

Clearly,

$$\Lambda_\gamma(t_0, t_1, ..., t_{n-1}, t, t_n) \geq \Lambda_\gamma(\Pi) > \tilde{\phi}(\bar{t}) - \frac{\varepsilon}{2},$$

$$\Lambda_\gamma(t_0, t_1, ..., t_{n-1}, t, t_n) = \Lambda_\gamma(t_0, t_1, ..., t_{n-1}, t) + ||\gamma(\bar{t}) - \gamma(t)||,$$

hence

$$\tilde{\phi}(\bar{t}) \geq \tilde{\phi}(t) \geq \Lambda_\gamma(t_0, t_1, ..., t_{n-1}, t) = \Lambda_\gamma(t_0, t_1, ..., t_{n-1}, t, t_n) - ||\gamma(\bar{t}) - \gamma(t)||$$

$$> \tilde{\phi}(\bar{t}) - \frac{\varepsilon}{2} - ||\gamma(\bar{t}) - \gamma(t)||.$$

Now, as γ is continuous, there exists $\tilde{t} < \bar{t}$ such that, if $\tilde{t} < t < \bar{t}$ then $||\gamma(\bar{t}) - \gamma(t)|| < \frac{\varepsilon}{2}$. Hence, if $\max\{\tilde{t}, t_{n-1}\} < t < \bar{t}$, then

$$\tilde{\phi}(\bar{t}) \geq \tilde{\phi}(t) > \tilde{\phi}(\bar{t}) - \varepsilon$$

and $\tilde{\phi}$ is continuous at $\bar{t}$ on the left. We now prove that $\tilde{\phi}$ is continuous on the right at any point $\bar{t} < b$. The proof is similar noting that

$$\tilde{\phi}(t) = L(\gamma) - L_{t,b}(\gamma) \quad \forall\, t \in [a, b]. \tag{3.1}$$

Let $\varepsilon > 0$. We find $\Pi \in \mathcal{P}_{\bar{t},b}$ so that

$$\Lambda_\gamma(\Pi) > L_{\bar{t},b}(\gamma) - \frac{\varepsilon}{2}.$$

We write $\Pi = (\bar{t} = t_0, t_1, ..., t_n = b)$. Let

$$t \in]t_0, t_1[.$$

Clearly,

$$\Lambda_\gamma(t, t_1, ..., t_n) = \Lambda_\gamma(t_0, t, t_1, ..., t_n) - ||\gamma(t) - \gamma(\bar{t})|| \geq \Lambda_\gamma(\Pi) - ||\gamma(t) - \gamma(\bar{t})||.$$

We find $\tilde{t} > \bar{t}$ such that, if $\tilde{t} > t > \bar{t}$ then $||\gamma(\bar{t}) - \gamma(t)|| < \frac{\varepsilon}{2}$. We conclude like before that, if $\bar{t} < t < \min\{\tilde{t}, t_1\}$, then

$$L_{t,b}(\gamma) \geq \Lambda_\gamma(t, t_1, ..., t_n) > L_{\overline{t},b}(\gamma) - \varepsilon$$

so that, using (3.1)

$$\tilde{\phi}(\overline{t}) \leq \tilde{\phi}(t) < \tilde{\phi}(\overline{t}) + \varepsilon$$

and $\tilde{\phi}$ is continuous at $\overline{t}$ on the right. ■

We now would like to reparametrize γ using the arcleght ϕ, but as previously observed, this is not a strictly increasing function, so we need a modification of it. First we need the following remark.

Remark 3.6. If $\gamma \in \mathcal{C}_{a,b,N}$ and $c, d \in [a, b]$, then $||\gamma(c) - \gamma(d)|| \leq L(\gamma)$. Indeed, if $c = d$ this is trivial, if not we can for example suppose $c < d$. Then one of the following $(a, c, d, b), (a, c, d), (c, d, b), (c, d)$ is in $\mathcal{P}_{a,b}$, depending of what of the inequalities $a \leq c$, $d \leq b$ are strict. Let Π be such an element of $\mathcal{P}_{a,b}$. Then, $L(\gamma) \geq \Lambda_\gamma(\Pi) \geq ||\gamma(c) - \gamma(d)||$. ■

