UNIVERSIDAD COMPLUTENSE DE MADRID

Departamento de Análisis Matemático y Matemática Aplicada

Master Course: Geometric Analysis on $\mathbb{R}^n$

This course was intended for Master students on Mathematics during the years 2023-2026 developed by

Miguel García Bravo

Last update: March 2026

Introduction

During the spring semesters of four consecutive years (2023-2026) I have been teaching at University Complutense of Madrid a Master Course called *Técnicas de Análisis Geométrico*. Part of this course considers advanced analysis tools and explores geometric and measure related properties of the Euclidean space $\mathbb{R}^n$. Fundamental results in the area are considered, like, for instance, the Whitney extension theorem, Rademacher's theorem, the Morse-Sard theorem , or the Alexandrov theorem among others. As an effort of making all this content available and enyoable to undergraduate students, and trying to provide all needed details in the proofs, I have developed the following notes. One big premise that I tried to keep is that these notes are as self-contained as possible and that no big results could be used without being proved before. Moreover, at the end of each chapter, the reader will find a set of didactic exercises designed to explore some of the key ideas and subtle aspects underlying this theory.

It must be mentioned that there exist similar works considering this very same topic, or close ones. Indeed, Hajłasz, and Kinunen's notes were extremely useful during the development of this course. Still, I wanted to add my own flavor in some of the main results.

Lastly, but not least, I thank my students for important suggestions and comments that helped clarifying technical steps. In particular, I thank Pablo Mateo Torrejón, whose final degree project helped a lot to develop the chapter about Alexandrov theorem.

Index

Chapter 1

Extension theorems: McShane and Whitney

Let X be a set and $A \subset X$ a subset. The problem of extending functions $f : A \to \mathbb{R}$ to $F : X \to \mathbb{R}$ preserving some kind of regularity of the function f has been widely studied in the mathematical history. Sometimes the function f is not assumed to have any regularity property a priori, but one aims for the extension F to satisfy one.

The most basic setting could be the case of $X = T$ being a topological space, $A \subset T$ any subset and $f : A \to \mathbb{R}$ a continuous function. Question: When is it possible to find a continuous function $F : T \to \mathbb{R}$ so that $F|_A = f$.?

Theorem 1.1 (Tietze, 1925)**.** *Let T be a normal topological space (that is, every two disjoint closed sets can be included within two disjoint open sets) and let $A \subset T$ be closed subset. Then if $f : A \to \mathbb{R}$ is a continuous function, there exists $F : T \to \mathbb{R}$ continuous as well with $F|_A = f$. Moreover one can choose F so that* $\sup\{F(x) : x \in T\} = \sup\{f(a) : a \in A\}$.

Just to mention, originally Lebesgue and Brouwer proved the case $T = \mathbb{R}^n$, then Tietze in 1925 proved the case when $T = X$ is a metric space and thereafter Uryshon proved the case T is a topological space the very same year 1925. The result is also called the Uryshon-Brouwer-Tietze lemma. The proof of Theorem 1.1 requires another result due to Uryshon that states that a topological space is normal if and only if for any two disjoint closed sets A_1 and A_2 there exists a continuous function taking the value 0 on A_1 and the value 1 on A_2.

Note as well that $A \subset T$ must be a closed set, since otherwise there could be no continuous extension. Think for instance about $A = (0, +\infty)$ and $f : A \to \mathbb{R}$ defined as $f(x) = 1/x$.

For us, the interest relies on the extension problem for Lipschitz functions and for C^1 functions. To define Lipschitzianity or differentiability we need the notion of distance, so it is natural to work with metric spaces X rather than topological spaces T. In any case, the reader may have in mind that the metric space X we are dealing with is just $\mathbb{R}^n$ endowed with the euclidean distance that we write as $d(x, y) := |x - y|$.

Definition 1.2. *A set X is said to be a metric space if there exists a function $d : X \times X \to [0, \infty)$, called distance*[1]*, so that*

1. $d(x, x) = 0$.
2. $d(x, y) > 0$ *for all* $x \neq y$. *(Positivity)*
3. $d(x, y) = d(y, x)$ *for all* $x, y \in X$. *(Symmetry)*

[1] Some mathematicians allow distances to take infinity values $d : X \times X \to [0, \infty]$. This is just a matter of taste, but most of the mathematics coming thereafter are basically the same.

4. *$d(x,z) \leq d(x,y) + d(y,z)$ for all $x, y, z \in X$. (Triangle inequality).*

If only $(1), (3), 4)$ hold we say that d is a pseudometric and X a pseudometric space. A metric space X can be indistinctly denoted by $X = (X,d) = (X, d_X)$.

We want to give an answer to the following two simple questions:

1. Let $A \subset \mathbb{R}^n$ be an arbitrary subset and an L-Lipschitz function $f : A \to \mathbb{R}$. Is there an L-Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}$ with $F|_A = f$?

2. Given a closed set $C \subset \mathbb{R}^n$ and a function $f : C \to \mathbb{R}$. When can we find a function $F : \mathbb{R}^n \to \mathbb{R}$ of class C^1 so that $F|_C = f$?

1.1 Extension of Lipschitz functions

Definition 1.3. *Let (X, d_X) and (Y, d_Y) be metric spaces and $L \geq 0$. A function $f : X \to Y$ is said to be L-Lipschitz if*

$$d_Y(f(x_1), f(x_2)) \leq L d_X(x_1, x_2), \quad \forall x_1, x_2 \in X. \tag{1.1}$$

The least constant $L \geq 0$ satisfying (1.1) is denoted by Lip(f). Namely,

$$\mathit{Lip}(f) = \sup \left\{ \frac{d_Y(f(x_1), f(x_2))}{d_X(x_1, x_2)} \,:\, x_1, x_2 \in X,\, x_1 \neq x_2 \right\}.$$

The set of Lipschitz functions $f : X \to \mathbb{R}$ is denoted by Lip(X). The set of bounded Lipschitz functions $f : X \to \mathbb{R}$ is written Lip$_\infty(X)$ and when endowed with the norm

$$\|f\|_{\mathit{Lip}_\infty(X)} = \mathit{Lip}(f) + \|f\|_\infty.$$

becomes a Banach space.

Remark 1.4. The space $(\mathrm{Lip}_\infty(X), \|\cdot\|_\infty)$ is not, in general, a Banach space. It is a normed space, but completeness may fail. In other words, if we call $B(X) = \{f : X \to \mathbb{R} \text{ bounded}\}$ the space of bounded functions, we are saying that $\mathrm{Lip}_\infty(X) \subset (B(X), \|\cdot\|_\infty)$ is not a closed subspace. For example, take the sequence of Lischitz functions $(f_n)_{n\geq 1} \subset \mathrm{Lip}_\infty([0,1])$ defined as

$$f_n(x) = \sqrt{x + 1/n}.$$

It is clear that this sequence converges uniformly to the function $f(x) = \sqrt{x}$. That is, $f_n \overset{\|\cdot\|_\infty}{\longrightarrow} f$. However, f is not a Lipschitz function.

Theorem 1.5 (McShane, 1934)**.** *Let $A \subset \mathbb{R}^n$ be an arbitrary set and $L \geq 0$. Then for every L-Lipschitz function $f : A \to \mathbb{R}$ there exists $F : \mathbb{R}^n \to \mathbb{R}$ with $F|_A = f$ and so that F is L-Lipschitz.*

Proof. Let $x \in \mathbb{R}^n$ and define the extension as

$$F(x) = \inf_{y\in A}\{f(y) + Ld(x,y)\} = \inf_{y\in A}\{f(y) + L|x-y|\}.$$

Let us check that $F : \mathbb{R}^n \to \mathbb{R}$ is well-defined, that is L-Lipschitz and that $F|_A = f$.

1. F is well-defined: Let $a \in A$ fixed. Then for every $y \in A$, $x \in \mathbb{R}^n$ we have that

$$f(y) + L|y-x| \geq f(y) + L|y-a| - L|x-a| \geq f(a) - L|x-a|.$$

   We have proved that for any $x \in \mathbb{R}^n$, the set $\{f(y) + L|x-y| \,:\, y \in \mathbb{R}^n\} \subset \mathbb{R}$ is bounded from below. Therefore by the supremum property of the real numbers there must exist an infimum of such set, that is, there exists $F(x) > -\infty$.

2. $F|_A = f$: Let $x, y \in A$. Then $f(x) \le f(y) + L|x - y|$, so
$$f(x) \le \inf_{y \in A}\{f(y) + L|x - y|\} \le f(x) + L|x - x| = f(x).$$
We get that $f(x) = F(x)$.

3. F is L-Lipschitz: Let $y \in A$. We define the function $g_y : \mathbb{R}^n \to \mathbb{R}$ by
$$g_y(x) = f(y) + L|y - x|.$$
We now easily verify that g_y is L-Lipschitz because
$$|g_y(x) - g_y(z)| = L\,||y - x| - |y - z|| \le L|x - z|, \quad \forall x, z \in \mathbb{R}^n.$$
It follows that for every $x, z \in \mathbb{R}^n$
$$\begin{cases} g_y(z) \ge g_y(x) - L|x - z| \\ g_y(x) \ge g_y(z) - L|x - z| \end{cases}.$$
The previous estimates, valid for every $x, z \in \mathbb{R}^n$ and $y \in A$ yield
$$g_y(z) \ge g_y(x) - L|x - z| \ge \inf_{y' \in A}\{g_{y'}(x) - L|x - z|\} = F(x) - L|x - z|.$$
Furthermore for every $x, z \in \mathbb{R}^n$
$$\inf_{y \in A}\{g_y(z)\} = F(z) \ge F(x) - L|x - z| \quad \Rightarrow \quad F(x) - F(z) \le L|x - z|.$$
One can analogously prove that $F(z) - F(x) \le L|x - z|$, so we conclude that
$$|F(x) - F(z)| \le L|x - z|, \quad \forall x, z \in \mathbb{R}^n.$$

□

Some final comments:

- McShane theorem works the same way for extensions of Lipschitz functions $f : A \to \mathbb{R}$ where $A \subset X$ is an arbitrary subset of a metric space X.
- McShane theorem provides an extension operator $T : \mathrm{Lip}(A) \to \mathrm{Lip}(\mathbb{R}^n)$ that is in general not linear. Even when a Lipschitz function $f : A \to \mathbb{R}$ is bounded, the McShane Lipschitz extension does not provide bounded Lipschitz extensions. Although for doubling metric spaces X and for any Banach space V, there alwaysd exists a bounded linear extension operator
$$T : \mathrm{Lip}_\infty(A; V) \to \mathrm{Lip}_\infty(V),$$
being A any subset of X. We refer to [18, Chapter 4] for a proof of this last result.
- For functions $f : A \subset X \to \mathbb{R}^m$ where $m > 1$ we can apply McShane theorem componentwise and we obtain a $\sqrt{m}L$-Lipschitz extension.
- There are examples of 1-Lipschitz functions $f : A \subset X \to \mathbb{R}^2$ where $A \subset X$ is a closed subset of a metric space that do not admit 1-Lipschitz extensions to the whole $\mathbb{R}^2$.

Example 1.6. Take $A = \{(1,-1), (-1,1), (1,1)\} \subset \mathbb{R}^2$ and let $f : A \to \mathbb{R}^2$ be defined as

$$\begin{cases} f(1,-1) = (1,0) \\ f(-1,1) = (-1,0) \\ f(1,1) = (0,\sqrt{3}) \end{cases}.$$

The reader may verify that $f : (A, \|\cdot\|_\infty) \to (\mathbb{R}^2, \|\cdot\|_2)$ is 1-Lipschitz but there is not 1-Lipschitz function $F : (\mathbb{R}^2, \|\cdot\|_\infty) \to (\mathbb{R}^2, \|\cdot\|_2)$ with $F|_A = f$.

- If one is interested in preserving the Lipschitz constant on the extensions one has the following important result which we shall not prove here.

Theorem 1.7 (Kirszbraun 1934, Valentine 1945)**.** *Let H_1, H_2 be Hilbert spaces and $A \subset H_1$ an arbitrary set. Then for every L-Lipschitz function $f : A \to H_2$ there exists an L-Lipschitz function $F : H_1 \to H_2$ with $F|_A = f$.*

1.2 Extension of C^1 functions

Let $C \subset \mathbb{R}^n$ be a closed set and let $f : C \to \mathbb{R}$. We look for necessary and sufficient conditions for the existence of a function $F : \mathbb{R}^n \to \mathbb{R}$ with $F|_C = f$ so that $F \in C^k(\mathbb{R}^n)$.

Theorem 1.8 (Whitney (case C^1), 1934)**.** *Let $f : C \to \mathbb{R}$ and $L : C \to \mathbb{R}^n$ continuous functions with $C \subset \mathbb{R}^n$ a closed set. Then there exists $F : \mathbb{R}^n \to \mathbb{R}$ of class C^1 with $F|_C = f$ and with $Df|_C = L$ if and only if*

$$\lim_{y\to x} \frac{f(y) - f(x) - L(x)\cdot(y-x)}{|y-x|} = 0, \quad \text{uniformly on compact sets of } C. \tag{1.2}$$

Let us begin with some previous comments:

1. The following conditions are equivalent to (1.2), which can be written as well with the notation

$$\lim_{\substack{|x-y|\to 0 \\ x,y\in K\ x\neq y}} \frac{f(y) - f(x) - L(x)\cdot(y-x)}{|y-x|} = 0.$$

   - For all compact sets $K \subset C$ we have

$$\lim_{y\to x} \frac{f(y) - f(x) - L(x)\cdot(y-x)}{|y-x|} = 0, \quad \text{uniformly on } K.$$

   - For all compact sets $K \subset C$ and every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon, K) > 0$ so that if $0 < |x-y| < \delta$, $x, y \in K$ then

$$\frac{f(y) - f(x) - L(x)\cdot(y-x)}{|y-x|} < \varepsilon.$$

   - For every compact set $K \subset C$ we have

$$\lim_{\delta\to 0^+} \left(\sup \left\{ \frac{f(y) - f(x) - L(x)\cdot(y-x)}{|y-x|} : 0 < |x-y| < \delta,\ x, y \in K \right\} \right) = 0.$$

2. If C is formed by isolated points (1.2) gives no information.

3. Consider now the case when $C \subset \mathbb{R}^n$ is not necessarily closed. Given $f : C \to \mathbb{R}$ and $L : C \to \mathbb{R}^n$ a sufficient condition for the existence of a continuous extension of f and L to $\overline{C}$ is that f and L are uniformly continuous. Then we would still need the assumption (1.2) to hold in order to apply the theorem.

4. From (1.2) it is deduced that f is uniformly continuous on compact subsets of C. The reason is that (1.2) implies that
$$\lim_{y \to x} |f(y) - f(x)| = 0$$
uniformly on compact subsets of C.

5. If on (1.2) we do not ask for uniform convergence we would only get the existence of Taylor polynomials of degree one on C, or what is the same, we would get an extension F that would be differentiable everywhere on $\mathbb{R}^n$, but with a noncontinuous derivative $DF : \mathbb{R}^n \to \mathbb{R}^n$. That is, F would not be a C^1 extension. We next give a precise example highlighting this possibility.

Example 1.9. Let $z_1 = 1/\sqrt{3}$ and define a decreasing sequence $(z_n)_{n\geq 1} \subset (0, +\infty)$ in such a way that
$$\frac{z_n^3 + z_{n+1}^3}{z_n - z_{n+1}} = 1 \tag{1.3}$$
We have that $\lim\limits_{n\to\infty} z_n = 0$. Now let $C = \{z_n\}_{n\geq 1} \cup \{0\}$, which is a closed subset of $\mathbb{R}$. Let $f : C \to \mathbb{R}$ given by
$$f(x) = \begin{cases} 0, & x = 0 \\ (-1)^{n+1} z_n^3, & x = z_n,\ n \in \mathbb{N} \end{cases}$$
And finally let $L : C \to \mathbb{R}$ to be $L(x) = 0$ for every $x \in C$.

We want to show that for every $x \in C$, even though we have
$$\lim_{\substack{y \to x \\ y \in C}} \frac{|f(y) - f(x)|}{|y - x|} = 0 \quad \text{pointwise,} \tag{1.4}$$
there is not C^1 function $F : \mathbb{R} \to \mathbb{R}$ so that $F(x) = f(x)$, $F'(x) = 0$ for every $x \in C$. The point is that we do not have uniform converge on compact sets in the limit (1.4).

(i) We assure that (1.4) holds pointwise for every $x \in C$. If $x = 0$, it is clear that
$$\lim_{y \to 0} \frac{|f(y) - 0|}{|y - 0|} = \lim_{n \to \infty} \frac{|(-1)^{n+1} z_n^3|}{z_n} = 0.$$
In the case $x \in C \setminus \{0\}$, x would be a singleton on C and (1.4) holds trivially.

(ii) We now check that there is no C^1 function $F : \mathbb{R} \to \mathbb{R}$ so that $F(x) = f(x)$ and $F'(x) = 0$ for every $x \in C$. Arguing by contradiction, if such extension exists by the mean value theorem and for every $n \in \mathbb{N}$ there would exist $u_n \in (z_{n+1}, z_n)$ so that
$$F'(u_n) = \frac{f(z_n) - f(z_{n+1})}{z_n - z_{n+1}} = \frac{(-1)^{n+1} z_n^3 - (-1)^{n+2} z_{n+1}^3}{z_n - z_{n+1}} = \begin{cases} 1, & \text{if } n \text{ odd} \\ -1, & \text{if } n \text{ even} \end{cases}.$$
Therefore $\lim\limits_{n\to\infty} F'(u_n) \neq 0 = F'(0) = L(0)$, contradicting the continuity of F' at zero.

(iii) There is not uniform convergence on compact sets in the limit (1.4). Indeed C itself is a compact set. Take $\delta > 0$ and observe that there exists $n_0 \in \mathbb{N}$ so that $z_n \leq \delta$ for every $n \geq n_0$ (because $\lim\limits_{n\to\infty} z_n = 0$). Then
$$\sup \left\{ \frac{|f(y) - f(x)|}{|y - x|} : 0 < |y - x| < \delta\, x, y \in C \right\} \geq \sup \left\{ \frac{|f(z_n) - f(z_{n+1})|}{|z_n - z_{n+1}|} : n \geq n_0 \right\} = 1$$

and hence

$$\lim_{\delta\to 0^+}\left(\sup\left\{\frac{|f(y)-f(x)|}{|y-x|}\,:\,0<|y-x|<\delta\,x,y\in C\right\}\right)\neq 0.$$

We next state the general version of the Whitney's extension theorem for C^k regularity, $k\geq 1$. Let us recall what is the notion of multiindex.

Definition 1.10. A multiindex α is any vector $\alpha=(\alpha_1,\dots,\alpha_n)\in(\mathbb{N}\cup\{0\})^n$. We also define

$$\begin{cases}|\alpha|=\alpha_1+\alpha_2+\cdots+\alpha_n \quad \text{(this is the order of } \alpha)\\ \alpha!=(\alpha_1!)\cdot(\alpha_2!)\cdots(\alpha_n!)\\ \text{For any } x=(x_1,\dots,x_n)\in\mathbb{R}^n \text{ define } x^\alpha=x_1^{\alpha_1}\cdot x_2^{\alpha_2}\cdots x_n^{\alpha_n}\end{cases}.$$

Theorem 1.11 (Whitney (case C^k))**.** *Given a closed set $C\subset\mathbb{R}^n$ and some $k\geq 1$, a necessary and sufficient condition for a function $f:C\to\mathbb{R}$ together with a family of functions $f_\alpha:C\to\mathbb{R}$, where α is a multiindex with $|\alpha|\leq k$, to admit an extension $F:C\to\mathbb{R}^n$ (that is $F|_C=f$) of class C^k and so that $D^\alpha|_C=f_\alpha$ for every $|\alpha|\leq k$ is that for each $|\alpha|\leq k$,*

$$\lim_{y\to x}\frac{f_\alpha(y)-\sum_{|\beta|\leq k-|\alpha|}\dfrac{f_{\alpha+\beta}(x)}{\beta!}(y-x)^\beta}{|y-x|^{k-|\alpha|}}=0,\quad \textit{uniformly on compact sets of } C. \tag{1.5}$$

Observe that the conditions (1.5) imply in particular that f and all functions f_α are continuous on C.

Proof. We refer to [16, Theorem 5.4 and 5.14]. □

In these notes we only give the full proof of the C^1 version of Whitney's theorem. that is Theorem 1.8. Before we begin with the proof we need to introduce smooth partitions of unity that are used to build up the extension.

Definition 1.12. Given an open set $U\subset\mathbb{R}^n$, a partition of unity on U is a family a functions $\{\varphi_j\}_{j=1}^\infty$, $\varphi_j:U\to[0,1]$ so that

1. For every $x\in U$, $\{j\in\mathbb{N}\,:\,\varphi_j(x)\neq 0\}$ is a finite set. (This means that the supports of the functions φ_j form a point-finite covering of U).
2. $\sum_{j=1}^\infty\varphi_j(x)=1$ for every $x\in U$.

Whenever the φ_j are continuous we refer to $\{\varphi_j\}_{j\geq 1}$ as a continuous partition of unity, and whenever the φ_j are of class C^k, for some $k\geq 1$, we we say $\{\varphi_j\}_{j\geq 1}$ is a C^k partition of unity.
Given an open covering $\{U_j\}_{j\geq 1}\subset U$ of U, a given partition of unity $\{\varphi_j\}_{j\geq 1}$ is said to be subordinated to the covering $\{U_j\}_{j\geq 1}$ if for every $j\in\mathbb{N}$,

$$\text{supp}(\varphi_j):=\{x\in\mathbb{R}^n:\varphi_j(x)\neq 0\}\subset U_j.$$

We state in Lemma 1.14 a useful result about the existence of Whitney-type smooth partitions of unity on open subsets of $\mathbb{R}^n$. For that, a necessary tool is the next lemma.

Lemma 1.13 (Vitali's covering lemma / $5r$-covering lemma)**.** *Let $\mathcal{F}$ be a collection of open balls (or closed) in $\mathbb{R}^n$ with $\sup\{diam(B)\,:\,B\in\mathcal{F}\}<\infty$. Then there exists a countable family $\mathcal{G}$ of disjoint balls from $\mathcal{F}$ in such a way that*

$$\bigcup_{B\in\mathcal{F}}B\subset\bigcup_{B\in\mathcal{G}}5B,$$

where $5B$ denotes the ball with same center as B and radius five times bigger than that of B.

Proof. This is proven for instance in [12]. □

Notation: Given a closed set $C \subset \mathbb{R}^n$ and $U = \mathbb{R}^n \setminus C$, for each $x \in U$ we define

$$r(x) := \min\{1, \operatorname{dist}(x, C)\} \cdot \frac{1}{20}, \quad \text{where } \operatorname{dist}(x, C) = \inf\{|y - x| : y \in C\}.$$

Lemma 1.14 (Whitney type partition of unity). *Let $C \in \mathbb{R}^n$ be a closed set and $U = \mathbb{R}^n \setminus C$. Then*

1. *There exists an open covering of $U = \bigcup_{j=1}^{\infty} B(x_j, 5r(x_j))$ so that $\{B(x_j, r(x_j))\}$ are pairwise disjoint. (Denote from now on $r_j = r(x_j)$).*
2. *There exists a partition of unity $\{\varphi_j\}_{j\geq 1}$ on U of class C^∞ that is subordinated to the covering $\{B(x_j, 10r_j)\}_{j=1}^{\infty}$ so that*

$$|D\varphi_j(x)| \leq \frac{C(n)}{r_j}, \quad \forall x \in B(x_j, 10r_j). \tag{1.6}$$

where $C(n)$ denotes a positive constant only depending on n, the dimension of the space[a].

[a] We warn the reader that the appearance of $C(n)$ in some estimates may vary from line to line.

Proof. Take the following open cover of U,

$$U = \bigcup_{x \in U} B(x, r(x)).$$

By Vitali's covering Lemma 1.13 there exists $\{x_j\}_{j\geq 1} \subset U$ so that $U = \bigcup_{j\geq 1} B(x_j, 5r_j)$ and $\{B(x_j, r_j)\}_{j\geq 1}$ are pairwise disjoint. For a given $x \in U$ define the set

$$A_x := \{j \in \mathbb{N} : B(x, 10r(x)) \cap B(x_j, 10r_j) \neq \emptyset\}.$$

We have the following properties.

(a) For every $j \in A_x$ we have $1/3 \leq r(x)/r_j \leq 3$.

Indeed if $j \in A_x$,

$$\begin{aligned} |r(x) - r_j| &= \frac{1}{20}|\min\{1, \operatorname{dist}(x, C)\} - \min\{1, \operatorname{dist}(x_j, C)\} \\ &\leq \frac{1}{20}|x - x_j| \leq \frac{1}{20}(10r(x) + 10r_j) = \frac{1}{2}r(x) + r_j. \end{aligned}$$

Then $r(x) \leq 3r_j$ and $r_j \leq 3r(x)$.

(b) $\#(A_x) = \operatorname{card}(A_x) \leq C(n) = 129^n$ for every $x \in U$.

From the previous property (a), given some $j \in A_x$,

$$|x - x_j| + r_j \leq 10(r(x) + r_j) + r_j = 10r(x) + 11r_j \leq 10r(x) + 33r(x) = 43r(x). \tag{1.7}$$

Take $x \in U$. For every $j \in A_x$, since $r(x) \leq 3r_j$, we get that

$$\mathcal{L}^n\left(\bigcup_{j \in A_x} B\left(x_j, \frac{r(x)}{3}\right)\right) \leq \mathcal{L}^n\left(\bigcup_{j \in A_x} B(x_j, r_j)\right) \leq \mathcal{L}^n(B(x, 43r(x))) \tag{1.8}$$

The last inequality follows from the fact that if $j \in A_x$ and we take $y \in B(x_j, r_j)$ then, using (1.7), we have $|y - x| \leq |y - x_j| + |x_j - x| \leq r_j + |x_j - x| \leq 43r(x)$. Next, since $\{B(x_j, r_j)\}_{j=1}^{\infty}$

are pairwise disjoint and $r(x) \leq 3r_j$ for every $j \in A_x$, then $\{B(x_j, r(x)/3)\}_{j\in A_x}$ are pairwise disjoint too and from (1.8) we conclude that for every $x \in U$ (call $\omega_n = \mathcal{L}^n(B(0,1))$)

$$\text{card}(A_x) \cdot \omega_n \cdot \left(\frac{r(x)}{3}\right)^n \leq \omega_n \cdot (43r(x))^n \quad \Rightarrow \quad \text{card}(A_x) \leq 129^n$$

Let us begin to build our partition of unity. Start by taking a C^∞ smooth function $h : \mathbb{R} \to [0,1]$ so that $h(t) = 1$ for every $|t| \leq 1$ and $h(t) = 0$ for every $|t| \geq 2$. Let $M = \max_{t\in\mathbb{R}} |h'(t)|$. For every $j \in \mathbb{N}$ define

$$h_j(x) = h\left(\frac{|x - x_j|}{5r_j}\right), \quad x \in \mathbb{R}^n.$$

We now enumerate some properties:

- $h_j \in C^\infty(U)$.
- $h_j = 1$ on $B(x_j, 5r_j)$ and $h_j = 0$ on $\mathbb{R}^n \setminus B(x_j, 10r_j)$.
- Given $x \in U$, $h_j(x) = 0$ for every $j \notin A_x$. In particular $\{\text{supp}(h_j)\}_{j\geq 1}$ is a locally finite open covering of U.
- We have the following growth control on the bumps h_j. For every $x \in B(x_j, 10r_j)$,

$$|\nabla h_j(x)| = \left|h'\left(\frac{|x - x_j|}{5r_j}\right)\frac{1}{5r_j}|\cdot|'(x - x_j)\right| \leq \frac{M}{5r_j}\left|\frac{(x - x_j)}{|x - x_j|}\right| = \frac{M}{5r_j}.$$

Note at this point that the family $\{h_j\}_{j\geq 1}$ does not yet form a partition of unity. The reason is that in general, for a given $x \in U$ we have

$$H(x) \colon = \sum_{j=1}^{\infty} h_j(x) \geq 1.$$

To solve this situation we need to define our final partition of unity $\{\varphi_j\}_{j\geq 1}$ as

$$\varphi_j(x) = \frac{h_j(x)}{H(x)}.$$

It is easy to check now that $\{\varphi_j\}_{j\geq 1}$ forms a partitions of unity subordinated to the covering $\{B(x_j, 10r_j)\}_{j\geq 1}$ and so that

- $\varphi_j \in C^\infty(U)$.
- $\sum_{j=1}^{\infty} h_j(x) = 1$ for every $x \in U$.
- $\varphi_j = 1$ on $B(x_j, 5r_j)$ and $\varphi_j = 0$ on $\mathbb{R}^n \setminus B(x_j, 10r_j)$.
- Given $x \in U$, $\varphi_j(x) = 0$ for every $j \notin A_x$. In particular $\{\text{supp}(h_j)\}_{j\geq 1}$ is a locally finite open covering of U.
- Given $x \in U$ we have

$$|\nabla H(x)| \leq \sum_{j\in A_x} |h_j'(x)| \leq \sum_{j\in A_x} \frac{M}{5r_j} \leq \sum_{j\in A_x} \frac{3M}{5r(x)} = \text{card}(A_x)\frac{3M}{5r(x)} = \frac{C(n)}{r(x)}.$$

  Consequently for every $x \in \text{supp}(\varphi_j)$ (in particular $j \in A_x$)

$$|\nabla\varphi_j'(x)| = \left|\frac{\nabla h_j(x)}{H(x)} - \frac{h_j(x)\nabla H(x)}{H^2(x)}\right| \leq |\nabla h_j(x)| + |\nabla H(x)| \leq \frac{M}{5r_j} + \frac{C(n)}{r(x)} \leq \frac{C(n)}{r_j}.$$

□

Proof of Whitney's theorem 1.8. Let $U = \mathbb{R}^n \setminus C$ be the complement of C, which is an open set of $\mathbb{R}^n$. Let $\{\varphi_j\}_{j\geq 1}$ the C^∞ partition of unity of U subordinated to $\{B(x_j, 10r_j)\}_{j\geq 1}$ given by Lemma 1.14. For each $j \in \mathbb{N}$ let $\widetilde{x_j} \in C$ such that $\text{dist}(x_j, C) = |x_j - \widetilde{x_j}|$ (note that $\widetilde{x_j}$ may not be unique).

We define the extension $F : \mathbb{R}^n \to \mathbb{R}$ as

$$F(x) = \begin{cases} f(x), & \text{if } x \in C \\ \sum_{j=1}^{\infty} \varphi_j(x)(f(\widetilde{x_j}) + L(\widetilde{x_j}) \cdot (x - \widetilde{x_j}), & \text{if } x \in U \end{cases} \tag{1.9}$$

The first thing to note is that $F \in C^\infty(U)$ because U is open and every $x \in U$ has a neighborhood $U_x = B(x, 10r(x))$ whose points satisfy

$$F(y) = \sum_{j \in A_x} \varphi_j(y)(f(\widetilde{x_j}) + L(\widetilde{x_j}) \cdot (y - \widetilde{x_j})), \quad y \in U_x.$$

This means that locally we have a finite sum of $C^\infty(U)$ functions. Moreover we have that for every $x \in U$,

$$\nabla F(x) = \sum_{j \in A_x} \nabla\varphi_j(x)(f(\widetilde{x_j}) + L(\widetilde{x_j}) \cdot (x - \widetilde{x_j})) + \varphi_j(x)L(\widetilde{x_j}).$$

We now divide the proof into two steps.

(**Step 1**) F is differentiable at every $x \in C$ with $\nabla F(x) = L(x)$. (Together with the fact that $F \in C^\infty(U)$ we get that F is differentiable everywhere on $\mathbb{R}^n$).

We need to check that for a given $x \in C$,

$$\lim_{y \to x} \frac{F(y) - F(x) - L(x) \cdot (y - x)}{|y - x|} = 0.$$

Since (1.2) implies

$$\lim_{\substack{y \to x \\ y \in C}} \frac{F(y) - F(x) - L(x) \cdot (y - x)}{|y - x|} = \lim_{\substack{y \to x \\ y \in C}} \frac{f(y) - f(x) - L(x) \cdot (y - x)}{|y - x|} = 0$$

we only need to prove that

$$\lim_{\substack{y \to x \\ y \notin C}} \frac{F(y) - F(x) - L(x) \cdot (y - x)}{|y - x|} = 0. \tag{1.10}$$

Let $y \notin C$ with $|y - x| \leq 1$. We have

$$\begin{aligned} |F(y) - F(x) - L(x) \cdot (y - x)| &= \left| \sum_{j \in A_y} \varphi_j(y) \left(f(\widetilde{x_j}) + L(\widetilde{x_j}) \cdot (y - \widetilde{x_j}) - f(x) - L(x) \cdot (y - x) \right) \right| \\ &\leq \sum_{j \in A_y} \varphi_j(y) \left(|f(\widetilde{x_j}) - f(x) - L(x) \cdot (\widetilde{x_j} - x)| + |(L(x) - L(\widetilde{x_j}) \cdot (\widetilde{x_j} - y)| \right) \end{aligned}$$

We note that for every $j \in A_y$, recalling that $|x_j - \widetilde{x_j}| = d(x_j, C)$,

$$\begin{aligned} |\widetilde{x_j} - x| &\leq |x - x_j| + |x_j - \widetilde{x_j}| \leq 2|x - x_j| \leq 2(|x - y| + |y - x_j|) \\ &\leq 2(|x - y| + 10(r(y) + r_j)) \leq 2|x - y| + 20r(y) + 20r_j \leq 2|x - y| + 80r(y) \\ &\leq 2|x - y| + 4|x - y| \leq 6|x - y| \end{aligned} \tag{1.11}$$

and

$$|\widetilde{x_j} - y| \leq |\widetilde{x_j} - x| + |x - y| \leq 7|x - y|. \tag{1.12}$$

By joining everything we get that

$$\frac{|F(y)-F(x)-L(x)\cdot(y-x)|}{|y-x|} \leq \sum_{j\in A_y} \varphi_j(y) \frac{|f(\widetilde{x}_j)-f(x)-L(x)\cdot(\widetilde{x}_j-x)|}{(1/6)|\widetilde{x}_j-x|} + \sum_{j\in A_y} \varphi_j(y) \frac{|L(x)-L(\widetilde{x}_j)|\cdot|\widetilde{x}_j-y|}{(1/7)|\widetilde{x}_j-y|}.$$

Therefore for any $\varepsilon>0$ by the continuity of L and by (1.2), there exists $\delta\in(0,1]$ so that if $0<|z-x|<\delta$ with $z\in C$ then

$$\frac{|f(z)-f(x)-L(x)\cdot(z-x)|}{|z-x|} < \frac{\varepsilon}{13} \quad \text{and} \quad |L(x)-L(z)| < \frac{\varepsilon}{13}.$$

Now if we take $0<|y-x|\leq\frac{\delta}{7}$ with $y\notin C$ we have that for every $j\in A_y$, $|\widetilde{x}_j-x|\leq 6|x-y|<\delta$ and hence

$$\frac{|F(y)-F(x)-L(x)\cdot(y-x)|}{|y-x|} \leq \sum_{j\in A_y}\varphi_j(y)\left(6\frac{\varepsilon}{13}\right)+\sum_{j\in A_y}\varphi_j(y)\left(7\frac{\varepsilon}{13}\right)=\varepsilon\sum_{j\in A_y}\varphi_j(y)=\varepsilon.$$

We have proved (1.10).

