The role of measure zero in analysis

Hester Pieters

Bachelor's thesis for mathematics
Supervisor: Dr. M. Müger

January 26, 2009

1 Introduction

The aim of this thesis is to prove some theorems in analysis where the notion of measure zero is important without using Lebesgue integration. In the first chapter we will prove some properties of sets of measure zero. The other three chapters will eacht treat one theorem. The first is Lebesgue's Criterion, this states that a function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if it is bounded and continuous almost everywhere, i.e. the set of points where f is discontinuous has measure zero. The second is a theorem by Lebesgue about differentiablility. It states that a continuous monotonuous function $f : [a, b] \to \mathbb{R}$ is differentiable almost everywhere. In the last chapter we will prove Sard's Theorem. This theorem is very important in differential topology and differential geometry. It states the following: If $f : \mathbb{R}^m \to \mathbb{R}^n$ is a smooth function then the set $C \subset \mathbb{R}^n$ of critical values of f has measure zero.

2 Sets of Measure Zero

Definition 1. *A subset $A \subset \mathbb{R}$ has* ***measure zero*** *if for every $\epsilon > 0$, there exists a countable covering of A by (open, bounded) intervals (a_i, b_i) such that the sum of the lengths of these intervals is less than ϵ.*

We can extend this to higher dimensions.

Definition 2. *A* ***cube*** *of edge length l in $\mathbb{R}^n$ is a product $C = \prod_{i=1}^n [a_i, b_i]$ of n intervals in $\mathbb{R}$ with $|a_i - b_i| = l$ for all i. The volume of the cube C is denoted by $|C|$ and defined to be $|C| = l^n$.*

Definition 3. *A subset $A \subset \mathbb{R}^n$ has* ***measure zero*** *if for every $\epsilon > 0$, there exists a countable covering of A by cubes C_i such that the sum of the volumes of C_i is less than ϵ. We arrive at the same notion of measure zero if we replace closed by open cubes or boxes.*

Definition 4. *If a property holds at all points of a set X except possibly the points of a set of measure zero, we say that this property holds* ***almost everywhere*** *on X or at* ***at almost every point*** *of X.*

To prove Sard's theorem we will need some properties of sets of measure zero:

Lemma 1. *A countable union of sets of measure zero has measure zero.*

Proof. Let $(U_i \subset R^n)_{i \in \mathbb{N}}$ be a sequence of sets of measure zero and let $\epsilon > 0$. Since U_i has measure zero it can be covered by a sequence $\{C_i^j ; j \in \mathbb{N}\}$ of cubes such that $\sum_{j=1}^{\infty} |C_i^j| < 2^{-i}\epsilon$. Then $\{C_i^j \mid i, j \in \mathbb{N}\}$ is a countable cover of $\bigcup_{i=1}^{\infty} U_i$ and $\sum_{i,j} |C_i^j| < \epsilon \sum_i 2^{-i} = \epsilon$. □

So, in particular, all countable sets have measure zero. There are however also uncountable sets of measure zero. An example is the Cantor set. Let $C_0 = [0, 1]$ and let C_1 denote the set obtained from deleting the middle third open interval from [0,1], that is,

$$C_1 = [0, 1/3] \cup [2/3, 1].$$

Next, we repeat this procedure for each sub-interval of C_1 and obtain

$$C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1].$$

Repeating this procedure yields a sequence $C_k, k = 0, 1, 2, ...$ of compact sets with

$$C_0 \supset C_1 \supset C_2 \supset ....$$

The Cantor set $\mathcal{C}$ is by definition the intersection of all C_k's:

$$\mathcal{C} = \bigcap_{k=0}^{\infty} C_k.$$

The set $\mathcal{C}$ is not empty, since all end-points of the intervals C_k belong to $\mathcal{C}$. It can even be mapped to the interval $[0,1]$, so the Cantor set has the cardinality of the continuum and is therefore definitely not countable. From the construction of $\mathcal{C}$, we know that $\mathcal{C} \subset C_k$ for all k. Each C_k is a disjoint union of 2^k closed intervals, each of length 3^{-k}. So $|C_k| = (2/3)^k$. Since $(2/3)^k \to 0$ as k tends to infinity, we conclude that the Cantor set can be covered by intervals of arbitrary small total length and hence it has measure zero.

Lemma 2. *Let $U \subset \mathbb{R}^m$ be open and $f : U \to \mathbb{R}^m$ a C^1 map. If $A \subset U$ has measure zero then $f(A) \subset \mathbb{R}^m$ has measure zero.*

Proof. Since A has measure zero there exists a sequence of cubes $\{C_i \subset \mathbb{R}^m\}_{i\in\mathbb{N}}$ such that $A \subset \bigcup_{i=1}^{\infty} C_i$. If we now show that $f(A \cap C_i)$ has measure zero, then $f(A)$ is a countable union of sets of measure zero and hence by the preceding lemma has measure zero itself. Let $\|\cdot\|$ be the euclidean norm on $\mathbb{R}^m$. Since C_i is compact, there exists a $M > 0$ such that $\|Df(x)\| \leq M$ for all $x \in C_i$. Then

$$\|f(x) - f(y)\| \leq M\|x - y\|,$$

for all $x, y \in C_i$. So if C_i has edge length l then the image of C_i is contained in a m-cube of edge length $l\sqrt{m}M$. It follows that $f(C_i)$ has measure zero if C_i has measure zero. □

Lemma 3. *If $U \subset \mathbb{R}^n$ has measure zero then any $V \subset U$ has measure zero.*

Proof. Since U has measure zero we can pick a sequence $C_i, i \in \mathbb{N}$ of cubes such that $U \subset \bigcup_{i=1}^{\infty} C_i$ and $\sum_{i=1}^{\infty} |C_i| < \epsilon$. Since V is contained in U it can be covered by the same C_i, so V also has measure zero. □

Lemma 4. *If $m < n$ then $\mathbb{R}^m \cong \mathbb{R}^m \times \{0\} \subset \mathbb{R}^n$ has measure zero.*

Proof. Let $\epsilon > 0$ and let $I_{(x_1,\dots,x_m)} = [x_1, x_1 + 1] \times \dots \times [x_m, x_m + 1]$, where $(x_1, \dots, x_m) \in \mathbb{Z}^m$. Then $I_{(x_1,..,x_m)} \times \{0\}$ is covered by $I_{(x_1,\dots,x_m)} \times [-\epsilon, \epsilon]^{n-m}$. Since $|I_{(x_1,..,x_m)} \times [-\epsilon, \epsilon]^{n-m}| = \epsilon^{n-m}$ and ϵ is arbitrary it follows that $I_{(x_1,..,x_m)} \times \{0\}$ has measure zero. We conclude that $\mathbb{R}^m \times \{0\} \subset \bigcup_{x\in\mathbb{Z}^m} I_x \times \{0\}$ is a subset of a countable union of sets of measure zero and therefore has measure zero itself. □

Lemma 5. *Let $U \subset \mathbb{R}^m$ be open and $f : U \to \mathbb{R}^n$ a C^1 map, where $n > m$. Then $f(U) \subset \mathbb{R}^n$ has measure zero.*

Proof. Define $\tilde{f} : U \times \mathbb{R}^{n-m} \to \mathbb{R}^n$ by $\tilde{f}(x, y) = f(x)$. Since $U \times \{0\} \subset \mathbb{R}^n$ has measure zero, Lemma 2 implies that $f(U) = \tilde{f}(U \times \{0\}) \subset \mathbb{R}^n$ has measure zero. □

3 Riemann Integrability

Theorem 1. *(Lebesgue's Criterion) A function $f : [a,b] \to \mathbb{R}$ is Riemann integrable iff f is bounded and continuous almost everywhere.*

For $f : [a,b] \to \mathbb{R}$ we define

$$S(f) = \{x \in [a,b] \mid f \text{ is not continuous at x}\}.$$

So Lebesgue's criterion says that f is Riemann integrable iff S(f) has measure zero. Before we prove this, we first recall some definitions.