Since the lenght of a curve in invariant up to a reparametrization we can and do assume that

$$[a, b] = [0, 1].$$

We now want to reparametrize in a Lipshitzian way a continuous curve from $[0, 1]$ to $\mathbb{R}^N$ of finite lenght. We modify the arclenght function in the following way. Let

$$\phi(t) = \frac{L_{0,t}(\gamma) + t}{L(\gamma) + 1}.$$

In such a definition, we add t to $L_{0,t}(\gamma)$ in order to have a *strictly* increasing function, and divide by $L(\gamma) + 1$ in order that ϕ map $[0, 1]$ onto $[0, 1]$. We easily verify that in fact ϕ is a continuous strictly increasing function from $[0, 1]$ onto itself, so that the curve $\tilde{\gamma}$ defined by

$$\tilde{\gamma} = \gamma \circ \phi^{-1}$$

is a reparametrization of γ. Note now that, if $0 \leq \tau_1 < \tau_2 \leq 1$ then

$$L_{\tau_1,\tau_2}(\gamma) = L_{0,\tau_2}(\gamma) - L_{0,\tau_1}(\gamma) < \frac{L_{0,\tau_2}(\gamma) - L_{0,\tau_1}(\gamma) + \tau_2 - \tau_1}{L(\gamma) + 1}(L(\gamma) + 1)$$
$$= \big(\phi(\tau_2) - \phi(\tau_1)\big)(L(\gamma) + 1) \qquad (3.2)$$

so that *if* $0 \leq t_1 < t_2 \leq 1$, using (3.2) with $\tau_1 = \phi^{-1}(t_1)$, $\tau_2 = \phi^{-1}(t_2)$,

$$||\tilde{\gamma}(t_2) - \tilde{\gamma}(t_1)|| \leq L_{\phi^{-1}(t_1),\phi^{-1}(t_2)}(\gamma) \leq (L(\gamma) + 1)(t_2 - t_1),$$

$$||\tilde{\gamma}(t_2) - \tilde{\gamma}(t_1)|| \leq (L(\gamma) + 1)|t_2 - t_1| \tag{3.3}$$

where in the first inequality we have used Remark 3.6 with $\gamma|_{[\phi^{-1}(t_1),\phi^{-1}(t_2)]}$ in place of γ. However, (3.3) holds for every $t_1, t_2 \in [0,1]$, as, if $t_1 > t_2$, we obtain (3.3), by changing t_1 with t_2, and (3.3) is obvious if $t_1 = t_2$. In conclusion, $\tilde{\gamma}$ is Lipshitzian with constant $L(\gamma) + 1$.

Proof of Theorem 3.1. Let $\hat{\gamma}$ be a continuous curve in A, defined on $[0,1]$ having finite lenght $\overline{L}$ with $\hat{\gamma}(0) = P$, $\hat{\gamma}(1) = Q$. Let

$$X = \left\{\gamma \in \mathcal{C}_{0,1;N} : \gamma(t) \in A \quad \forall\, t \in [0,1],\ \gamma(0) = P, \gamma(1) = Q,\ L(\gamma) \leq \overline{L}\right\}.$$

Since $\hat{\gamma} \in X$, then $X \neq \emptyset$. Let $\gamma_n \in X$ be such that

$$L(\gamma_n) \underset{n\to+\infty}{\longrightarrow} \inf_{\gamma \in X} L(\gamma). \tag{3.4}$$

Let $\tilde{\gamma}_n$ be a reparametrization of γ_n obtained as above, in particular using a function ϕ (which of course usually depends on n) strictly increasing so that $\tilde{\gamma}_n(0) = \gamma_n(0) = P$, $\tilde{\gamma}_n(1) = \gamma_n(1) = Q$, and $\tilde{\gamma}_n$ Lipshitzian with constant $L(\gamma_n) + 1$. Since $L(\gamma_n) \leq \overline{L}$, we easily see that all $\tilde{\gamma}_n$ are Lipshitzian with constant $\overline{L} + 1$, hence $\tilde{\gamma}_n$ are equilipshitzian. Also, since $L(\tilde{\gamma}_n) = L(\gamma_n)$, $\tilde{\gamma}_n$ being a reparametrization of γ_n, we easily see that $\tilde{\gamma}_n \in X$. Moreover, in view of Remark 3.6, for every $t \in [0,1]$, we have