(**Step 2**) $\nabla F(x)$ is continuous at every $x\in C$. For that we need to take $x\in C$ and check that $\lim_{y\to x}\nabla F(y)=L(x)$. Observe that this finishes the proof because ∇F was already continuous on U.

Note that by the continuity of L on C we have

$$\lim_{\substack{y\to x\\ y\in C}}|\nabla F(y)-L(x)| = \lim_{\substack{y\to x\\ y\in C}}|L(y)-L(x)|=0$$

so we only need to prove

$$\lim_{\substack{y\to x\\ y\notin C}}|\nabla F(y)-L(x)|=0 \tag{1.13}$$

Let $y\notin C$ with $|y-x|\leq 1/2$ and let $\widetilde{y}\in C$ a point so that $d(y,C)=|y-\widetilde{y}|$. We write

$$|\nabla F(y)-L(x)|\leq|\nabla F(y)-L(\widetilde{y})|+|L(\widetilde{y})-L(x)|.$$

Since we have that $|\widetilde{y}-x|\leq|\widetilde{y}-y|+|y-x|\leq 2|y-x|$ it is clear that by the continuity of L on C

$$\lim_{\substack{y\to x\\ y\notin C}}|L(\widetilde{y})-L(x)|=0.$$

We focus our attention on the term $|\nabla F(y)-L(\widetilde{y})|$. By using that for every $y\in U$, $\sum_{j\geq 1}\nabla\varphi_j(y)=0$

we have

$$
\begin{aligned}
|\nabla F(y) - L(\widetilde{y})| &= \left| \sum_{j \in A_y} \nabla \varphi_j(y) \left(f(\widetilde{x_j}) + L(\widetilde{x_j}) \cdot (y - \widetilde{x_j}) \right) + \varphi_j(y)(L(\widetilde{x_j}) - L(\widetilde{y})) \right| \\
&= \left| \sum_{j \in A_y} \nabla \varphi_j(y) \left(f(\widetilde{x_j}) - f(\widetilde{y}) + L(\widetilde{x_j}) \cdot (\widetilde{y} - \widetilde{x_j}) + (L(\widetilde{y}) - L(\widetilde{x_j}) \cdot (\widetilde{y} - y) \right) \right. \\
&\qquad \left. + \sum_{j \in A_y} \varphi_j(y)(L(\widetilde{x_j}) - L(\widetilde{y})) \right| \\
&\leq \sum_{j \in A_y} |\nabla \varphi_j(y)| \cdot |f(\widetilde{y}) - f(\widetilde{x_j}) - L(\widetilde{x_j}) \cdot (\widetilde{y} - \widetilde{x_j})| \\
&\qquad + \sum_{j \in A_y} |\nabla \varphi_j(y)| \cdot |L(\widetilde{y}) - L(\widetilde{x_j}| \cdot |\widetilde{y} - y)| + \sum_{j \in A_y} \varphi_j(y) |L(\widetilde{x_j}) - L(\widetilde{y})| \\
&\leq \sum_{j \in A_y} \frac{C(n)}{r_j} |f(\widetilde{y}) - f(\widetilde{x_j}) - L(\widetilde{x_j}) \cdot (\widetilde{y} - \widetilde{x_j})| \\
&\qquad + \sum_{j \in A_y} \frac{C(n)}{r_j} \cdot |L(\widetilde{y}) - L(\widetilde{x_j}| \cdot |\widetilde{y} - y)| + \sum_{j \in A_y} \varphi_j(y) |L(\widetilde{x_j}) - L(\widetilde{y})|
\end{aligned}
$$

To estimate these three summands, note that since $j \in A_y$,

$$
\begin{aligned}
|\widetilde{y} - \widetilde{x_j}| &\leq |\widetilde{y} - y| + |y - x_j| + |x_j - \widetilde{x_j}| = 20r(y) + |y - x_j| + 20r_j \\
&\leq 20r(y) + 10r(y) + 10r_j + 20r_j \leq \begin{cases} 120r_j \\ 120r(y) = 6|y - \widetilde{y}| \end{cases} .
\end{aligned}
$$

Moreover, we have $|\widetilde{y} - y| = 20r(y) \leq 60r_j$. All of this information allows us to write

$$
\begin{aligned}
|\nabla F(y) - L(\widetilde{y})| &\leq \sum_{j \in A_y} C(n) \frac{|f(\widetilde{y}) - f(\widetilde{x_j}) - L(\widetilde{x_j}) \cdot (\widetilde{y} - \widetilde{x_j})|}{|\widetilde{y} - \widetilde{x_j}|} \\
&\qquad + \sum_{j \in A_y} C(n) \cdot |L(\widetilde{y}) - L(\widetilde{x_j}| + \sum_{j \in A_y} \varphi_j(y) |L(\widetilde{x_j}) - L(\widetilde{y})|
\end{aligned}
$$

Using that (1.2) holds uniformly on compact sets of C and since L is uniformly continuous on compact sets of C we have that for any given $\varepsilon > 0$ and given the compact set $K_x = \overline{B(x, 1/2)} \cap C$, there is $\delta \in (0, 1/2)$ so that if $0 < |a - b| < \delta$ with $a, b \in K_x \cap C$, then

$$
\begin{cases} \dfrac{f(a) - f(b) - L(b) \cdot (a - b)|}{|a - b|} < \varepsilon \\ |L(a) - L(b)| < \varepsilon \end{cases} .
$$

So if we let $0 < |y - x| < \delta/6$ with $y \notin C$, then for every $j \in A_y$ we have $|\widetilde{x_j} - \widetilde{y}| < 6|y - \widetilde{y}| \leq 6|y - x| < \delta$ and hence

$$
|\nabla F(y) - L(\widetilde{y})| \leq \sum_{j \in A_y} C(n)(2\varepsilon) + \sum_{j \in A_y} \varphi_j(y)\varepsilon = \mathrm{card}(A_y) C(n) 2\varepsilon + \varepsilon \leq C(n)\varepsilon.
$$

We have now proved (1.13). $\square$

Exercises

1. Prove that the space of real-valued bounded Lipschitz functions $\mathrm{Lip}_\infty(X)$, where X is a metric space, must be a Banach space.

2. Consider the set $A = \{(1,-1),(-1,1),(1,1)\} \subset \mathbb{R}^2$ and define $f : (A, \|\cdot\|_\infty) \to (\mathbb{R}^2, \|\cdot\|_2)$ by

$$\begin{cases} f(1,-1) = (1,0) \\ f(-1,1) = (-1,0) \\ f(1,1) = (0,\sqrt{3}) \end{cases}.$$

   Prove that f is $1-$Lipschitz but there does not exist a function $F : (\mathbb{R}^2, \|\cdot\|_\infty) \to (\mathbb{R}^2, \|\cdot\|_2)$ so that $F|_A = f$ and being $1-$Lipschitz.

3. We take $\mathbb{R}^2$ endowed with the norm $\|(x,y)\|_1 = |x|+|y|$. Consider the function $f(x,y) = |x|-|y|$, defined only on $S^1 = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 = 1\}$.

   (a) Prove that $f : S^1 \to \mathbb{R}$ is 1-Lipschitz.

   (b) Prove that any extension $F : \overline{B(0,1)} \to \mathbb{R}$, which is 1-Lipschitz satisfies $F(0,0) \le 0$ and that $F(x,0) = |x|$ for all $x \in [-1,1]$.

   (c) Conclude that any extension $F : \overline{B(0,1)} \to \mathbb{R}$, that is 1-Lipschitz cannot be differentible at $(0,0)$.

4. Prove that McShane theorem also works by defining the extension of $f : A \to \mathbb{R}$ as

$$G(x) = \sup_{y\in A}\{f(y) - L|x-y|\}, \quad x \in \mathbb{R}^n.$$

   Moreover, prove that G is the ”smallest” Lipschitz possible extension of f (that is, for any other L-Lipschitz extension $h : \mathbb{R}^n \to \mathbb{R}$ we have $G(x) \le h(x)$ for all $x \in \mathbb{R}^n$).

   Analogously, prove that the extension $F(x) = \inf_{y\in A}\{f(y) + L|x-y|\}$ defines the ”biggest” L-Lipschitz extension of f.

5. Let $F \subset \mathbb{R}^n$ be a closed set and define its distance to a point $x \in \mathbb{R}^n$ as

$$d(x,F) = \inf\{|x-y| : y \in F\}.$$

   (a) Prove that $x \to d(x,F)$ defines a 1-Lipschitz function.

   (b) Prove that there exists $y \in F$ so that $d(x,F) = |x-y|$, or what is the same

$$d(x,F) = \min\{|x-y| : y \in F\}.$$

   (c) Given two closed sets $F_1, F_2 \subset \mathbb{R}^n$ we define their distance as $d(F_1,F_2) = \inf\{|x-y| : x \in F_1,\ y \in F_2\}$. Provide an example of two closed sets $F_1, F_2 \subset \mathbb{R}^n$ so that $d(F_1,F_2) = 0$ but $F_1 \cap F_2 = \emptyset$.

6. Find good examples as required in the following items.

   (a) A Banach space V so that there exist points $x, y \in V$ such that

$$\mathrm{int}\{z \in V : ||z-x|| = ||z-y||\} \neq \emptyset.$$

   (b) A metric space (X,d), such that for every $x \in X$ we have

$$\partial\{y \in X : d(x,y) < 1\} \neq \{y \in X : d(x,y) = 1\}.$$

7. Let $A \subset \mathbb{R}^n$ be an arbitrary set. We say that $f : A \to \mathbb{R}$ is Hölder continuous with constants $C \geq 0$, $\alpha \in (0,1)$ if $|f(x) - f(y)| \leq C|x - y|^\alpha$ for all $x, y \in A$.

   Prove that a "slight modification" of McShane's proof allows to extend Hölder continuous functions $f : A \to \mathbb{R}$ to the whole $\mathbb{R}^n$ (having the extension the same Hölder constants), being A an arbitray set.

8. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function and let $\varepsilon : \mathbb{R}^n \to (0, \infty)$ be a positive continuous function. By using the technique of partitions of unity give an explicit example of a function $F : \mathbb{R}^n \to \mathbb{R}$ so that $F \in C^\infty(\mathbb{R}^n)$ and $|F(x) - f(x)| \leq \varepsilon(x)$ for all $x \in \mathbb{R}^n$.

9. Let $C \subset \mathbb{R}^n$ be a closed set. Given a $L-$Lipschitz function $f : C \to \mathbb{R}$, use the Whitney partitions of unity from the proof of Whitney's extension theorem to show that

$$F(x) = \begin{cases} f(x), & x \in C \\ \sum_{j \geq 1} \varphi_j(x) f(\tilde{x}_j), & x \notin C \end{cases}$$

   defines a cL-Lipschitz extension of f with $c > 0$ a constant. (Here we use the same notation as in the proof of the Whiteny's extension theorem for φ_j and $\tilde{x}_j$).

10. By a dyadic cube we understand a set of the form $Q = [0, 2^{-k}]^n + \mathtt{j} \subset \mathbb{R}^n$ for some $k \in \mathbb{Z}$ and $\mathtt{j} \in 2^{-k}\mathbb{Z}^n$. Given an open set $U \subset \mathbb{R}^n$ we say that the family of cubes $\mathcal{W} = \{Q_i\}_{i \in \mathbb{N}}$ is a Whitney decomposition of U if the following properties hold.

    (W1) Each Q_i is a dyadic cube inside U.

    (W2) $U = \bigcup_i Q_i$ and for all $i \neq j$ we have $\mathrm{int}(Q_i) \cap \mathrm{int}(Q_j) = \emptyset$.

    (W3) For every i we have $\sqrt{n}\ell(Q_i) \leq \mathrm{dist}(Q_i, \partial U) \leq 4\sqrt{n}\ell(Q_i)$.

    (The existence of such decompositions appears in [25]).

    Prove that

    (a) If $Q_i \cap Q_j \neq \emptyset$, then $\frac{1}{4}\ell(Q_i) \leq \ell(Q_j) \leq 4\ell(Q_i)$.

    (b) For every $i \in \mathbb{N}$

$$\mathrm{card}\{j \in \mathbb{N} : Q_j \cap Q_i \neq \emptyset\} \leq C(n).$$

    Find, if possible, the best constant $C(n)$ for which the above inequality holds.

11. Prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is a C^1 function, then for every compact set $K \subset \mathbb{R}^n$,

$$\lim_{y \to x,\, y \in K} \frac{f(y) - f(x) - Df(x) \cdot (y - x)}{|y - x|} = 0 \quad \text{uniformly on } K.$$

    This means that given a compact set $K \subset \mathbb{R}^n$ y $\varepsilon > 0$ there exists $\delta > 0$ so that if $0 < |x - y| < \delta$ with $x, y \in K$ then

$$\frac{f(y) - f(x) - Df(x) \cdot (y - x)}{|y - x|} < \varepsilon.$$

Chapter 2

Rademacher and Stepanov theorem

In this chapter, we will give the proof of the classical Rademacher theorem about differentiability almost everywhere of Lipschitz functions. Thereafter, we will present one generalization: the Stepanov theorem. In Chapter 3 we further show another generalization of the Rademacher theorem for Sobolev functions $W^{1,p}(\mathbb{R}^n)$, $p > n$.

2.1 Classical result

We aim to prove the following result.

Theorem 2.1 (Rademacher, 1919). *Let $U \subset \mathbb{R}^n$ be an open set. If $f : U \to \mathbb{R}$ is Lipschitz then f is differentiable $\mathcal{L}^n$-almost everywhere on U.*

Remark 2.2. Observe that for locally Lipschitz functions the result still holds because one can always cover the set U by a countable number of balls where the function will be Lipschitz, then apply Theorem 2.1, and conclude by using that a countable number of negligible sets is negligible.

We first give the proof of the case $n = 1$ and then we consider $n \geq 2$. For the case $n = 1$ we need to introduce absolutely continuous functions.

Definition 2.3. A function $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ is absolutely continuous function if for every $\varepsilon > 0$ there exists $\delta > 0$ so that if $(x_1, x_1 + h_1), (x_2, x_2 + h_2) \ldots, (x_k, x_k + h_k)$ are disjoint subintervals of $[a, b]$ such that $\sum_{i=1}^{k} h_i < \delta$ then $\sum_{i=1}^{k} |f(x_i + h_i) - f(x_i)| < \varepsilon$.

Theorem 2.4. *If $f : [a, b] \to \mathbb{R}$ is absolutely continuous, then f is differentiable at $\mathcal{L}^1$-almost every point $x \in [a, b]$ with $f' \in L^1([a, b])$ and so that*

$$f(x) = f(a) + \int_a^x f'(t)\,dt \quad \textit{for every } x \in [a, b].$$

Indeed f is absolutely continuous if and only if there is some $g \in L^1([a, b])$ so that

$$f(x) = f(a) + \int_a^x g(t)\,dt \quad \textit{for every } x \in [a, b].$$

Proof. A very instructive proof can be found in Rudin's book [23, Chapter 7]. □

In general almost everywhere differentiability does not imply absolutely continuity. For instance Cantor staircase function is continuous and almost everywhere differentiable but it is not absolutely continuous.

Corollary 2.5. *If $f : [a,b] \subset \mathbb{R} \to \mathbb{R}$ is Lipschitz then f is differentiable $\mathcal{L}^1$-almost everywhere.*

Proof. Just observe that Lipschitz functions are absolutely continuous and apply Theorem 2.4. □

Note that there are absolutely continuous functions, like $f(x) = \sqrt{x}$, $x \in [0,1]$ that are absolutely continuous but not Lipschitz. Before we give the proof of Rademacher's theorem for higher dimensions $n \geq 2$ we need the following lemma.

Lemma 2.6. *Let $f \in L^1_{loc}(U)$ with $U \subset \mathbb{R}^n$ open so that*

$$\int_U f(x)\varphi(x)\,dx = 0 \quad \textit{for every } \varphi \in C_0^\infty(U)$$

where $C_0^\infty(U) = \{\varphi \in C^\infty(U) : \overline{supp(f)} = \overline{\{x \in U : f(x) \neq 0\}} \subset U$ is compact$\}$. Then $f(x) = 0$ for $\mathcal{L}^n$-almost every $x \in U$.

Proof. We argue by contradiction. Assume that $\mathcal{L}^n(\{x \in U : f(x) \neq 0\}) > 0$. Without loss of generality, we suppose that $\mathcal{L}^n(\{x \in U : f(x) > 0\}) > 0$ (otherwise take $-f$).

By the interior regularity of the Lebesgue measure $\mathcal{L}^n$ there exists $\varepsilon > 0$ and a compact set $K \subset U$ so that

$$\begin{cases} \mathcal{L}^n(K) > 0 \\ K \subset \{x \in U : f(x) \geq \varepsilon\} \end{cases}.$$

Note that, whenever $U \neq \mathbb{R}^n$, and due to the compactness of K we have $D :=$ dist$(K, \partial U) > 0$. In the case $U = \mathbb{R}^n$ let $D = 1$. We define the open sets

$$G_i := \left\{x \in U : \text{dist}(x,K) < \frac{D}{2i}\right\}, \quad i \in \mathbb{N}.$$

We have $\{G_i\}_{i\geq 1} \subset U$ forming a decreasing sequence of open sets

$$K \subset \overline{G_{i+1}} \subset G_i \subset \overline{G_i} \subset \cdots \subset \overline{G_1} \subset U,$$

where $\overline{G_i}$ is a compact set (in particular $\mathcal{L}^n(G_i) < \infty$ for every $i \in \mathbb{N}$). Let us take functions $\varphi_i \in C_0^\infty(G_i)$ with $0 \leq \varphi_i(x) \leq 1$ for all $x \in G_i$ and $\varphi_i = 1$ on K. Then for every $i \in \mathbb{N}$,

$$\begin{aligned} 0 = \int_U f(x)\varphi_i(x)\,dx &= \int_K f(x)\varphi_i(x)\,dx + \int_{U\setminus K} f(x)\varphi_i(x)\,dx \\ &= \int_K f(x)\,dx + \int_{G_i\setminus K} f(x)\varphi_i(x)\,dx \\ &\geq \varepsilon\mathcal{L}^n(K) - \int_{G_i\setminus K} |f(x)|\,dx. \end{aligned}$$

Now, since $\mathcal{L}^1(G_1) < \infty$ and $\bigcap_{i\geq 1} G_i \setminus K = \emptyset$ we get

$$\lim_{i\to\infty} \mathcal{L}^n(G_i \setminus K) = \mathcal{L}^n\left(\bigcap_{i\geq 1} G_i \setminus K\right) = 0.$$

This fact, together with the absolute continuity of the integral[1], and the assumption that $|f| \in L^1(G_1)$, implies that

$$\lim_{i\to\infty} \int_{G_i\setminus K} |f(x)|\,dx = 0.$$

[1] For a (Lebesgue) measurable set $F \subset \mathbb{R}^n$ let $f : F \to [0,\infty)$ be such that $f \in L^1(F)$. Then for every $\varepsilon > 0$ there exists $\delta > 0$ so that if $A \subset F$ is measurable with $\mathcal{L}^n(A) < \delta$ then $\int_A f(x)\,dx < \varepsilon$.

Hence

$$0 \geq \varepsilon \mathcal{L}^n(K) + \lim_{i\to\infty} \int_{G_i\setminus K} |f(x)|\, dx = \varepsilon \mathcal{L}^n(K) > 0.$$

This is a contradiction, so the lemma is proved. □

Let us now prove the main result of this chapter.

Proof of Rademacher theorem (Theorem 2.1). For each $v \in S^{n-1} = \{v \in \mathbb{R}^n : |v| = 1\}$ and each $x \in U$, we define directional derivative of f at x in the direction v, whenever it exists, by

$$D_v f(x) = \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t} = \frac{d}{dt}|_{t=0} f(x+tv).$$

Our first objective is to prove that once $v \in S^{n-1}$ is fixed, then $D_v f(x)$ exists for $\mathcal{L}^n$-almost every $x \in U$. Fix $v \in S^{n-1}$ and let

$$A_v = \{x \in U : D_v f(x) \text{ does not exist}\}.$$

We have the following properties:

- The set A_v is measurable. To prove this, consider

$$\begin{cases} \overline{D_v} f(x) = \limsup_{t\to 0} \dfrac{f(x+tv)-f(x)}{t} = \lim_{k\to\infty} \left(\sup_{\substack{0<|t|<1/k \\ t \text{ rational}}} \dfrac{f(x+tv)-f(x)}{t} \right) \\ \underline{D_v} f(x) = \liminf_{t\to 0} \dfrac{f(x+tv)-f(x)}{t} = \lim_{k\to\infty} \left(\inf_{\substack{0<|t|<1/k \\ t \text{ rational}}} \dfrac{f(x+tv)-f(x)}{t} \right) \end{cases}.$$

  We observe that, since the supremum (respectively, infimum) over a countable family of measurable functions is measurable, and since the pointwise limit of a sequence of measurable functions is measurable, both $\overline{D_v} f, \underline{D_v} f : U \to [-\infty, +\infty]$ are measurable functions. Finally, note that

$$A_v = \{x \in U : \overline{D_v} f(x) > \underline{D_v} f(x)\} = (\overline{D_v} f - \underline{D_v} f)^{-1}((0, +\infty])$$

  and is therefore a measurable set.

- For every line L parallel to the vector v, we have $\mathcal{H}^1(A_v \cap L) = 0$. Indeed, any such line can be written as $L = L_x = \{x + tv : t \in \mathbb{R}\}$ for some $x \in \mathbb{R}^n$. Then we define the function $f_x : \mathbb{R} \to \mathbb{R}$ as $f_x(t) = f(x+tv)$. Since f is Lipschitz we have that f_x is Lipschitz and hence absolutely continuous[2]. We then apply Corollary 2.5 and it is clear that f_x is differentiable at $\mathcal{H}^1$-almost every $t \in \mathbb{R}$. This means that $D_v f(x+tv)$ exists for $\mathcal{H}^1$-almost every $t \in \mathbb{R}$. Call $N \subset \mathbb{R}$ the null set such that $D_v f(x+tv)$ exists for all $t \in \mathbb{R} \setminus N$.
  Define the function $h : \mathbb{R} \to \mathbb{R}^n$ given by $h(t) = x + tv$. We have that $h(\mathbb{R}) = L_x$ and

$$D_v f(y) \text{ exists for all } y \in h(\mathbb{R} \setminus N).$$

  Since h is bijective, $h(\mathbb{R} \setminus N) = h(\mathbb{R}) \setminus h(N) = L_x \setminus h(N)$. Also, using the fact that h is Lipschitz and the properties of the Hausdorff measure, we have that $\mathcal{H}^1(N) = 0$ implies that $\mathcal{H}^1(h(N)) = 0$. Therefore we conclude that

$$D_v f(y) \text{ exists for all } y \in L_x \setminus h(N),$$

  so $D_v f$ exists at $\mathcal{H}^1$-almost every point of L_x. That is $\mathcal{H}^1(A_v \cap L_x) = 0$, and we are done.

[2] Depending on the shape of U, the function f_x does not need to be defined on an interval or on $\mathbb{R}$, but it is for sure defined on at most a countable union of open intervals. So we can proceed analogously to reach the conclusion that $D_v f(y)$ exists at $\mathcal{H}^1$ almost every $y \in L_x \cap U$.

We now apply Fubini's theorem to conclude that $\mathcal{L}^n(A_v) = 0$. Namely, for a given $v \in S^{n-1}$ call W_v its normal hyperplane. We have $\mathbb{R}^n = W_v \oplus \operatorname{span}\{v\}$ and we know that $\mathcal{L}^n = \mathcal{H}^{n-1}|_{W_v} \times \mathcal{H}^1|_{\operatorname{span}\{v\}}$. For $x \in W_v$, the intersection of A_v with the line L_x passing through x with direction v produces a section in A_v that we call

$$B_x = A_v \cap L_x$$

Since $A_v \subset \mathbb{R}^n$ is measurable, we have, by Fubini's Theorem

$$\mathcal{L}^n(A_v) = \int_{W_v} \mathcal{H}^1(B_x)\, d\mathcal{H}^{n-1}(x)$$

so, recalling that $\mathcal{H}^1(B_x) = 0$ for every $x \in W_v$, we get $\mathcal{L}^n(A_v) = 0$.

In particular, for every $i = 1, \dots, n$ and for the canonical vectors $v_i = (0, \dots, 0, 1^{i)}, 0, \dots, 0)$, there exists the directional derivatives at almost every $x \in U$,

$$\nabla f(x) :\, = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) = \left(D_{e_1} f(x), \dots D_{e_n} f(x) \right).$$

We next prove the following facts:

(i) For every $v \in S^{n-1}$ there exists a null set $N_v \subset \mathbb{R}^n$ so that $\nabla f(x) \cdot v = D_v f(x)$ for every $x \in \mathbb{R}^n \setminus N_v$.

(ii) f is differentiable at almost every $x \in U$.

(i): Let $v \in S^{n-1}$ be fixed. Let also $\varphi \in C_0^\infty(U)$. Then take $t_0 > 0$ be sufficiently small so that whenever $\varphi(x) \neq 0$ then $x + tv \in U$ for all $0 \leq |t| < 2t_0$. In this way $f(x+tv)\varphi(x)$ is well defined for every $x \in U$ and every $0 \leq |t| < 2t_0$. We can write

$$\int_U \frac{f(x+tv) - f(x)}{t} \varphi(x)\, dx = \int_U \frac{f(x+tv)}{t} \varphi(x)\, dx + \int_U -\frac{f(x)}{t} \varphi(x)\, dx.$$

Making a change of variables $x + tv \to y$ in the first term of the sum we get

$$\begin{aligned} \int_U \frac{f(x+tv) - f(x)}{t} \varphi(x)\, dx &= \int_U \frac{f(y)}{t} \varphi(y - tv)\, dy + \int_U -\frac{f(x)}{t} \varphi(x)\, dx \\ &= -\int_U f(x) \frac{\varphi(x) - \varphi(x - tv)}{t}\, dx \end{aligned} \tag{2.1}$$

In order to take limits $t \to 0$, we will apply the Dominated Convergence Theorem, which requires to check the integrability of the above functions. Indeed, this is true because for every $0 < t < t_0$

- $\left| \frac{f(x+tv) - f(x)}{t} \varphi(x) \right| \leq L|\varphi(x)|$, where the last term is integrable because it is continuous with compact support in U.

- $\left| f(x) \frac{\varphi(x - tv) - \varphi(x)}{t} \right| \leq C|f(x)|$, where the last term is integrable on $(\overline{\operatorname{supp}(\varphi)} + t_0 \overline{B(0,1)}) \subset U$ because it is a Lipschitz function (in particular continuous) on a compact set[3].

As previously announced, we apply now the Dominated Convergence Theorem in both sides of (2.1) and we get

$$\lim_{t \to 0} \int_U \frac{f(x+tv) - f(x)}{t} \varphi(x)\, dx = \int_U D_v f(x) \varphi(x)\, dx = -\int_U f(x) D_v \varphi(x)\, dx.$$

[3] Given any two sets $A, B \subset \mathbb{R}^n$ we write its sum as $A + B = \{a + b :\, a \in A,\, b \in B\}$.

And this holds for every $|v| = 1$. By choosing $v = e_i$ for every $i = 1, \dots, n$ we have

$$\int_U \frac{\partial f}{\partial x_i}(x)\varphi(x)\,dx = -\int_U f(x)\frac{\partial \varphi}{\partial x_i}(x)\,dx. \tag{2.2}$$

Therefore for every $v \in S^{n-1}$, and every $\varphi \in C_0^\infty(U)$,

$$\int_U D_v f(x)\varphi(x)\,dx = -\int_U f(x) D_v\varphi(x)\,dx = -\int_U f(x)\nabla\varphi(x)\cdot v\,dx = -\sum_{i=1}^n \int_U f(x)\frac{\partial\varphi}{\partial x_i}(x)v_i\,dx$$
$$= \sum_{i=1}^n \int_U \frac{\partial f}{\partial x_i}(x)\varphi(x)v_i\,dx = \int_U (\nabla f(x)\cdot v)\varphi(x)\,dx$$

Applying Lemma 2.6 we conclude that $D_v f(x) = \nabla f(x)\cdot v$ for almost every $x \in U$ and we are done.

(ii): Let us finally check that f is differentiable at almost every $x \in U$ to conclude the proof.

Take $\{v_k\}_{k=1}^\infty \subset S^{n-1}$ a dense subset and for every $k \in \mathbb{N}$ let

$$A_k = \{x \in U : \nabla f(x) \text{ exists, } D_{v_k} f(x) \text{ exists and } D_{v_k} f(x) = \nabla f(x)\cdot v_k\}$$

It is clear that by the subadditivity of the Lebesgue measure for each $k \in \mathbb{N}$ we have $\mathcal{L}^n(U \setminus A_k) = 0$ and then $\mathcal{L}^n\left(U \setminus \bigcap_{k=1}^\infty A_k\right) = 0$. Denote $A = \bigcap_{k=1}^\infty A_k$. We now aim to show that f is differentiable at every point $x \in A$. Observe that if $x \in A$ then

$$D_{v_k} f(x) = \nabla f(x)\cdot v_k, \quad \text{for every } k \in \mathbb{N}. \tag{2.3}$$

Warning: Even if for a point $x \in U$ there exists $D_v f(x)$ for all directions $v \in S^{n-1}$ with $D_v f(x) = \nabla f(x)\cdot v$, we do not necessarily have the differentiability of f at x. Indeed, there are examples of Gateaux differentiable functions at a point x (i.e. there exists all directional derivatives at x) whose Gateaux derivative is linear (hence continuous), but nonetheless f is not (Fréchet) differentiable at x. Such examples cannot be Lipschitz in any neighborhood of x.

Continuing with the proof, for a given $x \in A$, $v \in S^{n-1}$ and $t > 0$ define

$$Q(x, v, t) = \frac{f(x+tv) - f(x)}{t} - \nabla f(x)\cdot v.$$

We assert that it is enough to prove that for every $x \in A$ and for every $\varepsilon > 0$ there exists $\delta > 0$ so that $|Q(x, v, t)| < \varepsilon$ for all $0 < t < \delta$ and all $v \in S^{n-1}$. Indeed, if this is the case, we readily obtain the differentiability of f at every $x \in A$: For a given $x \in A$ and $\varepsilon > 0$ take $\delta > 0$ as mentioned above. Then for every $0 < |y - x| < \delta$ we can write $y = x + t_y v_y$ for some $0 < t_y < \delta$ and some $v_y \in S^{n-1}$, and we obtain that

$$\left|\frac{f(y) - f(x) - \nabla f(x)\cdot(y-x)}{|y-x|}\right| = \left|\frac{f(x+tv) - f(x) - \nabla f(x)\cdot(t_y v_y)}{|t_y v_y|}\right|$$
$$= \left|\frac{f(x+tv) - f(x)}{t_y} - \nabla f(x)\cdot v_y\right| = |Q(x, v_y, t_y)| < \varepsilon,$$

which yields to

$$\lim_{y\to x} \frac{f(y) - f(x) - \nabla f(x)\cdot(y-x)}{|y-x|} = 0.$$

So, to finish the proof let us fix $x \in A$ and $\varepsilon > 0$, and let us try to find $\delta > 0$ so that $|Q(x, v, t)| < \varepsilon$ for all $0 < t < \delta$ and all $v \in S^{n-1}$. We need to quote some properties:

(a) For every $x \in A$ and $i = 1, \dots, n$ using that f is L-Lipschitz we have

$$\left|\frac{\partial f}{\partial x_i}(x)\right| = \left|\lim_{t\to 0}\frac{f(x+te_i)-f(x)}{t}\right| \leq L.$$

(b) For every $v, w \in S^{n-1}$, $t > 0$ and $x \in A$, by using (a) and again the L-Lipschitzianity of f,

$$|Q(x,v,t) - Q(x,w,t)| \leq (\sqrt{n}+1)L|v-w|.$$

(c) Given $\varepsilon > 0$, since $\{v_k\}_{k\geq 1} \subset S^{n-1}$ is dense and S^{n-1} is compact, we can choose $p \in \mathbb{N}$ sufficiently large so that for any $v \in S^{n-1}$ there exists $j \in \{1, \dots, p\}$ so that

$$|v - v_j| \leq \frac{\varepsilon}{2(\sqrt{n}+1)L}.$$

(d) For any $x \in A$ we have that for all $k \in \mathbb{N}$,

$$\lim_{t\to 0^+} Q(x, v_k, t) = D_{v_k}f(x) - \nabla f(x)\cdot v_k = 0.$$

Let us finish the argument. Take $x \in A$ and $\varepsilon > 0$. Next, choose $p \in \mathbb{N}$ as in (c) and observe that by (d) we have

$$\lim_{t\to 0^+} Q(x, v_k, t) = 0 \quad \text{for all } k = 1, \dots, p.$$

By definition of limit there exists $\delta > 0$ so that if $0 < t < \delta$ we have

$$|Q(x, v_k, t)| < \frac{\varepsilon}{2} \quad \text{for all } k = 1, \dots, p. \tag{2.4}$$

Hence, given any $0 < t < \delta$ and $v \in S^{n-1}$, by choosing $j \in \{1, \dots, p\}$ as in (c), and by using (b) and (2.4)

$$|Q(x,v,t)| \leq |Q(x,v_j,t)| + |Q(x,v_j,t) - Q(x,v,t)| \leq \frac{\varepsilon}{2} + (\sqrt{n}+1)L|v-v_j| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

2.2 Stepanov theorem

The main goal of this section is to prove the following generalization of Rademacher's theorem due to Stepanov.