Definition 5. *A* ***partition*** *P of $[a,b]$ is a finite system of points $x_0, \dots, x_n$ such that $a = x_0 < x_1 < \dots < x_n = b$. We write $\Delta_i = [x_{i-1}, x_i]$ and $\Delta x_i = x_{i-1} - x_i$. The* ***mesh*** *$\lambda(P)$ of a partition is the largest of the lengths of the Δ_i, where $i = 1, \dots, n$.*

Definition 6. *A function $f : [a, b] \to \mathbb{R}$ is* ***Riemann integrable*** *(over the interval $[a, b]$) if there exists $V \in \mathbb{R}$ (easily seen to be unique) such that for every $\epsilon > 0$ there is a $\delta > 0$ such that for any partition P with $\lambda(P) < \delta$ and any $\xi_j \in \Delta_j$:*

$$\left| \sum_{j=1}^{n} f(\xi_j)\Delta x_j - V \right| < \epsilon,$$

In this case we write $\int_a^b f(x)dx = V$.

Definition 7. *For $U \subset \mathbb{R}$ the* ***oscillation*** *of $f : U \to \mathbb{R}$ on U is defined as*

$$\omega(f, U) = \sup_{x,y \in U} |f(x) - f(y)|.$$

For $f : [a, b] \to \mathbb{R}$ the oscillation at the point $x \in [a, b]$ is defined as

$$\omega(f, x) = \inf_{\epsilon > 0} \omega(f, [a, b] \cap (x - \epsilon, x + \epsilon)).$$

Furthermore, we define

$$S_\epsilon(f) = \{x \in [a, b] \mid \omega(f, x) > \epsilon\}.$$

Note that $S_0(f) = S(f)$.

The above definition for Riemann integrability is equivalent to another criterion given below.

Proposition 1. *A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable iff*

$$\lim_{\lambda(P) \to 0} \sum_{i=1}^{n} \omega(f, \Delta_i)\Delta x_i = 0$$

Proof. $\Rightarrow$: We remark that $\omega(f, \Delta_i) = \sup_{x \in \Delta_i} f(x) - \inf_{x \in \Delta_i} f(x)$. Since f is Riemann integrable

$$\lim_{\lambda(P) \to 0} \sum_{j=1}^{n} f(\xi_j)\Delta x_j = \int_a^b f(x)dx = V,$$

for any $\xi_j \in \Delta_j$. We immediately see that we must have

$$\lim_{\lambda(P) \to 0} \sum_{j=1}^{n} \sup_{x \in \Delta_j} f(x)\Delta x_j = \lim_{\lambda(P) \to 0} \sum_{j=1}^{n} \inf_{x \in \Delta_j} f(x)\Delta x_j = V.$$

So $\lim_{\lambda(P) \to 0} \sum_{i=1}^{n} \omega(f, \Delta_i)\Delta x_i = \lim_{\lambda(P) \to 0} \sum_{i=1}^{n} \sup_{x \in \Delta_i} f(x) - \inf_{x \in \Delta_i} f(x) = 0$.
$\Leftarrow$: Denote the Riemann sum by $\sigma(f; P, \xi) := \sum_{i=1}^{n} f(\xi_i)\Delta x_i$. Let P be a partion of $[a, b]$ and let $\tilde{P}$ be a refinement of P. We denote the intervals of P by $\Delta_i = [x_{i-1}, x_i]$ and the intervals of the refinement that are included in this interval by $\Delta_{ij} := [x_{ij-1}, x_{ij}]$. Thus $\Delta x_i = \Delta x_{i1} + ... + \Delta x_{in_i}$. The difference between the Riemann sums $\sigma(f; \tilde{P}, \tilde{\xi}) - \sigma(f; P, \xi)$ can be estimated by

$$\begin{aligned}
|\sigma(f; \tilde{P}, \tilde{\xi}) - \sigma(f; P, \xi)| &= \left| \sum_{i=1}^{n} \sum_{j=1}^{n_i} f(\xi_{ij})\Delta x_{ij} - \sum_{i=1}^{n} f(\xi_i)\Delta x_i \right| \\
&= \left| \sum_{i=1}^{n} \sum_{j=1}^{n_i} (f(\xi_{ij}) - f(\xi_i))\Delta x_{ij} \right| \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n_i} |f(\xi_{ij}) - f(\xi_i)|\Delta x_{ij} \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n_i} \omega(f; \Delta_i)\Delta x_{ij} = \sum_{i=1}^{n} \omega(f; \Delta_i)\Delta x_i.
\end{aligned}$$

So for any $\epsilon > 0$ we can find a $\delta > 0$ such that if $\lambda(P) < \delta$ the difference between the Riemann sums must satisfy

$$|\sigma(f; \tilde{P}, \tilde{\xi}) - \sigma(f; P, \xi)| < \frac{\epsilon}{2}.$$

Let P', P'' be arbitrary partitions on $[a, b]$ whose meshes satisfy $\lambda(P') < \delta$ and $\lambda(P'') < \delta$. Then the partition $\tilde{P} = P' \cup P$ is a refinement of both of them so it must satisfy

$$|\sigma(f; \tilde{P}, \tilde{\xi}) - \sigma(f; P', \xi')| < \frac{\epsilon}{2},$$
$$|\sigma(f; \tilde{P}, \tilde{\xi}) - \sigma(f; P'', \xi'')| < \frac{\epsilon}{2}.$$

It follows that

$$|\sigma(f; P', \xi') - \sigma(f; P'', \xi'')| < \epsilon.$$

Therefore, by the Cauchy criterion, the limit of the Riemann sum exists:

$$\lim_{\lambda(P)\to 0} \sum_{j=1}^{n} f(\xi_j)\Delta x_j = \int_a^b f(x)dx = V.$$

□

Proposition 2. *A Riemann integrable function $f : [a, b] \to \mathbb{R}$ is bounded.*

Proof. Assume that f is not bounded on $[a, b]$. Then in every partition P there exists at least one interval $[x_{i-1}, x_i]$ on which f is not bounded. This implies that by choosing the point ξ_i in different ways, we can make $|f(\xi_i)\Delta x_i|$ as large as desired. It follows that the absolute value of the Riemann sum $\sum f(\xi_i)\Delta x_i$ can be made arbitrary large and therefore it cannot converge to a finite limit. □

So from now on we can always assume that f is bounded. To prove Lebesgue's criterion we will first show that it is equivalent to another criterion, the criterion of du Bois-Reymond. This states that every $S_\epsilon(f)$ admits a finite cover of open intervals that has arbitrary small length. Then we will prove that the criterion of du Bois-Reymond is indeed a criterion for Riemann integrability and therefore so is Lebesgue's criterion.

Proposition 3. *Let $f : [a, b] \to \mathbb{R}$ be bounded. Then, for every $\epsilon, \delta > 0$, there is a finite sequence of intervals $(a_i, b_i), i = 1, ..., m$ such that*

$$S_\epsilon(f) \subset \bigcup_{i=1}^{m} (a_i, b_i) \qquad \textit{and} \qquad \sum_{i=1}^{m} b_i - a_i < \delta$$

iff S(f) has measure zero.