$$||\tilde{\gamma}_n(t)|| \leq ||\tilde{\gamma}_n(t) - \tilde{\gamma}_n(0)|| + ||\tilde{\gamma}_n(0)|| \leq L(\tilde{\gamma}_n) + ||P|| \leq \overline{L} + ||P||$$

so that $\tilde{\gamma}_n$ are equibounded. We can thus use the Ascoli-Arzelà Theorem and deduce that a suitable subsequence $\tilde{\gamma}_{n_k}$ of $\tilde{\gamma}_n$ converges, for $k \to \infty$, to some $\tilde{\gamma} \in \mathcal{C}_{0,1;N}$ with respect to $||\ ||_\infty$, that is, uniformly. We now see that $\tilde{\gamma} \in X$. In fact, as $\tilde{\gamma}_{n_k}(t) \underset{k\to+\infty}{\longrightarrow} \tilde{\gamma}(t)$ for each $t \in [0,1]$, we have $\tilde{\gamma}(t) \in A$ for each $t \in [0,1]$ (recall that A is assumed to be closed). For the same reason, $\tilde{\gamma}(0) = P$, $\tilde{\gamma}(1) = Q$. Finally, since L is l.s.c. on $\mathcal{C}_{0,1;N}$, we have

$$L(\tilde{\gamma}) \leq \liminf_{k\to+\infty} L(\tilde{\gamma}_{n_k}) = \liminf_{k\to+\infty} L(\gamma_{n_k}) \leq \overline{L}, \tag{3.5}$$

and $\tilde{\gamma} \in X$ as claimed. Using (3.4) and (3.5) we get $L(\tilde{\gamma}) \leq \inf_{\gamma \in X} L(\gamma)$ (thus $L(\tilde{\gamma}) = \min_{\gamma \in X} L(\gamma)$). It remains to prove that $L(\tilde{\gamma}) \leq L(\gamma)$ when γ is a continuous curve in A connecting P and Q, but which is not an element of X. This is obvious, as in such a case $L(\gamma) > \overline{L} \geq L(\tilde{\gamma})$. ∎

The argument of Theorem 3.1 can be used to prove the following general theorem, which resembles the well-known Weierstrass Theorem on extrema of continuous functions defined on a compact set.

Theorem 3.7. *Let f be a l.s.c. function from a (nonempty) metric space X to $\mathbb{R} \cup \{+\infty\}$, not identically $+\infty$. Suppose X is compact, or more generally*
a) the set $\{x \in X : f(x) \leq L\}$ is compact for each real L.
Then f has a minimum on X.

Proof. Suppose a) holds. Let $\overline{x} \in X$ be so that $L := f(\overline{x}) < +\infty$. Let

$$Y = \{x \in X : f(x) \leq L\}.$$

Let $x_n \in Y$ be so that $f(x_n) \underset{n \to +\infty}{\longrightarrow} \inf_{x \in Y} f(x)$. Since Y is compact by our assumption, there exists a subsequence x_{n_k} of x_n and $\tilde{x} \in Y$ such that $x_{n_k} \underset{k \to \infty}{\longrightarrow} \tilde{x}$. Then, as f is l.s.c.,

$$f(\tilde{x}) \leq \liminf_{k \to \infty} f(x_{n_k}) = \inf_{x \in Y} f(x).$$

It follows that $f(\tilde{x}) \leq f(x)$ for each $x \in Y$, in particular $f(\tilde{x}) \leq f(\overline{x})$ but also $f(\tilde{x}) \leq f(x)$ for each $x \in X \setminus Y$, as in such a case $f(\tilde{x}) \leq f(\overline{x}) < f(x)$. In conclusion, f takes its minimum at $\tilde{x}$. Note that if X is compact, then a) holds since, as a consequence of the lower semicontinuity of f, the set $Z := \{x \in X : f(x) \leq L\}$ is closed (if $x \notin Z$, then $f(x) > L$, hence there exists U neighborhood of x such that $f(y) > L$ for each $y \in U$, and therefore the complement of Z is open) in the compact set X, hence it is compact. ∎