Theorem 2.7 (Stepanov, 1923)**.** *Let $U \subset \mathbb{R}^n$ be open. Then a given measurable function $f : U \to \mathbb{R}$ is differentiable $\mathcal{L}^n$-almost everywhere if and only if for almost every $x \in U$,*

$$\limsup_{y\to x}\frac{|f(y)-f(x)|}{|y-x|} < +\infty.$$

It is clear that all L-Lipschtiz functions satisfy that

$$\limsup_{y\to x}\frac{|f(y)-f(x)|}{|y-x|} \leq L < +\infty$$

for all $x \in U$. So Stepanov's theorem is indeed a generalization of Rademacher's theorem.

Before proceeding with the proof, we recall some preliminary notions and fundamental results. The first is the celebrated Lebesgue differentiation theorem.

Theorem 2.8 (Lebesgue differentiation)**.** *If $f \in L^1_{loc}(\mathbb{R}^n)$ then*

$$\lim_{r\to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int B(x,r) f(y)\, dy = f(x) \quad \textit{for almost every } x \in \mathbb{R}^n.$$

And if $f \in L^p_{loc}(\mathbb{R}^n)$ for some $1 \le p \le \infty$ then

$$\lim_{r\to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|^p\, dy = 0 \quad \textit{for almost every } x \in \mathbb{R}^n. \tag{2.5}$$

We say that $x \in \mathbb{R}^n$ is a Lebesgue L^p point of f whenever (2.5) holds.

Definition 2.9. Let $E \subset \mathbb{R}^n$ be a measurable set and $x \in \mathbb{R}^n$. We say that E has density 1 at x (or that x is of density 1 on E) if

$$\lim_{r\to 0} \frac{\mathcal{L}^n(B(x,r) \cap E)}{\mathcal{L}^n(B(x,r))} = 1.$$

And we say that E has density 0 at x (or that x is of density 0 on E) if

$$\lim_{r\to 0} \frac{\mathcal{L}^n(B(x,r) \cap E)}{\mathcal{L}^n(B(x,r))} = 0.$$

Example 2.10. The set $E = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le 1, -x^2 \le y \le x^2\}$ satisfies that $0 = (0,0)$ is of density 0 on E. Indeed

$$\lim_{r\to 0} \frac{\mathcal{L}^n(B(0,r) \cap E)}{\pi r^2} \le \lim_{r\to 0} \frac{r^3/3}{\pi r^2} = 0.$$

Lemma 2.11. *Let $E \subset \mathbb{R}^n$ be a measurable set. Then*

1. *Almost every point $x \in E$ is of density 1 on E.*
2. *Almost every point $x \in \mathbb{R}^n \setminus E$ is of density 0 on E.*

Proof. Let $f = \chi_E$ be the characteristic function of the set E, that is

$$\chi_E(x) = \begin{cases} 1, & \text{if } x \in E \\ 0, & \text{if } x \notin E \end{cases}.$$

We have that χ_E is a measurable function and $\chi_E \in L^1_{loc}(\mathbb{R}^n)$. By using the Lebesgue differentiation theorem (Theorem 2.8) we get that for almost every $x \in \mathbb{R}^n$,

$$\lim_{r\to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} \chi_E(y)\, dy = \chi_E(x).$$

In particular

1. For almost every $x \in E$, since $\chi_E(x) = 1$,

$$\lim_{r\to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} \chi_E(y)\, dy = \lim_{r\to 0} \frac{\mathcal{L}^n(B(x,r) \cap E)}{\mathcal{L}^n(B(x,r))} = 1.$$

2. For almost every $x \in \mathbb{R}^n \setminus E$, since $\chi_E(x) = 0$,

$$\lim_{r\to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} \chi_E(y)\, dy = \lim_{r\to 0} \frac{\mathcal{L}^n(B(x,r) \cap E)}{\mathcal{L}^n(B(x,r))} = 0.$$

□

Proof of Stepanov theorem (Theorem 2.7). We need to prove two implications:

$\Rightarrow$): This one is easy. If f is differentiable at almost every point $x \in U$, then for almost every $x \in U$ there exists $\nabla f(x) \in \mathbb{R}^n$ so that

$$\lim_{y\to x} \frac{|f(y) - f(x) - \nabla f(x)\cdot(y-x)|}{|y-x|} = 0.$$

Therefore

$$\begin{aligned}\limsup_{y\to x} \frac{|f(y)-f(x)|}{|y-x|} &= \limsup_{y\to x} \frac{|f(y) - f(x) - \nabla f(x)\cdot(y-x) + \nabla f(x)\cdot(y-x)|}{|y-x|}\\ &\le \limsup_{y\to x} \frac{|f(y) - f(x) - \nabla f(x)\cdot(y-x)|}{|y-x|} + \limsup_{y\to x}\frac{|\nabla f(x)\cdot(y-x)|}{|y-x|}\\ &= 0 + \limsup_{y\to x}\frac{|\nabla f(x)\cdot(y-x)|}{|y-x|}\\ &\le \limsup_{y\to x} |\nabla f(x)| = |\nabla f(x)| < +\infty.\end{aligned}$$

$\Leftarrow$): Let

$$A = \left\{ x \in U : \limsup_{y\to x} \frac{|f(y)-f(x)|}{|y-x|} < +\infty \right\}.$$

We have by assumptions that $\mathcal{L}^n(U \setminus A) = 0$ so our goal is to prove that f is differentiable at almost every $x \in A$. We split A as follows. For every $k \in \mathbb{N}$ define

$$E_k = \left\{ x \in A : \, |f(x)| \le k, \frac{|f(x)-f(y)|}{|x-y|} \le k \text{ if } |y-x| < \frac{1}{k} \right\}$$

Let us check that $A = \bigcup_{k\in\mathbb{N}} E_k$. Take $x \in A$. Then

$$\limsup_{y\to x} \frac{|f(y)-f(x)|}{|y-x|} = k_1 < \infty.$$

Then there exists $k_2 > 0$ so that whenever $0 < |y-x| < \frac{1}{k_2}$ then $\frac{|f(y)-f(x)|}{|y-x|} \le k_1 + 1$. Lastly, let $k_3 = |f(x)|$. By letting k be the closest upper integer to $\max\{k_1+1, k_2, k_3\}$ we have $x \in E_k$.

It is enough to prove that f is differentiable at almost every point $x \in E_k$ for every $k \in \mathbb{N}$. Let us then fix $k \in \mathbb{N}$.

Observe that $f|_{E_k}$ is Lipschitz because:

- If $x, y \in E_k$ with $|x-y| < \frac{1}{k}$ then $|f(y) - f(x)| < k|x-y|$.
- If $x, y \in E_k$ with $|x-y| \ge \frac{1}{k}$ then $|f(x) - f(y)| \le 2k \le 2k^2|x-y|$.

We now apply the McShane extension theorem (Theorem 1.5) to the Lipschitz function $f : E_k \to \mathbb{R}$ and find $F_k : \mathbb{R}^n \to \mathbb{R}$ Lipschitz with $F|_{E_k} = f$. By Rademacher theorem (Theorem 2.1), we have that F_k is differentiable $\mathcal{L}^n$ almost everywhere on $\mathbb{R}^n$.

For the rest of the proof we want to show that f is differentiable at every differentiable point of F_k that belongs to E_k and that moreover is a point of density 1 on E_k. Note that, since the set E_k is measurable (we leave this fact as an exercise for the reader), $\mathcal{L}^n$-almost every $x \in E_k$ is of density 1 in E_k. Therefore, let us call

$$A_k = \{x \in E_k : \, x \text{ is of density 1 on } E_k\} \cap \{x \in E_k : \, F_k \text{ is differentiable at } x\}.$$

and let us verify that f is differentiable at every $x \in A_k$ with $\nabla f(x) = \nabla F(x)$. With this, the proof will be complete.

Take $x \in A_k$. On the one hand

$$\lim_{\substack{y\to x\\ y\in E_k}} \frac{|f(y)-f(x)-\nabla F_k(x)\cdot(y-x)|}{|y-x|} = \lim_{\substack{y\to x\\ y\in E_k}} \frac{|F_k(y)-F_k(x)-\nabla F_k(x)\cdot(y-x)|}{|y-x|} = 0$$

by the differentiability of F_k at x. Consequently, one only needs to prove that

$$\lim_{\substack{y\to x\\ y\notin E_k}} \frac{|f(y)-f(x)-\nabla F_k(x)\cdot(y-x)|}{|y-x|} = 0.$$

To deal with that limit it is crucial to use the next key fact, that allow us to change points $y \notin E_k$ for points $\widetilde{y} \in E_k$ that converge to x faster than y converges to x. This will be possible to do because x has density 1 on E_k and, intuitively, this means that "most" of the surrounding of x are points of E_k.

Key fact: For every $y \notin E_k$ there exists $\widetilde{y} = \widetilde{y}(x,y) \in E_k$ so that $\lim_{y\to x} \frac{|y-\widetilde{y}|}{|y-x|} = 0$.

Proof of the **Key Fact**: For every $r > 0$ define

$$h(r) = \left(1 - \frac{\mathcal{L}^n(B(x,4r)\cap E_k)}{\mathcal{L}^n(B(x,4r))}\right)^{1/n} 4r$$

- If there exists $r_0 > 0$ so that $h(r_0) = 0$, then $\mathcal{L}^n(B(x,4r_0)\cap E_k) = \mathcal{L}^n(B(x,4r_0))$. Then, if $y \in B(x,4r_0)$ we take $\widetilde{y} \in B(y,|y-x|^2)\cap E_k \neq \emptyset$. And if $y \notin B(x,4r_0)$ we take any $\widetilde{y} \in E_k$. In this way,
$$\lim_{y\to x} \frac{|y-\widetilde{y}|}{|y-x|} \le \lim_{y\to x} \frac{|y-x|^2}{|y-x|} = 0.$$
- If $h(r) > 0$ for every $r > 0$ then one can check that by the density 1 of x on E_k
$$\lim_{r\to 0} \frac{h(r)}{r} = 0, \tag{2.6}$$
and moreover
$$\mathcal{L}^n(B(0,h(r))) = \mathcal{L}^n(B(x,4r)) - \mathcal{L}^n(B(x,4r)\cap E_k) = \mathcal{L}^n(B(x,4r)\setminus E_k). \tag{2.7}$$
Hence, for every $y \notin E_k$, if we call $r = |y-x|$, and assuming that $r > 0$ is sufficiently small for $h(r) < r$ to hold (this is possible due to (2.6)) we have
$$B(y,2h(r))\cap E_k \neq \emptyset.$$
Otherwise, since $h(r) < r$, then $B(y,2h(r)) \subset B(x,4r)\setminus E_k$ and therefore
$$\mathcal{L}^n(B(x,4r)\setminus E_k) \ge \mathcal{L}^n(B(y,2h(r))) > \mathcal{L}^n(B(0,h(r)).$$
which contradicts (2.7). This means that for every given $y \notin E_k$, whenever $r > 0$ is small enough so that $h(r) < r$ we have $B(y,2h(r))\cap E_k \neq \emptyset$ and we may take $\widetilde{y} \in B(y,2h(r))\cap E_k$. If $r > 0$ is big we take any $\widetilde{y} \in E_k$. We conclude the proof of the **Key fact** by writing
$$\lim_{y\to x} \frac{|y-\widetilde{y}|}{|y-x|} \le \lim_{y\to x} \frac{2h(|y-x|)}{|y-x|} = 0.$$

We can now finish the proof of Stepanov's theorem. Take $x \in A_k$, $y \notin E_k$ and choose $\widetilde{y} \in E_k$ as in the **Key fact**. We have

$$\begin{aligned}\frac{|f(y)-f(x)-\nabla F_k(x)\cdot(y-x)|}{|y-x|} &= \frac{|f(y)-F_k(x)-\nabla F_k(x)\cdot(y-x)|}{|y-x|}\\ &\le \frac{|f(y)-F_k(x)-\nabla F_k(x)\cdot(\widetilde{y}-x)|}{|y-x|} + \frac{|\nabla F_k(x)\cdot(\widetilde{y}-y)|}{|y-x|}\\ &\le \frac{|f(y)-f(\widetilde{y})|}{|y-x|} + \frac{|f(\widetilde{y})-F_k(x)-\nabla F_k(x)\cdot(\widetilde{y}-x)|}{|y-x|} + |\nabla F_k(x)|\frac{|\widetilde{y}-y|}{|y-x|}.\end{aligned}$$

We analyse each of these three summands separately. For the first one, by the definition of E_k we know that for every $\widetilde{y} \in E_k$ with $|y-\widetilde{y}| < 1/k$ we have $|f(y) - f(\widetilde{y})| \le k|y-\widetilde{y}|$, so the **Key fact** gives

$$\lim_{y\to x} \frac{|f(y) - f(\widetilde{y})|}{|y-x|} \le k \lim_{y\to x} \frac{|y-\widetilde{y}|}{|y-x|} = 0.$$

For the second term note that $f(\widetilde{y}) = F_k(\widetilde{y})$ so we can write

$$\frac{|f(\widetilde{y}) - F_k(x) - \nabla F_k(x)\cdot(\widetilde{y}-x)|}{|y-x|} = \frac{|F_k(\widetilde{y}) - F_k(x) - \nabla F_k(x)\cdot(\widetilde{y}-x)|}{|\widetilde{y}-x|}\frac{|\widetilde{y}-x|}{|y-x|}.$$

Since $y \to x$ implies $\widetilde{y} \to x$, by the differentiability of F_k at x and since

$$\limsup_{y\to x} \frac{|\widetilde{y}-x|}{|y-x|} \le \limsup_{y\to x} \frac{|\widetilde{y}-y|}{|y-x|} + 1 = 1$$

(again using the **Key fact**), yields to

$$\lim_{y\to x} \frac{|f(\widetilde{y}) - F_k(x) - \nabla F_k(x)\cdot(\widetilde{y}-x)|}{|y-x|} = 0.$$

Lastly, for the third term, by a direct application of the **Key fact** we get

$$\lim_{y\to x} |\nabla F_k(x)| \frac{|\widetilde{y}-y|}{|y-x|} = 0.$$

The proof of Stepanov's theorem is finished.

□

Last comment: For a point x of density 1 in a measurable set $E \subset \mathbb{R}^n$ we know that we can define for every $y \notin E$ some $\widetilde{y}$ so that

$$\lim_{y\to x} \frac{|y-\widetilde{y}|}{|y-x|} = 0. \tag{2.8}$$

It is interesting to note that for a given $\alpha > 1$, and some fixed $x \in \mathbb{R}^n$, one can build examples of measurable sets $E \subset \mathbb{R}^n$ so that x has density 1 on E but with the property that for some sequence $y_n \to x$ we have that

$$\lim_{n\to\infty} \frac{\mathrm{dist}(y_n, E)}{|y_n - x|^\alpha} \neq 0.$$

In other words, the convergence in (2.8) can not be improved to have a bigger exponent $\alpha > 1$ in the denominator.

For an specific example think simply about the real line, where we remove small intervals

$$I_n = (y_n - y_n^\alpha, y_n + y_n^\alpha)$$

with $(y_n)_{n\ge1} \subset (0, +\infty)$ converging to zero and with $1 < \alpha' < \alpha$. We do this in such a way that 0 will have density one on $E := \mathbb{R} \setminus \bigcup_{n\ge1} I_n$ but so that

$$\lim_{n\to\infty} \frac{\mathrm{dist}(y_n, E)}{|y_n - 0|^\alpha} = \lim_{n\to\infty} \frac{y_n^{\alpha'}}{y_n^\alpha} = +\infty$$

Exercises

1. Prove the following statements:
   (a) If $f : [a, b] \to \mathbb{R}$ is absolutely continuous then f is continuous.
   (b) $f(x) = \sqrt{x}$, $x \in [0, 1]$ is absolutely continuous but not Lipschitz.
   (c) Given an example of a Hölder continuous function of exponent α for all $\alpha \in (0, 1)$ but not Lipschitz.

2. Prove the existence of functions $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(0, 0) = 0$, so that are Lispchitz on the unit ball, Gateaux differentiable (in every direction) at $(0, 0)$, but are not diffeentiable at $(0, 0)$. Recall that f is said to be Gateaux differentiable at a point $x_0 \in \mathbb{R}^2$ if for every unit vector $v \in S^{n-1}$ all directional derivatives exist
$$D_v f(x_0) = \lim_{t\to 0} \frac{f(x + tv) - f(x)}{t}.$$

3. Do there exist Lipschitz functions $f : \mathbb{R} \to \mathbb{R}$ which are not differentiable at the rationals $\mathbb{Q} \subset \mathbb{R}$? If the answer is yes, provide an example.

4. Given two open sets $U, V \subset \mathbb{R}^n$ with $U \subset \overline{U} \subset V$, define explicitly a function $h : \mathbb{R}^n \to [0, 1]$ of class C^∞ with $h = 0$ on $\mathbb{R}^n \setminus V$ and $h = 1$ nn $\overline{U}$.

5. (**Absolutely continuity of the integral**) Let $F \subset \mathbb{R}^n$ be a (Lebesgue) measurable set. Let also $f : F \to [0, \infty)$ a (Lebesgue) measurable set so that $f \in L^1(F)$. Prove that for all $\varepsilon > 0$ there exists $\delta > 0$ so that if $A \subset F$ is measurable with $\mathcal{L}^n(A) < \delta$ then $\int_A f(x)\, dx < \varepsilon$.

6. Prove that there exist functions $f : \mathbb{R} \to \mathbb{R}$ that are differentiable almost everywhere for which there do not exist weak derivatives. That is, there does not exist any integrable function $g \in L^1_{loc}(\mathbb{R})$ so that
$$\int_{\mathbb{R}} g(x)\varphi(x)\, dx = \int_{\mathbb{R}} f(x)\varphi'(x)\, dx \quad \forall \varphi \in C_0^\infty(\mathbb{R}).$$
Do there exist such kind of examples for continuous functions? Do there exist such kind of examples for Lipschitz functions?

7. Let $f : \mathbb{R}^n \to \mathbb{R}$. Prove that
   - If f is measurable and $k \in \mathbb{N}$, then the set
   $$E_k = \{x \in \mathbb{R}^n : |f(x)| \le k,\ |f(y) - f(x)| \le k|y - x| \,\forall |y - x| < 1/k\}$$
   is measurable.
   - If f is differentiable almost everywhere then f is measurable. Moreover, prove that the differential ∇f (extended by 0 where it does not exist) is a measurable function too.

8. Let $f : \mathbb{R} \to \mathbb{R}$ be a function so that for all $x \in \mathbb{R}$,
$$\limsup_{y\to x} \frac{|f(y) - f(x)|}{|y - x|} < +\infty$$
Prove that f is a continuous function (in particular measurable) and that, if $\mathcal{L}(N) = 0$, then $\mathcal{L}(f(N)) = 0$.

9. Let Ω be a domain of $\mathbb{R}^n$ so that there exists $c > 0$ so that for all $r \in (0, 1]$ and all $x \in \Omega$,
$$\mathcal{L}^n(\Omega \cap B(x, r)) \ge c\mathcal{L}^n(B(x, r)).$$
(Domains with this property are called *Alhfors n-regular* or domains with the measure density condition). Prove that $\mathcal{L}^n(\partial\Omega) = 0$. (Hint: Use the Lebesgue differentiation theorem).

Chapter 3

Rademacher theorem for Sobolev functions

Let $\Omega \subset \mathbb{R}^n$ be an open set and let $\mathcal{L}^n$ denote the Lebesgue measure. The goal of this chapter is to prove the next result.

Theorem 3.1. *Let $f \in W^{1,p}(\mathbb{R}^n)$ for some $n < p < \infty$. Then there exists a representative of f that is Hölder continuous with exponent $\alpha(p) = 1 - \frac{n}{p}$ which moreover is differentiable at $\mathcal{L}^n$ almost every point $x \in \mathbb{R}^n$.*

Before going into the details of the proof, we need to make a quick introduction to Sobolev spaces and some of their properties. Finally, we aim to state the important Morrey's inequality, which will be the essential tool for the proof of Theorem 3.1.

3.1 Introduction to Sobolev spaces

Sobolev spaces are subspaces of L^p spaces where there exists weak derivatives up to some order belonging as well to L^p. These spaces arise naturally in the theory of Partial Differential Equations. Namely, it is the correct setting to consider solutions of boundary value problems of differential equations, because strong tools from Functional Analysis may be used. Another idea behind Sobolev spaces is that they provide a space where there is an equilibrium between smooth and rough functions.

Recall that for us $C_0^\infty(\Omega)$ denotes the space of test functions given by

$$\{\varphi \in C^\infty(\Omega) : \overline{\text{supp}(\varphi)} := \overline{\{x \in \Omega : \phi(x) \neq 0\}} \subset \Omega \text{ is compact}\}.$$

Weak derivatives arise naturally after the following observation: For every $f \in C^1(\Omega)$ if we take $\varphi \in C_c^\infty(\Omega)$, by integration by parts we get

$$\int_\Omega f \frac{\partial \varphi}{\partial x_i}\, dx = -\int_\Omega \frac{\partial f}{\partial x_i} \varphi\, dx \tag{3.1}$$

for each $i = 1, \ldots, n$ (the integral over $\partial\Omega$ does not appear since φ has compact support on Ω). More generally if $k \in \mathbb{N}$, $f \in C^k(\mathbb{R}^n)$, $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multiindex of order $|\alpha| = \alpha_1 + \cdots + \alpha_n = k$, then if $D^\alpha := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$,

$$\int_\Omega f D^\alpha \varphi\, dx = (-1)^{|\alpha|} \int_\Omega D^\alpha f \varphi\, dx. \tag{3.2}$$

We obtain this by applying (3.1) $k = |\alpha|$ times. Regarding these expressions, in (3.2), we note the that left-hand side term is well defined whenever $f \in L^1_{loc}(\Omega)$. The expression at the right-hand side yields to the definition of weak derivative of functions in $L^1_{loc}(\Omega)$.

Definition 3.2. Let $f, g \in L^1_{loc}(\Omega)$ and α a multiindex. We say that g is the $\alpha^{th}-$weak partial derivative of f, written $D^\alpha f = g$, if

$$\int_\Omega f D^\alpha \varphi \, dx = (-1)^{|\alpha|} \int_\Omega g\varphi \, dx, \quad \forall \varphi \in C_c^\infty(\Omega).$$

If the $\alpha^{th}-$weak partial derivative of f exists then, by Lemma 2.6, it is unique up to a measure zero set. And, in the case that $f \in C^1(\Omega)$, the derivative in the usual sense coincides with the derivative up to a measure zero set.

Definition 3.3. Let $1 \leq p \leq \infty$ and $k \in \mathbb{N}$. We define the Sobolev space $W^{k,p}(\Omega) = W^{k,p}(\Omega;\mathbb{R})$ to be the set if functions $f : \Omega \to \mathbb{R}$ in $L^1_{loc}(\Omega)$ such that for every multiindex α with $|\alpha| \leq k$, $D^\alpha f$ exists in the weak sense and belongs to $L^p(\Omega)$. The space $W^{k,p}_{loc}(\Omega)$ consists of those functions that belong to $W^{k,p}(V)$ for every $V \subset \Omega$ open with $\overline{V} \subset \Omega$ compact.

Notation: We write $W^{0,p}(\Omega) = L^p(\Omega)$. And for the special case $p = 2$ we usually write $H^k(\Omega) = W^{k,2}(\Omega), \quad k \in \mathbb{N} \cup \{0\}$. The letter H comes from Hilbert, because these spaces, with an appropriate norm, are Hilbert.

As in L^p, we identify functions in $W^{k,p}(\Omega)$ that are equal almost everywhere. We endow these spaces with the norms[1]

$$\|f\|_{W^{k,p}(\Omega)} := \left(\sum_{0\leq|\alpha|\leq k} \int_\Omega |D^\alpha f|^p \, dx \right)^{1/p}, \quad 1 \leq p < \infty$$
$$\|f\|_{W^{k,\infty}(\Omega)} := \sum_{0\leq|\alpha|\leq k} ess \sup_\Omega |D^\alpha f|.$$

where the essential supremum is defined as $ess \sup_\Omega f := \inf \{\lambda \in \mathbb{R} : \mathcal{L}^n(\{x : f(x) > \lambda\}) = 0\}$.

We highlight the following important result.

Theorem 3.4. *For each $k \geq 0$, $1 \leq p \leq \infty$, the Sobolev space $W^{k,p}(\Omega) = \left(W^{k,p}(\Omega), \|\cdot\|_{W^{k,p}(\Omega)}\right)$ is Banach. Moreover, $W^{k,p}(\Omega)$ is reflexive for $1 < p < \infty$ and separable for $1 \leq p < \infty$. Additionally, $W^{k,2}(\Omega) = H^k(\Omega)$ is a Hilbert space.*

Proof. One may find a proof in classical books like [11]. □

We also note that the Sobolev spaces $W^{1,p}(\mathbb{R}^n)$, $1 < p < \infty$, are isomorphic to $L^p(\mathbb{R}^n)$ (see [21, Theorem 11]).

Now we move to the question whether these functions admit approximations by smooth ones, and if so, in which sense. We need to talk about convolution and mollifiers.

Definition 3.5. Let $\delta : \mathbb{R}^n \to \mathbb{R}$ be the next C^∞ function

$$\delta(x) := \begin{cases} Ce^{\frac{1}{|x|^2-1}} & \text{si } |x| < 1 \\ 0 & \text{si } |x| \geq 1 \end{cases}$$

where $C >$ is chosen so that $\int_{\mathbb{R}^n} \delta(x) \, dx = 1$. We define now for every $\varepsilon > 0$ the mollifier

$$\delta_\varepsilon(x) = \frac{1}{\varepsilon^n} \delta\left(\frac{x}{\varepsilon}\right).$$

Note that δ_ε is of class C^∞ satisfying $\displaystyle\int_{\mathbb{R}^n} \delta_\varepsilon(x) = 1$ and $\text{supp}(\delta_\varepsilon) = B(0, \varepsilon)$.

[1]The given norm for $W^{k,p}(\Omega)$ with $1 \leq p < \infty$ is equivalent to $\|f\|_{W^{k,p}(\Omega)} := \sum_{0\leq|\alpha|\leq k} \|D^\alpha f\|_{L^p(\Omega)}$.

Definition 3.6. If $f \in L^1_{loc}(\Omega)$ we define $f^\varepsilon := \delta_\varepsilon * f : \Omega_\varepsilon : \ = \{x \in \Omega : \ dist\,(x, \partial\Omega) > \varepsilon\}$ as

$$f^\varepsilon(x) = \int_\Omega \delta_\varepsilon(x-y) f(y)\, dy = \int_{B(0,\varepsilon)} \delta_\varepsilon(y) f(x-y)\, dy = \int_{B(0,1)} \delta(y) f(x - \varepsilon z)\, dz.$$

Mollifiers are very useful to prove the next result, about approximating Sobolev functions by smooth ones in different ways.

Theorem 3.7. *Let $f \in L^1_{loc}(\Omega)$. Then the following is satisfied:*

1. *For each $\varepsilon > 0$, $f^\varepsilon \in C^\infty(\Omega_\varepsilon)$.*
2. *If $f \in C(\Omega)$ then $f^\varepsilon \to f$ uniformly on compact subsets of Ω.*
3. *If $f \in L^p_{loc}(\Omega)$ for some $1 \leq p < \infty$ then $f^\varepsilon \to f$ in $L^p_{loc}(\Omega)$, and moreover $f^\varepsilon(x) \to f(x)$ for every Lebesgue point x of f.*
4. *If $f \in W^{k,p}_{loc}(\Omega)$ for some $1 \leq p < \infty$ and $k \geq 1$ then $f^\varepsilon \to f$ in $W^{k,p}_{loc}(\Omega)$.*
5. *If $f \in W^{k,p}(\mathbb{R}^n)$ then $f^\varepsilon \to f$ in $W^{k,p}(\mathbb{R}^n)$. In this case $f^\varepsilon \in C^\infty_0(\mathbb{R}^n)$.*

Proof. (1) Fix $x \in \Omega_\varepsilon$, $1 \leq i \leq n$ and h sufficiently small so that $x + he_i \in \Omega_\varepsilon$ (e_i denotes i-th vector of the canonical basis of $\mathbb{R}^n$). Then

$$\begin{aligned} \frac{f^\varepsilon(x+he_i) - f^\varepsilon}{h} &= \int_\Omega \left[\frac{\delta_\varepsilon(x+he_i-y) - \delta_\varepsilon(x-y)}{h}\right] f(y)\, dy = \\ &= \frac{1}{\varepsilon^n}\int_\Omega \left[\frac{\delta\left(\frac{x+he_i-y}{\varepsilon}\right) - \delta\left(\frac{x-y}{\varepsilon}\right)}{h}\right] f(y)\, dy \\ &= \frac{1}{\varepsilon^n}\int_V \left[\frac{\delta\left(\frac{x+he_i-y}{\varepsilon}\right) - \delta\left(\frac{x-y}{\varepsilon}\right)}{h}\right] f(y)\, dy \end{aligned} \tag{3.3}$$

form some open set $V \subset\subset \Omega$ with $\overline{V} \subset \Omega$ compact. Being able to restrict the integral to certain V is due to the fact of how the supports of the functions δ act on the integrals. Now, by the regularity properties of δ,

$$\lim_{h\to 0} \frac{1}{h}\left[\delta\left(\frac{x-y}{\varepsilon} + \frac{he_i}{\varepsilon}\right) - \delta\left(\frac{x-y}{\varepsilon}\right)\right] = \frac{1}{\varepsilon}\frac{\partial \delta}{\partial x_i}\left(\frac{x-y}{\varepsilon}\right) = \varepsilon^n \frac{\partial \delta_\varepsilon}{\partial x_i}(x-y)$$

uniformly on $y \in V$.
Moreover, aiming to use the Dominated Convergence Theorem, we observe that the absolute value of the integrand is bounded

$$\begin{aligned} \frac{1}{h}\left|\delta\left(\frac{x+he_i-y}{\varepsilon}\right) - \delta\left(\frac{x-y}{\varepsilon}\right)\right| |f(y)| &\leq \frac{1}{h}|D\delta(\xi)| \left|\frac{x+he_i-y}{\varepsilon} - \frac{x-y}{\varepsilon}\right| |f(y)| \leq \\ &\leq \frac{1}{h}\|D\delta\|_{L^\infty}\left|\frac{he_i}{\varepsilon}\right| |f(y)| \leq \frac{1}{\varepsilon}\|D\delta\|_{L^\infty}|f(y)| \in L^1(V) \end{aligned}$$

where ξ is some point in the segment joining $\frac{x-y}{\varepsilon}$ and $\frac{x+he_i-y}{\varepsilon}$. We can use the Dominated Convergence Theorem in (3.3) to conclude that

$$\frac{\partial f^\varepsilon}{\partial x_i}(x) = \lim_{h\to 0}\frac{f^\varepsilon(x+he_i) - f^\varepsilon(x)}{h} = \int_\Omega \frac{\partial \delta_\varepsilon}{\partial x_i}(x-y) f(y)\, dy.$$

A similar argument proves that all partial derivatives of f^ε exist and are continuous everywhere on Ω_ε.

(2) Let $V \subset W \subset \Omega$ be open sets with $\overline{V} \subset W$ and $\overline{W} \subset \Omega$ compact sets. For each $y \in V$,

$$f^\varepsilon(x) = \frac{1}{\varepsilon^n}\int_{B(x,\varepsilon)} \delta\left(\frac{x-y}{\varepsilon}\right) f(y)\,dy = \int_{B(0,1)} \delta(z) f(x-\varepsilon z)\,dz.$$

Since $\displaystyle\int_{B(0,1)} \delta(z)\,dz = 1$, we have

$$|f^\varepsilon(x) - f(x)| \le \int_{B(0,1)} \delta(z)|f(x-\varepsilon z) - f(x)|\,dz.$$

Moreover, f is uniformly continuous on W ($f \in C(\Omega)$) and because $\varepsilon > 0$ is sufficiently small, $f(x - \varepsilon z) \in W$ ($x \in V$, $z \in B(0,1)$) we get that $f^\varepsilon \to f$ uniformly on V.