Proof. $\Rightarrow$: Let $\delta > 0$. For every $n \in \mathbb{N}$ there are intervals $(a_{n,1}, b_{n,1}), ..., (a_{n,m(n)}, b_{n,m(n)})$ such that:

$$S_{2^{-n}}(f) \subset \bigcup_{i=1}^{m(n)} (a_{n,i}, b_{n,i}) \qquad \text{and} \qquad \sum_{i=1}^{m(n)} b_{n,i} - a_{n,i} < 2^{-n}\delta.$$

Since $S(f) = S_0(f) = \bigcup_{n\in\mathbb{N}} S_{2^{-n}}(f)$, we find that S(f) is contained in $\{(a_{n,i}, b_{n,i}) \mid n \in \mathbb{N}, i = 1, ..., m(n)\}$, where the total length of this last set is bounded by $\sum_{n=1}^{\infty} 2^{-n}\delta = \delta$. We conclude that S(f) has measure zero.
$\Leftarrow$: Let $\delta > 0$, since S(f) has measure zero we can pick a sequence of intervals $\{(a_i, b_i) \mid i \in \mathbb{N}\}$ such that

$$S(f) \subset \bigcup_{i=1}^{\infty} (a_i, b_i) \qquad \text{and} \qquad \sum_{i=1}^{\infty} b_i - a_i < \delta.$$

We are done if we can show that $S_\epsilon(f)$ is compact for any $\epsilon > 0$, since by $S_\epsilon(f) \subset S_0(f) = S(f)$ it is then covered by a finite subfamily of $\{(a_i, b_i) \mid i \in \mathbb{N}\}$. For this we need to prove that $S_\epsilon(f)$ is closed. (It is clearly bounded.)
To this end, let $x \in [a, b]$ be a limit point of $S_\epsilon(f)$. This means that every neighborhood of x contains a x' with $\omega(f, x') > \epsilon$. Thus, by definition, every neighborhood of x' contains an x'' such that$|f(x') - f(x'')| > \epsilon$. It follows that $\omega(f, x) > \epsilon$, thus $x \in S_\epsilon(f)$ and we conclude that $S_\epsilon(f)$ is closed. □

For the next proposition we define a function $\theta : \mathbb{R} \to \mathbb{R}$ by

$$\theta(x) = \begin{cases} 0, \text{ if } x \leq 0 \\ 1, \text{ if } x > 0 \end{cases}$$

Proposition 4. *Let $f : [a, b] \to \mathbb{R}$ be bounded. Then f is Riemann integrable iff for every $\epsilon, \delta > 0$ there is a finite sequence of intervals $(a_i, b_i), i = 1, ..., m$ such that*

$$S_\epsilon(f) \subset \bigcup_{i=1}^{m} (a_i, b_i) \qquad \textit{and} \qquad \sum_{i=1}^{m} b_i - a_i < \delta \tag{1}$$

Proof. ⇒: Let $\epsilon, \delta > 0$. We use the criterion for Riemann integrability of Proposition 1:
$\lim_{\lambda(P)\to 0} \sum_{i=1}^{n} \omega(f, \Delta_i)\Delta x_i = 0$. Considering now the contribution to $\sum \omega(f, \Delta_i)\Delta x_i$ of the intervals Δ_i on which the oscillation of f is larger then ϵ, we obtain

$$\epsilon \sum_{i=1}^{n} \theta(\omega(f, \Delta_i) - \epsilon)\Delta x_i \leq \sum_{i=1}^{n} \omega(f, \Delta_i)\Delta x_i.$$

We see that we can make $\epsilon \sum_i \theta(\omega(f, \Delta_i) - \epsilon)\Delta x_i$ arbitrary small, i.e. there exists a partition P such that

$$\sum_{i=1}^{n} \theta(\omega(f, \Delta_i) - \epsilon)\Delta x_i < \delta.$$

Let $I = \{i \in \{1, ..., n\} \mid \omega(f, \Delta_i) > \epsilon\}$. Then $\{(x_{i-1}, x_i) \mid i \in I\}$ is a finite family of open intervals and by the above inequality we see that its total length is less than δ. For $i \notin I$ we have $\omega(f, \Delta_i) \leq \epsilon$, so $\omega(f, x) \leq \epsilon$ for all $x \in \Delta_i$. It follows that if $i \notin I$ and $x \in \Delta_i$, then $x \notin S_\epsilon(f)$. We see that

$$S_\epsilon(f) \subset \bigcup_{i\in I} (x_{i-1}, x_i) \qquad \text{and} \qquad \sum_{i\in I} x_{i-1} - x_i < \delta.$$

⇐: Let $\epsilon, \delta > 0$. We have

$$\sum_{i=1}^{n} \omega(f, \Delta_i)\Delta x_i = \sum_{i=1}^{n} \theta(\epsilon - \omega(f, \Delta_i))\omega(f, \Delta_i)\Delta x_i + \sum_{i=1}^{n} \theta(\omega(f, \Delta_i) - \epsilon)\omega(f, \Delta_i)\Delta x_i. \tag{2}$$

If $\theta(\epsilon - \omega(f, \Delta_i)) \neq 0$ then $\omega(f, \Delta_i) < \epsilon$ so the first term is smaller than $\sum_i \epsilon\Delta x_i = \epsilon \cdot (b - a)$. Since f is bounded we always have $\omega(f, \Delta_i) \leq \omega(f, [a, b]) < \infty$, so the second term is bounded by

$$\sum_{i=1}^{n} \theta(\omega(f, \Delta_i) - \epsilon)\omega(f, [a, b])\Delta x_i.$$

Pick now a sequence $(a_1, b_1), ..., (a_m, b_m)$ such that (1) holds for $S_{\frac{\epsilon}{2}}$. We may assume all intervals to be disjoint and having no boundary points in common (if we have (a,b) and (b,c) then replacing these by (a,c) leaves the total length unchanged). Let $J = [a, b] - \bigcup_i (a_i, b_i)$. For $x \in J$ we have $\omega(f, x) \leq \frac{\epsilon}{2}$. It follows by the definition of $\omega(f, x)$ that there exists an interval $I_x = (x - \alpha_x, x + \alpha_x)$ such that $\omega(f, [a, b] \cap I_x) \leq \epsilon$. J is compact, so it can be covered by a finite number of such intervals: $J \subset \bigcup_{i=1}^{k} I_{x_i}$. We conclude that there exists a partition $P = \{a = x_0 < x_1 < ... < x_n = b\}$ that contains the intervals (a_i, b_i) and such that $\omega(f, \Delta_i) \leq \epsilon$ for the remaining intervals of the partition. We see that these remaining intervals do not contribute to the second term in (2). It follows that this term is less than $\delta \cdot \omega(f, [a, b])$. We conclude

$$\sum_{i=1}^{n} \omega(f, \Delta_i)\Delta x_i < \epsilon \cdot (b - a) + \delta \cdot \omega(f, [a, b]),$$

which proves the proposition, since ϵ, δ were arbitrary. □

Lebesgue's criterion also holds in $\mathbb{R}^n$. The proof is quite analogous to the one above, see for example [8].

4 Differentiability

Theorem 2. *If $f : [a, b] \to \mathbb{R}$ is continuous and monotonic, then f is differentiable almost everywhere (on $[a, b]$).*

We begin by proving two useful lemmas about open subsets of the real line.

Lemma 6. *Every open subset $U \subset \mathbb{R}$ can be written uniquely as a countable union of disjoint open intervals.*

Proof. Choose any $x \in U$. Since U is open, x is contained in some interval, and therefore if

$$a_x = \inf\{a < x \mid (a, x] \subset U\} \qquad \text{and} \qquad b_x = \sup\{b > x \mid [x, b) \subset U\},$$

the point a_x cannot be in U. Since if it was, there would be a neighborhood of a_x contained in U and we could find an $a < a_x$ for which $(a, x] \subset U$. The point b_x cannot be in U either, we must have $a_x < x < b_x$ (with possibly $a_x = -\infty$ and/or $b = \infty$) and we see that $I_x = (a_x, b_x)$ is the largest open interval containing x and contained in U. Hence

$$U = \bigcup_{x \in U} I_x.$$

If x and y are two points in U, then either $I_x = I_y$ or $I_x \cap I_y = \emptyset$, so U is a union of disjoint open intervals. It is easy to see that there are at most countably many open intervals, since in every open interval I_x we can find a distinct rational number. □

Lemma 7. *Every open subset $U \subset \mathbb{R}$ can be written as a countable union of (almost) disjoint closed intervals.*

Proof. We can construct such a union by the following procedure. Subdivide the real line in closed intervals of length 1. If an interval is entirely contained in U accept it as part of the union. If an interval is entirely in U^c reject it. The remaining ones, which intersect both U and U^c, are divided into two intervals of half the length. We repeat this procedure indefinitely. By contruction, the remaining set of intervals is countable and contained in U. To see that all of U is contained in this set, we not that for $x \in U$ there exists an interval of length $2^{-N}, N \in \mathbb{N}$ that contains x and is entirely contained in U. Either this interval has been accepted or it is contained in an interval that has been previously accepted. □

Before we can prove Theorem 2 we will need to develop a bit more of measure theory than just the notion of measure zero.