The fact that a continuous function on $\mathbb{R}$ with values in $\mathbb{R}$ which tends to $+\infty$ both at $-\infty$ and at $+\infty$ has a minimum can be seen as a particular case of Theorem 3.7, as in such a case a) in Theorem 3.7 holds. In general, a function f satisfying a) in Theorem 3.7, is said to be coercive. The coerciveness condition is important in many problems in calculus of variations as usually the domain of a functional is not compact, but in many examples the functional is coercive.

Exercise 3.1. Prove that, if A is a nonempty *open* subset of $\mathbb{R}^N$, then any two points in A are connected by a continuous curve of minimum lenght if and only if A is convex. It follows that the hypothesis that A is closed cannot be removed in Theorem 3.1.

4. Geodesics on surfaces

In the previous section, we proved Theorem 3.1. In the following, we will call geodesic in a subset A of $\mathbb{R}^N$ a curve satisfying a) of Theorem 3.1 for some $P, Q \in A$. Note that in differential geometry the term *geodesic* is used in a slightly different sense. The geodesic connecting two given points is not unique as for example any (verse-preserving) reparametrization of it satisfies the same property. Further, in some cases, we can have essentially different geodesics (i.e., having a different image) connecting two given points. For example, if A is a sphere in $\mathbb{R}^3$, and the two points are the north and the south pole, then every meridian is such a geodesic. If the closed subset A of $\mathbb{R}^N$ has no further properties we cannot in general find a *regular* (say C^1) geodesic connecting two given points. An example is when $N = 2$ and A is the complement of an open square, and the two points lie on different edges of the square. This fact is highly intuitive and will not prove this fact in detail.

Suppose now we have a surface S in $\mathbb{R}^3$ and we will assume it is compact, connected and, for simplicity, of class C^∞. In this case it is possible to prove that given two points there exists a regular geodesic in S connecting them, that can be parametrized by arclenght. We recall that a C^1 curve γ is parametrized by arclenght if the vector $||\gamma'||$ is identically 1. The purpose of this section is to give a characterization of (sufficiently smooth) geodesics on S. We recall that a (nonempty) subset S of $\mathbb{R}^3$ is said to be a surface if

a) For every $\overline{P} \in S$ *there exists* U *open neighborhood of* $\overline{P}$ *and* $g : U \to \mathbb{R}$ *of class* C^∞ *such that* $\text{grad} g \neq 0$ *on* U *and* $S \cap U = \Big\{P \in U : g(P) = 0\Big\}$,

or equivalently

b) For every $\overline{P} \in S$ *there exists* U *open neighborhood of* $\overline{P}$ *and* V *open set in* $\mathbb{R}^2$*, and* $\phi : V \to \mathbb{R}^3$ *of class* C^∞ *such that the rank of the Jacobian matrix of* ϕ *equals* 2 *on* V *and* $S \cap U = \phi(V)$,

or also

c) S *can be locally represented as the graph of a* C^∞ *function* ψ *of the form* $z = \psi(x, y)$ *or* $y = \psi(z, x)$ *or* $x = \psi(y, z)$*, more precisely, for every* $\overline{P} \in S$ *there exists* U *open neighborhood of* $\overline{P}$ *and* V *open in* $\mathbb{R}^2$*, and* $\psi : V \to \mathbb{R}$ *of class* C^∞ *such that one of the following holds*
c_1) $S \cap U = \big\{(x, y, z) : (x, y) \in V, z = \psi(x, y)\big\}$,
c_2) $S \cap U = \big\{(x, y, z) : (z, x) \in V, y = \psi(z, x)\big\}$,
c_3) $S \cap U = \big\{(x, y, z) : (y, z) \in V, x = \psi(y, z)\big\}$.