(3) Take open sets $V \subset W \subset \Omega$ with $\overline{V} \subset W$ and $\overline{W} \subset \Omega$ being compact. We assert that for $\varepsilon > 0$ sufficiently small

$$\|f^\varepsilon\|_{L^p(V)} \le \|f\|_{L^p(W)}. \tag{3.4}$$

Firstly, if $1 < p < \infty$ and $x \in V$, we have

$$|f^\varepsilon(x)| = \left|\int_{B(0,1)} \delta(z) f(x-\varepsilon z)\,dz\right| \le \int_{B(0,1)} \delta(z)|f(x-\varepsilon z)|\,dz =$$
$$= \int_{B(0,1)} \delta(z)^{1-\frac{1}{p}}\delta(z)^{\frac{1}{p}}|f(x-\varepsilon z)|\,dz \le \left(\int_{B(0,1)} \delta(z)\,dz\right)^{1-\frac{1}{p}}\left(\int_{B(0,1)} \delta(z)|f(x-\varepsilon z)|^p\,dz\right)^{1/p}.$$

Next, taking $1 \le p < \infty$, we have

$$\int_V |f^\varepsilon(x)|^p\,dx \le \int_V \left(\int_{B(0,1)} \delta(z)|f(x-\varepsilon z)|^p\,dz\right)dx =$$
$$= \int_{B(0,1)} \delta(z)\left(\int_V |f(x-\varepsilon z)|^p\,dx\right)dz \le \int_{B(0,1)} \delta(z)\left(\int_W |f(y)|^p\,dy\right)dz = \int_W |f(y)|^p\,dy$$

for $\varepsilon > 0$ sufficiently small. We get hence (3.4).
Let now $\delta > 0$. Since $f \in L^p(W)$, we can choose $g \in C(W)$ so that $\|f - g\|_{L^p(W)} < \delta$ (for the density of continuous functions on the spaces L^p, $1 \le p < \infty$ we refer to [23, Page 69]). We can write the following.

$$\begin{aligned}\|f^\varepsilon - f\|_{L^p(V)} &\le \|f^\varepsilon - g^\varepsilon\|_{L^p(V)} + \|g^\varepsilon - g\|_{L^p(V)} + \|g - f\|_{L^p(V)} \le \\ &\le \|f - g\|_{L^p(W)} + \|g^\varepsilon - g\|_{L^p(V)} + \|g - f\|_{L^p(V)} \le \\ &\le 2\delta + \|g^\varepsilon - g\|_{L^p(V)}.\end{aligned}$$

Thanks to the fact that $g \in C(W)$, by (2), we have that $g^\varepsilon \to g$ uniformly on $V \subset\subset W$, and then $g^\varepsilon \to g$ in $L^p(V)$. Therefore, $\limsup_{\varepsilon\to 0}\|f^\varepsilon - f\|_{L^p(V)} \le 2\delta$ and being $\delta > 0$ arbitrary we conclude that $f^\varepsilon \to f$ en $L^p_{loc}(\Omega)$.

Let us show that $\lim_{\varepsilon\to 0} f^\varepsilon(x) = f(x)$ for every Lebesgue point x of f. Hence, let $x \in \mathbb{R}^n$ be a Lebesgue point of f. Recalling that $\int_{B(x,\varepsilon)} \delta_\varepsilon(x-y)\,dy = 1$, we have

$$\begin{aligned}|f^\varepsilon(x) - f(x)| &= \left|\int_{B(x,\varepsilon)} \delta_\varepsilon(x-y)(f(y) - f(x))\,dy\right| \le \int_{B(x,\varepsilon)} \delta_\varepsilon(x-y)|f(y) - x(x)|\,dy \\ &= \frac{1}{\varepsilon^n}\int_{B(x,\varepsilon)} \delta((x-y)/\varepsilon)|f(y) - f(x)|\,dy \le \|\delta\|_{L^\infty(\mathbb{R}^n)}\frac{1}{\varepsilon^n}\int_{B(x,\varepsilon)} |f(y) - f(x)|\,dy \\ &= \|\delta\|_{L^\infty(\mathbb{R}^n)}\omega_n\frac{1}{\mathcal{L}^n(B(x,\varepsilon))}\int_{B(x,\varepsilon)} |f(y) - f(x)|\,dy\end{aligned}$$

Taking limits $\varepsilon \to 0$, and using that x is a Lebesgue point of f, we conclude the argument.

(4) Let us first check that

$$(D^\alpha f)^\varepsilon = \delta_\varepsilon * D^\alpha f \ \text{ in } \ \Omega_\varepsilon \ \text{ if } \ 0 \le |\alpha| \le k. \tag{3.5}$$

Indeed, if we take $x \in \Omega_\varepsilon$, using that $\delta_\varepsilon(x-y) \in C_c^\infty(\Omega)$ for every $y \in \Omega$ and integrating by parts

$$\begin{aligned}(D^\alpha f)^\varepsilon(x) &= D^\alpha \int_\Omega \delta_\varepsilon(x-y) f(y)\,dy = \int_\Omega D_x^\alpha \delta_\varepsilon(x-y) f(y)\,dy = \\ &= (-1)^{|\alpha|} \int_\Omega D_y^\alpha \delta_\varepsilon(x-y) f(y)\,dy = (-1)^{|\alpha|}(-1)^{|\alpha|} \int_\Omega \delta_\varepsilon(x-y) D^\alpha f(y)\,dy = \\ &= (\delta_\varepsilon * D^\alpha f)(x).\end{aligned}$$

Take now $V \subset \Omega$ open with $\overline{V} \subset \Omega$ compact. Since $D^\alpha f \in L^p(V)$ (note that $f \in L^p_{loc}(\Omega)$), using (3),

$$(D^\alpha f)^\varepsilon = \delta_\varepsilon * D^\alpha f \underset{\varepsilon \to 0}{\longrightarrow} D^\alpha f \ \text{ en } \ L^p(V)$$

for all $0 \le |\alpha| \le k$. Subsequently

$$\|f^\varepsilon - f\|^p_{W^{k,p}(V)} = \sum_{0 \le |\alpha| \le k} \|D^\alpha f^\varepsilon - D^\alpha f\|^p_{L^p(V)} \underset{\varepsilon \to 0}{\longrightarrow} 0.$$

(5) This is left as an exercise for the reader. The proof uses similar techniques as those in (4).

□

A stronger result about smooth approximations in $W^{k,p}(\Omega)$, not only in $W^{k,p}_{loc}(\Omega)$, for an arbitrary open set $\Omega \subset \mathbb{R}^n$ is the next one, due to Meyers and Serrin.

Theorem 3.8 (Meyers-Serrin,)**.** *Let $f \in W^{k,p}(\Omega)$ for some $1 \le p < \infty$. Then, there exists a sequence $\{f_i\}_{i=1}^\infty \subset W^{k,p}(\Omega) \cap C^\infty(\Omega)$ so that*

$$f_i \underset{i \to \infty}{\longrightarrow} f \ \text{ en } \ W^{k,p}(\Omega).$$

Proof. We follow [12, Page 125] and [11, Page 251].

We define

$$\begin{cases} \Omega_k = \left\{x \in \Omega : \ dist(x, \partial\Omega) > \frac{1}{k}\right\} \cap B(0,k), \quad k = 1, 2, \dots \\ \Omega_0 = \emptyset \end{cases}$$

and we write $V_k = \Omega_{k+1} \setminus \overline{\Omega}_{k-1}$, $k = 1, 2 \dots$, whose closure is compact in Ω. Fix $\varepsilon > 0$ and let $\{\xi_k\}_{k=1}^\infty$ a smooth partition of unity subordinated to the open covering $\{V_k\}_{k=1}^\infty$, that is,

$$\begin{cases} \xi_k \in C_c^\infty(V_k), \quad k = 1, 2 \dots \\ 0 \le \xi_k \le 1, \quad k = 1, 2, \dots \\ \sum_{k=1}^\infty \xi_k = 1 \ \text{ in } \ \Omega \end{cases}$$

For each $k = 1, 2 \dots$, we have $f\xi_k \in W^{k,p}(\Omega)$. Indeed, one can check this fact directly by definition, first for the case of multiindexes of order $|\alpha| = 1$ and then applying induction (see [11, Page 261]). Moreover $sop(f\xi_k) \subset V_k$, so using the previous result about convergence in $W^{k,p}_{loc}(\Omega)$ (Theorem 3.7) there exists $\varepsilon_k > 0$ so that

$$\begin{cases} sop(\delta_{\varepsilon_k} * (f\xi_k)) \subset V_k \\ \|\delta_{\varepsilon_k} * (f\xi_k) - f\xi_k\|_{W^{k,p}(\Omega)} \le \frac{\varepsilon}{2^k} \end{cases}$$

Define $f_\varepsilon = \sum_{k=1}^\infty \delta_{\varepsilon_k} * (f\xi_k)$.

- $f_\varepsilon \in C^\infty(\Omega)$ because for every point $x \in \Omega$ there is only a finite number of terms in the infinite sum that are not zero.
- Since $f = \sum_{k=1}^\infty f\xi_k$, we have

$$\|f_\varepsilon - f\|_{W^{k,p}(\Omega)} = \|\sum_{k=1}^{\infty}(\delta_{\varepsilon_k} * (f\xi_k) - f\xi_k)\|_{W^{k,p}(\Omega)} \leq \sum_{k=1}^{\infty}\frac{\varepsilon}{2^k} = \varepsilon.$$

We get the result from here. This is, we have proved that $W^{k,p}(\Omega) \cap C^\infty(\Omega)$ is dense in $W^{k,p}(\Omega)$.

□

3.2 Sobolev embeddings. Morrey's inequality

We now turn to talk about the well-known Sobolev embeddings. We distinguish different cases:

1. Case $1 \leq p < n$: We have the Gagliardo-Niremberg-Sobolev inequality, which states that for every $f \in W^{1,p}(\mathbb{R}^n)$,
$$\|f\|_{L^{p*}(\mathbb{R}^n)} \leq C(n,p)\|Df\|_{L^p(\mathbb{R}^n)},$$
where $p* = \dfrac{np}{n-p}$. In particular, the embedding $W^{1,p}(\mathbb{R}^n) \subset L^{p*}(\mathbb{R}^n)$ is continuous.
2. Case $p = n$: In this case, we only mention that $W^{1,n}(\mathbb{R}^n) \subset L^p_{loc}(\mathbb{R}^n)$ for every $1 \leq p < \infty$, and that there exists functions in $W^{1,n}$ that do not belong to L^∞. For counterexamples, we refer to the exercise sheet at the end of this chapter, or to Example 6.38 of [9], or to [17, Pages 123-125].
3. Case $p < n < \infty$: In this situation we have **Morrey's inequality**, stated below in Theorem 3.9, and to which we dedicate more time. But, as a consequence, we get
$$W^{1,p}(\mathbb{R}^n) \subset C^{0,1-n/p}(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n).$$

Theorem 3.9 (Morrey inequality)**.** *Let $\Omega \subset \mathbb{R}^n$ be open and suppose $n < p < \infty$. Let $B(x,r) \subset \Omega$ be any open ball. Then, for $\mathcal{L}^n$-almost every $y,\ z \in B(x,r)$, there exists a constant $C > 0$, depending on n and p, so that*

$$|f(y) - f(z)| \leq Cr^{1-\frac{n}{p}}\left(\int_{B(x,r)} |Df(w)|^p\, dw\right)^{\frac{1}{p}} \quad \forall f \in W^{1,p}(\Omega). \tag{3.6}$$

For the proof we follow [12, 19]. We need first a simple preliminary lemma.

Lemma 3.10. *Let $f : \mathbb{R}^n \to \mathbb{R}$ with $f \in C^1(B(x,r))$. For each $1 \leq p < \infty$ there exists $C(n,p) > 0$ so that*

$$\int_{B(x,r)} |f(y) - f(z)|^p\, dy \leq Cr^{n+p-1}\int_{B(x,r)} |Df(y)|^p|y-z|^{1-n}\, dy$$

for all $B(x,r) \subset \mathbb{R}^n$ and $z \in B(x,r)$.

Proof. Name $B = B(x,r)$. For $y,\ z \in B$, using that $f \in C^1(B)$, we have

$$f(y)-f(z) = \int_0^1 \frac{d}{dt}f(z+t(y-z))\, dt = \int_0^1 Df(z+t(y-z))\cdot(y-z)\, dt = \int_0^1 Df(z+t(y-z))\, dt\cdot(y-z).$$

Hence,

$$|f(y)-f(z)|^p \le |y-z|^p \int_0^1 |Df(z+t(y-z))|^p\,dt,$$

and for every $s>0$,

$$\begin{aligned}\int_{B\cap\partial B(z,s)} |f(y)-f(z)|^p\,d\mathcal{H}^{n-1}(y) &\le \int_{B\cap\partial B(z,s)} |y-z|^p \left(\int_0^1 |Df(z+t(y-z))|^p\,dt\right) d\mathcal{H}^{n-1}(y)\\ &= s^p \int_{B\cap\partial B(z,s)} \left(\int_0^1 |Df(z+t(y-z))|^p\,dt\right) d\mathcal{H}^{n-1}(y).\end{aligned}$$

Performing the change of variables $w=z+t(y-z)$ ($|w-z|=ts$), we have $d\mathcal{H}^{n-1}(y)=\frac{1}{t^{n-1}}d\mathcal{H}^{n-1}(w)$, and we obtain

$$\begin{aligned}\int_{B\cap\partial B(z,s)} |f(y)-f(z)|^p\,d\mathcal{H}^{n-1}(y) &\le s^p \int_0^1 \frac{1}{t^{n-1}} \left(\int_{B\cap\partial B(z,ts)} |Df(w)|^p\,d\mathcal{H}^{n-1}(w)\right) dt\\ &= s^{n+p-1}\int_0^1 \left(\int_{B\cap\partial B(z,ts)} |Df(w)|^p |w-z|^{1-n}\,d\mathcal{H}^{n-1}(w)\right) dt.\end{aligned}$$

Passing to polar coordinates (see Evans-Gariepy [12], pp. 118) it follows that

$$\int_{B\cap\partial B(z,s)} |f(y)-f(z)|^p\,d\mathcal{H}^{n-1}(y) = s^{n+p-2}\int_{B\cap B(z,s)} |Df(w)|^p|w-z|^{1-n}\,dw.$$

Integrating again in polar coordinates and using that $s>0$ was arbitrary, we get that

$$\int_B |f(y)-f(z)|^p\,dy \le C(n,p) r^{n+p-1}\int_B |Df(w)|^p |w-z|^{1-n}\,dw.$$

□

Proof of Morrey's inequality (Theorem 3.9). Suppose first that $f\in C^1(\Omega)$. The general case will follow by an approximation argument.

Let y, $z\in B(x,r)$. Since $f\in C^1(\Omega)$, by applying Lemma 3.10 for the case $p=1$ we have

$$\begin{aligned}|f(y)-f(z)| &= \fint_{B(x,r)} |f(y)-f(z)|\,dw \le \fint_{B(x,r)} |f(y)-f(w)|+|f(w)-f(z)|\,dw\\ &\le \frac{Cr^n}{\mathcal{L}^n(B(x,r))}\int_{B(x,r)} |Df(w)|\left(|y-w|^{1-n}+|w-z|^{1-n}\right)\,dw.\end{aligned}$$

Recall that $\mathcal{L}^n(B(x,r)) = Vol(B(x,r)) = \frac{\pi^{n/2}r^n}{\Gamma(\frac{n}{2}+1)} = C(n,p)r^n$, where Γ is the well-known gamma function. Then,

$$\begin{aligned}|f(y)-f(z)| &\le C\int_{B(x,r)} |Df(w)|\left(|y-w|^{1-n}+|w-z|^{1-n}\right)\,dw\\ &\underbrace{\le}_{\text{Hölder}} C\left(\int_{B(x,r)}\left(|y-w|^{1-n}+|w-z|^{1-n}\right)^{\frac{p}{p-1}}\,dw\right)^{\frac{p-1}{p}}\left(\int_{B(x,r)} |Df(w)|^p\,dw\right)^{\frac{1}{p}}\end{aligned}$$

Observe now that since y, $z\in B(x,r)$, we have $|y-w|\le 2r$, $|z-w|\le 2r$. A direct computation shows that

$$\begin{aligned}\int_{B(x,r)}\left(|y-w|^{1-n}+|z-w|^{1-n}\right)^{\frac{p}{p-1}}\,dw &\le \int_{B(x,r)}\left((2r)^{1-n}+(2r)^{1-n}\right)^{\frac{p}{p-1}}\,dw\\ &= \left(2(2r)^{1-n}\right)^{\frac{p}{p-1}}\int_{B(x,r)} dw\\ &= 2^{(2-n)\frac{p}{p-1}} r^{(1-n)\frac{p}{p-1}}\mathcal{L}^n(B(x,r))\\ &= C(n,p) r^{n-(n-1)\frac{p}{p-1}},\end{aligned}$$

Going back to our previous expression, we can write

$$|f(y)-f(z)| \leq Cr^{\left(n-(n-1)\frac{p}{p-1}\right)\frac{p-1}{p}} \left(\int_{B(x,r)} |Df(w)|^p\, dw\right)^{\frac{1}{p}} = Cr^{1-\frac{n}{p}} \left(\int_{B(x,r)} |Df(w)|^p\, dw\right)^{\frac{1}{p}},$$

finishing the proof for the case $f \in C^1(\Omega)$.

Consider now the case when $f \in W^{1,p}(\Omega)$. By Theorem 3.8, we can approximate f (because $p < \infty$) by functions of class C^1 converging in the norm of $W^{1,p}$. That is, there exists $\{f_i\}_{i=1}^{\infty} \subset C^1(\Omega)$ so that $f_i \xrightarrow{i\to\infty} f$ in $W^{1,p}(\Omega)$. These functions f_i do satisfy Morrey's inequality, as we already proved.

$$|f_i(y)-f_i(x)| \leq Cr^{1-\frac{n}{p}} \left(\int_{B(x,r)} |Df_i(w)|^p\, dw\right)^{\frac{1}{p}} = Cr^{1-\frac{n}{p}} \|Df_i\|_{L^p(B(x,r))}$$

Moreover, recalling the expression of the functions f_i, and applying the basic properties of convolution together with (3.5), we have

$$Df_i = D\left(\delta_{\varepsilon_i} * f\right) = \delta_{\varepsilon_i} * Df \;\; \text{donde} \;\; \varepsilon_i \to 0$$

and

$$\|Df_i\|_{L^p(B(x,r))} \leq \|\delta_{\varepsilon_i}\|_{L^1(\mathbb{R}^n)} \|Df\|_{L^p(B(x,r))} = \|Df\|_{L^p(B(x,r))}.$$

Therefore,

$$|f_i(y)-f_i(x)| \leq Cr^{1-\frac{n}{p}} \|Df\|_{L^p(B(x,r))}.$$

On the other hand, $\{f_i\}$ is an equicontinuous family of functions, so we can apply Ascoli-Arzelà Theorem to find the existence of a subsequence $\left\{f_{i_j}\right\}$ that converges uniformly to some $\tilde{f}$. Uniform convergence implies convergence in L^p_{loc} and by the uniqueness of limits in L^p_{loc} we conclude that $f = \tilde{f}$ almost everywhere.

We then get the validity of the result for the space $W^{1,p}(\Omega)$. □

Let us state some consequences:

(1) The next result is a slight modification of Theorem 3.9.

Corollary 3.11. *Let $\Omega \subset \mathbb{R}^n$ be open, $n < p < \infty$ and some $R > 0$. Then for every ball $B(x,r) \subset B(0,R)$ we have that for $\mathcal{L}^n$-almost every $y, z \in B(x,r)$ there exists a constant $C > 0$, only depending on n and p, so that*

$$|f(y)-f(z)| \leq C|y-z|^{1-\frac{n}{p}} \|Df\|_{L^p(B(x,r))} \quad \forall f \in W^{1,p}(B(0,R)). \tag{3.7}$$

Proof. Take a countable dense set $\{x_m\}_{m\geq 1}$ over $B(0,R)$. For every $m \in \mathbb{N}$, let $r_{m,k} := \{q \in Q : B(x_m, r_{m,k}) \subset B(0,R)\}$. By using Theorem 3.9, each ball $B(x_m, r_{m,k})$ has an associated null set $N_{m,k}$ so that for all $y, z \in B(x_m, r_{m,k}) \setminus N_{m,k}$ we have (3.6) holding. By letting $N = \bigcup_{m,k} N_{m,k}$, we have $\mathcal{L}^n(N) = 0$ and for every $x, y \notin N$ so that $y, z \in B(x_m, r_{m,k})$ for some $k, m \in \mathbb{N}$ we have

$$|f(y)-f(z)| \leq C(n,p)(r_{m,k})^{1-n/p} \|Df\|_{L^p(B(x,r_{m,k}))}.$$

Fix an arbitrary $B(x,r) \subset B(0,R)$ and take $y, z \in B(x,r) \setminus N$. Choose $\widetilde{x} \in B(0,R)$ so that $B(\widetilde{x}, |y-z|) \subset B(x,r)$ and $y, z \in B(\widetilde{x}, |y-z|)$. Next, define a sequence of points $(x_m)_{m\geq 1}$ and a sequence of radii $(r_{m,k(m)})_{m\geq 1}$ so that $x_m \to \widetilde{x}$ and $r_{m,k(m)} \to |y-z|$ and so that $y, z \in B(x_m, r_{m,k(m)})$. Therefore, we have

$$|f(y)-f(z)| \leq C(n,p)(r_{m,k(m)})^{1-n/p} \|Df\|_{L^p(B(x_m, r_{m,k(m)}))},$$

and by taking limits $m \to \infty$ we conclude that

$$|f(y) - f(z)| \leq C(n,p)|y-z|^{1-n/p}\|Df\|_{L^p(B(\widetilde{x},|y-z|)} \leq C(n,p)|y-z|^{1-n/p}\|Df\|_{L^p(B(x,r)}$$

□

(2) If we have $f \in W^{1,p}(\mathbb{R}^n)$, then for every $j \in \mathbb{N}$ we have $f \in W^{1,p}(B(0,j))$. Call $N_j \subset \mathbb{R}^n$ the negligible exceptional sets where (3.7) does not hold. In particular, $N = \bigcup_{j\geq 1} N_j$ satisfies $\mathcal{L}^n(N) = 0$. We can say that for all $x, y \in \mathbb{R}^n \setminus N$ we have

$$|f(y) - f(z)| \leq C|y-z|^{1-\frac{n}{p}}\|Df\|_{L^p(\mathbb{R}^n)}, \tag{3.8}$$

because $x, y \in B(0,j)$ for some $j \in \mathbb{N}$ big enough. This means that f is Hölder continuous of exponent $\alpha = 1 - n/p$ on $\mathbb{R}^n \setminus N$. Since f is Hölder continuous on a dense subset of $\mathbb{R}^n$, it has a unique extension to a Hölder continuous function with same exponent defined everywhere on $\mathbb{R}^n$. This extension is a Hölder continuous representative of exponent $1 - n/p$ of the function f and satisfies (3.8). Call f (abusing of notation) this Hölder representative. We have

$$\|f\|_{C^{0,1-n/p}(\mathbb{R}^n)} \leq C(n,p)\|f\|_{W^{1,p}(\mathbb{R}^n)}.$$

Moreover, we also have that $f \in L^\infty(\mathbb{R}^n)$ with $\|f\|_{L^\infty(\mathbb{R}^n)} \leq C(n,p)\|f\|_{W^{1,p}(\mathbb{R}^n)}$. Indeed, for every $x \in \mathbb{R}^n \setminus N$, the inequality (3.8) and the Hölder inequality leads to

$$\begin{aligned}
|f(x)| &= \left|\frac{1}{\mathcal{L}^n(B(x,1))}\int_{B(x,1)} f(x)\,dy\right| \\
&\leq \frac{1}{\mathcal{L}^n(B(x,1))}\left(\int_{B(x,1)} |f(x)-f(y)|\,dy + \int_{B(x,1)} |f(y)|\,dy\right) \\
&\leq C(n,p)\frac{1}{\mathcal{L}^n(B(x,1))}\int_{B(x,1)} |x-y|^{1-n/p}\|Df\|_{L^p(\mathbb{R}^n)}\,dy + C(n,p)\|f\|_{L^p(\mathbb{R}^n)} \\
&\leq C(n,p)\|f\|_{W^{1,p}(\mathbb{R}^n)}.
\end{aligned}$$

We then conclude that

$$W^{1,p}(\mathbb{R}^n) \subset C^{0,1-n/p}(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n) \quad \text{continuously.}$$

3.3 Proofs of Rademacher theorem for Sobolev functions

With the help of Morrey's inequality, we can now prove the main theorem of this chapter, which is the Rademacher theorem for $W^{1,p}$ functions where $n < p < \infty$. Let us recall its statement.

Theorem 3.1. *Let $f \in W^{1,p}(\mathbb{R}^n)$ for some $n < p < \infty$. Then there exists a representative of f that is Hölder continuous with exponent $\alpha(p) = 1 - \frac{n}{p}$ which moreover is differentiable at $\mathcal{L}^n$ almost every point $x \in \mathbb{R}^n$.*

Proof. By the last comments (1) and (2) from the previous section, we can assume without loss of generality that f is a bounded Hölder continuous function with exponent $\alpha = 1 - n/p$. Let us consider the weak derivative of f,

$$Df(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right) \in L^p(\mathbb{R}^n; \mathbb{R}^n).$$

By the Lebesgue differentiation theorem (see Theorem 2.8), we have that

$$\lim_{r\to 0}\frac{1}{\mathcal{L}^n(B(x,r))}\int_{B(x,r)}|Df(z)-Df(x)|^p\,dz=0 \quad \text{for almost every } x\in\mathbb{R}^n. \tag{3.9}$$

We will show that f is differentiable at every Lebesgue point of Df (that is, exactly at the points where (3.9) holds), and the differential will be exactly $Df(x)$. Observe that $Df(x)\in\mathbb{R}^n$ is well defined because we have

$$Df(x)=\lim_{r\to 0}\frac{1}{\mathcal{L}^n(B(x,r))}\int_{B(x,r)}Df(y)\,dy.$$

Fix $x\in\mathbb{R}^n$ to be one of those points. Let also $g:\mathbb{R}^n\to\mathbb{R}$ be defined by

$$g(y):=f(y)-f(x)-Df(x)\cdot(y-x),\quad y\in\mathbb{R}^n.$$

Observe that $g\in W^{1,p}(B(0,R))$ for every $R>0$ (however, $g\notin L^p(\mathbb{R}^n)$). Fix $R=|x|+2$ and let $y\in B(x,1)$ and $r=|x-y|$. By using the Morrey's inequality (3.7) with the function $g\in W^{1,p}(B(0,R))$ and since $x,y\in B(x,2r)\subset B(0,R)$,

$$\begin{aligned}|g(y)-g(x)|=|g(y)|&\le C(n,p)|x-y|^{1-n/p}\|Dg\|_{L^p(B(x,2r))}\\ &=C(n,p)|x-y|\left(\frac{1}{\mathcal{L}^n(B(x,2r))}\int_{B(x,2r)}|Df(z)-Df(x)|^p\,dz\right)^{1/p}.\end{aligned}$$

Therefore, if $y\in B(x,1)$ we have

$$\frac{|f(y)-f(x)-Df(x)\cdot(y-x)|}{|y-x|}\le C(n,p)\left(\frac{1}{\mathcal{L}^n(B(x,2r)}\int_{B(x,2r))}|Df(z)-Df(x)|^p\,dz\right)^{1/p}.$$

By letting $y\to x$, and hence $r\to 0$, and using (3.9), we conclude that at every Lebesgue point x of Df

$$\lim_{y\to x}\frac{|f(y)-f(x)-Df(x)\cdot(y-x)|}{|y-x|}=0.$$

□

Corollary 3.11. *1. Let $f\in W^{1,p}_{loc}(\mathbb{R}^n)$ with $n<p<\infty$. Then f has a representative that is differentiable $\mathcal{L}^n$-almost everywhere.*

2. Let $f\in W^{1,\infty}_{loc}(\mathbb{R}^n)$. Then f has a representative which is differentiable $\mathcal{L}^n$-almost everywhere.

Proof. 1. We have $f\in W^{1,p}(B(0,j))$ for every $j\in\mathbb{N}$. Let $\varphi_j\in C_0^\infty(\mathbb{R}^n)$ with $\varphi_j=1$ on $B(0,j)$. Then it is an easy exercise to verify that $f\cdot\varphi_j\in W^{1,p}(\mathbb{R}^n)$. Now apply Theorem 3.1 to get that $f\cdot\varphi_j$ is differentiable $\mathcal{L}^n$-almost everywhere. Therefore, since $f=f\cdot\varphi_j$ on $B(0,j)$, f is differentiable $\mathcal{L}^n$-almost everywhere on $B(0,j)$. By covering $\mathbb{R}^n=\bigcup_{j\ge 1}B(0,j)$ and doing the same argument for every $j\in\mathbb{N}$, and by the subadditivity of the Lebesgue measure, we conclude that f is differentiable $\mathcal{L}^n$-almost everywhere on $\mathbb{R}^n$.

2. It is enough to observe that $W^{1,\infty}_{loc}(\mathbb{R}^n)\subset W^{1,p}_{loc}(\mathbb{R}^n)$ for every $1\le p<\infty$ and apply item (1).

□

We have just seen that Rademacher theorem for $W^{1,\infty}$ functions relies heavily on the case $W^{1,p}$, $n<p<\infty$, where we did the hard work. But an interesting consequence of having the Rademacher theorem for $W^{1,\infty}$ functions is that it gives an alternative proof of the classical Rademacher theorem for

Lipschitz functions (Theorem 2.1) once we prove that locally Lipschitz functions are functions belonging to $W^{1,\infty}_{loc}$.

Theorem 3.12. *Every locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies that $f \in W^{1,\infty}_{loc}(\mathbb{R}^n)$. Moreover, the same is true if we replace $\mathbb{R}^n$ with an open set $\Omega \subset \mathbb{R}^n$.*

Remark 3.13. The converse is also true, in the sense that every $f \in W^{1,\infty}_{loc}(\mathbb{R}^n)$ has a locally Lipschitz representative. Moreover one can prove that $\mathrm{Lip}(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n) = W^{1,\infty}(\mathbb{R}^n)$. It is important to stress out that the previous fact is not necessarily true for other domains $\Omega \subset \mathbb{R}^n$. Indeed, for a bounded domain $\Omega \subset \mathbb{R}^n$ we have $\mathrm{Lip}(\Omega) \cap L^\infty(\Omega) = W^{1,\infty}(\Omega)$ if and only if Ω is quasiconvex[2] (This was proven by Hajłasz, Koskela and Tuominen in 2008).

Proof. We mainly follow the proof given in [12]. The case $\Omega \subset \mathbb{R}^n$ is an open set is left for the reader.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. That is, for every open set $V \subset \mathbb{R}^n$ with $\overline{V}$ compact we have that $f|_{\overline{V}}$ is Lipschitz. Let us take then an open subset V with $\overline{V}$ compact and the theorem will be proved if we check that $f \in W^{1,\infty}(V)$.

Firstly, using that f is continuous and $\overline{V}$ compact we have

$$\|f\|_{L^\infty(V)} = \sup_{x\in V} |f(x)| \leq \sup_{x\in \overline{V}} |f(x)| < \infty.$$

Secondly we aim to show that f has weak partial derivatives

$$\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}$$

and each of them belong to the space L^∞. For that take $t > 0$ and fix $i = 1, \dots, n$. Define now

$$g_i^t := \frac{f(x + te_i) - f(x)}{t}, \quad x \in V.$$

We have

$$\sup_{0<t<1} \|g_i^t\|_{L^\infty(V)} \leq \mathrm{Lip}(f|_W) < \infty,$$

where $W = V + B(0,1)$ and $\mathrm{Lip}(f|_W)$ is the least Lipschitz constant of f on W (note that $\overline{W}$ is compact).

Take next a decreasing sequence $(t_j)_{j\geq 1} \subset (0,1)$ with $\lim_{j\to\infty} t_j = 0$. We have that

$$(g_i^{t_j})_{j\geq 1} \subset L^\infty(V)$$

is a uniformly bounded sequence. Therefore, since every closed ball of $L^\infty(V)$ is weak* sequentially compact, there exists a subsequence $(g_i^{t_j})_{j\geq 1}$, which we denote the same way, that weak* converges to some function $g_i \in L^\infty(V)$. Let us explain briefly why every closed ball of $L^\infty(V)$ is weak* sequentially compact.

- By the Banach-Alaouglu theorem we have that the every closed ball of $L^\infty(V)$ is compact in the weak* topology, just because it is the dual of a Banach space.
- Since $L^1(V)$ is separable, we have that the unit ball of the dual endowed with the weak* topology is metrizable. Indeed, for a Banach space X, we have that $(\overline{B_{X^*}}, w^*)$ is metrizable as a topological space if and only if X is separable.
- In a metric space, a subset $A \subset X$ is compact if and only if it is sequentially compact.

[2]A domain $\Omega \subset \mathbb{R}^n$ is said to be quasiconvex if and only if there exists a constant $C \geq 1$ so that for every $x, y \in \Omega$ there exists a rectifiable curve γ joining x with y such that $\ell(\gamma) \leq C|x - y|$.

A proof of the Banach-Alouglu theorem can be found in [14, Theorem 3.37]. And, a proof of the second fact follows from [14, Propositions 3.103 and 3.101].