Definition 8. *Let $E \subset \mathbb{R}$ and let $\mathcal{C}$ be the collection of all countable covers of S consisting of pairwise disjoint open intervals. The* ***Lebesgue outer measure*** *of S, $m_e(E)$, is the infimum over $C \in \mathcal{C}$ of the sum of the lengths of the pairwise disjoint open intervals that constitute C,*

$$m_e(E) = \inf_{C \in \mathcal{C}} m(C).$$

Equivalently, we can define the outer measure to be the infimum over $C \in \mathcal{C}$ of the sum of the lengths of the pairwise disjoint closed intervals that constitute C.

Definition 9. *The **distance** between two sets E and F is defined by*

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

Lemma 8. *The outer measure has the following properties:*

1. *(Monotonicity of outer measure.) If $E_1 \subset E_2 \subset \mathbb{R}$, then $m_e(E_1) \leq m_e(E_2)$.*
2. *(Sub-additivity of outer measure.) If $E = \bigcup_{i=1}^{\infty} E_i$, then $m_e(E) \leq \sum_{i=1}^{\infty} m_e(E_i)$.*
3. *(Outer regularity of outer measure.)If $E \subset \mathbb{R}$, then $m_e(E) = \inf m_e(\mathcal{O})$, where the infimum is taken over all open sets $\mathcal{O}$ containing E.*
4. *If $E = E_1 \cup E_2$ and $d(E_1, E_2) > 0$, then $m_e(E) = m_e(E_1) + m_e(E_2)$.*

Proof. 1. This follows immediately from the definition of outer measure, since any countable covering of E_2 is also a covering of E_1.

2. If there is an i such that $m_e(E_i) = \infty$, then the inequality clearly holds. So we may assume $m_e(E_i) < \infty$ for all i. Let $\epsilon > 0$. For each i there exists a covering of E_i by open intervals $(a_{i,j}, b_{i,j})_{j \in \mathbb{N}}$ such that

$$\sum_{j=1}^{\infty} b_{i,j} - a_{i,j} \leq m_e(E_i) + \frac{\epsilon}{2^i}.$$

Then, $E \subset \bigcup_{i,j=1}^{\infty}(a_{i,j}, b_{i,j})$ is a countable covering of E and we have

$$\begin{aligned} m_e(E) &\leq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} b_{i,j} - a_{i,j} \\ &\leq \sum_{i=1}^{\infty}(m_e(E_i) + \frac{\epsilon}{2^i}) \\ &= \sum_{i=1}^{\infty} m_e(E_i) + \epsilon. \end{aligned}$$

3. Since $E \subset \mathcal{O}$, monotonicity immediately implies $m_e(E) \leq \inf(\mathcal{O})$. For the reverse inequality, let $\epsilon > 0$. We can choose intervals $(a_i, b_i)_{i \in \mathbb{N}}$ such that $E \subset \bigcup_i (a_i, b_i)$ and

$$\sum_i b_i - a_i \leq m_e(E) + \epsilon.$$

The outer measure of (a_i, b_i) is cleary $b_i - a_i$ so, by countable sub-additivity, we have

$$\inf m_e(\mathcal{O}) \leq m_e(\bigcup_i (a_i, b_i)) \leq \sum_i b_i - a_i \leq m_e(E) + \epsilon.$$

4. From property 2 we immediately see that $m_e(E) \leq m_e(E_1) + m_e(E_2)$. For the reverse inequality, let $\epsilon > 0$ and select $\delta > 0$ such that $d(E_1, E_2) > \delta$. Pick a covering $[a_i, b_i]_{i \in \mathbb{N}}$ of E such that $\sum_i b_i - a_i \leq m_e(E) + \epsilon$. We can subdivide these intervals into intervals of length less than δ. It follows that each of these intervals can intersect at most one of the two sets E_1, E_2. So there exist two disjoint sets I_1, I_2 such that

$$E_1 \subset \bigcup_{i \in I_1} [a_i, b_i] \qquad \text{and} \qquad E_2 \subset \bigcup_{i \in I_2} [a_i, b_i].$$

We find

$$\begin{aligned} m_e(E_1) + m_e(E_2) &\leq \sum_{i \in I_1} b_i - a_i + \sum_{i \in I_2} b_i - a_i \\ &\leq \sum_{i=1}^{\infty} b_i - a_i \\ &\leq m_e(E) + \epsilon. \end{aligned}$$

□

Definition 10. *A subset S of $\mathbb{R}$ is* ***Lebesgue measurable****, or simply* ***measurable****, if for any $\epsilon > 0$ there exists an open set $\mathcal{O}$ with $S \subset \mathcal{O}$ and*

$$m_e(\mathcal{O} - S) \leq \epsilon.$$

If S is measurable, we define its ***Lebesgue measure*** *(or* ***measure****) $m(S)$ by*

$$m(S) = m_e(S).$$

From the definition we immediately see that every open set is measurable. Furthermore, the Lebesgue measure clearly inherits the properties of the outer measure. Additional properties are:

Lemma 9. *The Lebesgue measure has the following properties:*

1. *If $m_e(E) = 0$, then E is measurable.*
2. *A countable union of measurable sets is measurable.*
3. *Closed sets are measurable.*
4. *The complement of a measurable set is measurable.*
5. *A countable intersection of measurable sets is measurable.*

Proof. 1. Let $\epsilon > 0$. There exists a countable cover of E of open intervals $(a_i, b_i)_{i \in \mathbb{N}}$ such that for $\mathcal{O} = \bigcup_i (a_i, b_i)$, $m_e(\mathcal{O}) \leq \epsilon$. Since $(\mathcal{O} - E) \subset \mathcal{O}$, monotonicity of the outer measure implies $m_e(\mathcal{O} - E) \leq \epsilon$. $\mathcal{O}$ is an open set, so we conclude that E is measurable.

2. Let $\epsilon > 0$, suppose $E_1, E_2, ...$ are measurable sets and $E = \bigcup_{i=1}^{\infty} E_i$. For each i we can pick an open set $\mathcal{O}_i$ with $E_i \subset \mathcal{O}_i$ and $m_e(\mathcal{O}_i - E_i) \leq \frac{\epsilon}{2^i}$. Then $\mathcal{O} = \bigcup_{i=1}^{\infty} \mathcal{O}_i$ is open, $E \subset \mathcal{O}$, and $(\mathcal{O} - E) \subset \bigcup_i (\mathcal{O}_i - E_i)$. We see that monotonicity and sub-additivity of the outer measure imply that

$$m_e(\mathcal{O} - E) \leq \sum_{i=1}^{\infty} m_e(\mathcal{O}_i - E_i) \leq \sum_{i=1}^{\infty} \frac{\epsilon}{2^i} = \epsilon.$$

3. Let F be a closed set. We can write $F = \bigcup_{k=1}^{\infty} F \cap [-k, k]$. It follows that we can assume that F is compact, property 2 will then imply that an arbitrary closed set is measurable. Suppose F is compact, so that in particular $m_e(F) < \infty$. By property 3 of the outer measure, we can choose an open set U with $F \subset U$ and $m_e(U) \leq m_e(F) + \epsilon$. Since $U - F$ is open we can write it as a countable union of disjoint closed intervals $[a_i, b_i]_{i \in \mathbb{N}}$. For a fixed N, $K = \bigcup_{i=1}^{N} [a_i, b_i]$ is compact. F closed and K compact implies $d(K, F) > 0$ (without proof, see for example [6]). Since $(K \cup F) \subset U$, we find

$$m_e(K) = \sum_{i=1}^{N} m_e([a_i, b_i]) \leq m_e(U) - m_e(F) \leq \epsilon.$$

We conclude

$$m_e(U - F) \leq \sum_i m_e([a_i, b_i]) \leq \epsilon.$$

4. Let E be a measurable set. Then for every $n \in \mathbb{N}$ we may choose an open set U_n with $E \subset U_n$ and $m_e(U_n - E) \leq \frac{1}{n}$. The complement of U_n is closed, hence measurable. So, by property 2, $S = \bigcup_n U_n^c$ is measurable. We have $S \subset E^c$ and $(E^c - S) \subset (U_n - E)$. It follows that $m_e(E^c - S) \leq \frac{1}{n}$, thus $m_e(E^c - S) = 0$. We find, by property 1, that $m_e(E^c - S)$ is measurable. Since $E^c = S \cup (E^c - S)$, we conclude that E^c is measurable.