We shortly recall how it can be proved that a) b) and c) are equivalent: If a) holds, then one of the derivatives $\frac{\partial g}{\partial x}$, $\frac{\partial g}{\partial y}$, $\frac{\partial g}{\partial z}$ is different from 0 at $\overline{P}$, thus in a neighborhood of $\overline{P}$. Suppose for example $\frac{\partial g}{\partial z} \neq 0$ at $\overline{P}$. Then, as a consequence of the implicit function Theorem, we can represent S as in c_1), and c) holds. If b) holds, then one of the three submatrices of 2×2 of the jacobian matrix of $\phi = (\phi_1, \phi_2, \phi_3)$ is nonsingular at $\overline{Q}$ with $\phi(\overline{Q}) = \overline{P}$. Suppose for example it is the matrix relative to (ϕ_1, ϕ_2). Then, writing ϕ as $\phi(u, v)$, by the inverse function Theorem we can express locally (u, v) as a C^∞ function h of (x, y). Thus, letting $\psi = \phi_3 \circ h$, we can represent S as in c_1). On the other hand if c), for example c_1), holds, then we get a) putting $g(x, y, z) = z - \psi(x, y)$, and b) putting $\phi(u, v) = \big(u, v, \psi(u, v)\big)$. Now, in c), for example c_1), we can assume

$$\overline{P} = 0, \quad \frac{\partial \psi}{\partial x}(0, 0) = \frac{\partial \psi}{\partial y}(0, 0) = 0 \tag{4.1}$$

in the sense that if (4.1) does not hold, then the image of S via a suitable affine isometry satisfies (4.1). Once we realize that the second condition in (4.1) means that the tangent plane Π to S at $\overline{P}$ is the plane Π' of equation $z = 0$, this can be seen, observing that there exists an affine isometry that carries $\overline{P}$ into 0 and Π into Π'. By this point of view, the assumption (4.1) is valid when we treat properties invariant whith respect to affine isometries. Finally, we recall, that given a C^2 curve γ in $\mathbb{R}^3$, parametrized by arclenght, then the principal normal to γ at the point $\gamma(t)$ is $\gamma''(t)$.

Remark 4.1. We explicitly note that if S is a compact and connected surface in $\mathbb{R}^3$ of class C^∞, then any two points Q and P can be connected by a piecewise C^1, hence of finite lenght, curve in S, thus S satisfies the hypothesis of Theorem 3.1. The argument of the proof resembles that used for proving that, in an open connected set U in $\mathbb{R}^N$, any two points can be connected by a polygonal that remains in U. Fixing the point Q, let

$$A = \left\{\overline{P} \in S : Q \text{ and } \overline{P} \quad \text{are connected by a piecewise } C^1 \text{ curve in } S\right\}.$$

Then $Q \in A$ so that $A \neq \emptyset$. We will see that A is closed and open, hence by the assumption that S is connected, $A = S$. Let $\overline{P} \in S$. Let us represent S using c), for example c_1, with $\overline{P} = (\overline{x}, \overline{y}, \psi(\overline{x}, \overline{y}))$. Consider an open ball B in $\mathbb{R}^2$ with center at $(\overline{x}, \overline{y})$ contained in V and let

$$W := \left\{(x, y, z) : (x, y) \in B, z = \psi(x, y)\right\}.$$

Since $W = S \cap U \cap \pi^{-1}(B)$, where the continuous function $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ is defined by $\pi(x, y, z) = (x, y)$, then W is open in S. Moreover, any $P \in W$ is connected to $\overline{P}$ by a C^1 curve γ in S. In fact, let $P = (x, y, \psi(x, y))$, let $\tilde{\gamma}$ be the segment line connecting $(\overline{x}, \overline{y})$ to (x, y) (which lies in B). Then, it suffices to take γ defined by $\gamma(t) = (\tilde{\gamma}(t), \psi(\tilde{\gamma}(t))$. It easily follows that, if $\overline{P} \in A$ then $W \subseteq A$, while if $\overline{P} \in S \setminus A$, then $W \subseteq S \setminus A$. It follows that in fact both A and $S \setminus A$ are open in S, hence A is closed and open in S. ■