In summary, we find that for every $i = 1, \dots, n$ the sequence $(g_i^{t_j})_{j\geq 1}$ is weak* convergent to some $g_i \in L^\infty(V)$. In particular, for every $\varphi \in C_0^\infty(V) \subset L^1(V)$ we have that

$$\lim_{j\to\infty} \int_V g_i^{t_j}(x)\varphi(x)\,dx = \int_V g_i(x)\varphi(x)\,dx. \tag{3.10}$$

On the other hand,

$$\int_V g_i^{t_j}(x)\varphi(x)\,dx = \int_V \frac{f(x+t_je_i)-f(x)}{t_j}\varphi(x)\,dx = -\int_V f(x)\frac{\varphi(x)-\varphi(x-t_je_i)}{t_j}\,dx,$$

so taking $j \to \infty$ and using the Dominated Convergence Theorem together with (3.10) we get

$$\int_V g_i(x)\varphi(x)\,dx = -\int_V f(x)\frac{\partial\varphi}{\partial x_i}(x)\,dx.$$

The latter equality holds for every $i = 1, \dots, n$ and for every $\varphi \in C_0^\infty(V)$. It follows that the functions $g_i \in L^\infty(V)$ are the weak partial derivatives of f on V. So we are done.

□

Before finishing this chapter let us give a final version of Rademacher theorem for $W^{1,1}$ functions $f : \mathbb{R} \to \mathbb{R}$. This is the simplest case, where we can prove the following.

Theorem 3.14. *Let $f \in W^{1,1}_{loc}(\mathbb{R})$. Then f has an absolutely continuous representative in every interval $[a,b] \subset \mathbb{R}$ that is differentiable $\mathcal{L}^1$-almost everywhere.*

Proof. Once we prove that f is equal almost everywhere to an absolutely continuous function, by applying Theorem 2.4, we find that f is differentiable almost everywhere. Precisely, we will prove that the precise representative of f, defined as

$$f^*(x) := \begin{cases} \lim\limits_{r\to 0} \dfrac{1}{\mathcal{L}^n(B(x,r))}\displaystyle\int_{B(x,r)} f(y)\,dy & \text{if the limit exists} \\ 0 & \text{otherwise} \end{cases}$$

is absolutely continuous on every $[a,b] \subset \mathbb{R}$. Note that if $x_0 \in (a,b)$ is a Lebesgue point of f, then $f(x_0) = f^*(x_0)$.

Fix an interval $[a,b] \subset \mathbb{R}$ and a Lebesgue point $x_0 \in (a,b)$ of f. Then for every $n \in \mathbb{N}$ and letting $\varepsilon_n = 1/n > 0$ consider $f_n = \delta_{\varepsilon_n} * f$, where δ_{ε_n} is the standard mollifier from Definition 3.5. By the smoothness of f_n and the fundamental theorem of calculus, we have

$$f_n(y) = f_n(x) + \int_x^y (f_n)'(t)\,dt, \quad \forall [x,y] \subset \mathbb{R}. \tag{3.11}$$

Moreover, for every $n, m \in \mathbb{N}$ and $x \in [a,b]$,

$$|f_n(x) - f_m(x)| = \left| f_n(x_0) + \int_{x_0}^x (f_n)'(t)\,dt - f_m(x_0) - \int_{x_0}^x (f_m)'(t)\,dt \right| \tag{3.12}$$

$$\leq \int_{x_0}^x |(f_n)'(t) - (f_m)'(t)|\,dt + |f_n(x_0) - f_m(x_0)| \tag{3.13}$$

Observe that

- Since $x_0 \in [a,b]$ is a Lebesgue point of f, by Theorem 3.7 (3), we have that

$$\lim_{n\to\infty} f_n(x_0) = f^*(x_0).$$

- Since $f \in W^{1,1}_{loc}(\mathbb{R})$, by Theorem 3.7 (4), we know that $f_n \to f$ in $W^{1,1}([a,b])$. In particular,

$$(f_n)' \to Df \quad \text{in } L^1([a,b])$$

From this, we get that $(f_n)_{n\geq 1}$ defines a Cauchy sequence in $(C([a,b]), \|\cdot\|_\infty)$, that is, the space of continuous function with the supremum norm in $[a,b]$. Indeed, given $\varepsilon > 0$, by the above facts, there exists $n_0 \in \mathbb{N}$ big enough so that for every $n, m \geq n_0$ we have

$$|f_n(x_0) - f_m(x_0)| < \frac{\varepsilon}{2}$$

and

$$\int_{x_0}^{x} |(f_n)'(t) - (f_m)'(t)|\, dt \leq \int_a^b |(f_n)'(t) - Df(t)|\, dt + \int_a^b |Df(t) - (f_m)'(t)|\, dt < \frac{\varepsilon}{2}, \quad \forall x \in [a,b].$$

Hence, for every $x \in [a,b]$ and for every $n, m \geq n_0$, recalling (3.12),

$$|f_n(x) - f_m(x)| < \varepsilon.$$

This proves that $(f_n)_{n\geq 1} \subset (C([a,b]), \|\cdot\|_\infty)$ is a Cauchy sequence. Therefore, the sequence $(f_n)_{n\geq 1}$ is uniformly convergent to a continuous function $g : [a,b] \to \mathbb{R}$. In particular, $f_n(x) \to g(x)$ everywhere.

On the other hand, it is also true that $f_n(x) \to f(x)$ for almost every $x \in [a,b]$ (again by Theorem 3.7), so we must have $g = f = f^*$ almost everywhere.

Furthermore, by letting $n \to \infty$ in (3.11), we can write

$$g(y) = g(x) + \int_x^y (Df^*)(t)\, dt \quad \forall [x,y] \subset [a,b].$$

Here we are using the pointwise convergence of f_n to g, the L^1-convergence of $(f_n)'$ to Df on compact intervals, and the fact that $Df = Df^*$ almost everywhere.

Finally, observe that the continuity of g and the fact that $f = g$ almost everywhere gives that for every $x \in [a,b]$,

$$\lim_{r\to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} f(y)\, dy = \lim_{r\to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} g(y)\, dy = g(x),$$

therefore $f^* = g$ everywhere. In particular,

$$f^*(x) = f^*(x_0) + \int_{x_0}^{x} D(f^*)(t)\, dt.$$

By applying Theorem 2.4, we conclude that f^* is absolutely continuous.

□

Exercises

1. Prove the following:

   (a) For $n < p \leq \infty$, if $f \in W^{1,p}(\mathbb{R}^n)$, the following limit exists for all $x \in \mathbb{R}^n$.

   $$\lim_{r \to 0} \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} f(y)\, dy.$$

   (b) If for a function $f : \mathbb{R}^n \to \mathbb{R}$ there exists $\alpha > 1$ and $C \geq 0$ so that

   $$|f(x) - f(y)| \leq C|x - y|^\alpha, \quad \forall x, y \in \mathbb{R}^n,$$

   then f is a constant function.

2. Prove that the space of Hölder continuous functions with exponent $\alpha \in (0,1)$ and bounded, endowed with the norm

   $$\|f\| = \|f\|_{L^\infty(\mathbb{R}^n)} + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}$$

   forms a Banach space.

3. Prove that the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

   $$f(x) = \begin{cases} \log\left(\log\left(1 + \frac{1}{|x|}\right)\right), & x \neq 0 \\ 0, \quad x = 0 \end{cases}$$

   belongs to $W^{1,2}(B(0,1))$, but $f \notin L^\infty(B(0,1))$.

4. Let $\eta > 0$. Prove that $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) = |x|^{-\eta}$ satisfies $f \notin W^{1,p}(B(0,1))$ when $p \geq 2$ and that, whenever $p < 2$ and $\eta < \frac{2}{p} - 1$, then $f \in W^{1,p}(B(0,1))$.

5. Given a couple of functions $f \in W^{1,p}(\mathbb{R}^n)$ and $\varphi \in C_0^\infty(\mathbb{R}^n)$, prove that $f\varphi \in W^{1,p}(\mathbb{R}^n)$ by estimating its Sobolev norm.

6. Prove that the $f : [0, 1/2] \to \mathbb{R}$ defined by

   $$f(x) = \begin{cases} \frac{1}{\log x}, & x \in (0, 1/2] \\ 0, \quad x = 0 \end{cases}$$

   is not Hölder continuous for any exponent $\alpha \in (0,1]$, though it is absolutely continuous. Moreover, give an example of Hölder continuous function which is not absolutely continuous.

7. Prove that the domain $\Omega \subset \mathbb{R}^2$ given by

   $$\Omega = B(0,1) \setminus \{(x,y) \in \mathbb{R}^2 : \; 0 \leq x \leq 1, \; |y| \leq x^2\}.$$

   is not quasiconvex. Latter, give an example of a function $f : \Omega \to \mathbb{R}$ which satisfies $f \in W^{1,\infty}(\Omega)$ but so that f is not Lipschitz on Ω.

8. Prove that if a function $f : \mathbb{R}^n \to \mathbb{R}$, $n \geq 2$, is locally integrable ($f \in L^1_{loc}(\mathbb{R}^n)$), then

   $$\mathcal{H}^{n-1}\left(\left\{x \in \mathbb{R}^n : \; \limsup_{r \to 0} \frac{1}{r^{n-1}} \int_{B(x,r)} |f(y)|\, dy > 0\right\}\right) = 0$$

Chapter 4

Hausdorff measures

This chapter is intended to be a brief introduction to Hausdorff measures. We basically follow [12].

Definition 4.1. *Let $A \subset \mathbb{R}^n$ be a set, $0 \leq s < \infty$ and $0 < \delta \leq \infty$. We define the s-dimensional δ-Hausdorff content of A as*

$$\mathcal{H}^s_\delta(A) = \inf\left\{\sum_{j=1}^{\infty} diam(C_j)^s : A \subset \bigcup_{j\geq 1} C_j,\ diam(C_j) \leq \delta\right\}$$

We define the s-dimensional Hausdorff measure of A as

$$\mathcal{H}^s(A) = \lim_{\delta\to 0^+} \mathcal{H}^s_\delta = \sup_{\delta>0} \mathcal{H}^s_\delta(A)$$

(observe that the mapping $\delta \to \mathcal{H}^s_\delta(A)$ is decreasing).

There are other definitions of Hausdorff measures. For instance one could sum $\sum_{j=1}^{\infty} \alpha(s)(\mathrm{diam}(C_j)/2)^s$ instead of $\sum_{j=1}^{\infty} \mathrm{diam}(C_j)^s$, where $\alpha(s) = \dfrac{\pi^{s/2}}{\Gamma(s/2+1)}$ and $\Gamma(t) = \int_0^\infty e^{-x}x^{t-1}\,dx$ is the Gamma function. The reason is to have $\mathcal{H}^n = \mathcal{L}^n$ on $\mathbb{R}^n$ as will be shown in Theorem 4.5.

Theorem 4.2. *For every $0 \leq s < \infty$ we have that $\mathcal{H}^s$ is a Borel regular measure (exterior) on $\mathbb{R}^n$. (If $s < n$ it is not a Radon measure)*

Proof. We start by showing that $\mathcal{H}^s_\delta$ is an exterior measure for every $\delta > 0$. First, it is clear that $\mathcal{H}^s_\delta(\emptyset) = 0$ (we understand $\mathrm{diam}(\emptyset) = 0$. Second, we take $A \subset \mathbb{R}^n$ and $\{A_k\}_{k\geq 1} \subset \mathbb{R}^n$ so that $A \subset \bigcup_{k\geq 1} A_k$, and we want to show that

$$\mathcal{H}^s_\delta(A) \leq \sum_{k=1}^{\infty} \mathcal{H}^s_\delta(A_k) \tag{4.1}$$

Indeed, for every $k \in \mathbb{N}$ we take a covering $A_k \subset \bigcup_{j\geq 1} C_{j,k}$ so that $\mathrm{diam}(C_{j,k}) \leq \delta$. Then

$$A \subset \bigcup_{k\geq 1} A_k \subset \bigcup_{j,k=1}^{\infty} C_{j,k}$$

and therefore

$$\mathcal{H}^s_\delta(A) \leq \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \mathrm{diam}(C_{j,k})^s.$$

By taking infimums over all possible coverings $\{C_{j,k}\}_{j\geq 1}$ of A_k we easily conclude (4.1).

The second step of the proof is to show that $\mathcal{H}^s$ is also an exterior measure. Again $\mathcal{H}^s(\emptyset) = 0$ is clear. For the subadditivity we take $\{A_k\}_{k\geq 1} \subset \mathbb{R}^n$. For every $\delta > 0$, since $\mathcal{H}^s_\delta$ is a measure we have

$$\mathcal{H}^s_\delta\left(\bigcup_{k\geq 1} A_k\right) \leq \sum_{k=1}^{\infty} \mathcal{H}^s_\delta(A_k) \leq \sum_{k=1}^{\infty} \sup_{\delta'>0} \mathcal{H}^s_{\delta'}(A_k) = \sum_{k=1}^{\infty} \mathcal{H}^s(A_k).$$

Now, by taking limits $\delta \to 0^+$ we conclude that $\mathcal{H}^s\left(\bigcup_{k\geq 1} A_k\right) \leq \sum_{k=1}^{\infty} \mathcal{H}^s(A_k)$.

The next step of the proof is to show that $\mathcal{H}^s$ is Borel. For that we will use Caratheodory criterion (see [12, Theorem 1.9]) asserting that a measure μ is Borel on $\mathbb{R}^n$ is for every two sets $A, B \subset \mathbb{R}^n$ with $\text{dist}(A, B) > 0$ we have $\mu(A \cup B) = \mu(A) + \mu(B)$. Indeed, for A, B subsets of $\mathbb{R}^n$ with $\text{dist}(A, B) > 0$ let $0 < \delta < \text{dist}(A, B)/4$ and let $A \cup B \bigcup_{k\geq 1} C_k$ be covering such that $\text{diam}(C_k) \leq \delta$. Define the families of indexes

$$\mathcal{A} = \{j \in \mathbb{N} : C_j \cap A \neq \emptyset\} \quad ; \quad \mathcal{B} = \{j \in \mathbb{N} : C_j \cap B \neq \emptyset\}.$$

It is clear that $A \subset \bigcup_{j\in\mathcal{A}} C_j$, $B \subset \bigcup_{j\in\mathcal{B}} C_j$ and that $C_i \cap C_j = \emptyset$ whenever $i \in \mathcal{A}$, $j \in \mathcal{B}$. Hence

$$\sum_{j=1}^{\infty} \text{diam}(C_j)^s \geq \sum_{j\in\mathcal{A}} \text{diam}(C_j)^s + \sum_{j\in\mathcal{B}} \text{diam}(C_j)^s \geq \mathcal{H}^s_\delta(A) + \mathcal{H}^s_\delta(B).$$

By taking the infimum over all possible coverings $A \cup B \subset \bigcup_{k\geq 1} C_k$ with $\text{diam}(C_k) \leq \delta$ we have

$$\mathcal{H}^s_\delta(A \cup B) \geq \mathcal{H}^s_\delta(A) + \mathcal{H}^s_\delta(B).$$

Letting $\delta \to 0^*$ we get $\mathcal{H}^s(A\cup B) \geq \mathcal{H}^s(A)+\mathcal{H}^s(B)$ and by the subadditivity of $\mathcal{H}^s$ we get $\mathcal{H}^s(A\cup B) \leq \mathcal{H}^s(A) + \mathcal{H}^s(B)$, so we are done.

The final step of the proof is to show that $\mathcal{H}^s$ is Borel regular, that is for every $A \subset \mathbb{R}^n$ we must find a Borel set $B \supset A$ with $\mathcal{H}^s(A) = \mathcal{H}^s(B)$. We will use that for every set $C \subset \mathbb{R}^n$ we have $\text{diam}(C) = \text{diam}(\overline{C})$, which in particular allows us to write

$$\mathcal{H}^s_\delta(A) = \inf\left\{\sum_{j=1}^{\infty} \text{diam}(C_j)^s : A \subset \bigcup_{j\geq 1} C_j,\ \text{diam}(C_j) \leq \delta,\ C_j \text{ closed}\right\}$$

Take $A \subset \mathbb{R}^n$ with $\mathcal{H}^s(A) < \infty$ (otherwise we can let $B = \mathbb{R}^n$). Then $\mathcal{H}^s_\delta(A) < \infty$ for every $\delta > 0$. For each $k \in \mathbb{N}$ we will choose a family $\{C_{j,k}\}_{j\geq 1}$ of closed sets with $\text{diam}(C_{j,k}) \leq 1/k$, with $A \subset \bigcup_{j\geq 1} C_{j,k}$ and such that

$$\sum_{j=1}^{\infty} \text{diam}(C_{j,k})^s \leq \mathcal{H}^s_{1/k}(A) + \frac{1}{k}.$$

Define $A_k = \bigcup_{j\geq 1} C_{j,k}$ and $B = \bigcap_{k\geq 1} A_k$. As $C_{j,k}$ are closed sets, it is clear that A_k are Borel sets for every k and hence B is a Borel set too. Moreover, we have $B \subset A_k$ for every $k \in \mathbb{N}$ so

$$\mathcal{H}^s_{1/k}(B) \leq \mathcal{H}^s_{1/k}(A_k) \leq \sum_{j=1}^{\infty} \text{diam}(C_{j,k})^s \leq \mathcal{H}^s_{1/k}(A) + \frac{1}{k}.$$

Let $k \to \infty$ to get $\mathcal{H}^s(B) \leq \mathcal{H}^s(A)$, and since $A \subset B$ we also get $\mathcal{H}^s(A) \leq \mathcal{H}^s(B)$. The proof is finished. □

We shall now state more basic, but important, properties of Hausdroff measures.

Proposition 4.3. *Let* $A \subset \mathbb{R}^n$.

(1) $\mathcal{H}^0$ *is the counting meausure. That is,* $\mathcal{H}^0(A) = \begin{cases} \#A & \text{if } A \text{ finite} \\ \infty & \text{if } A \text{ infinite} \end{cases}.$

(2) $\mathcal{H}^s = 0$ *on* $\mathbb{R}^n$ *for every* $s > n$*.*

(3) If $\mathcal{H}^s_\delta(A) = 0$ *for some* $\delta \in (0, \infty]$*, then* $\mathcal{H}^s(A) = 0$*.*

(4) Let $0 \leq s < t < \infty$*. Then*

$$\mathcal{H}^s(A) < \infty \quad \Rightarrow \quad \mathcal{H}^t(A) = 0 \quad ; \quad \mathcal{H}^t(A) > 0 \quad \Rightarrow \quad \mathcal{H}^s(A) = \infty.$$

(5) Let $A \subset \mathbb{R}^n, \lambda > 0$ *and denote* $\lambda A = \{\lambda x : x \in A\}$*. Then* $\mathcal{H}^s(\lambda A) = \lambda^s \mathcal{H}^s(A)$ *for all* $s \in [0, \infty)$*.*

(6) If $f : \mathbb{R}^n \to \mathbb{R}^m$ *is Lipschitz, then* $\mathcal{H}^s(f(A)) \leq (Lip(f))^s \mathcal{H}^s(A)$*. In particular* $dim_{\mathcal{H}}(f(A)) \leq dim_{\mathcal{H}}(A)$*.*

Proof. The proof of (1) is easy and the proof of (5)-(6) is left as an exercise. So we will only prove (2), (3) and (4). For further details we refer to [12, Chapter 2].

(2): Let $m \in \mathbb{N}$ and split the unit cube $Q = [0,1]^n \subset \mathbb{R}^n$ in m^n subcubes of side-length $1/m$ and diameter

$$\operatorname{diam}(Q) = \sqrt{n} \cdot \ell(Q) = \frac{\sqrt{n}}{m}.$$

Then,

$$0 \leq \mathcal{H}^s_{\sqrt{n}/m} \leq \sum_{i=1}^{m^n} \left(\frac{\sqrt{n}}{m}\right)^s = n^{s/2} m^{n-s}$$

Using that $s > n$, and taking limits $m \to \infty$ we get that $\mathcal{H}^s(Q) = 0$. Since we can write $\mathbb{R}^n = \bigcup_{j \geq 1} Q_j$ as a countable union of cubes of side-length one, we conclude that

$$\mathcal{H}^s(\mathbb{R}^n) = \mathcal{H}^s\left(\bigcup_{j \geq 1} Q_j\right) \leq \sum_{j \geq 1} \mathcal{H}^s(Q_j) = 0,$$

and $\mathcal{H}^s(A) \leq \mathcal{H}^s(\mathbb{R}^n) = 0$ for every subset $A \subset \mathbb{R}^n$.

(3) If $s = 0$ it is clear. Assume $s > 0$ and let $\varepsilon > 0$. By assumptions, we can find a covering $\{C_j\}_{j \geq 1}$ of the set A so that

$$\sum_{j=1}^{\infty} \operatorname{diam}(C_j)^s \leq \varepsilon.$$

In particular, for every $j \in \mathbb{N}$,

$$\operatorname{diam}(C_j) \leq \varepsilon^{1/s}.$$

Therefore, if we define $\delta(\varepsilon) = \varepsilon^{1/s}$ we have $\mathcal{H}^s_{\delta(\varepsilon)} \leq \varepsilon$. Taking limits $\varepsilon \to 0^+$ we conclude that

$$\mathcal{H}^s(A) = \lim_{\varepsilon \to 0^+} \mathcal{H}^s_{\delta(\varepsilon)}(A) \leq \lim_{\varepsilon \to 0^+} \varepsilon = 0$$

(4) Let $A \subset \mathbb{R}^n$ and let $0 \leq s < t < \infty$. We first assume that $\mathcal{H}^s(A) < \infty$. Let $\delta > 0$ and let $\{C_j\}_{j \geq 1}$ a covering of A so that $\operatorname{diam}(C_j) \leq \delta$ and

$$\sum_{j \geq 1} \operatorname{diam}(C_j)^s \leq \mathcal{H}^s_\delta(A) + 1 \leq \mathcal{H}^s(A) + 1.$$

Then,

$$0 \leq \mathcal{H}^t_\delta(A) \leq \sum_{j=1}^{\infty} \operatorname{diam}(C_j)^t = \sum_{j=1}^{\infty} \operatorname{diam}(C_j)^s \operatorname{diam}(C_j)^{t-s} \leq \delta^{t-s}(\mathcal{H}^S(A) + 1).$$

Since $t - s > 0$, if we let $\delta \to 0$ we conclude that

$$0 \leq \lim_{\delta \to 0^+} \mathcal{H}^t_\delta(A) = \mathcal{H}^t(A) = 0.$$

In the case that $\mathcal{H}^t(A) > 0$ we prove that $\mathcal{H}^s(A) = \infty$ by contradiction. Assume $\mathcal{H}^s(A) < \infty$, then we apply the first part and we have $\mathcal{H}^t(A) = 0$, a contradiction. We have finished.

□

Definition 4.4. *Let $A \subset \mathbb{R}^n$. We define the Hausdorff dimension of A as*

$$dim_{\mathcal{H}}(A) = \inf\{s \in [0, \infty) : \mathcal{H}^s(A) = 0\}.$$

Note that we also have $dim_{\mathcal{H}}(A) = \sup\{t \in [0, \infty) : \mathcal{H}^t(A) > 0\} = \sup\{t \in [0, \infty) : \mathcal{H}^t(A) = \infty\}$.

Observe that we always have $\dim_{\mathcal{H}}(A) \leq n$ for all $A \subset \mathbb{R}^n$ and that if we call $s = \dim_{\mathcal{H}}(A)$ then $\mathcal{H}^t(A) = 0$ for all $t > s$, $\mathcal{H}^t(A) = \infty$ for all $t < s$ and $\mathcal{H}^s(A)$ could take any value between 0 and $+\infty$ (both values included). We state next the following interesting result that appears in Falconer's book [13]: *For every $n \in \mathbb{N}$, every $s \in [0, n]$ and every $t \in [0, \infty]$ there exists a set $A_{t,s} \subset \mathbb{R}^n$ with $dim_{\mathcal{H}}(A) = s$ and $\mathcal{H}^s(A_{t,s}) = t$.*

We finish this chapter by explaining that $\mathcal{H}^n \sim \mathcal{L}^n$ on $\mathbb{R}^n$. This means that there exists a constant $C > 0$ (depending on n) so that

$$C^{-1}\mathcal{L}^n(A) \leq \mathcal{H}^n(A) \leq C\mathcal{L}^n(A) \quad \forall A \subset \mathbb{R}^n.$$

Recall that the Lebesgue measure is defined as

$$\mathcal{L}^n(A) = \inf\left\{\sum_{j=1}^{\infty} \text{vol}(Q_i)^n : Q_i \text{ cubes}, \ A \subset \bigcup_{j \geq 1} Q_j\right\}, \quad A \text{ measurable}.$$

Theorem 4.5. *On $\mathbb{R}^n$ we have $\mathcal{H}^n \sim \mathcal{L}^n$.*

Exercises

1. Let
$$A = \{(x, x\sin(1/x)) : x \in (0,1)\} \subset \mathbb{R}^2$$
Prove that $\mathcal{H}^1(A) = \infty$, but $\mathcal{H}^2(A) = 0$.

2. Prove that for $0 \leq s < n$ then H^s is not a Radon measure in $\mathbb{R}^n$.

3. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be $L-$Lipschitz, $A \subset \mathbb{R}^n$ and $0 \leq s < \infty$. Prove that $\mathcal{H}^s(f(A)) \leq L^s \mathcal{H}^s(A)$.

4. Prove that the ternary Cantor set is a perfect set, that is, all its points are accumulation points.

Chapter 5

Morse-Sard theorem

Whitney in 1934 gave his very important extension theorem. We already saw in detail the case C^1 in the first chapter of these notes. Relying on this result, in 1935 ([27]) Whitney built a function $f : \mathbb{R}^2 \to \mathbb{R}$ of class C^1 so that $\mathcal{L}(f(C_f)) > 0$, where $C_f = \{x \in \mathbb{R}^2 : Df(x) = 0\}$ denotes the set of critical points of f. Namely, the function f is nonconstant on a nonrectifiable curve $\Gamma \subset \mathbb{R}^2$ where $Df(x) = 0$ for all $x \in \Gamma$. Precisely $f(\Gamma) = [0, 1]$.

Why such a "weird" example is possible?

Observe that if we deal with a rectifiable curve $\gamma : [0, 1] \to \mathbb{R}^2$, by taking a Lipschitz parametrization that we name the same way and calling $\Gamma = \{\gamma(t) : t \in [0, 1]\}$, we have that any C^1 function $f : \mathbb{R}^2 \to \mathbb{R}$ so that $Df(x) = 0$ for all $x \in \Gamma$ must be constant on Γ. This follows from the fact that $(f \circ \gamma) : [0, 1] \to \mathbb{R}$ is a Lipschitz function, hence differentiable almost everywhere and the Fundamental Theorem of Calculus holds, that is,

$$
\begin{aligned}
f(\gamma(x)) &= f(\gamma(0)) + \int_0^x (f \circ \gamma)'(t)\, dt = f(\gamma(0)) + \int_0^x \nabla f(\gamma(t)) \cdot \gamma'(t)\, dt \\
&= f(\gamma(0)) + \int_0^x 0 \cdot \gamma'(t)\, dt = f(\gamma(0)) \quad \forall x \in (0, 1].
\end{aligned}
$$

In the years 1939 and 1942, Morse and Sard respectively gave a first explanation of what was going on with this example.

Theorem 5.1 (Morse-Sard, 1942)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function of class C^k, where $k \geq \max\{n - m + 1, 1\}$, then $\mathcal{L}^n(f(C_f)) = 0.$, where $C_f = \{x \in \mathbb{R}^n :$ rank$(Df(x))$ is not maximum$\}$ denotes the set of critical points. (The set $f(C_f)$ is known as the critical values of f.)*

We aim to give a complete proof of this result. However, we prefer to start by explaining its possible refinements and generalizations, as well to some applications to other branches of mathematics.

5.1 Introduction and comments

First thing to mention is that the Morse-Sard theorem is optimal in the scale of spaces C^j. We have already explained Whitney's example, where a function $f : \mathbb{R}^2 \to \mathbb{R}$ is given of class C^1 so that $\mathcal{L}(f(C_f)) > 0$. Due to its importance let us give some details of how this example is constructed.

Example 5.2. For the full construction we refer to original paper of Whitney from 1935, [27].

The idea is to define a nonrectifiable curve (that is a closed set $C \subset \mathbb{R}^2$) and a function $f : C \to \mathbb{R}$ so that

1. $f(C)$ has positive measure. In particular we will get $f(C) = [0,1]$.
2. $\lim_{y\to x} \frac{|f(y)-f(x)-0|}{|y-x|} = 0$ uniformly. In particular f is uniformly continuous in C.

Using (ii) and thinking of the "derivative" of f on C to be equal to the linear map $L = 0$, by applying the Whitney's extension theorem 1.8 there exists a function $F : \mathbb{R}^2 \to \mathbb{R}$ of class C^1 such that $F|_C = f$ and $Df|_C = L = 0$. Hence $C \subset C_F$ and therefore

$$\mathcal{L}(F(C_F)) \geq \mathcal{L}(F(C)) \overset{(i)}{=} \mathcal{L}([0,1]) = 1.$$

After Whitney's example, many mathematicians have tried to give *easier* examples showing this type of pathological behaviour on a function. For instance, Gringberg presented in 1985 another example of this kind (see [15]).

Observe that for any function $f : \mathbb{R}^n \to \mathbb{R}^m$ for which the More-Sard theorem applies $\mathcal{L}^m$-almost every point $y \in \mathbb{R}^m$ satisfies that $f^{-1}(y)$ consist of regular values (those are points $x \in \mathbb{R}^n$ where $Df(x)$ has maximum rank). There exist many generalizations of the Morse-Sard theorem to other classes of functions. Let us enumerate some of these generalizations:

• **[Bates, 1993]**: If $f \in C^{k-1,1}(\mathbb{R}^n;\mathbb{R}^m)$ with $k \geq \max\{n-m+1,1\}$ then $\mathcal{L}^m(f(C_f)) = 0$.

Note that for any $\alpha \in (0,1)$ there exist functions $f : \mathbb{R}^n \to \mathbb{R}^m$ of class $C^{k-1,\alpha}$ with $k \geq \max\{n-m+1,1\}$ but $\mathcal{L}^m(f(C(f)) > 0$.

• **[De Pascale, 2001]**: If $f \in W^{k,p}(\mathbb{R}^n;\mathbb{R}^m)$ with $p > n$ and $k \geq \max\{n-m+1,1\}$ then $\mathcal{L}^m(f(C_f)) = 0$.

We need to precise how one should understand the set of critical points of a Soblev function. In general in these context one defines the set of critical points as

$$C_f = \{x \in \mathbb{R}^n : \ Df(x) \text{ exists and } Df(x) \text{ has not maximum rank}\}.$$

Still one may wonder what happens with the set of points where f is not differentiable. Is the set of non-differentiability points sent to a $\mathcal{L}^m$-null set? Because if that is the case one can still assure that for $\mathcal{L}^m$-almost every point $y \in \mathbb{R}^m$, $f^{-1}(y)$ consist of regular values. In the conditions of De Pascale theorem this is the case. Thanks to the Morrey inequality, if $p > n$,

$$W^{k,p}(\mathbb{R}^n;\mathbb{R}^m) \subset C^{k-1,1-n/p}(\mathbb{R}^n;\mathbb{R}^m).$$

In the case $k \geq 2$ we have $Df(x)$ is well defined everywhere. In the case $k = 1$ we have that functions in $W^{1,p}(\mathbb{R}^n;\mathbb{R}^m)$ satisfy what is called the Lusin N-property, that is if $\mathcal{L}^n(N) = 0$ then $\mathcal{H}^n(f(N)) = 0$. Using that for the case $k = 1$ we have $m \geq n$, we conclude that $\mathcal{L}^m(f(N)) = 0$.

•**[Bourgain, Korobkov, Kristensen, 2014]**: If $f \in W^{n,1}(\mathbb{R}^n;\mathbb{R})$ then f is differentiable at $\mathcal{H}^1$-almost every point and $\mathcal{L}(f(C_f)) = 0$ where $C_f = \{x \in \mathbb{R}^n : \ Df(x) \text{ exists and } Df(x) = 0\}$.

Again in this case it is known that functions $f \in W^{n,1}(\mathbb{R}^n;\mathbb{R})$ have the Lusin N property with respect to the $\mathcal{H}^1$ measure. That is if $\mathcal{H}^1(N) = 0$ then $\mathcal{H}(f(N)) = 0$.

5.2 Proofs of the Morse-Sard theorem

We now give the proof of the Morse-Sard theorem by distinguishing three cases

$$n = m \quad ; \quad n < m \quad ; \quad n > m.$$

We start with the case $n = m$, which follows as a corollary from the next more general result (recall that C^1 functions are locally Lipschitz).

Theorem 5.3. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz. Then $\mathcal{L}^n(f(C_f)) = 0$ where $C_f = \{x \in \mathbb{R}^n : \ Df(x)$ exists and rank$(D(f(x)) \leq n-1\}$.*

Remark 5.4. A locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable $\mathcal{L}^n$-almost everywhere by Rademacher theorem. Then if

$$G = \{x \in \mathbb{R}^n : Df(x) \text{ does not exist}\}$$

we have $\mathcal{L}^n(G) = 0$. Moreover the local Lipschitzianity of f also implies that $\mathcal{L}^n(f(G)) = 0$. This is usually referred as the Lusin N-property, and it is easily valid for locally Lipschitz functions by dividing G into countable many pieces G_i where f is L_i-Lipschitz and using that $\mathcal{L}^n(f(G_i)) \leq L_i^n \mathcal{L}^n(G_i) = 0$.

Proof. Let $\mathbb{R}^n = \bigcup_{j=1}^{\infty} Q_j$, where Q_j are open cubes of sidelenght 1, not necessarily disjoint. We have that for every $j \in \mathbb{N}$, $f|_{5Q_j}$ is L_j-Lipschitz for some $L_j > 0$. We will prove that for every $j \in \mathbb{N}$ we have

$$\mathcal{L}^n(f(C_f \cap Q_j)) = 0$$

and this will enough to conclude the proof by the subadditivity of the Lebesgue measure $\mathcal{L}^n$.