5. This follows from properties 2 and 4 since

$$\bigcap_{i=1}^{\infty} E_i = \left(\bigcup_{i=1}^{\infty} E_i^c \right)^c.$$

□

Note that by property 1 we arrive at the same definition of measure zero as we had before.

Proposition 5. *(Countable Additivity) Let $(E_i)_{i \in \mathbb{N}}$ be a family of pairwise disjoint measurable sets whose union has finite outer measure. Then*

$$m\left(\bigcup_{i=1}^{\infty} E_i \right) = \sum_{i=1}^{\infty} m(E_i).$$

Proof. From sub-additivity we have $m(E) \leq \sum_i m(E_i)$. For the inequality in the reverse direction we first assume that each E_i is bounded. Since E_i^c is measurable there exists an open set U_i with $E_i^c \subset U_i$ such that $m_e(U_i - E_i^c) \leq \frac{\epsilon}{2^i}$. We have $E_i - U_i^c = U_i - E_i^c$, so $U_i^c = F_i \subset E_i$ such that $m_e(E_i - F_i) \leq \frac{\epsilon}{2^i}$. Since E_i is bounded, the sets $F_1, ..., F_N$ are compact, thus $m\left(\bigcup_{i=1}^{N} F_i \right) = \sum_{i=1}^{N} m(F_i)$. We have $\bigcup_{i=1}^{N} F_i \subset E$, it follows that

$$m(E) \geq \sum_{i=1}^{N} m(F_i) \geq \sum_{i=1}^{N} m(E_i) - \epsilon.$$

Letting N tend to infinity we find $m(E) \geq \sum_{i=1}^{\infty} m(E_i)$. For the general case let $S_k = [-k, -k+1] \cup [-k+1, k-1]$ and $E_{i,k} = E_i \cap S_k$. Then the $E_{i,k}$ are disjoint, bounded and $E_i = \sum_k E_{i,k}$. Using the above we obtain

$$m(E) = \sum_i \sum_k m(E_{i,k}) = \sum_i m(E_i).$$

□

Corollary 1. *If $U_1, U_2, ...$ are measurable sets such that:*

$$U_1 \subset U_2 \subset U_3 \subset ..., \quad \bigcup_{i=1}^{\infty} U_i = U,$$

then U is measurable and

$$m(U) = \lim_{i \to \infty} m(U_i).$$

Similarly, if $V_1 \supseteq V_2 \supseteq V_3 \supseteq ...$ are measurable, V_1 has finite measure, and

$$\bigcap_{i=1}^{\infty} V_i = V,$$

then V is measurable and

$$m(V) = \lim_{i \to \infty} m(V_i).$$

Proof. The sets $U_2 - U_1, U_3 - U_2, ...$ are pairwise disjoint. If we define $U_0 = \emptyset$, then we can write

$$\begin{aligned}\lim_{i\to\infty} m(U_i) &= \lim_{i\to\infty} \sum_{j=1}^{i} (m(U_j) - m(U_{j-1})) \\ &= \lim_{i\to\infty} \sum_{j=1}^{i} m(U_j - U_{j-1}) \\ &= m\left(\bigcup_{j=1}^{\infty} U_j - U_{j-1}\right) = m(U).\end{aligned}$$

For the second part we set $U_k = V_k - V_{k+1}$, so $V_1 = V \cup \bigcup_{k=1}^{\infty} U_k$ is a disjoint union of measurable sets. We find that

$$\begin{aligned}m(V_1) &= m(V) + \lim_{i\to\infty} \sum_{k=1}^{i-1} (m(V_k) - m(V_{k+1})) \\ &= m(V) + m(V_1) - \lim_{i\to\infty} m(V_i).\end{aligned}$$

We can conclude, since $m(V_1) < \infty$, that $m(V) = \lim_{i\to\infty} m(V_i)$. □

Definition 11. *The function $f : [a, b] \to \mathbb{R}$ is* ***measurable*** *if, for all $c \in \mathbb{R}$, the set $\{x \in [a, b] \mid f(x) > c\}$ is measurable.*

Lemma 10. *If f and g are measurable functions on $[a, b]$ and if k is any constant, then kf and $f + g$ are also measurable on $[a, b]$.*

Proof. If $k = 0$, then $kf = 0$ and any constant function is clearly measurable. If $k > 0$, then

$$\{x \in [a, b] \mid kf(x) > c\} = \{x \in [a, b] \mid f(x) > \frac{c}{k}\}.$$

If $k < 0$, then

$$\begin{aligned}\{x \in [a, b] \mid kf(x) > c\} &= \{x \in [a, b] \mid f(x) < \frac{c}{k}\} \\ &= [a, b] - \{x \in [a, b] \mid f(x) \geq \frac{c}{k}\} \\ &= [a, b] - \bigcap_{n=1}^{\infty} \{x \in [a, b] \mid f(x) > \frac{c}{k} - \frac{1}{n}\}.\end{aligned}$$

We conclude that kf is measurable. For $q \in \mathbb{Q}$, the set

$$E_q = \{x \in [a, b] \mid f(x) > q\} \cap \{x \in [a, b] \mid g(x) > c - q\}$$

is measurable. Since clearly $\{x \in [a, b] \mid f(x) + g(x) > c\} = \bigcup_{q\in\mathbb{Q}} E_q$, we conclude that f+g is measurable. □

Lemma 11. *If $(f_n)_{n\in\mathbb{N}}$ is a sequence of measurable functions on $[a, b]$, then $\limsup_{n\to\infty} f_n(x)$ is a measurable function.*

Proof. We have

$$\{x \in [a, b] \mid (\inf_{n\geq 1} f_n(x)) \geq c\} = \bigcap_{n=1}^{\infty} \{x \in [a, b] \mid f_n(x) \geq c\},$$

$$\text{and } \{x \in [a, b] \mid (\sup_{n\geq m} f_n(x)) > c\} = \bigcup_{n=m}^{\infty} \{x \in [a, b] \mid f_n(x) > c\}.$$

Since
$$\limsup_{n\to\infty} f_n(x) = \inf_{n\geq 1}\big(\sup_{m\geq n} f_m(x)\big),$$
we can conclude that $\limsup_{n\to\infty} f_n(x)$ is a measurable function. □

For the proof of Theorem 2 we define the four Dini Derivatives.

Definition 12. *The four **Dini derivatives** of f at c are*

$$D^+(f)(x) = \limsup_{h\to\ 0^+} \frac{f(x+h)-f(x)}{h}, \quad D_+f(x) = \liminf_{h\to\ 0^+} \frac{f(x+h)-f(x)}{h},$$
$$D^-(f)(x) = \limsup_{h\to\ 0^-} \frac{f(x+h)-f(x)}{h}, \quad D_-f(x) = \liminf_{h\to\ 0^-} \frac{f(x+h)-f(x)}{h}.$$

We now recall Theorem 2 : If $f : [a,b] \to \mathbb{R}$ is continuous and monotonic, then f is differentiable almost everywhere (on $[a,b]$).