Theorem 4.2. *Suppose S is a compact and connected surface in $\mathbb{R}^3$ of class C^∞, and let $\gamma : [a, b] \to S$ be a geodesic of class C^2 in S connecting two given points $P, Q \in S$, parametrized by arclenght. Then, for any $\overline{t} \in]a, b[$, the principal normal to γ at $\gamma(\overline{t})$ is normal to the surface at $\gamma(\overline{t})$.*

Proof. Put $\overline{P} := \gamma(\overline{t})$. Then we can assume that c_1) holds, and also, (4.1) holds, as the properties we use are invariant with respect to an affine isometry. As $\gamma(\overline{t}) \in U$, by a continuity argument, there exist c, d with $a \leq c < \overline{t} < d \leq b$ such that $\gamma(t) \in U$ for each $t \in [c, d]$. Then $\gamma|_{[c,d]}$ minimizes the lenght of the curves connecting $\gamma(c)$ to $\gamma(d)$ in $S \cap U$ as if there would be a curve η connecting $\gamma(c)$ to $\gamma(d)$ in $S \cap U$, with $L(\eta) < L(\gamma|_{[c,d]})$, then, replacing in γ the piece from c to d by η we would obtain a curve connecting P, Q in S of lenght less than $L(\gamma)$, a contradiction. Since any $\eta : [c, d] \to S \cap U$ has the form

$$\eta(t) = \Big(y_1(t), y_2(t), \psi(y_1(t), y_2(t))\Big),$$

it follows that (γ_1, γ_2) minimizes the integral

$$\int_c^d \sqrt{y_1'(t)^2 + y_2'(t)^2 + \frac{d}{dt}\big(\psi(y(t))\big)^2} \tag{4.2}$$

among the C^1 $y : [c,d] \to V$ satisfying the conditions

$$y(c) = (\gamma_1(c), \gamma_2(c)), \quad y(d) = (\gamma_1(d), \gamma_2(d)).$$

Thus, $\overline{y} = (\gamma_1, \gamma_2)$ has to satisfy the Euler equation

$$\frac{d}{dt}\left(\frac{y_1'(t) + y_1'(t)\left(\frac{\partial\psi}{\partial y_1}(y(t))\right)^2 + y_2'(t)\,\frac{\partial\psi}{\partial y_1}(y(t))\,\frac{\partial\psi}{\partial y_2}(y(t))}{\sqrt{y_1'(t)^2 + y_2'(t)^2 + \frac{d}{dt}\big(\psi(y(t))\big)^2}}\right)$$

$$= \frac{\partial}{\partial y_1}\left(\sqrt{y_1'(t)^2 + y_2'(t)^2 + \frac{d}{dt}\big(\psi(y(t))\big)^2}\right)$$

for every $t \in [c,d]$, in particular for $t = \overline{t}$. Now, by our assumption, for $y = \overline{y}$, the denominator is constant in t, as it represents $||\gamma'(t)||$ and γ is parametrized by arclenght. Hence, the derivative of the numerator is 0, but for $t = \overline{t}$, this amounts to $\overline{y}_1''(\overline{t}) = 0$, as

$$\frac{\partial\psi}{\partial y_1}(\overline{y}(\overline{t})) = \frac{\partial\psi}{\partial y_2}(\overline{y}(\overline{t})) = 0 \tag{4.3}$$

and $\overline{P} = \gamma(\overline{t}) = (0,0,0)$. Similarly, we get $\overline{y}_2''(\overline{t}) = 0$, thus $\gamma''(\overline{t})$ is a multiple of the vector $(0,0,1)$ which is, in view of (4.1), normal to S at $\overline{P}$. ∎

Remark 4.3. In the previous proof, we have used the representation of S in c_1), but such a representation is only valid in a neighborhood of $\overline{P}$, and as there is no reason that the curve γ lies in such a neighborhood, we have to work on the restriction of γ to a suitable neighborhood of $\overline{t}$. This is the reason for which we have not considered the integral on all of the interval $[a,b]$, but we have restricted it to $[c,d]$. We also remark that the considerations in previous proof are only valid at $\overline{t}$ as at the other points (4.3) does not (necessarily) hold, but once t is given we can use a suitable affine isometry (depending on t) for which (4.3) holds at t. ∎