Let us fix $j \in \mathbb{N}$ and let us take $x \in C_f \cap Q_j$. By the differentiability of f at x, for a given $\varepsilon \in (0, L_j)$ there exists $r_x > 0$ so that

$$\begin{cases} B(x, r_x) \subset Q_j \\ |f(y) - f(x) - Df(x)(y - x)| < \varepsilon r_x, \quad \forall y \in B(x, 5r_x) \end{cases}.$$

Call $k = \text{rank}(Df(x)) \leq n - 1$ the rank of the matrix $Df(x)$ and denote by

$$W_x = f(x) + Df(x)(\mathbb{R}^n)$$

the k-dimensional affine subspace passing through $f(x)$ and generated by the subspace $Df(x)(\mathbb{R}^n)$. By the previous properties we have

$$\text{dist}(f(y), W_x) < \varepsilon r_x, \quad \forall y \in B(x, 5r_x). \tag{5.1}$$

By using (5.1) and assuming that $f|_{5Q_j}$ is L_j-Lipschitz (we call $L = L_j$ from now on to ease notation) we have that

$$f(B(x, 5r_x)) \subset B(f(x), 5Lr_x) \cap \{z \in \mathbb{R}^n : \text{dist}(z, W_x) < \varepsilon r_x\}. \tag{5.2}$$

Fact: For some given radii $0 < r < R$, the k-dimensional ball $B(0, R) \subset \mathbb{R}^k$ can be covered by less than $C(k)(R/r)^k$ balls of radius r, where $C(k)$ is a constant only depending on the dimension k. (We leave the proof of this fact as an exercise for the reader).

Using the previous fact, the k-dimensional ball $B(f(x), 5Lr_x) \cap W_x$ can be covered by $C(k)\left(\frac{5Lr_x}{\varepsilon r_x}\right)^k$ balls of radius $\varepsilon r_x > 0$. Take $C(n) = \max\{C(1), \dots, C(n-1)\}$. Since $k \in \{1, \dots, n-1\}$ and $\varepsilon < L$, hence $5L/\varepsilon > 1$, we can assure that

$$B(f(x), 5Lr_x) \cap W_x \text{ can be covered by } C(n)\left(\frac{5L}{\varepsilon}\right)^{n-1} \text{ balls of radius } \varepsilon r_x.$$

If we double the radii of those balls and we see them with the same centers but now living in $\mathbb{R}^n$ instead of living on a k-dimensional subspace, we can conclude that

$$B(f(x), 5Lr_x) \cap \{z \in \mathbb{R}^n : \text{dist}(z, W_x) < \varepsilon r_x\}$$

can be covered by $C(n)\left(\frac{5L}{\varepsilon}\right)^{n-1}$ balls of radius $2\varepsilon r_x$. And the same covering works for the set $f(B(x, 5r_x))$ by (5.2). We get that

$$\mathcal{H}^n_\infty(f(B(x, 5r_x)) \leq C(n)\left(\frac{5L}{\varepsilon}\right)^{n-1} \cdot \omega_n 2^n \varepsilon^n r_x^n = C(n)\varepsilon r_x^n. \tag{5.3}$$

We now write $C_f \cap Q_j = \bigcup_{x \in C_f \cap Q_j} B(x, r_x)$ and by using Vitali's covering lemma (Theorem 1.13) there exists a subfamily $\{B(x_j, r_j)\}_{j \geq 1}$ of disjoint balls so that

$$C_f \cap Q_j \subset \bigcup_{j \geq 1} B(x_j, 5r_j).$$

We can write, by using (5.3) and that $\bigcup_{j \geq 1} B(x_j, r_j) \subset Q_j$ is a disjoint union,

$$\begin{aligned}\mathcal{H}^n_\infty(f(C_f \cap Q_j)) &\leq \sum_{j=1}^{\infty} \mathcal{H}^n_\infty(f(B(x_j, 5r_j))) \leq \sum_{j=1}^{\infty} C(n)\varepsilon r_j^n = C(n)\varepsilon \sum_{j=1}^{\infty} \omega_n r_j^n \\ &= C(n)\varepsilon \sum_{j=1}^{\infty} \mathcal{L}^n(B(x_j, r_j)) = C(n)\varepsilon \mathcal{L}^n\left(\bigcup_{j \geq 1} B(x_j, r_j)\right) \\ &\leq C(n)\varepsilon \mathcal{L}^n(Q_j) = C(n)\varepsilon.\end{aligned}$$

Since $\varepsilon \in (0, L)$ was arbitrary we have that $\mathcal{H}^n_\infty(f(C_f \cap Q_j)) = 0$. In particular this implies that $\mathcal{H}^n(f(C_f \cap Q_j)) = 0$ and since $\mathcal{H}^n = \mathcal{L}^n$ on $\mathbb{R}^n$ we conclude that $\mathcal{L}^n(f(C_f \cap Q_j)) = 0$. We are done. □

The case $n < m$ is much easier.

Theorem 5.5. *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz with $n < m$ then $\mathcal{L}^m(f(A)) = 0$ for every set $A \subset \mathbb{R}^n$. In particular, $\mathcal{L}^m(f(C_f)) = 0$ where $C_f = \{x \in \mathbb{R}^n : Df(x)$ exists and rank$(Df(x)) \leq n - 1\}$.*

Proof. Let $A \subset \mathbb{R}^n$. We cover A by balls $A \subset \bigcup_{j \geq 1} B_j$ and we let $L_j \geq 0$ be the Lipschitz constant of f in B_j. We have that for every $j \in \mathbb{N}$ and $0 \leq s < \infty$,

$$\mathcal{H}^s(f(A \cap B_j)) \leq (L_j)^s \mathcal{H}^s(A \cap B_j).$$

Then if $d = \dim_{\mathcal{H}}(A) \leq n$ we have that $\mathcal{H}^s(A) = 0$ for every $s > d$. Recall that $\dim_{\mathcal{H}}(A) = \inf\{s \geq 0 : \mathcal{H}^s(A) = 0\}$. Hence for every $s > d$,

$$\mathcal{H}^s(f(A)) \leq \sum_{j=1}^{\infty} \mathcal{H}^s(f(A \cap B_j)) \leq \sum_{j=1}^{\infty} (L_j)^s \mathcal{H}^s(A \cap B_j) = 0.$$

Then $\dim_{\mathcal{H}}(f(A)) \leq d = \dim_{\mathcal{H}}(A) \leq n$. We conclude that $\mathcal{H}^m(f(A)) = 0$ and since $\mathcal{H}^m = \mathcal{L}^m$ on $\mathbb{R}^m$, $\mathcal{L}^m(f(A)) = 0$. □

The case $n > m$ is more involved. For the proof an essential tool is the following result, which is known in the literature as *critically Morse lemma*.

Lemma 5.6 (Critically Morse lemma)**.** *Let $k \in \mathbb{N}$, $A \subset \mathbb{R}^n$. Then we can write $A = \bigcup_{j=0}^{\infty} A_j$ so that A_0 is countable, each A_j with $j \geq 1$ is bounded and without isolated points, and if $g : \mathbb{R}^n \to \mathbb{R}$ is a function of class C^k such that $A \subset C_g = \{x \in \mathbb{R}^n : \nabla g(x) = 0\}$ then for every $j \geq 1$ and every $x \in A_j$,*

$$\lim_{\substack{y \to x \\ y \in A_j}} \frac{|g(y) - g(x)|}{|y - x|^k} = 0. \tag{5.4}$$

Proof. For this proof we refer to [22]. □

Now we prove the next result that only deals with image of the set of critical points where the differential has rank zero. As a corollary one gets the classical Morse theorem from 1939 for real-valued functions $f : \mathbb{R}^n \to \mathbb{R}$ of class C^n.

Theorem 5.7. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ with $n > m$ of class C^k where $k = n - m + 1 \geq 2$. Let $A = \{x \in \mathbb{R}^n : Df(x) = 0\} = \{x \in \mathbb{R}^n : rank(D(f(x)) = 0\}$. Then $\mathcal{L}^m(f(A)) = 0$.*

Proof. By the critically Morse Lemma 5.6 we decompose the set A as $A = \bigcup_{j=0}^{\infty} A_j$ with A_j having the aforementioned properties. It is enough to check that

$$\mathcal{L}^m(f(A_j)) = 0 \quad \text{for every} j \geq 0.$$

First, as A_0 is countable, $f(A_0)$ is countable and then $\mathcal{L}^m(f(A_0)) = 0$. Fix now $j \geq 1$. By the boundedness of A_j there exists $R > 0$ so that $A_j \subset B(0, R)$ for some $R > 0$. Take $\varepsilon > 0$ arbitrary. By (5.4), for each $x \in A_j$ and since every component f_i, $i = 1, \dots, m$, is of class $C^k(\mathbb{R}^n; \mathbb{R})$ we have

$$\lim_{\substack{y \to x \\ y \in A_j}} \frac{|f_i(y) - f_i(x)|}{|y - x|^k} = 0.$$

Then

$$\lim_{\substack{y \to x \\ y \in A_j}} \frac{|f(y) - f(x)|}{|y - x|^k} = 0.$$

Hence, for the given $\varepsilon > 0$ there exists $0 < r_x < 1$ so that

$$|f(y) - f(x)| \leq \varepsilon |y - x|^k, \quad \forall y \in B(x, r_x) \cap A_j.$$

If we now take $y, z \in B(x, r_x) \cap A_j$ and use the triangle inequality

$$|f(y) - f(z)| \leq \varepsilon |y - x|^k + \varepsilon |y - z|^k \leq 2\varepsilon (r_x)^k.$$

This yields that $\operatorname{diam}(f(A_j \cap B(x, r_x))) \leq 2\varepsilon (r_x)^k$. And by covering the set by just one single cube with sidelenght the diameter of the set itself we get

$$\mathcal{H}^m_\infty(f(A_j \cap B(x, r_x))) \leq 2^m \varepsilon^m (r_x)^{km} \tag{5.5}$$

Since we can write $A_j \subset \bigcup_{x \in A_j} B(x, r_x)$ by Vitali's Lemma 1.13 there is a subfamily $\{B(x_i, r_i)\}_{i \in \mathbb{N}}$ so that $A_j \subset \bigcup_{i \in \mathbb{N}} B(x_i, r_i)$ and moreover the balls $\{B(x_i, r_i/5)$ are pairwise disjoint. Hence by using (5.5), that $(n - m + 1)m \geq n$ and noting that $\bigcup_{i \geq 1} B(x_i, r_i/5) \subset B(0, R + 1/5)$,

$$\begin{aligned}
\mathcal{H}^m_\infty(f(A_j)) &\leq \sum_{i=1}^{\infty} \mathcal{H}^m_\infty(f(B(x_i, r_i)) \leq \sum_{i=1}^{\infty} 2^m \varepsilon^m (r_i)^{km} \leq 2^m \varepsilon^m \sum_{i=1}^{\infty} (r_i)^n = C(n) \varepsilon^m \sum_{i=1}^{\infty} \omega_n \left(\frac{r_i}{5}\right)^n \\
&= C(n,m) \varepsilon^m \sum_{i=1}^{\infty} \mathcal{L}^n \left(B(x_i, r_i/5)\right) = C(n,m) \varepsilon^m \mathcal{L}^n \left(\bigcup_{i \geq 1} B(x_i, r_i/5)\right) \\
&\leq C(n,m) \varepsilon^m \mathcal{L}^n(B(0, R + 1/5)) = C(n,m,R) \varepsilon^m.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary we get that $\mathcal{H}^m_\infty(f(A_j)) = 0$. Then $\mathcal{H}^m(f(A_j)) = 0$ and since $\mathcal{H}^m = \mathcal{L}^m$ on $\mathbb{R}^m$ we conclude that $\mathcal{L}^m(f(A_j)) = 0$ and the proof is complete.

□

Finally it remains to proof the general case, where we deal with the total set of critical of points. We follow the classical approach of Sard in 1942, which heavily relies on Theorem 5.7.

Theorem 5.8. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ with $n > m$ of class C^k where $k = n-m+1 \geq 2$. Then the for the sets $C_j = \{x \in \mathbb{R}^n : , rank(Df(x)) = j\}$, $j = 0, 1 \ldots, m-1$ satisfy that $\mathcal{L}^m(f(C_j)) = 0$. In particular since $C_f = \bigcup_{j=0}^{m-1} C_j$ we get that $\mathcal{L}^m(f(C_f)) = 0$.*

Proof. The fact that $\mathcal{L}^m(f(C_0)) = 0$ follows directly from Theorem 5.7. For the rest of the cases we will perform a kind of reduction in order to use Theorem 5.7 again.

Let us fix $j = 1, \ldots, m-1$. We leave as an exercise to check that C_j is a measurable set (indeed an intersectipn between a closed and open set). Fix $x_0 \in C_j$ and we will prove the existence of some $r_0 > 0$ so that $\mathcal{L}^m(f(B(x_0, r_0) \cap C_j))) = 0$. This will enough by using the separability of $\mathbb{R}^n$ and the countable subadditivity of the Lebesgue measure $\mathcal{L}^m$.

Since $x_0 \in C_j$ we have that rank$(Df(x_0)) = j \in \{1, \ldots, m-1\}$, so there is a minor of order j which is non null. Reordering the variables and components of f if necessary we can write

$$f(x) = (f_1(x_1, \ldots, x_n), \ldots f_m(x_1, \ldots, x_n))$$

such that

$$det\left(\frac{\partial(f_1, \ldots, f_j)}{\partial(x_1, \ldots, x_j)}(x_0)\right) = det\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_j}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_j}{\partial x_1}(x_0) & \cdots & \frac{\partial f_j}{\partial x_j}(x_0) \end{pmatrix} \neq 0.$$

We define the function $h : \mathbb{R}^n \to \mathbb{R}^n$ as

$$h(x) = h(x_1, \ldots, x_n) = (f_1(x), \ldots, f_j(x), x_{j+1}, \ldots, x_n).$$

We have that $h \in C^k(\mathbb{R}^n; \mathbb{R}^n)$ and moreover $\det(Dh(x_0)) \neq 0$, so by using the inverse function theorem there exists $r_0 > 0$ so that $h|_{B(x_0,r_0)}$ is a C^k-diffeomorphism. Let us call $V = h(B(x_0, r_0)) \subset \mathbb{R}^n$, which is an open set. It is clear that $h(B(x_0, r_0) \cap C_j) = V \cap h(C_j)$. Define now the function $F = f \circ h^{-1} : V \to \mathbb{R}^m$, which can be written as

$$F(y) = F(y_1, \ldots, y_n) = (y_1, \ldots, y_j, g(y_1, \ldots, y_n)), \quad \forall y \in V.$$

for some function $g : V \subset \mathbb{R}^n \to \mathbb{R}^{m-j}$ of class C^k.

Observe that we are done if we prove that $\mathcal{L}^m(F(V \cap h(C_j))) = 0$ because

$$\mathcal{L}^m(f(B(x_0, r_0) \cap C_j)) = \mathcal{L}^m(f(h^{-1}(V \cap h(C_j))) = \mathcal{L}^m(F(V \cap h(C_j))) = 0.$$

For the functions F and g we have the following properties:

1. By the chain rule, for every $y \in V$,

$$DF(y) = Df(h^{-1}(y))Dh^{-1}(y) = \begin{pmatrix} Id_j & 0 \\ * & D(g|_{(y_1,\ldots,y_j)})(y_{j+1}, \ldots, y_n) \end{pmatrix},$$

where $g|_{(y_1,\ldots,y_j)} : V_{(y_1,\ldots,y_j)} \to \mathbb{R}^{m-j}$ is defined by $g|_{(y_1,\ldots,y_j)}(y_{j+1}, \ldots, y_n) = g(y_1, \ldots, y_n)$, with $V_{(y_1,\ldots,y_j)} = \{(z_1, \ldots, z_{n-j}) \in \mathbb{R}^{n-j} : (y_1, \ldots, y_j, z_1, \ldots, z_{n-j}) \in V\}$.
Moreover we fix some more notation: $\pi_j : \mathbb{R}^n \to \mathbb{R}^j$ denotes the projection onto the first j-coordinates and $\pi^{n-j} : \mathbb{R}^n \to \mathbb{R}^{n-j}$ denotes the projection onto the last $(n-j)$ coordinates.

2. h is a diffeomorphism with $Dh^{-1}(y) = (Dh(x))^{-1}$ for every $y = h(x) \in V$, and by item (1) we have that

$$DF(y) = Df(h^{-1}(y))Dh^{-1}(y), \quad \forall y \in V \quad \Rightarrow \quad Df(x) = DF(h(x))Dh(x), \quad \forall x \in B(x_0, r_0).$$

Since $\det(Dh(x)) \neq 0$ we have that the rank of $DF(h(x))$ is the same as the rank of $Df(x)$ for every $x \in B(x_0, r_0)$. In particular, if $x \in B(x_0, r_0) \cap C_j$ then $h(x) = y$ is a critical point of F with $DF(y)$ having rank j. And by the expression of $DF(y)$ in item (1) it is clear that for every $y \in V \cap h(C_j)$

$$\text{rank}(DF(y)) = j \quad \Leftrightarrow \quad (D(g|_{(y_1,\ldots,y_j)})(y_{j+1}, \ldots, y_n) = 0.$$

In other words for a given fixed $(y_1, \ldots, y_j) \in \pi_j(V \cap h(C_j))$ we have that

$$(Dg|_{(y_1,\ldots,y_j)})(y_{j+1}, \ldots, y_n) = 0 \quad \Leftrightarrow \quad (y_{j+1}, \ldots, y_n) \in V_{(y_1,\ldots,y_j)} \cap \pi^{n-j}(h(C_j)).$$

And the previous situation occurs if and only if $x = h^{-1}(y) \in B(x_0, r_0)$ is a critical point of f with $Df(x)$ having rank j.

Due to the fact that for any given $(y_1, \ldots, y_j) \in \pi_j(V \cap h(C_j))$ the function

$$g|_{(y_1,\ldots,y_j)} : V_{(y_1,\ldots,y_j)} \subset \mathbb{R}^{n-j} \to \mathbb{R}^{m-j}$$

is of class C^k and

$$V_{(y_1,\ldots,y_j)} \cap \pi^{n-j}(V \cap h(C_j)) = \{(y_{j+1}, \ldots, y_n) \in V_{(y_1,\ldots,y_j)} : (Dg|_{(y_1,\ldots,y_j)}(y_{j+1}, \ldots, y_n) = 0\},$$

then by applying Theorem 5.7 we obtain that

$$\mathcal{L}^{m-j}(g|_{(y_1,\ldots,y_j)})(V_{(y_1,\ldots,y_j)} \cap \pi^{n-j}(h(C_j))) = 0 \tag{5.6}$$

for every $(y_1, \ldots, y_j) \in \pi_j(V \cap h(C_j))$. We finish the proof by using Fubini's theorem (note that the set $F(V \cap h(C_j))$ is measurable and that F fixes the first j coordinates).

$$\mathcal{L}^m\left(F(V \cap h(C_j))\right) = \int_{F(V \cap h(C_j))} 1 \, dy_1 \ldots dy_j dz_{j+1} \ldots dz_m =$$
$$= \int_{\pi_j(V \cap h(C_j))} \left(\int_{g|_{(y_1,\ldots,y_j)}\left(V_{(y_1,\ldots,y_j)} \cap \pi^{n-j}(h(C_j))\right)} 1 \, dz_{j+1} \ldots dz_m \right) dy_1 \ldots dy_j =$$
$$= \int_{\pi_j(V \cap h(C_j))} \mathcal{L}^{m-j}(g|_{(y_1,\ldots,y_j)})(V_{(y_1,\ldots,y_j)} \cap \pi^{n-j}(h(C_j))) \, dy_1 \ldots dy_j = 0.$$

□

It is clear now that Theorem 5.3, Theorem 5.5 and Theorem 5.8 give a complete proof of the classical Morse-Sard theorem (Theorem 5.1).

5.3 Morse-Sard theorem in infinite-dimensional Banach spaces

The next question is:

What happens if we work with infinite dimensional Banach spaces X?

Here, unfortunately one does not have the validity of the Morse-Sard theorem. That is, there exists Banach spaces X and C^∞ smooth functions $f : X \to \mathbb{R}$ whose set of critical values has positive measure. The firs one to give such an example was Kupka in [20]. We present the following *easier* example from Bates and Moreira, contained in [8].

Example 5.9. Let $f : \ell_2 \to \mathbb{R}$ be the following polynomial of degree 3 (hence of class C^∞).

$$f\left(\sum_{n=1}^{\infty} x_n e_n\right) = \sum_{n=1}^{\infty}(3 \cdot 2^{-\frac{n}{3}} x_n^2 - 2x_n^3).$$

The function f satisfies the next properties, whose verification is left to the reader.

- For a given $x \in \ell_2$, $Df(x) \in \ell_2^*$ is written as

$$Df(x) = \sum_{n=1}^{\infty} \left(6 \cdot 2^{-\frac{n}{3}} x_n - 6x_n^2\right) e_n.$$

- We have that $Df(x) = 0$ if and only if $x_n(2^{-\frac{n}{3}} - x_n) = 0$ for all $n \in \mathbb{N}$. Therefore we have

$$C_f = \{\sum_{n=1}^{\infty} x_n e_n : x_n \in \{0, 2^{-\frac{n}{3}}\}\}$$

- We have that $f(C_f) = [0, 1]$. On the one hand given $x = (x_n)_{n\geq 1} \in C_f$ clearly $f(x)$ is an infinite sum of positive terms and

$$f(x) = \sum_{n=1}^{\infty}(3 \cdot 2^{-\frac{n}{3}} x_n^2 - 2x_n^3) \leq \sum_{n=1}^{\infty}(3 \cdot 2^{-\frac{n}{3}} 2^{-2n/3} - 22^{-n}) = \sum_{n=1}^{\infty} 2^{-n} = 1.$$

  On the other hand let $t \in [0, 1]$. Then there exists a unique sequence $(y_n)_{n\geq 1} \subset \{0, 1\}$ such that $t = \sum_{n\geq 1} y_n 2^{-n}$. Now let $x = (x_n)_{n\geq 1}$ be defined as $x_n = 0$ if $y_n = 0$ and $x_n = 2^{-n/3}$ if $y_n = 1$. In this way we have $x \in C_f$ and

$$f(x) = \sum_{n=1}^{\infty}(3\cdot 2^{-\frac{n}{3}} x_n^2 - 2x_n^3) = \sum_{\{n\in\mathbb{N}:\, y_n=1\}} (3\cdot 2^{-\frac{n}{3}} 2^{-2n/3} - 22^{-n}) = \sum_{\{n\in\mathbb{N}:\, y_n=1\}} 2^{-n} = \sum_{n=1}^{\infty} y_n 2^{-n} = t.$$

- Finally we have that $f \in C^\infty(\ell_2; \mathbb{R})$.

Exercises

1. Related to Grinberg's example [15]. Let $C \subset [0,1]$ be the ternary Cantor set. Build explicitly a function $f : \mathbb{R} \to \mathbb{R}$ that is of class C^1 and so that $C \subset f(C_f)$. Justify rigorously the C^1 smoothness of f once given.

2. Prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is of class C^n then f is constant on every connected component of $C_f = \{x \in \mathbb{R}^n : \nabla f(x) = 0\}$.

3. Let $B(0,R) \subset \mathbb{R}^n$ be the n-dimensional open ball of radius $R > 0$. Prove that given $0 < r < R$, there exists a constant $C(n)$ that only depends on the dimension n in such a way that with a number $C(n)(R/r)^n$ of open balls of radii $r > 0$ we can cover the whole ball $B(0,R)$.

4. Related to the Morse-Sard theorem, prove that if $n > m$, $f : \mathbb{R}^n \to \mathbb{R}^m$ is of class C^q and for some $s < m$ we have $qs \geq n$ then $\mathcal{H}^s(f(A)) = 0$, where $A = \{x \in \mathbb{R}^n : Df(x) = 0\}$.

5. Given the function $f(x) = |x|$, $x \in \mathbb{R}$, is it possible to find a function $g : \mathbb{R} \to \mathbb{R}$ of class C^1 with $|g(x) - f(x)| < 1/2$ for all $x \in \mathbb{R}^2$ and g not having critical points? Secondly, provide a function $h \in C^\infty(\mathbb{R};\mathbb{R})$ with only one critical point so that $|h(x) - f(x)| < 1/2$ for all $x \in \mathbb{R}$.

6. Let $\Gamma = \{t(\cos(t), \sin(t)) : t \in [0, 2\pi]\}$. Prove that it is not possible to define a function $f : \mathbb{R}^2 \to \mathbb{R}$ of class C^1 so that $\nabla f(x) = 0$ for all $x \in \Gamma$ and with $\mathcal{L}^1(f(\Gamma)) > 0$.

7. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be of class C^1 with $m \leq n$. prove that for every $j = 0, 1 \ldots, m-1$ the sets $\{x \in \mathbb{R}^n : \text{rank}(Df(x)) \leq j\}$ are closed.

Chapter 6

Alexandrov theorem

The objective of this chapter is to prove the following result:

Theorem 6.1 (Alexandrov, 1939)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then f is differentiable at almost every point $x \in \mathbb{R}^n$, and at almost every point $x \in \mathbb{R}^n$ where f is differentiable there exists a symmetric matrix, which we denote by $D^2 f(x)$, such that*

$$\lim_{y \to x} \frac{f(y) - f(x) - Df(x)(y-x) - \frac{1}{2}(y-x)^T D^2 f(x)(y-x)}{|y-x|^2} = 0. \tag{6.1}$$

This result was originally proved in [2] (1939) by A. D. Alexandrov, a Russian mathematician and physicist. To this day, there exist several alternative methods for its proof. We just mention [6], where one can consult a description of different approaches that exist in the literature. The proof we include here is very recent, published in 2023 by Daniel Azagra, Anthony Capello, and Piotr Hajłasz in [3]. This proof is purely geometric and, in a certain sense, is expected to be relatively accessible to anyone with a basic background in Analysis. The method followed in [3] makes crucial use of the fact that convex functions satisfy Lusin properties of class $C^{1,1}$. This is a parallel result, which is of interest in its own right, and whose original proof appears in [5]. More specifically, we will prove that for every convex function $f : \mathbb{R}^n \to \mathbb{R}$, every $R > 0$, and every $\varepsilon > 0$, there exists a convex function $g \in C^{1,1}(B(0,R))$ such that

$$\mathcal{L}^n\big(\{x \in B(0,R) : f(x) \neq g(x)\}\big) < \varepsilon.$$

The study of Lusin properties for Lipschitz functions or convex functions could very well be the subject of another chapter. Here we only mention that it has long been known (see [1]) that for every convex function $f : \mathbb{R}^n \to \mathbb{R}$ and every $\varepsilon > 0$ there exists $g \in C^2(\mathbb{R}^n)$ such that

$$\mathcal{L}^n\big(\{x : f(x) \neq g(x)\}\big) < \varepsilon.$$

However, until 2024, it was not known whether the function g could also be chosen to be convex. This question is answered affirmatively by the recent article [4] for strongly convex functions f. As we already indicated, for the purposes of this work we only need that g can be taken in $C^{1,1}$.

All convex functions are locally Lipschitz; hence the first part of the statement of Alexandrov's theorem follows from Rademacher's theorem. On the other hand, we must point out that (6.1) does not imply that the function f is twice differentiable in the usual sense, since our function need not have a differential defined at every point. What is true is that if f is convex and differentiable at every point, then f will indeed be twice differentiable at almost every point (see Corollary 6.23).

The proof of the Alexandrov theorem and some corollaries appear in Section 5.3. The proof relies heavily on Lemma 6.9 and Theorem 6.10. The first result is developed in Section 5.1 and the second one is developed in Section 5.2.

6.1 Some topics about convex functions and subdifferentials

Let us start by introducing four key notions of the theory of Convex Analysis; that is, convex sets, convex hulls, epigraph of a function, and the subdifferential.

Definition 6.2. *1. A set $C \subset \mathbb{R}^n$ is called a convex set if for every $x, y \in C$ and every $t \in [0,1]$ we have that $tx + (1-t)y \in C$.*

2. Given a set $A \subset \mathbb{R}^n$, its convex hull $co(A)$ is defined as the intersection of all convex sets containing A. Equivalently,

$$co(A) = \{\lambda x + (1-\lambda)y \in \mathbb{R}^n \ : \ x, y \in A, \ \lambda \in [0,1]\}.$$

3. A function $f : C \subset \mathbb{R}^n \to \mathbb{R}$, where C is convex, is called convex if for every $x, y \in C$ and $t \in [0,1]$ we have that

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y).$$

4. Given a function $f : U \subset \mathbb{R}^n \to \mathbb{R}$, its epigraph is defined as the set

$$epi(f) := \{(x,t) \in U \times \mathbb{R} : x \in U, \ t \ge f(x)\}.$$

Lemma 6.3. *Let $C \subset \mathbb{R}^n$ be a nonempty convex set so that $C \neq \mathbb{R}^n$. Then, for every $x \in \partial C$ there exists $v \in \mathbb{R}^n$ and $b \in \mathbb{R}$ so that either*

$$C \subset \{y \in \mathbb{R}^n : \ \langle y, v\rangle \ge b\} \quad or \quad C \subset \{y \in \mathbb{R}^n : \ \langle y, v\rangle \le b\}$$

and moreover $\langle x, v\rangle = b$.

From now on, we will call $T_x = C \subset \{y \in \mathbb{R}^n : \ \langle y, v\rangle = b\}$ a supporting hyperplane of C at x.

Proof. The proof is left as an exercise to the reader. □

Lemma 6.4. *If $U \subset \mathbb{R}^n$ is a convex set, the epigraph $epi(f)$ of a convex function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is a convex subset of $\mathbb{R}^{n+1}$.*

Proof. Let $\lambda \in [0,1]$ and let $(x_1, t_1), (x_2, t_2) \in \text{epi}(f)$. Since U is convex,

$$\lambda x_1 + (1-\lambda)x_2 \in U,$$

and since f is convex on U,

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda)f(x_2) \le \lambda t_1 + (1-\lambda)t_2.$$

Therefore,

$$\lambda(x_1, t_1) + (1-\lambda)(x_2, t_2) = (\lambda x_1 + (1-\lambda)x_2, \lambda t_1 + (1-\lambda)t_2) \in \text{epi}(f).$$

□

Definition 6.5. *Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}$. The (Fréchet) subdifferential of f at a point $x \in U$ is defined as the set of vectors $v \in \mathbb{R}^n$ so that*

$$\liminf_{y \to x} \frac{f(y) - f(x) - \langle v, y - x\rangle}{|y - x|} \ge 0.$$

Each element $v \in \partial f(x)$ is called a subgradient of f at x. Observe that for any open neighborhood $V \subset U$ of x we have $\partial f(x) = \partial(f|_V)(x)$.

The notion of subdifferential generalizes the usual concept of differential, and furthermore, in the case where a convex function f is differentiable, the gradient is the only subgradient.

Proposition 6.6. *If $U \subset \mathbb{R}^n$ is an open set and $f : U \to \mathbb{R}$ is differentiable at $x \in U$, then $\partial f(x) = \{\nabla f(x)\}$.*

Proof. By the definition of differential, it is obvious that $\nabla f(x) \in \partial f(x)$. On the other hand, pick $v \in \mathbb{R}^n$ so that

$$\liminf_{y \to x} \frac{f(y) - f(x) - \langle v, y - x\rangle}{|y - x|} \geq 0.$$

Then, it is easy to check that

$$\liminf_{y \to x} \frac{\langle v - \nabla f(x), y - x\rangle}{|y - x|} \geq 0.$$

We will prove that $w := v - \nabla f(x) = 0$ by contradiction. In the case $w \neq 0$,

$$\liminf_{t \to 0^+} \frac{\langle v - \nabla f(x), -tw\rangle}{|-tw|} = \liminf_{t \to 0^+} \frac{\langle w, -tw\rangle}{|-tw|} = \liminf_{t \to 0^+} \frac{-t}{|t|} = -1 \geq 0,$$

which is absurd. □

For the particular case of convex functions we have the following result.

Lemma 6.7. *If $U \subset \mathbb{R}^n$ is open and convex and $f : U \to \mathbb{R}$ is a convex function, we have*

$$\partial f(x) := \{v \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x\rangle \text{ for all } y \in U\}.$$

Hence, if $v \in \partial f(x)$ then $s = (v, 1) \in \mathbb{R}^{n+1}$ determines a supporting hyperplane of epi(f) at $(x, f(x))$, which means that

$$\begin{aligned} \text{epi}(f) &\subset \{(y,t) \in U \times \mathbb{R} : \langle (y,t), (v,-1)\rangle \geq \langle (x, f(x)), (v, -1)\rangle\} \\ &= \{(y,t) \in U \times \mathbb{R} : t \leq f(x) + \langle y - x, v\rangle\} \end{aligned}$$

Proof. First, take $v \in \mathbb{R}^n$ so that

$$f(y) \geq f(x) + \langle v, y - x\rangle \quad \forall y \in U.$$

Then, it is clear that

$$\liminf_{y \to x} \frac{f(y) - f(x) - \langle v, y - x\rangle}{|y - x|},$$

hence $v \in \partial f(x)$.

The proof of the reverse inclusion

$$\partial f(x) \subset \{v \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x\rangle \text{ for all } y \in U\}$$

is left as an exercise to the reader.

□

It may happen that $\partial f(x) = \emptyset$. For example, the function : $\mathbb{R}^3 \to \mathbb{R}$ defined as $f(x) = x^3$, has no supporting hyperplane to epi(f) at $x = 0$, and therefore $\partial f(0) = \emptyset$. However, for convex functions $f : U \to \mathbb{R}$ we always have $\partial f(x) \neq \emptyset$ for every $x \in U$.