In other words, the quotient
$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$
exists for almost every $x \in [a,b]$.

We can assume, without loss of generality, that f is monotonically increasing. In this case, the Dini derivatives are never negative. The derivative of f exists at almost every point if the Dini derivatives are finite and equal at almost every point. Clearly, one has $D_+f(x) \leq D^+f(x)$ and $D_-f(x) \leq D^-f(x)$ for all $x \in [a,b]$. To prove the theorem it suffices to show that

$$D^+f(x) \leq D_-f(x) \text{ almost everywhere, and} \tag{3}$$
$$D^+f(x) < \infty \text{ almost everywhere.} \tag{4}$$

Since if we can establish that $D^+f(x) \leq D_-f(x)$ holds almost everywhere (provided that f is continuous and monotonically increasing), then this inequality also holds almost everywhere on $[-b,-a]$ for $k(x) = -f(-x)$: $D^+k(x) \leq D_-k(x)$. But we have $D_-k(-x) = D_+f(x)$ and $D^+k(-x) = D^-f(x)$, for example:

$$\begin{aligned} D_-k(-x) &= \liminf_{h\to 0^-} \frac{k(-x+h)-k(-x)}{h} \\ &= \liminf_{h\to 0^-} \frac{-f(x-h)+f(x)}{h} \\ &= \liminf_{h\to 0^-} \frac{f(x-h)-f(x)}{-h} \\ &= \liminf_{-h\to 0^+} \frac{f(x+(-h))-f(x)}{-h} = D_+f(x). \end{aligned}$$

It follows that $D^-f(x) \leq D_+f(x)$ almost everywhere. Therefore, equations (3),(4) imply

$$D^+f(x) \leq D_-f(x) \leq D^-f(x) \leq D_+f(x) \leq D^+f(x) < \infty \text{ almost everywhere,}$$

which will prove the theorem.

For each pair of rational numbers $0 < r < R < \infty$ we define $E_r^R = \{x \in [a,b] \mid D_-f(x) < r < R < D^+f(x)\}$. Since the number of such pairs is countable, we have proven (3) if E_r^R has measure zero for each pair. The set E_r^R is the intersection of $E^R = \{x \in [a,b] \mid D^+f(x) > R\}$ and

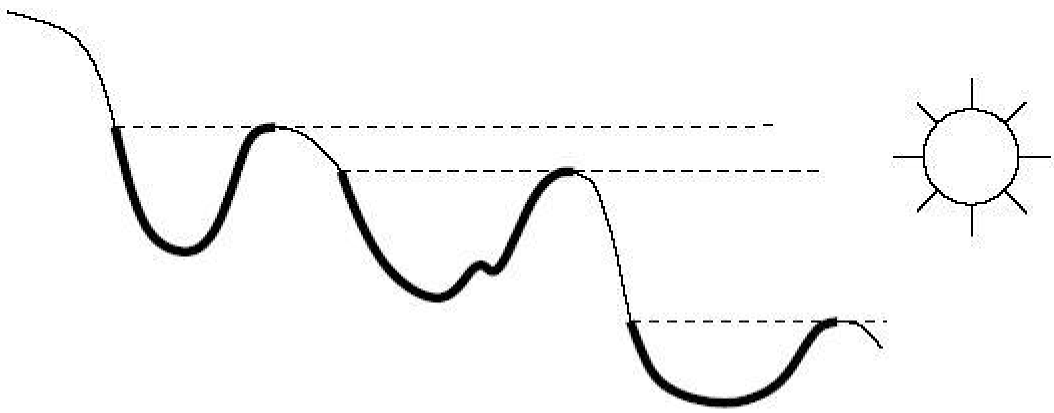

Figure 1: Rising sun lemma

$E_r = \{x \in [a,b] \mid D_-f(x) < r\}$. We need to bound the size of these sets. For $l(x) = f(-x)$, we have

$$\begin{aligned} D^+l(x) &= \limsup_{h\to 0^+} \frac{l(x+h)-l(x)}{h} \\ &= \limsup_{h\to 0^+} \frac{f(-x-h)-f(-x)}{h} \\ &= -\liminf_{h\to 0^+} \frac{f(-x-h)-f(-x)}{-h} \\ &= -\liminf_{-h\to 0^-} \frac{f(-x+(-h))-f(-x)}{-h} = -D_-f(-x). \end{aligned}$$

We can translate any result for E^R into a similar one for E_r, thus our strategy is to first limit the size of E^R.

Definition 13. *Given a continous function $g : [a,b] \to \mathbb{R}$, we say that $x \in [a,b]$ is a **shadow point** of g if there exists $h > 0$, such that $x + h \in [a,b]$ and $g(x+h) > g(x)$.*

For $x \in E^R$, we can find a $z(= x+h) > x$ such that

$$\frac{f(z)-f(x)}{z-x} > R,$$

or equivalently

$$f(z) - Rz > f(x) - Rx.$$

Let $g(x) = f(x) - Rx$, then E^R is contained in the set of shadowpoints of g. If one thinks of the sun rising from the east with the rays of light parallel to the $x-$axis, then the shadowpoints x of g correspond to the points $(x, g(x))$ in the shadow of the rising sun (see figure 1). This is why the next lemma often carries the name "rising sun lemma".

Lemma 12. *(Rising Sun Lemma) Let $g : [a,b] \to \mathbb{R}$ be continuous. The set of shadow points of g that lie in $[a,b]$ is a countable union of pairwise disjoint open intervals (a_k, b_k) for which*

$$\begin{cases} g(a_k) = g(b_k) & \text{if } a_k \neq a \\ g(a_k) \leq g(b_k) & \text{if } a_k = a \end{cases}$$

for all k.

Proof. Let E be the set of shadow points of g. Since g is continuous, for $x \in E$ we can find a neighborhood of x left of z over which the value of the function g stays less than $g(z)$. We find that E is open, and with Lemma 6 it follows that it can be written as a disjoint union of countably many open intervals. If (a_k, b_k) is an interval in this union, then $a_k \notin E$. Therefore we cannot have $g(b_k) > g(a_k)$.
Assume now $g(b_k) < g(a_k)$. By continuity of g, there exists $a_k < c < b_k$ such that

$$g(c) = \frac{g(a_k) + g(b_k)}{2}.$$

We can pick c farthest to the right in the interval (a_k, b_k). Because $g(b_k) < g(a_k)$, we have $g(c) > g(b_k)$. Since c is an element of E, there exists a $d > c$ such that $g(d) > g(c)$. Also, because $b_k \notin E$, we have $g(x) \leq g(b_k)$ for all $x \geq b_k$. It follows that $g(d) > g(b_k)$, and therefore $d < b_k$. Since $g(d) > g(c) > g(b_k)$ and $c < d < b_k$, there exists, by continuity of g, a c' with $d < c' < b_k$ and $g(c') = g(c)$. This contradicts the fact that c was chosen farthest to the right in (a_k, b_k). We conclude that we must have $g(a_k) = g(b_k)$. For $a_k = a$, we can have $a_k \in E$, so the first part of the proof is not valid. We conclude that in this case $g(a_k) \leq g(b_k)$. □

We can now use the rising sun lemma to limit the size of E^R.

Lemma 13. *If $R > 0$ and f is a monotonically increasing continuous function on $[a, b]$, then E^R is contained in a countable union of pairwise disjoint intervals, (a_k, b_k), for which*

$$\sum_k (b_k - a_k) \leq \frac{f(b) - f(a)}{R}.$$

Proof. The idea of the proof is to apply the rising sun lemma to the function $g(x) = f(x) - Rx$. We find that the set of shadow points of g is a countable union of pairwise disjoint open intervals (a_k, b_k), and therefore E^R is contained in such a union. Furthermore, we have $g(a_k) \leq g(b_k)$, so

$$f(a_k) - Ra_k \leq f(b_k) - Rb_k \tag{5}$$

$$\Rightarrow b_k - a_k \leq \frac{1}{R}(f(b_k) - f(a_k)). \tag{6}$$

Since f is monotonically increasing, we find

$$\sum_k (b_k - a_k) \leq \sum_k \frac{1}{R}(f(b_k) - f(a_k)) \leq \frac{f(b) - f(a)}{R}.$$

□

Corollary 2. *If $f : [a, b] \to \mathbb{R}$ is monotonically increasing and continuous, then $D^+ f(x) < \infty$ almost everywhere.*

Proof. First we notice that

$$\{x \in [a, b] \mid D^+ f(x) = \infty\} = \bigcap_{R \in \mathbb{N}} E^R.$$

Furthermore, we have $E^1 \supseteq E^2 \supseteq E^3 \supseteq ....$ The set E^R is measurable, since $D^+ f(x) = \limsup_{h \to 0^+} \frac{f(x+h) - f(x)}{h}$ is measurable. To show this, first note that

$$\big\{x \in [a, b] \mid f(x) > c\big\} = f^{-1}\big((c, \infty)\big)$$

It follows that every continuous function is measurable. Define $g_n = n(f(x + \frac{1}{n}) - f(x))$, this function is measurable for any constant n. Because of the continuity of f we can restrict to countably many h, and hence n, in taking the limsup in $D^+ f(x)$. We have

$$D^+ f(x) = \limsup_{n \to \infty} g_n.$$

So indeed we have that $D^+f(x)$ is measurable. We can now apply Corollary 1, since E^R measurable for all R, and E^1 has measure $\leq (b-a) < \infty$. It follows that

$$m(\{x \in [a,b] \mid D^+f(x) = \infty\}) = \lim_{R\to\infty} m(E^R) \leq \lim_{R\to\infty} \frac{f(b)-f(a)}{R} = 0.$$

□

The next lemma concerns the size of E_r.

Lemma 14. *If f is a monotonically increasing continuous function on $[a,b]$, then E_r is contained in a countable union of pairwise disjoint open intervals, (a_k, b_k), for which*

$$f(b_k) - f(a_k) \leq r(b_k - a_k)$$

Proof. We follow the proof of Lemma 13 with f replaced by $l(x) = f(-x)$ and R replaced by $-r$. Since $D^+l(-x) = D_-f(x)$, we have

$$D^+l(-x) > -r \iff D_-f(x) < r,$$

so

$$E_r = \{x \in [a,b] \mid D_-f(x) < r\} = \{x \mid -x \in [-b,-a], D^+l(-x) > -r\}.$$

We define

$$-E_r = \{x \in [-b,-a] \mid D^+l(x) > -r\}.$$

Following the proof of Lemma 13 up to inequality (5), we see that $-E_r \subset \bigcup_k(-b_k, -a_k)$ and

$$l(-b_k) - (-r)(-b_k) \leq l(-a_k) - (-r)(-a_k).$$

This is equivalent to $E_r \subset \bigcup_k(a_k, b_k)$ and

$$f(b_k) - f(a_k) \leq r(b_k - a_k).$$

□

The next lemma will complete the proof of Theorem 2.

Lemma 15. *let $0 < r < R < \infty$. Let $f : [a,b] \to \mathbb{R}$ be a monotonically increasing continuous function. Then*

$$m(\{x \in [a,b] \mid D_-f(x) < r < R < D^+f(x)\}) = 0.$$

Proof. Let $(\alpha, \beta) \subset [a,b]$ be an open interval and consider its intersection with E_r^R:

$$E_r^R \cap (\alpha,\beta) = (E_r \cap (\alpha,\beta)) \cap E^R.$$

By applying the previous Lemma, we find that $E_r \cap (\alpha,\beta)$ is contained in a countable union of disjoint open intervals (a_k, b_k) for which

$$f(b_k) - f(a_k) \leq r(b_k - a_k).$$

Now we can apply Lemma 13 to each of the intersections $(a_k, b_k) \cap E^R$, it follows that $(a_k, b_k) \cap E^R$ is contained in a countable union of open intervals $(a_{k,l}, b_{k,l})$ for which

$$\sum_l (b_{k,l} - a_{k,l}) \leq \frac{f(b_k) - f(a_k)}{R}.$$

This tells us that $E_r^R \cap (\alpha,\beta)$ is contained in a countable union of disjoint open intervals $\bigcup_{k,l}(a_{k,l}, b_{k,l})$ for which

$$\sum_k \sum_l (b_{k,l} - a_{k,l}) \leq \sum_k \frac{f(b_k) - f(a_k)}{R} \leq \sum_k \frac{r(b_k - a_k)}{R} \leq \frac{r}{R}(\beta - \alpha). \tag{7}$$

We now apply this result to $E_r^R \cap (a_{k,l}, b_{k,l})$ and find that this set is contained in a countable union of disjoint open intervals $(a_{k,l,m}, b_{k,l,m})$, for which

$$\sum_m (b_{k,l,m} - a_{k,l,m}) \leq \frac{r}{R}(b_{k,l} - a_{k,l}). \tag{8}$$

Combining equations (7) and (8), we see that $E_r^R \cap (\alpha, \beta)$ is contained in $\bigcup_{k,l,m}(a_{k,l,m}, b_{k,l,m})$ and

$$\sum_{k,l,m} (b_{k,l,m} - a_{k,l,m}) \leq \frac{r}{R} \sum_{k,l} (b_{k,l} - a_{k,l}) \leq \frac{r^2}{R^2}(\beta - \alpha)$$

We now can repeat this procedure and apply equation (7) to $E_r^R \cap (a_{k,l,m}, b_{k,l,m})$. This yields a countable cover of $E_r^R \cap (\alpha, \beta)$ by disjoint open intervals $(a_{k,l,m,n}, b_{k,l,m,n})$ for which

$$\sum_{k,l,m,n} (b_{k,l,m,n} - a_{k,l,m,n}) \leq \frac{r^3}{R^3}(\beta - \alpha).$$

By induction, we conclude that for any $N \in \mathbb{N}$, we can put E_r^R inside a countable collection of disjoint open intervals for which the total length is less than

$$\frac{r^N}{R^N}(\beta - \alpha)$$

Since $\frac{r}{R} < 1$, we can conclude that the outer measure of E_r^R, and thus the measure, is zero.

□

The conclusion of Theorem 2 still holds if we drop the continuity condition on f, see for example [1].

5 Sard's Theorem

We first recall some definitions.

Definition 14. *The Jacobian matrix of* $f : \mathbb{R}^n \supset E \to \mathbb{R}^m$ *is*

$$Df(x) = \left(\frac{\partial f^i}{\partial x^j}(x) \right)_{ij}, \text{ with } 1 \leq i \leq n \text{ and } 1 \leq j \leq m.$$

We also write $D_j f_i = \frac{\partial f^i}{\partial x^j}$.

Definition 15. *Let* $U \subset \mathbb{R}^m$ *and let* $f : U \to \mathbb{R}^n$. *The set of* ***critical points*** *of f is the set of points in* $\mathbb{R}^m$ *at which the Jacobian matrix of f has rank* $< n$. *The set of* ***critical values*** *is the image of set of critical points under f.*

Theorem 3. *(Sard's Theorem) Let* $U \subset \mathbb{R}^m$ *be open and let* $f : U \to \mathbb{R}^n$ *a smooth map. Then the set of critical values of f has measure zero in* $\mathbb{R}^n$.

Definition 16. *Let* $U, V \subset \mathbb{R}^n$ *open subsets and let* $f : U \to V$. *Then f is a* C^p***-diffeomorphism*** *if f is bijective and both* $f, f^{-1} \in C^p$.