Proposition 6.8. *If $U \subset \mathbb{R}^n$ is an open set and $f : U \to \mathbb{R}$ is a convex function, then for all $x \in U$ we have $\partial f(x) \neq \emptyset$.*

Proof. We start by fixing a ball $B := B(x,r) \subset U$. By Lemma 6.4, epi$(f|_B)$ is a convex set, and we observe that $(x, f(x)) \in \partial(\text{epi}(f|_B))$.

As epi$(f|_B)$ is a convex set, there must exists a supporting hyperplane of epi$(f|_B)$ at $(x, f(x))$. That is, there must exist a vector $s = (v, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ so that

$$\langle v, y\rangle + \alpha t \geq \langle v, x\rangle + \alpha f(x) \quad \forall (y,t) \in \text{epi}(f|_B) = \{(y,t) \in B \times \mathbb{R} : \, t \geq f(y)\}.$$

The previous fact follows from Lemma 6.3.

By the definition of epi$(f|_B)$, for a fixed $y \in B$, one can take $t \geq f(y)$ arbitrarily large, so necessarily $\alpha \geq 0$. Furthermore, $\alpha > 0$ because if $\alpha = 0$ then we would be obtaining a supporting hyperplane of B at x, but $x \in B = B(x,r)$, so this is absurd. We can then divide by α to assume that $\alpha = 1$. Therefore, $v \in \mathbb{R}^n$ satisfies

$$\langle v, x\rangle + f(x) \leq \langle v, y\rangle + f(y) \text{ for all } y \in B,$$

because if $y \in B$ clearly $(y, f(y)) \in \text{epi}(f)$. Then for all $y \in B$

$$f(y) \geq f(x) + \langle v, x - y\rangle = f(x) + \langle -v, y - x\rangle,$$

and, making use of Lemma 6.7, we have $-v \in \partial(f|_B)(x) = \partial f(x)$.

□

Next, we prove the main result of this section, which is one of the two main ingredients of the proof of Alexandrov theorem.

Lemma 6.9. *Let $f, g : B(0,R) \subset \mathbb{R}^n \to \mathbb{R}$ be convex functions such that $f \leq g$, $g \in C^{1,1}(B(0,R))$. Then, f is differentiable at every point of $\{f = g\}$ with $\nabla f(x) = \nabla g(x)$, and for almost every $x_0 \in \{f = g\}$*

$$\lim_{x \to x_0} \frac{f(x) - \Big(f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T D^2 g(x_0)(x - x_0)\Big)}{|x - x_0|^2} = 0.$$

Proof. Recall that since f is convex, its subdifferential is non-empty at every point $x \in B(0,R)$. Take $x \in \{f = g\}$ and let $v \in \partial f(x)$. It is clear that $v \in \partial g(x)$ (since $f(x) = g(x)$ and $f \leq g$), and using that g is differentiable at x, Proposition 6.6 gives us that $v = \nabla g(x)$. Then $\partial f(x) = \{\nabla g(x)\}$ so f is differentiable at x with $\nabla f(x) = \nabla g(x)$.

On the other hand, the differential Dg is Lipschitz, so by Rademacher's Theorem Dg, is differentiable almost everywhere. We denote the second derivative as $D^2 g(x)$, whenever exists.

The functions f, g are convex, in particular continuous, so $\{f = g\}$ is measurable. By the Lebesgue differentiation theorem, in particular by Lemma 2.11, we know that almost every point of $\{f = g\}$ is a point of density one of the set, so it suffices to prove the result for the density points of $\{f = g\}$ such that Dg is differentiable at x_0. Let $x_0 \in \{f = g\}$ be one of these points, without loss of generality and in order to simplify the notation we will assume $x_0 = 0$. To prove that the result holds we have to see that

$$\lim_{x \to 0} \frac{f(x) - f(0) - Df(0)x - \frac{1}{2}x^T D^2 g(0)x}{|x|^2} = 0 \tag{6.2}$$

We know that $f(0) = g(0)$ and $Df(0) = Dg(0)$, and by *Taylor's Theorem* with the Peano remainder applied to g and centered at 0

$$\lim_{x \to 0} \frac{g(x) - \big(g(0) + Dg(0)x + \frac{1}{2}x^T D^2 g(0)x\big)}{|x|^2} = 0$$

then (6.2) is equivalent to

$$\lim_{x\to 0}\frac{(f(x)-g(x))+\left(g(x)-g(0)-Dg(0)x-\frac{1}{2}x^TD^2g(0)x\right)}{|x|^2}=\lim_{x\to 0}\frac{f(x)-g(x)}{|x|^2}.$$

We are going to prove that the limit of the term on the right is 0. Since 0 is a density point of $\{f=g\}$, by making use of the **Key fact** from the proof of Stepanov Theorem 2.7, we obtain that for each x sufficiently close to 0, there is some $y\in\{f=g\}$ such that

$$\lim_{x\to 0}\frac{|x-y|}{|x|}=0. \tag{6.3}$$

Since $f(y)=g(y)$, we recall that f is differentiable at y with $Df(y)=Dg(y)$. Then by Proposition 6.6, $\partial f(y)=\{Df(y)\}$ and by definition of subdifferential

$$f(x)\geq f(y)+Df(y)(x-y)=g(y)+Dg(y)(x-y).$$

Using that $g\geq f$ and the previous inequality we can write

$$0\leq\frac{g(x)-f(x)}{|x|^2}\leq\frac{g(x)-g(y)-Dg(y)(x-y)}{|x|^2} \tag{6.4}$$

We will prove next that there exists some $M\geq 0$ (the Lipschitz constant of Dg works) so that

$$|g(x)-g(y)-Dg(y)(x-y)|\leq M|y-x|^2. \tag{6.5}$$

Hence, from (6.3), (6.4) and (6.5), taking limit $x\to 0$ we conclude that

$$\lim_{x\to 0}\frac{g(x)-f(x)}{|x|^2}=0$$

and this concludes the proof.

To prove (6.4), since $g\in C^1(U)$, the Mean Value Theorem ensures the existence of some ξ within the segment joining x to y such that

$$f(y)-f(x)=Df(\xi)(y-x).$$

Then,

$$f(y)-f(x)-Df(x)(y-x)=Df(\xi)(y-x)-Df(x)(y-x)=(Df(\xi)-Df(x))(y-x),$$

and taking norms and calling $M\geq 0$ the Lipschitz constant of Dg

$$|f(y)-f(x)-Df(x)(y-x)|\leq|(Df(\xi)-Df(x))||(y-x)|\leq M|\xi-x|\cdot|y-x|\leq M|x-y|^2.$$

□

6.2 Convex bodies and $C^{1,1}$-Lusin property for convex functions

The goal of this section is to show that convex functions admit a $C^{1,1}$ Lusin-type approximation by convex functions. Namely we aim to prove that.

Theorem 6.10. *Let $f:\mathbb{R}^n\to\mathbb{R}$ be a convex function. Then for each $R>0$ and $\varepsilon>0$, there exists a convex function $g\in C^{1,1}(B^n(0,R))$ such that $g\geq f$ and*

$$\mathcal{L}^n\big(\{x\in B^n(0,R):f(x)\neq g(x)\}\big)<\varepsilon. \tag{6.6}$$

When a class of functions $\mathcal{A}$ defined on $\Omega \subset \mathbb{R}^n$ satisfies that for every $f \in \mathcal{A}$ and for every $\varepsilon > 0$ there exists $g \in C^k(\Omega)$ such that $\mathcal{L}^n(\{f \neq g\}) < \varepsilon$, we say that the class of functions $\mathcal{A}$ has Lusin's property of class C^k. This nomenclature comes from the well-known result proven by Lusin which guarantees that

Lusin Theorem (1912): *For every measurable $f : \mathbb{R}^n \to \mathbb{R}$ and for every $\varepsilon > 0$ there exists $g \in C(\mathbb{R}^n)$ such that $\mathcal{L}^n(\{f \neq g\}) < \varepsilon$.*

That is to say, the class of measurable functions $f : \mathbb{R}^n \to \mathbb{R}$ has Lusin's property of class $C(\mathbb{R}^n)$. Following this terminology, Theorem 6.10 proves that for any $R > 0$, the class of convex functions $f : B(0, R) \to \mathbb{R}$ has Lusin's property of class $C^{1,1}(B(0, R))$, and furthermore that the approximating functions can be chosen to be convex. A few years ago, Azagra, Darke and Hajłasz have proved the stronger result that convex functions also have Lusin's property of class C^2, with the approximating function also being a convex function.

Before entering the proof of Theorem 6.10, which is postponed until the end of this section, we need a series of lemmas; Lemmas 6.11, 6.13, 6.15, 6.16 and 6.19. Among these, we highlight Lemma 6.19, where boundaries of a *special class* of convex bodies are shown to be of class $C^{1,1}$.

Lemma 6.11. *Let $W \subset \mathbb{R}^n$ be a non-empty, convex and closed set.*

(a) For every $x \in \mathbb{R}^n$ there exists a unique point, which we denote $\pi_W(x)$, such that

$$\pi_W(x) \in W \text{ and } |\pi_W(x) - x| = dist(x, W).$$

(b) If $x \notin W$ then $\pi_W(x) \in \partial W$.

(c) $\pi_W(x)$ is the only point so that for all $y \in W$,

$$\langle y - \pi_W(x), x - \pi_W(x)\rangle \leq 0 \tag{6.7}$$

Proof. (a) If $x \in W$, then it is trivial by taking $\pi_W(x) = x$. Suppose then that $x \notin W$. Let $d = dist(x, W) = \inf_{v\in W}\{|v - x|\}$, then there exists a sequence $(v_k)_{k\geq 1} \subset W$ such that $|v_k - x| \to d$ as $k \to \infty$. We will prove that the limit of this sequence is the sought $\pi_W(x)$:
By the Parallelogram Law, we have that for $k, m \geq 1$

$$2(|x - v_m|^2 + |x - v_k|^2) = |2x - (v_m + v_k)|^2 + |v_m - v_k|^2. \tag{6.8}$$

Since W is convex, the point $\frac{1}{2}(v_m + v_k)$ belongs to W, so

$$|2x - (v_m + v_k)| = 2|x - \frac{1}{2}(v_m + v_k)| \geq 2d$$

thus, solving for $|v_m - v_k|^2$ in (6.8), we have

$$\begin{aligned} |v_m - v_k|^2 &= 2(|x - v_m|^2 + |x - v_k|^2) - |2x - (v_m + v_k)|^2 \\ &\leq 2(|x - v_m|^2 + |x - v_k|^2) - 4d^2 \xrightarrow[k,m\to\infty]{} 2(d^2 + d^2) - 4d^2 = 0. \end{aligned}$$

Therefore, the sequence $(v_k)_{k\geq 1}$ is Cauchy, so it converges to some vector v, which belongs to W because it is closed. We then take $\pi_W(x) = v$ and by continuity of the norm we have

$$d = \lim_{k\to\infty} |v_k - x| = |\pi_W(x) - x|.$$

It remains to prove that it is indeed unique. Suppose there exist $v, w \in W$ such that

$$d = dist(x, W) = |x - v| = |x - w|$$

then $\frac{1}{2}(v+w) \in W$, so

$$d^2 \leq |x - \frac{1}{2}(v+w)|^2 = |\frac{1}{2}(x-v) + \frac{1}{2}(x-w)|^2$$

and this implies, again by the Parallelogram Law

$$\begin{aligned} d^2 \leq |\frac{1}{2}(x-v) + \frac{1}{2}(x-w)|^2 = 2\big(|\frac{1}{2}(x-v)|^2 + |\frac{1}{2}(x-w)|^2\big) - |\frac{1}{2}(x-v) - \frac{1}{2}(x-w)|^2 \\ = d^2 - \frac{1}{4}|v-w|^2. \end{aligned}$$

Then $|v-w| \leq 0$ and therefore $v = w$. Thus, we have uniqueness.

(b) The fact that if $x \notin W$ then $\pi_W(x) \in \partial W$ is obvious, because otherwise we could take points of W in some ball $B(\pi_W(x); r) \subset W$ closer to x; it would suffice to take them on the segment joining x and $\pi_W(x)$.

(c) Let us first prove that $\pi_W(x)$ satisfies (6.7). Let $y \in W$ be arbitrary. Define for $t \in \mathbb{R}$, $y_t := ty + (1-t)\pi_W(x)$. Then

$$\begin{aligned} |x - y_t|^2 &= |tx + (1-t)x - ty - (1-t)\pi_W(x)|^2 = |t(x-y) + (1-t)(x - \pi_W(x))|^2 \\ &= t^2|x-y|^2 + (1-t)^2|x - \pi_W(x)|^2 + 2t(1-t)\langle x-y, x-\pi_W(x)\rangle. \end{aligned}$$

Thus, if we consider the function for $t \in \mathbb{R}$ such that $t \mapsto |x-y_t|^2$, it is clearly differentiable and therefore

$$\begin{aligned} &\frac{d}{dt}|x-y_t|^2\Big|_{t=0} = \Big(2t|x-y|^2 - 2(1-t)|x-\pi_W(x)|^2 + (2-4t)\langle x-y, x-\pi_W(x)\rangle\Big)\Big|_{t=0} \\ &= -2|x-\pi_W(x)|^2 + 2\langle x-y, x-\pi_W(x)\rangle = -2\langle x-\pi_W(x), x-\pi_W(x)\rangle + 2\langle x-y, x-\pi_W(x)\rangle \\ &= -2\langle y-\pi_W(x), x-\pi_W(x)\rangle. \end{aligned}$$

If we take $t \in [0,1]$, since the projection satisfies $|\pi_W(x) - x| = dist(x, W)$, then it minimizes $\{|v-x| : v \in W\}$. Thus, since for $t \in [0,1]$, $y_t \in W$ because W is convex, then $\pi_W(x) = y_0$ minimizes $\{|x-y_t| : t \in [0,1]\}$ and so $t = 0$ minimizes the function $|x-y_t|^2$ on the interval $[0,1]$ and therefore

$$\frac{d}{dt}|x-y_t|^2\Big|_{t=0} \geq 0.$$

Then $-2\langle y-\pi_W(x), x-\pi_W(x)\rangle \geq 0$ and since we took $y \in W$ arbitrarily, we conclude that for all $y \in W$

$$\langle y-\pi_W(x), x-\pi_W(x)\rangle \leq 0.$$

Suppose now that for some $z \in W$ it holds that $\langle y-z, x-z\rangle \leq 0$ for all $y \in W$, and let us show that necessarily $z = \pi_W(x)$. Indeed, for such z we would have that for any $y \in W$

$$\begin{aligned} |y-x|^2 &= |(y-z) - (x-z)|^2 = |y-z|^2 + |x-z|^2 - 2\langle y-z, x-z\rangle \\ &\geq |z-x|^2 - 2\langle y-z, x-z\rangle, \end{aligned}$$

and therefore using our hypothesis

$$|y-x|^2 \geq |z-x|^2 \text{ for all } y \in W.$$

This inequality shows that $z \in W$ minimizes $\{dist(x,y) : y \in W\}$, so $z = \pi_W(x)$.

□

Definition 6.12. *The mapping $\pi_W : \mathbb{R}^n \to W$ is called the projection onto W. It is also often called orthogonal projection.*

Lemma 6.13. *If $W \subset \mathbb{R}^n$ is closed and convex, then the function $\pi_W : \mathbb{R}^n \to W$ is 1-Lipschitz.*

Proof. Let $x_1, x_2 \in \mathbb{R}^n$ be arbitrary. If in Lemma 6.11(3) we take $x = x_1,\ y = \pi_W(x_2)$ we obtain

$$\langle \pi_W(x_2) - \pi_W(x_1), x_1 - \pi_W(x_1)\rangle \le 0$$

and taking $x = x_2,\ y = \pi_W(x_1)$

$$\langle \pi_W(x_1) - \pi_W(x_2), x_2 - \pi_W(x_2)\rangle \le 0.$$

Adding both inequalities

$$\langle \pi_W(x_1) - \pi_W(x_2), x_2 - x_1 + \pi_W(x_1) + \pi_W(x_2)\rangle \le 0.$$

Then

$$|\pi_W(x_1) - \pi_W(x_2)|^2 = \langle \pi_W(x_1) - \pi_W(x_2), \pi_W(x_1) - \pi_W(x_2)\rangle \le \langle \pi_W(x_1) - \pi_W(x_2), x_1 - x_2\rangle.$$

Applying the Cauchy-Schwarz inequality to the previous inequality, we have

$$|\pi_W(x_1) - \pi_W(x_2)|^2 \le |\pi_W(x_1) - \pi_W(x_2)||x_1 - x_2|$$

and therefore

$$|\pi_W(x_1) - \pi_W(x_2)| \le |x_1 - x_2|$$

for all $x_1, x_2 \in \mathbb{R}^n$. Then π_W is Lipschitz with constant $L \le 1$. That the Lipschitz constant is, in general, $L = 1$ is easy to see, since if we take $x, z \in W$ then $\pi_W(x) = x,\ \pi_W(z) = z$ and therefore

$$|\pi_W(x) - \pi_W(z)| = |x - z|$$

and thus the best Lipschitz constant we can obtain is 1.

□

Definition 6.14. *We say that $K \subset \mathbb{R}^n$ is a* convex body *if it is convex, compact, and has nonempty interior. For a convex body K and $r > 0$ we define the* inner parallel convex body *as*

$$K_r := \{x \in K : \operatorname{dist}(x, \partial K) \ge r\}.$$

and

$$K(r) := \bigcup \{\, \overline{B}(x, r) : \overline{B}(x, r) \subset K\,\}.$$

Lemma 6.15. *Let $K \subset \mathbb{R}^n$ be a convex body and assume that K contains a ball of radius $r_0 > 0$. Then, for every $r \in (0, r_0)$:*

(a) The set K_r is a nonempty convex body.

(b) The set $K(r)$ is a nonempty convex body.

(c) $\mathcal{H}^{n-1}(\partial K_r) \le \mathcal{H}^{n-1}\big(\partial K \cap \partial K(r)\big)$.

Proof. (a) It is easy to see that $K_r \ne \emptyset$ for all $r \in (0, r_0]$ and that K_r has nonempty interior if $r \in (0, r_0)$. Moreover, K_r is a closed subset of K, which is compact, hence K is always a compact set. It only remains to prove that K_r is convex.

Let $x, y \in K_r$. We must show that $[x, y] \subset K_r$. Observe that $\overline{B}(x, r), \overline{B}(y, r) \subset K$. Let $z \in [x, y]$, so that $z = (1 - \lambda)x + \lambda y$ for some $\lambda \in [0, 1]$. We claim that

$$\overline{B}(z, r) \subset \operatorname{co}\big(\overline{B}(x, r) \cup \overline{B}(y, r)\big).$$

Indeed, if $a \in \overline{B}(z,r)$, then there exists a vector w such that $a = z + w$ with $|w| \le r$, and hence

$$a = z + w = (1-\lambda)x + \lambda y + (1-\lambda)w + \lambda w = (1-\lambda)(x+w) + \lambda(y+w). \tag{6.9}$$

Since $|w| \le r$, we have $x + w \in \overline{B}(x,r)$ and $y + w \in \overline{B}(y,r)$. Therefore, (6.9) shows that a is a convex linear combination of points in $\overline{B}(x,r) \cup \overline{B}(y,r)$, and thus $a \in \mathrm{co}\,\big(B(x,r) \cup B(y,r)\big)$. Consequently,

$$B(z,r) \subset \mathrm{co}\,\big(B(x,r) \cup B(y,r)\big).$$

Since the convex hull of a set is the intersection of all convex sets containing it, we have

$$\mathrm{co}\,\big(B(x,r) \cup B(y,r)\big) \subset K.$$

In particular, $\overline{B}(z,r) \subset K$, and hence $\mathrm{dist}(z, \partial K) \ge r$. This shows that $z \in K_r$, and therefore K_r is convex.

(b) It is clear that $K(r) \neq \emptyset$ and has nonempty interior for all $r \in (0, r_0]$.

To see that $K(r)$ is compact, it suffices to show that it is closed, since $K(r)$ is contained in a compact set K. Let $(x_n)_{n\ge1} \subset K(r)$ be a sequence such that $x_n \to x$. It is enough to prove that $x \in K(r)$, which is equivalent to showing that $\overline{B}(x,r) \subset K$. Suppose this is not the case. Then there exists $y \in \overline{B}(x,r) \setminus K$. Since $\overline{B}(x,r) \setminus K$ is an open set, we can assume without loss of generality that $y \in B(x,r) \setminus K$. That is, there exists $0 < \rho < r$ such that $|x - y| = \rho$. Then, for each $n \in \mathbb{N}$,

$$|x_n - y| \le |x - x_n| + |x - y| < |x - x_n| + \rho,$$

and therefore

$$\limsup_{n\to\infty} |x_n - y| < \rho.$$

This implies that there exists some $n_0 \in \mathbb{N}$ such that $|x_{n_0} - y| < \rho < r$. Hence $y \in \overline{B}(x_{n_0}, r) \setminus K$, which implies that $x_{n_0} \notin K(r)$. This contradicts the fact that the sequence is contained in $K(r)$. Therefore $x \in K(r)$, and thus $K(r)$ is closed.

Finally, we show that $K(r)$ is convex. Let $x, y \in K(r)$. Then there exist two balls such that $x \in \overline{B}(x', r) \subset K$ and $y \in \overline{B}(y', r) \subset K$. Let $\lambda \in [0,1]$ and define $z' = (1-\lambda)x' + \lambda y'$. Then

$$\overline{B}(z', r) = (1-\lambda)\overline{B}(x', r) + \lambda \overline{B}(y', r) := \{(1-\lambda)\omega_1 + \lambda\omega_2 : \omega_1 \in \overline{B}(x', r),\ \omega_2 \in \overline{B}(y', r)\}.$$

The point $z = (1-\lambda)x + \lambda y$ belongs to $\overline{B}(z', r)$. Since the points of $B(z', r)$ are convex linear combinations of points in K, and K is convex, we have $\overline{B}(z', r) \subset K$. Hence $z \in K(r)$ by definition, since it is contained in a closed ball contained in K. As this holds for all $\lambda \in [0,1]$, we conclude that $K(r)$ is convex.

(c) We already know that K_r is a convex body. Consider the projection onto K_r and observe that

$$\pi_{K_r}\big(\partial K \cap \partial K(r)\big) = \partial K_r. \tag{3.2}$$

Indeed, if $z \in \partial K_r$, then $\mathrm{dist}(z, \partial K) = r$. Since ∂K is compact, there exists $x \in \partial K$ that minimizes the distance to z, that is, $r = |x - z|$. Therefore $x \in B(z,r) \subset K$, and hence by definition $x \in K(r)$. Clearly, if $x \in \partial K$ and $x \in K(r)$, then necessarily $x \in \partial K(r)$, and thus $x \in \partial K \cap \partial K(r)$.

Since $x \in \partial K$, by the definition of K_r we have $\mathrm{dist}(x, K_r) = r$. Moreover, since $\mathrm{dist}(x, K_r) = |x - z|$, by Lemma 6.11(a) it follows that $\pi_{K_r}(x) = z$, and hence $z \in \pi_{K_r}\big(\partial K \cap \partial K(r)\big)$. The reverse inclusion is trivial, since by Lemma 6.11(b) we already proved that the projection of a point onto a closed convex set belongs to the boundary.

The conclusion of the theorem follows immediately from (3.2) and Proposition 4.3(5), since by Lemma 6.13 the projection π_{K_r} is 1-Lipschitz.

□

Lemma 6.16. *If $K \subset \mathbb{R}^n$ is a convex body, then*

$$\lim_{r \to 0^+} \mathcal{H}^{n-1}(\partial K \setminus \partial K(r)) = 0.$$

Proof. By the definition of a convex body, the interior of K is non-empty, so there exists $r_0 > 0$ such that there is some ball of radius r_0 contained in K. Since the Hausdorff measure is invariant under translations, translating K does not alter the result, so we can assume without loss of generality that this ball is centered at 0, that is, $\overline{B}(0, r_0) \subset K$. This implies that for all $r \in (0, r_0)$, the point 0 belongs to the interior of K_r. For $\lambda > 0$ we consider $\lambda K_r = \{\lambda z : z \in K_r\}$, and we define from these sets, for each $r \in (0, r_0)$,

$$\lambda(r) := \inf\{\lambda > 0 : K \subset \lambda K_r\}.$$

Observe that $\lambda(r)$ is well-defined because since $0 \in \text{int}(K_r)$, K_r contains a ball $B(0, \gamma)$, and since K is compact, it is bounded, so there exists some λ such that $K \subset \lambda B(0, \gamma) \subset \lambda K_r$ and therefore the infimum is finite. We will continue the proof by showing that $K \subset \lambda(r) K_r$.

By contradiction, suppose that $K \not\subset \lambda(r) K_r$ and let $x \in K \setminus \lambda(r) K_r$. We take a decreasing sequence $(\lambda_n)_{n \geq 1}$ such that $\lambda_n \to \lambda(r)$ and $K \subset \lambda_n K_r$ for all $n \in \mathbb{N}$. We will reach a contradiction by arguing that for some $N \in \mathbb{N}$, $x \notin \lambda_N K_r$.

Since $0 \in K_r$, $0 \in \lambda(r) K_r$, if we consider the ray starting from 0 and passing through x, that is

$$l = \{tx : x \geq 0\},$$

we can take the point $z \in l \cap \lambda(r) K_r$ closest to x, which is easily seen to be given by $z = t_0 x$ such that

$$t_0 = \max\{t \geq 0 : tx \in \lambda(r) K_r\}.$$

We observe that $t_0 < 1$, because otherwise $x \in [0, z] \subset \lambda(r) K_r$, which is a contradiction since $x \notin \lambda(r) K_r$. Furthermore, by the choice of z, t_0 satisfies

$$t > t_0 \Longrightarrow tx \notin \lambda(r) K_r. \tag{6.10}$$

Now let $y \in K_r$ such that $z = \lambda(r) y$, and we define for each $n \in \mathbb{N}$ the point $z_n = \lambda_n y \in \lambda_n K_r$, it is immediate that $z_n \in l$, and therefore we take for each $n \in \mathbb{N}$ the corresponding $t_n \geq 0$ such that $z_n = t_n x$. We now observe that since $\lambda_n \to \lambda(r)$, then $z_n \to z$, and therefore looking at their expression as points of l, necessarily $t_n \to t_0$. From this fact we will find a contradiction. Since $t_0 < 1$ there exists $N \in \mathbb{N}$ such that $t_N < 1$. Since $z_N = \lambda_N y = t_N x$ we have

$$x = \frac{\lambda_N}{t_N} y. \tag{6.11}$$

We can now find the contradiction we seek, since $1/t_N > 1$ then $\frac{t_0}{t_N} > t_0$ and by (6.10) $\frac{t_0}{t_N} x \notin \lambda(r) K_r$. Taking into account that $z = t_0 x$ and that $z = \lambda(r) y$, it follows that

$$\frac{1}{t_N} z \notin \lambda(r) K_r \Longrightarrow \frac{\lambda(r)}{t_N} y \notin \lambda(r) K_r \Longrightarrow \frac{1}{t_N} y \notin K_r \Longrightarrow \frac{\lambda_N}{t_N} y \notin \lambda_N K_r \overset{(6.11)}{\Longrightarrow} x \notin \lambda_N K_r,$$

which is a contradiction since λ_N is, by choice, such that $x \in \lambda_N K_r$. The contradiction comes from assuming that $K \not\subset \lambda(r) K_r$, and therefore it is proved that $K \subset \lambda(r) K_r$.

The function $r \mapsto \lambda(r)$ is non-decreasing, because if $0 < r < s \leq r_0$ then it is clear that $K_s \subset K_r$, so for any $\lambda > 0$, $\lambda K_s \subset \lambda K_r$ and thus if $K \subset \lambda K_s$ then $K \subset \lambda K_r$. Therefore $\{\lambda > 0 : K \subset \lambda K_s\} \subset \{\lambda > 0 : K \subset \lambda K_r\}$ so

$$\lambda(r) = \inf\{\lambda > 0 : K \subset \lambda K_r\} \leq \inf\{\lambda > 0 : K \subset \lambda K_s\} = \lambda(s).$$

To finish the proof, it suffices to show that

(a) $\lim_{r\to 0}\lambda(r) = 1$

(b) $\pi_K(\partial\lambda(r)K_r) = \partial K$.

Once these two equalities are proven, the proof will be complete, because since π_K is 1-Lipschitz,

$$\mathcal{H}^{n-1}(\pi_K(\partial\lambda(r)K_r)) \leq \mathcal{H}^{n-1}(\partial\lambda(r)K_r),$$

then by Lemma 6.15 (c)

$$\mathcal{H}^{n-1}(\partial K) \overset{\text{(b)}}{=} \mathcal{H}^{n-1}(\pi_K(\partial\lambda(r)K_r)) \leq \mathcal{H}^{n-1}(\partial\lambda(r)K_r) =$$
$$\lambda(r)^{n-1}\mathcal{H}^{n-1}(\partial K_r) \overset{6.15(c)}{\leq} \lambda(r)^{n-1}\mathcal{H}^{n-1}(\partial K \cap \partial K(r)) \leq \lambda(r)^{n-1}\mathcal{H}^{n-1}(\partial K) \xrightarrow[r\to 0^+]{} \mathcal{H}^{n-1}(\partial K).$$

where in the limit we used (a). From the chain of inequalities we deduce that $\mathcal{H}^{n-1}(\partial K \cap \partial K(r)) \to \mathcal{H}^{n-1}(\partial K)$ as $r \to 0^+$, this completes the proof of the lemma because

$$\mathcal{H}^{n-1}\big(\partial K \setminus \partial K(r)\big) = \mathcal{H}^{n-1}\big(\partial K \setminus (\partial K \cap \partial K(r))\big) = \mathcal{H}^{n-1}\big(\partial K\big) - \mathcal{H}^{n-1}\big(\partial K \cap \partial K(r)\big) \to 0.$$

Proof of (a): That $\lambda(r) \geq 1$ is immediate. Given $\varepsilon > 0$, $(1+\varepsilon)^{-1}K \subset \text{int}(K)$ so $dist((1+\varepsilon)^{-1}K, \partial K) > 0$. We take $\delta = dist((1+\varepsilon)^{-1}K, \partial K)$ and we have that for all $r \in (0,\delta]$

$$(1+\varepsilon)^{-1}K \subset K_r \Longrightarrow K \subset (1+\varepsilon)K_r \Longrightarrow (1+\varepsilon) \in \{\lambda > 0 : K \subset \lambda K_r\}$$

so $\lambda(r) \leq (1+\varepsilon)$. We have proved that if $\varepsilon > 0$ there exists $\delta > 0$ such that for all $r \in (0,\delta]$, $1 \leq \lambda(r) \leq 1+\varepsilon$. This shows that

$$\lim_{r\to 0^+}\lambda(r) = 1.$$

Proof of (b): That $\pi_K(\lambda(r)K_r) \subset \partial K$ is already known by Lemma 6.11 because if $x \in \partial\lambda(r)K_r$ then either $x \notin K$ or $x \in \partial K$. Let's see the other inclusion. Let $x \in \partial K$, by Lemma 6.3 there exists a vector $v \in \mathbb{R}^n$, which we can assume to be unitary, such that v and x define a hyperplane that supports K at x (this vector is commonly called the *outer unit normal vector at* x), that is,

$$\langle y, v\rangle \leq \langle x, v\rangle \tag{6.12}$$

for all $y \in K$. Let $t \geq 0$, the vector $x + tv \notin K$ since $\langle x+tv, v\rangle = \langle x, v\rangle + t|v|^2 > \langle x, v\rangle$ so it does not satisfy (6.12). Therefore, since $K \subset \lambda(r)K_r$, and $\lambda(r)K_r$ is compact, there exists $T \geq 0$ such that $z := x + Tv \in \partial\lambda(r)K_r$. If we take the projection of z onto K, it is exactly x because for all $y \in K$

$$\langle y-x, z-x\rangle = \langle y-x, Tv\rangle = T(\langle y, v\rangle - \langle x, v\rangle) \overset{(6.12)}{\leq} 0$$

then by Lemma 6.11(c) $\pi_K(z) = x$, so $x \in \pi_K(\partial\lambda(r)K_r)$ and (b) is proved.

□

The proof of Lusin $C^{1,1}$-property for convex functions (i.e. Theorem 6.10) relies very much on the next final lemma, where convex bodies W for which $W = W(r)$ for some $r > 0$, are shown to have a boundary of class $C^{1,1}$.

Definition 6.17. *Given a function $f : U \to \mathbb{R}$, with $U \subset \mathbb{R}^n$, we say that $f \in C^{1,1}(U)$ when $f \in C^1(U)$ and the differential $Df : U \to \mathbb{R}^n$ is Lipschitz on U. We also say that $f \in C^{1,1}_{\text{loc}}(U)$ if the differential is locally Lipschitz on U.*

Definition 6.18. *We say that the boundary of a bounded set $U \subset \mathbb{R}^n$ is of class $C^{1,1}$ if locally it is the graph of a $C^{1,1}$ function. That is, for each point $p \in \partial U$ there exist a neighborhood $V \subset \mathbb{R}^n$ of p and an open set $W \subset \mathbb{R}^{n-1}$ such that*

$$\partial U \cap V = \{(x, f(x)) \in \mathbb{R}^n : x \in W\},$$

where $f : W \to \mathbb{R}$ satisfies $f \in C^{1,1}(W)$.

Lemma 6.19. *If W is a convex body such that, for some $r > 0$, $W = W(r)$, then the boundary of W is of class $C^{1,1}$.*

Proof. Our goal in this proof is to show that, locally, ∂W is the graph of a $C^{1,1}$ function. To do this, since $W = W(r)$, every point $p \in \partial W$ must belong to the boundary of some ball of radius r entirely contained within W. Taking the unique supporting hyperplane of W at p as the coordinate space, we will see how locally we can trap the boundary of W between the graphs of two differentiable functions, which lead to the fact that the parametrization of ∂W is differentiable (see Exercise 5). After this, it will only remain to prove that such differential is Lipschitz.

Since $W = W(r)$, we can assume that for all $p \in \partial W$ there exists an $h(p) \in W$ such that $p \in \overline{B}(h(p), r) \subset W$. By Lemma 6.3 there exists a supporting hyperplane T_p for W at p. Furthermore, we observe that the only possible supporting hyperplane T_p is exactly the tangent hyperplane to $\overline{B}(h(p), r)$ at p, since any other hyperplane passing through p is not a supporting hyperplane of $\overline{B}(h(p), r)$ because it cuts into the interior of the ball, and therefore it could not be one for W either because $\overline{B}(h(p), r) \subset W$.

Since $\overline{B}(h(p), r) \subset W$ and necessarily p belongs to the boundary of the ball, then $dist(h(p), \partial W) = r$, which implies that $h(p) \in W_r$, so $|p - h(p)| = r = dist(p, W_r)$ by the definition of W_r, and therefore by Lemma 6.11 $h(p) = \pi_{W_r}(p)$.

On the other hand, the inner unit normal vector of the hyperplane T_p, where by inner we refer to the one pointing towards the side of T_p where W lies, is the normal to $\overline{B}(h(p), r)$ at p, which is clearly

$$\nu(p) := \frac{h(p) - p}{r} = \frac{\pi_{W_r}(p) - p}{r}.$$

We define the function $\nu : \partial W \to \mathbb{S}^{n-1}$, which is a Lipschitz function because

$$|\nu(p) - \nu(q)| = \frac{|\pi_{W_r}(p) - p - (\pi_{W_r}(q) - q)|}{r} \leq \frac{|\pi_{W_r}(p) - \pi_{W_r}(q)| + |p - q|}{r} \leq \frac{2}{r}|p - q|$$

where in the last inequality we used that π_{W_r} is 1-Lipschitz.

We are now going to construct the function f such that locally ∂W is the graph of f, and then we prove that such f is of class $C^{1,1}$.

Let $p_0 \in \partial W$ be fixed for the whole argument. We want to see that in a neighborhood of p_0, ∂W is the graph of an $f \in C^{1,1}$. Let T_{p_0} be the tangent hyperplane to $\overline{B}(h(p_0), r)$ at p_0, which we have already seen is the unique hyperplane that supports W at p_0. We choose a Euclidean coordinate system such that $p_0 = (0, \ldots, 0)$ and $T_{p_0} = \{x_n = 0\}$, identifying $(x_1, \ldots, x_{n-1}, x_n) = (x', x_n)$. In this coordinate system $T_{p_0} = \{x \in \mathbb{R}^n : \langle x, v\rangle = 0\}$ for $v = (0, \ldots, 0, 1) \in \mathbb{R}^n$ and $h(p_0) = (0, \ldots, 0, r)$. We can assume without loss of generality that $W \subset \{x : \langle v, x\rangle \geq 0\} = \{(x', x_n) : x_n \geq 0\}$. Let

$$U = \{x' = (x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1} : dist((x_1, \ldots, x_{n-1}, 0), p_0) < r/2\}$$

which is an open $(n-1)$-dimensional ball (in the subspace topology of T_{p_0}) of radius $r/2$ and center p_0.

We now claim that there exists a function $f : U \to \mathbb{R}$ of class $C^{1,1}$, defined by

$$f(x') := \eta \in \mathbb{R} \text{ such that } \eta = \min\left\{t \geq 0 : (x', 0) + (0, \overset{n-1)}{\ldots}, 0, t) \in W\right\}.$$

so that

$$\partial W \cap (U \times [0, r/2)) = \{(x', f(x')) : x' \in U\}.$$

We will prove the following properties separately:

(i) f is well-defined.

(ii) We have $\partial W \cap (U \times [0, r/2)) = \{(x', f(x')) : x' \in U\}$.

(iii) f is differentiable at every $x'_0 \in U$.

(iv) ∇f is Lipschitz on U.

(i) **f is well-defined**:

We can see that for any $x' \in U$ there exists $t_0 \geq 0$ such that $(x', t_0) \in W$. Indeed, when taking (x', t) for $x' \in U$, there must necessarily exist some $t_0 \geq 0$ such that $(x', t_0) \in \overline{B}(h(p_0), r)$. Since $\overline{B}(h(p_0), r) \subset W$, it follows that $(x', t_0) \in W$. We observe that $f(x')$ is actually given by a minimum because W is closed, and also $(x', f(x')) \in \partial W$. Furthermore, by the definition of f it is clear that for all $x' \in U$

$$t < f(x') \Longrightarrow (x', t) \notin W. \tag{6.13}$$

(ii) **We have $\partial \mathbf{W} \cap (\mathbf{U} \times [\mathbf{0}, \mathbf{r}/\mathbf{2})) = \{(\mathbf{x}', \mathbf{f}(\mathbf{x}')) : \mathbf{x}' \in \mathbf{U}\}$**:

We first show that given $x' \in U$ we have $(x', f(x')) \in \partial W$ and that $f(x') \in [0, r/2)$. That $(x', f(x')) \in \partial W$ is immediate because by the very definition of f, $(x', f(x')) \in W$ and also for all $\varepsilon > 0$ we have that $(x', f(x') - \varepsilon) \notin W$, from which it follows that

$$B((x', f(x')), \varepsilon) \cap W \neq \emptyset, \quad B((x', f(x')), \varepsilon) \cap (\mathbb{R}^n \setminus W) \neq \emptyset,$$

therefore $(x', f(x')) \in \partial W$. However, by definition $f(x') \geq 0$, so we only need to show that $f(x') < \frac{r}{2}$. Let $\tau : U \to \mathbb{R}$ be the function that parametrizes $\partial \overline{B}(h(p_0), r) \cap (U \times [0, r))$, it is easy to verify that

$$\tau(x') = r - \sqrt{r^2 - |x'|^2}.$$

Since for $x' \in U$, $|x'| \in [0, \frac{r}{2})$, it follows that $0 \leq \tau(x') < (1 - \frac{\sqrt{3}}{2})r < r/2$. Since $(x', \tau(x')) \in W$, from (6.13) it is deduced that $f(x') \leq \tau(x')$, which implies that $f(x') < r/2$.

Next, for the reverse inclusion, let us show that given $y = (y_1, \ldots, y_n) \in \partial W \cap (U \times [0, r/2))$ there exists $y' \in U$ such that $y = (y', f(y'))$. Our candidate is $y' = (y_1, \ldots, y_{n-1}) \in U$, and we must show that $f(y') = y_n$. Let T_y be a supporting hyperplane of W at y, such that if $s = (s', s_n) \in \mathbb{R}^n$ determines such hyperplane, then

$$\langle s', y' \rangle + s_n y_n \leq \langle s, x \rangle \text{ for all } x \in W. \tag{6.14}$$

Since $(y', f(y')) \in W$, we deduce from (6.14)

$$\langle s', y' \rangle + s_n y_n \leq \langle s', y' \rangle + s_n f(y') \overset{(*)}{\Longrightarrow} y_n \leq f(y')$$

which by the definition of f and given that $(y', y_n) \in W$ implies that $f(y') = y_n$. For $(*)$ to be true, it would remain to see that $s_n > 0$. From what was observed at the beginning of the proof, T_y is the tangent hyperplane to a ball $B(h(y), r) \subset W$ at y, so the vector s, which is normal to the hyperplane, must be either the outer normal to $B(h(y), r)$ or the inner one, but from (6.14) it follows that s points towards

W, so it points towards $B(h(y), r) \subset W$ and thus s is the inner normal. This, together with the fact that $y_n < r/2$, which implies that y is in the lower half of $B(h(y), r)$ viewed with respect to $\{x_n = 0\}$, gives us that $s_n > 0$.

(iii) **f is differentiable at every $\mathbf{x}'_0 \in$ U**:

For the whole argument that follows, let $x'_0 \in U$ be fixed. We denote $p(x') := (x', f(x'))$. We have $p(x'_0) \in \partial W$ and we take the ball $\overline{B}' := \overline{B}(h(p(x'_0)), r) \subset W$ such that $p(x'_0) \in \partial \overline{B}'$. The boundary of $\overline{B}'$ can be expressed as the graph of a convex and differentiable function $g : V \to \mathbb{R}$ where V is an open neighborhood of x'_0 in U. Indeed,

$$\partial \overline{B}' = \{(y_1, \dots, y_n) \in \mathbb{R}^n : (y_1 - \alpha_1)^2 + \cdots + (y_n - \alpha_n)^2 = r^2\}$$

where $(\alpha_1, \dots, \alpha_n) = h(p(x'_0))$ is the center of the ball, so taking

$$g(y') = \alpha_n - \sqrt{r^2 - (y_1 - \alpha_1)^2 - \cdots - (y_{n-1} - \alpha_{n-1})^2}$$

for $y' = (y_1, \dots, y_{n-1}) \in V \subset U$, the boundary of $\overline{B}'$ in a neighborhood of $p(x'_0)$ is given by the graph of g, that is, in the form $(y', g(y'))$. The function g is clearly differentiable, and furthermore $f(y') \leq g(y')$ by (6.13) since $(y', g(y')) \in W$.

Now let $T_{p(x'_0)}$ be the tangent hyperplane to $\overline{B}'$ at $p(x'_0)$, which we have already seen is the unique supporting hyperplane of W at $p(x'_0)$. Let $s = (s_1, \dots, s_n) \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$ such that

$$T_{p(x'_0)} = \{x \in \mathbb{R}^n : \langle s, x \rangle = b\}. \tag{6.15}$$

In particular $\langle s, p(x'_0) \rangle = b$ and $\langle s, y \rangle \geq b$ for all $y \in W$. Let $s' := (s_1, \dots, s_{n-1})$ and define the function $\varphi : V \to \mathbb{R}$ as

$$\varphi(x') := \frac{b - \langle s', x' \rangle}{s_n}$$

then it is clearly differentiable. The case $s_n = 0$ is impossible because s is the normal vector to $T_{p(x'_0)}$, so if $s_n = 0$, $T_{p(x'_0)}$ is perpendicular to $\{x_n = 0\}$ and therefore since $x'_0 \in U$ and U being an $(n-1)$-dimensional ball in $\{x_n = 0\}$, it would mean that $T_{p(x'_0)}$ is not a supporting hyperplane of $\overline{B}(h(p_0), r) \subset W$, which is absurd.

Observe that by (6.15), $T_{p(x'_0)}$ is, in a neighborhood of $p(x'_0)$, equal to the graph of φ, that is, the points of the form $(x', \varphi(x'))$. Moreover, since $T_{p(x'_0)}$ is a supporting hyperplane of W, then $\varphi(x') \leq f(x')$ for all $x' \in V$.

In summary, we have that $\varphi, f, g : V \subset U \to \mathbb{R}$ are such that $\varphi(x') \leq f(x') \leq g(x')$, $\varphi(x'_0) = f(x'_0) = g(x'_0) = p(x'_0)$ and φ, g are differentiable at x'_0, therefore identifying the hyperplane $\{x_n = 0\}$ with $\mathbb{R}^{n-1}$, we can conclude (see the Exercise 5) that f is differentiable at x'_0 and that $\nabla f(x'_0) = \nabla \varphi(x'_0)$. Note that the equality between the gradients implies that the tangent hyperplane to f at x'_0 is $T_{p(x'_0)}$.

(iv) **∇f is Lipschitz on U**:

We will proceed by first proving that there exists $M > 0$ such that

$$|\nabla f(x')| \leq M \quad \text{for all } x' \in U. \tag{6.16}$$

It is sufficient to prove that for all $x' \in U$ and $j = 1, \dots, n-1$ we have $|(\partial f / \partial x_j)(x')| < 2$. If $(\partial f / \partial x_j)(x') = 0$ the bound is immediate, so let us assume without loss of generality that $(\partial f / \partial x_j)(x') \neq 0$. Given $(x', f(x'))$ we consider the tangent hyperplane $T_{p(x')}$, which we know has normal vector $(\nabla f(x'), 1)$, and we consider the line that passes through the point $(x', f(x'))$, lies in the hyperplane $T_{p(x')}$ and whose projection onto the first $(n-1)$ coordinates follows the direction e_j. This line is given by the direction vector

$$v = \left(0, \dots, 0, 1^{j)}, 0, \dots, \frac{\partial f}{\partial x_j}(x')\right)$$

and thus the aforementioned line can be written as

$$(x', f(x')) + \lambda \left(0, \ldots, 0, 1^{j)}, 0 \ldots, \frac{\partial f}{\partial x_j}(x')\right) \quad \lambda \in \mathbb{R}.$$

Let us now consider the intersection point of this line with the hyperplane $\{x_n = r\}$, which occurs when

$$\lambda = \widetilde{\lambda} := \frac{r - f(x')}{\frac{\partial f}{\partial x_j}(x')}.$$

Since the line cannot intersect $B((0, \ldots, 0, r), r)$ then it cannot intersect the set $B^{n-1}(0, r) \times \{r\}$ either, and in that case we must have that

$$|x' + \widetilde{\lambda} e_j| > r.$$

Now we can conclude that

$$\begin{aligned} &|\widetilde{\lambda} e_j| > r - |x'| \geq \frac{r}{2} \quad \Rightarrow \quad |\widetilde{\lambda}| > \frac{r}{2} \quad \Rightarrow \\ &\frac{|r - f(x')|}{\left|\frac{\partial f}{\partial x_j}(x')\right|} > \frac{r}{2} \quad \Rightarrow \quad \left|\frac{\partial f}{\partial x_j}(x')\right| < 2\frac{(r - f(x'))}{r} < 2 \end{aligned}$$

In the previous calculations recall that $f(x') \in [0, r)$ and that $|x'| < r/2$. We have already proved (6.16).

For any $x' \in U$, since $T_{p(x')}$ coincides with the tangent hyperplane to f, we have that the inner unit normal vector is given by

$$\nu(p(x')) = \frac{1}{\sqrt{1 + |\nabla f(x')|^2}} \left(- \frac{\partial f}{\partial x_1}(x'), \ldots, -\frac{\partial f}{\partial x_{n-1}}(x'), 1\right)$$

so if $\pi : \mathbb{R}^n \to \mathbb{R}^{n-1}$, $\pi(x', x_n) = x'$ is the orthogonal projection, we have

$$\pi(\nu(p(x'))) = \frac{-\nabla f(x')}{\sqrt{1 + |\nabla f(x')|^2}}.$$

Let the functions $\Phi, \Theta : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ be given by

$$\Phi(z) = \frac{-z}{\sqrt{1 + |z|^2}} \qquad \Theta(z) = \frac{-z}{\sqrt{1 - |z|^2}}$$

then $\Theta(\Phi(z)) = z$ for all $z \in \mathbb{R}^{n-1}$ and therefore

$$\nabla f(x') = \Theta(\Phi(\nabla f(x'))) = \Theta\left(\frac{-\nabla f(x')}{\sqrt{1 + |\nabla f(x')|^2}}\right) = \Theta(\pi(\nu(x', f(x')))).$$

The function Φ is clearly Lipschitz for being C^∞ and the function Θ is also Lipschitz on U because it is C^∞ on $\{|z| < 1\}$, therefore it has bounded partial derivatives in any compact subset of $\{|z| < 1\}$. In our case $z = \frac{-\nabla f(x')}{\sqrt{1+|\nabla f(x')|^2}}$ and therefore it is contained in a compact subset of $\{|z| < 1\}$ since

$$\left|\frac{-\nabla f(x')}{\sqrt{1 + |\nabla f(x')|^2}}\right| \leq \frac{M}{\sqrt{1 + M^2}} < 1.$$

Likewise, $x' \mapsto (x', f(x'))$ is also Lipschitz since we have already seen that $f \in C^1(U)$, and we already saw at the beginning of the proof that ν is in turn Lipschitz. Then $\nabla f(x')$ is a composition of four Lipschitz functions, and therefore is itself a Lipschitz function.

Putting together $(i) - (iv)$ we have thus proved that for all $p_0 \in \partial W$ there exists a neighborhood of p_0 where ∂W is the graph of a differentiable function f with Lipschitz differential, meaning that $f \in C^{1,1}$. Therefore ∂W is of class $C^{1,1}$.

□

The purpose of all these previous lemmas is to prove Theorem 6.10, which is essential for our proof of Alexandrov's Theorem.

Proof of Theorem 6.10. Let $\varepsilon > 0$. Let us take $M := \sup\{f(x) : x \in \overline{B}^n(0, 2R)\}$ and define

$$W := \{(x, y) \in \overline{B}^n(0, 2R) \times \mathbb{R} : f(x) \leq y \leq M + 2R\}.$$

Since $\overline{B}^n(0, 2R)$ is convex and f is convex, then $W \subset \mathbb{R}^{n+1}$ is a convex set. Furthermore, it is clearly closed, and has non-empty interior since $f(x) < M + 2R$ for all $x \in \overline{B}^n(0, 2R)$. Additionally, W is bounded, and therefore W is a convex body, as it is contained within the set bounded by the graph of f, the cylinder $\partial B^n(0, 2R) \times \mathbb{R}$ and the hyperplane $y = M + 2R$. By Lemma 6.16 there exists $0 < \delta < R$ such that

$$\mathcal{H}^n(\partial W \setminus \partial W(\delta)) < \varepsilon.$$

Let us now see that

$$\overline{B}^n(0, 2R) \times \{M + R\} \subset W(\delta).$$

Indeed, this is true because $W(\delta)$ is the union of closed balls of radius δ contained in W, and we have that

$$\overline{B}^n(0, 2R) \times \{M + R\} \subset \bigcup_a \overline{B}^{n+1}(a, \delta) \tag{6.17}$$

for $a \in \overline{B}^n(0, 2R - \delta) \times \{M + R\} \subset W$ (that these products are subsets of W is trivial since $f \leq M < M + R < M + 2R$). To see the inclusion (6.17), we take $(x, y) \in \overline{B}^n(0, 2R) \times \{M + R\}$ then $x \in \overline{B}^n(0, 2R)$ so

$$dist_{\mathbb{R}^n}(x, \overline{B}^n(0, 2R - \delta)) \leq \delta$$

and there exists $x' \in \overline{B}^n(0, 2R - \delta)$ such that $x \in \overline{B}^n(x', \delta)$. Then

$$(x, y) \in \overline{B}^n(x', \delta) \times \{M + R\} \subset \overline{B}^{n+1}((x', M + R), \delta),$$

where $(x', M + R) \in \overline{B}^n(0, 2R - \delta) \times \{M + R\}$.

Observe that (6.17) tells us that the intersection of $W(\delta)$ with the hyperplane $y = M + R$ is a closed n-dimensional ball of radius $2R$. Therefore, if $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is the orthogonal projection,

$$\pi(W(\delta)) = \overline{B}^n(0, 2R),$$

then for $x \in \overline{B}^n(0, 2R)$ we can define

$$g(x) := \inf\{y \in \mathbb{R} : (x, y) \in W(\delta)\}.$$

The function g is a function that parametrizes the lower part of $\partial W(\delta)$, where we refer to lower with respect to the hyperplane $\{(x, y) \in \mathbb{R}^n \times \mathbb{R} : y = 0\} \subset \mathbb{R}^{n+1}$. That g is convex is immediate by the convexity of $W(\delta)$. By Lemma 6.19, the boundary of $W(\delta)$ is of class $C^{1,1}$, and since g parametrizes such boundary this means that $g \in C^{1,1}_{loc}(B^n(0, 2R))$.

Note that $W(\delta) \subset W$ and by definition of W, $W \subset \text{epi}(f)$, therefore $W(\delta) \subset \text{epi}(f)$, and by definition of epi(f) this implies that $g \geq f$.

Finally, we will prove that (6.6) holds. Let $x \in B^n(0, R)$ such that $f(x) \neq g(x)$, and suppose that $x = \pi((x', y'))$ such that $(x', y') \in \partial W \cap \partial W(\delta)$, we arrive at a contradiction because if it belongs to ∂W then $y' = f(x)$, and if it belongs to $\partial W(\delta)$ then $y' = g(x)$, which means $f(x) = g(x)$. We conclude that

$$\{x \in B^n(0, R) : f(x) \neq g(x)\} \subset \pi(\partial W \setminus \partial W(\delta)),$$

then

$$\begin{aligned}\mathcal{L}^n(\{x \in B^n(0, R) : f(x) \neq g(x)\}) &\leq \mathcal{L}^n(\pi(\partial W \setminus \partial W(\delta))) = \mathcal{H}^n(\pi(\partial W \setminus \partial W(\delta))) \\ &\leq \mathcal{H}^n(\partial W \setminus \partial W(\delta)) < \varepsilon,\end{aligned}$$

where we used Theorem 4.5. □

6.3 Proof of Alexandrov's Theorem and some corollaries

Now, we have studied all the results that we are going to need to prove Alexandrov's Theorem.

Theorem 6.1 (Alexandrov's Theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then f is differentiable at almost every point, and at almost every point where f is differentiable there exists a symmetric matrix, which we denote $D^2 f(x)$, such that*

$$\lim_{y \to x} \frac{f(y) - f(x) - Df(x)(y-x) - \frac{1}{2}(y-x)^T D^2 f(x)(y-x)}{|y-x|^2} = 0. \tag{6.18}$$

We insist again that we cannot speak of f being twice differentiable in the usual sense, because we only guarantee the existence of the differential of f almost everywhere, and it would make no sense to say that a function that we only know is defined almost everywhere is differentiable. We will see in 6.23 that this is indeed true in the case where f is differentiable in the usual sense.

The main two ingredients of the proof are Lemma 6.9 and Theorem 6.10.

Proof of Theorem 6.1. Let $R > 0$ and $\varepsilon > 0$ and let g be the function obtained from Theorem 6.10, then g satisfies the hypotheses of Lemma 6.9 and

$$\mathcal{L}^n\big(\{f \neq g\}\big) < \varepsilon.$$

By Lemma 6.9, (6.18) is satisfied for almost every $x \in \{f = g\}$ by taking $D^2 f(x) := D^2 g(x)$. Then it is satisfied in $B(0, R)$ for almost every point of $\{f = g\}$, whose complement $\{f \neq g\}$ has measure less than ε. Since $\varepsilon > 0$ is arbitrary, taking $\varepsilon \to 0^+$ we conclude that (6.18) is true almost everywhere. □

For the case of convex and differentiable functions can indeed be guaranteed to be twice differentiable in the usual sense almost everywhere.

Corollary 6.22. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then, for every point $x \in \mathbb{R}^n$ where f satisfies* (6.18)*, the following holds*

$$\lim_{y \to x} \sup_{\sigma_y \in \partial f(y)} \frac{|\sigma_y - Df(x) - \langle D^2 f(x), y - x\rangle|}{|x-y|} = 0.$$

Proof. Let $x \in \mathbb{R}^n$ be a point where (6.18) is satisfied, and suppose without loss of generality and to simplify the notation, that $x = 0$. We need to show that

$$\lim_{x \to 0} \frac{\sigma_x - Df(0) - D^2 f(0) x}{|x|} = 0 \text{ for all } \sigma_x \in \partial f(x). \tag{6.19}$$

Thanks to (6.18), if $D^2 f(0)$ is the symmetric matrix given by Theorem 6.1 and we define

$$A = \frac{1}{2} D^2 f(0), \quad R(x) = f(x) - f(0) - Df(0)x - \langle Ax, x\rangle,$$

then $\lim_{x\to 0} \frac{|R(x)|}{|x|^2} = 0$. Therefore,

$$a(r) := \sup_{0 < |x| \leq 2r} \frac{|R(x)|}{|x|^2} \to 0 \text{ as } r \to 0^+. \tag{6.20}$$

Furthermore, it is immediate that

$$|R(x)| \leq a(|x|/2)|x|^2 \leq a(|x|)|x|^2 \tag{6.21}$$

Now we take $x, y \neq 0$, we can express $f(x)$ and $f(y)$ in the form

$$\begin{cases} f(x) = f(0) + Df(0)x + \langle Ax, x\rangle + R(x) \\ f(y) = f(0) + Df(0)y + \langle Ay, y\rangle + R(y) \end{cases},$$

and by the definition of the subdifferential, for all $\sigma_x \in \partial f(x)$ we know that $f(x) + \langle \sigma_x, y - x\rangle \leq f(y)$. It follows that for all $\sigma_x \in \partial f(x)$

$$\langle \sigma_x, y - x\rangle \leq f(y) - f(x) = Df(0)(y - x) + \langle A(x + y), y - x\rangle + R(y) - R(x), \tag{6.22}$$

where we use that since A is symmetric $\langle Ax, y\rangle = \langle Ay, x\rangle$, then $\langle A(x+y), y-x\rangle = \langle Ay, y\rangle - \langle Ax, x\rangle$. We will take in particular

$$y = x + w,$$

with $w = \sqrt{a(|x|)}|x|z$ for z with $|z| = 1$. Then from (6.22) it is deduced that for all $\sigma_x \in \partial f(x)$

$$\begin{aligned} \langle \sigma_x, w\rangle \leq Df(0)w + \langle A(2x + w), w\rangle + R(y) - R(x) \Longrightarrow \\ \langle \sigma_x - Df(0) - 2Ax, w\rangle \leq \langle Aw, w\rangle + R(y) - R(x) \Longrightarrow \\ \sqrt{a(|x|)}|x|\langle \sigma_x - Df(0) - 2Ax, z\rangle \leq a(|x|)|x|^2\langle Az, z\rangle + R(y) - R(x). \end{aligned} \tag{6.23}$$

Since we consider the limit $x \to 0$, we can take $|x|$ small enough so that (6.20) guaranties $a(|x|) \leq 1$, then $|w| \leq |x|$ and therefore $|y| \leq |x| + |w| \leq 2|x|$. Therefore, using (6.21) and the fact that the function a is clearly increasing,

$$\begin{aligned} |R(y)| \leq a\left(\frac{|y|}{2}\right)|y|^2 \leq a(|x|)(2|x|)^2 \leq 4a(|x|)|x|^2 \Longrightarrow \\ |R(y) - R(x)| \leq |R(y)| + |R(x)| \leq 5a(|x|)|x|^2. \end{aligned}$$

We now consider the following results from Functional Analysis, which generally refer to a Hilbert space but in our case we particularize them to $\mathbb{R}^n$:

(a) For all $v \in \mathbb{R}^n$ we have $|v| = \sup\{\langle v, u\rangle : |u| = 1\}$.

(b) If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a self-adjoint linear operator, then $\|T\| = \sup\{|\langle Tu, u\rangle| : |u| = 1\}$.

We do not include their proofs, which are available in most Functional Analysis textbooks. In our case, since A is a symmetric and real matrix, it is self-adjoint, so identifying A with the linear operator in $\mathbb{R}^n$ that it determines, we have by (b) that $\|A\| = \sup\{|\langle Au, u\rangle| : |u| = 1\}$. We can then take supremums in (6.23) with respect to all z with $|z| = 1$, first on the left side of the inequality and then on the right, to conclude that for all $\sigma_x \in \partial f(x)$

$$\sqrt{a(|x|)}|x||\sigma_x - Df(0) - 2Ax| \leq \|A\|a(|x|)|x|^2 + R(y) - R(x) \leq \|A\|a(|x|)|x|^2 + 5a(|x|)|x|^2.$$

This finishes the proof since $2A = D^2 f(0)$, so we have proven that for all $\sigma_x \in \partial f(x)$

$$\frac{|\sigma_x - Df(0) - D^2 f(0)x|}{|x|} \leq (\|A\| + 5)\sqrt{a(|x|)} \to 0 \text{ as } x \to 0.$$

□

The following corollary is immediately deduced from Theorem 6.22 using the fact that we know, by Proposition 6.6, that in the case that f is differentiable, $\partial f(x) = \{\nabla f(x)\}$.

Corollary 6.23. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable on $\mathbb{R}^n$, then it is twice differentiable in the usual sense at almost every point $x \in \mathbb{R}^n$.*

Exercises

1. Prove that if $C \subset \mathbb{R}^n$ is a convex set with nonempty interior, then
$$\text{int}\big(\overline{C}\big) = \text{int}(C).$$

2. Let $C \subset \mathbb{R}^n$ be a convex set so that $C \neq \mathbb{R}^n$. Prove that for every $x \in \partial C$, there exists a hyperplane T_x which supports C at x. That is, show that there is a hyperplane
$$T_x = \{y \in \mathbb{R}^n : \langle v, y\rangle = b\} \quad \text{for some } v \in \mathbb{R}^n, b \in \mathbb{R},$$
so that $x \in T_x \cap \overline{C}$ and either $C \subset \{y : \mathbb{R}^n : \langle v, y\rangle \le b\}$ or $C \subset \{y : \mathbb{R}^n : \langle v, y\rangle \le b\}$.

3. Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ so that $\partial f(0) = [-1, +\infty)$.

4. If $U \subset \mathbb{R}^n$ is open and convex and $f : U \to \mathbb{R}$ is a convex function, prove that
$$\partial f(x) \subset \{v \in \mathbb{R}^n : f(y) \ge f(x) + \langle v, y - x\rangle \text{ for all } y \in U\}.$$

5. Let $f, g, h : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, so that $g \le f \le h$ and let $x_0 \in \Omega$ so that $f(x_0) = g(x_0) = h(x_0)$ and g, h are differentiable at x_0. Then f is differentiable at x_0 and $\nabla f(x_0) = \nabla g(x_0) = \nabla h(x_0)$.

6. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a self-adjoint linear operator, then show that $\|T\| = \sup\{|\langle Tu, u\rangle| : |u| = 1\}$.

7. Give and example of a Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, differentiable everywhere, so that ∇f is nowhere differentiable.

Bibliography

[1] G. Alberti, *On the structure of singular sets of convex functions*, Calc. Var. Partial Differential Equations 2 (1994), 17-27.

[2] A. D. Aleksandrov, *Almost everywhere existence of the second differential of a convex functions and related properties of convex surfaces*, Uchenye Zapisky Leningrad. Gos. Univ. Math. Ser., 37 (1939), 3-35 (in russian).

[3] D. Azagra, A. Capello, P. Hajłasz, *A geometric approach to second order differentiability of convex functions*, Proc. Am. Math. Soc. Ser. B 10, 382-397 (2023).

[4] D. Azagra, M. Drake, P. Hajłasz, C^2*-Lusin approximation of strongly convex functions*, Inventiones Math. 236 (2024), 1055-1082.

[5] D. Azagra, P. Hajłasz, *Lusin-type properties of convex functions and convex bodies*, J. Geom. Anal. 31, 11685-11701 (2021).

[6] G. Bianchi, A. Colesanti, C. Pucci, *On the second differentiability of convex surfaces*, Geom. Dedicata 60 (1996), 39-48.

[7] S. M. Bates, *Toward a precise smoothness hypothesis in Sard's theorem*, Proc. Amer. Math. Soc. 117 (1993), no. 1, 279–283.

[8] S. M. Bates and C. G. Moreira, *De nouvelles perspectives sur le théorème de Morse-Sard*, C.R. Acad. Sci. Paris, t. 332, Série I (2001), 13–17.

[9] D. Cruz-Uribe and A. Fiorenza, *Variable Lebesgue Spaces: Foundations and Harmonic Analysis.* Springer (2013). `https://doi.org/10.1007/978-3-0348-0548-3`

[10] L. De Pascale, *The Morse-Sard theorem in Sobolev spaces*, Indiana Univ. Math. J. 50 (2001), 1371–1386.

[11] L. C. Evans, *Partial differential equations* (second edition) Gradiate Textd in Mathematics, Volume 19, Amer. Mtah. Soc. 2010.

[12] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[13] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*. Second edition, John Wiley & Sons, 2004.

[14] M. Fabian, P. Habala, P, Hájek, V. Montesinos and V. Zizler, *Banach space theory. The basis for linear and nonlinear analysis*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011.

[15] E.L. Grinberg, *On the Smoothness Hypothesis in Sard's Theorem*. The American Mathematical Monthly, 92(10), 733–734, 1985.

[16] P. Hajłasz, Notes *Geometric Analysis*, University of Pittsburgh. `https://sites.pitt.edu/~hajlasz/Notatki/Analysis_4.pdf`

[17] P. Hajłasz, Notes *Sobolev spaces*, University of Pittsburgh. `https://sites.pitt.edu/~hajlasz/Notatki/Sobolev%20Spaces.pdf`

[18] J. Heinonen, P. Koskela, N. Shanmugalingam, and J.T. Tyson, *Sobolev classes of Banach space valued functions and quasiconformal mappings*, J. Anal. Math., 85 (2001), 87-139.

[19] J. Kinunen, Notes *Sobolev spaces*, University of Aalto, 2020.

[20] I. Kupka, *Counterexample to the Morse-Sard theorem in the case of infinite-dimensional manifolds*, Proc. Amer. Math. Soc. 16 (1965), 954–957.

[21] A. Pelczyński and M. Wojciechowski, *Sobolev spaces, Handbook of the Geometry of Banach spaces*, Vol. 2, W.B. Johnson and J. Lindenstrauss editors, Elsevier Amsterdam (2003), 1361-1423.

[22] A. Morse, *The behavior of a function on its critical set*, Annals of Math. 40 (1939), 62–70.

[23] W. Rudin, *Real and Complex Analysis*, Mc Graw Hill International, third edition 1987 (first in 1966) Pp. 428).

[24] A. Sard, *The measure of the critical values of differentiable maps*, Bull. Amer. Math. Soc. 48 (1942), 883–890.

[25] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton Mathematical Series, vol. 30, Princeton University Press, Princeton, N.J., 1970.

[26] H. Whitney, *Analytic extensions of differentiable functions defined in closed sets*, Trans. Amer. Math. Soc. 36 (1934), 63–89.

[27] H. Whitney, *A function not constant on a connected set of critical points*, Duke Math. J. 1 (1935), 514–517.