Proposition 6. *(Inverse function theorem) Let* $U \subset \mathbb{R}^n$ *be an open subset and* $f : U \to \mathbb{R}^n$. *If*

1. $f \in C^p$, $p \geq 1$,
2. $f'(x_0)$ *is invertible at* $x_0 \in U$,

then there exist a neighborhood $U' \subset U$ *of* x_0 *and a neighborhood* V *of* $y_0 = f(x_0)$ *such that* $f : U' \to V$ *is a* C^p*-diffeomorphism.*

We start by treating two easy cases of Sard's theorem, the one where $m < n$ and the case $m = n = 1$. If $m < n$ then Lemma 5 immediately implies that $f(U)$ has measure zero. From this, we also see that it is enough for f to be C^1. The proof for $m = n = 1$ is still considerably easier than the proof of the general case:

Let $U \subset \mathbb{R}$, let $f : U \to \mathbb{R}$, and let C be the set of critical points of f. It suffices to show that for every compact interval $[a, b] \subset U$ the set $f(C \cap [a, b])$ has measure zero. Let M be an upper bound for f" on $[a, b]$, let $N \in \mathbb{N}$ and let $h = \frac{(b-a)}{N}$. For $l = 0, ..., N - 1$ consider the intervals $I_l = [a + lh, a + (l + 1)h]$. Let A be the set of those l for which $C \cap I_l \neq \emptyset$, so that $f(C \cap [a, b]) = \bigcup_{l \in A} f(C \cap I_l)$. Let $l \in A$. Then I_l contains a critical point, i.e. there is a point in I_l where f' vanishes. So by the Mean Value Theorem applied to f' we have $\|f'(x)\| \leq Mh$ for $x \in I_l$. Applying the Mean Value Theorem again, but this time to f, we see that the image of I_l under f is contained in an interval of length less than Mh^2. Thus $f(C \cap [a, b])$ is contained in an union of at most N intervals of length less than Mh^2. The total lenght of these intervals is therefore bounded above by

$$NMh^2 = \frac{M}{N}(b - a)^2,$$

which is arbitrary small. We conclude that $f(C \cap [a, b])$, and therefore $f(C)$, has measure zero. □

The proof of the general case will use Fubini's lemma, to be proven first. We define $\mathbb{R}^{n-1}_c := \mathbb{R}^{n-1} \times \{c\} \subset \mathbb{R}^n$.

Lemma 16. *An open cover of the interval $[a, b]$ contains a finite subcover consisting of intervals of total length $\leq 2(b - a)$.*

Proof. $[a, b]$ is compact, thus there exists a finite subcover. Pick such a cover: $[a, b] \subset \sum_{j=1}^{k} I_j$. We may assume that this cover is minimal, i.e. none of the intervals in the cover may be omitted. Then every point $p \in [a, b]$ is contained in at most two of the I_j: Assume $p \in I_1 \cap I_2 \cap I_3$ and let $s = \min(I_1 \cup I_2 \cup I_3)$, $t = \max(I_1 \cup I_2 \cup I_3)$. Now s and t are both in at least one of the intervals, say $s \in I_1$ and $t \in I_2$, then I_1 contains $[s, p]$ and I_2 contains $[p, t]$. So $I_1 \cup I_2 = [s, t]$ and the third interval can be omitted, contradicting the minimality of the cover. Therefore, the I_j cover $[a, b]$ at most twice, i.e. $\sum_{j=1}^{k} |I_j| \leq 2(b - a)$. □

Proposition 7. *(Fubini's Lemma) Let A be a countable union of compact sets in $\mathbb{R}^n$ such that $A_c = A \cap \mathbb{R}^{n-1}_c$ has measure zero in $\mathbb{R}^{n-1}_c \cong \mathbb{R}^{n-1}$ for all $c \in R$. Then A has measure zero.*

Proof. First assume that A itself is compact. This implies that there exists an interval $[a, b] \subset \mathbb{R}$ such that $A \subset [a, b]^n$. Since $A_c = A \cap \mathbb{R}^{n-1}_c$ has measure zero it can be covered by open cubes $K^i_c \subset [a, b]^{n-1}, i \in \mathbb{N}$, such that $\sum_{i=1}^{\infty} |K^i_c| < \epsilon$. Let $K_c = \bigcup_{i=1}^{\infty} |K^i_c| \subset [a, b]^{n-1}$. For any fixed $c \in \mathbb{R}$, the map $\lambda_c : \mathbb{R}^n \to \mathbb{R}, x \mapsto |x_n - c|$ is continuous and vanishes precisely on $\mathbb{R}^{n-1}_c$ and therefore on A_c. Since K_c is a union of open cubes, it is itself open. This implies that $([a, b]^{n-1} - K_c) \times [a, b]$ is closed. So $[a, b]^n - (K_c \times [a, b])$ is closed and also $A \subset [a, b]^n$ is compact, thus $A - (K_c \times [a, b])$ is compact, and therefore λ_c assumes a minimum $\alpha_c > 0$ on this set. We see that

$$\{x \in A \mid \lambda_c(x) < \alpha_c\} \subset K_c \times I_c \text{ where } I_c = (c - \alpha_c, c + \alpha_c).$$

These intervals I_c cover $[a, b]$, thus by Lemma 16 there exists a finite subcover $I_1, ..., I_k$, with

$1, ..., k \in [a, b]$, of volume $\leq 2(b - a)$. This implies

$$\begin{aligned}\{K_j^i \times I_j \mid i \in \mathbb{N}, j = 1, ..., k\} &\supset \bigcup_{j=1}^{k} \{x \in A \mid \lambda_j(x) < \alpha_j\} \\ &= \bigcup_{j=1}^{k} \{x \in A \mid x_n \in I_j\} \\ &= \{x \in A \mid x_n \in \bigcup_{j=1}^{k} I_j\} \\ &\supset \{x \in A \mid x_n \in [a, b]\} = A.\end{aligned}$$

So the boxes $\{K_j^i \times I_j \mid i \in \mathbb{N}, j = 1, ..., k\}$ cover A and their total volume is $\leq \epsilon 2(b-a)$. It follows that A has measure zero.
Now let A be a countable union of compact subsets of $\mathbb{R}^n$, $A = \bigcup_{j=1}^{\infty} A^j$. Then $A_c = A \cap \mathbb{R}_c^{n-1} = (\bigcup_{j=1}^{\infty} A^j) \cap \mathbb{R}_c^{n-1} = \bigcup_{j=1}^{\infty}(A^j \cap \mathbb{R}_c^{n-1})$. Since A_c has measure zero, $A^j \cap \mathbb{R}_c^{n-1}$ has measure zero for all j. With the above it follows that A^j itself has measure zero for all j. So A is a countable union of sets of measure zero and thus it follows with lemma 1 that A has measure zero. □

Now we can prove Sard's theorem. Let $U \subset \mathbb{R}^m$ be open, $f : U \to \mathbb{R}^n$ a smooth map and let $C \subset U$ be the set of critical points. We define

$$C_i := \{x \in U \mid D^j f(x) = 0 \text{ for } j = 1, ..., i\}.$$

The C_i form a descending sequence $C_1 \supset C_2 \supset C_3 \supset ...$ of closed sets, so we can write:

$$C = (C - C_1) \cup (C_1 - C_2) \cup ... \cup (C_{k-1} - C_k) \cup C_k$$

To prove that $f(C)$ has measure zero we will prove the following:

1. $f(C_k)$ has measure zero for sufficiently large k.
2. $f(C - C_1)$ has measure zero.
3. $f(C_i - C_{i+1})$ has measure zero for all $i \in \{1, ..., k - 1\}$.

The claim then follows with Lemma 1. Since $f(C) = f(C - C_1) \cup f(C_1 - C_2) \cup ... \cup f(C_{k-1} - C_k) \cup f(C_k)$ is a union of a countable number of sets of measure zero.
Proof of 1:
Let $k > \frac{m}{n} - 1$ and let $K \subset U$ be a cube of edge length a. It is enough to prove that $f(C_k \cap K)$ has measure zero since C_k is contained in a countable union of cubes. By Taylor's formula and the definition of C_k, we have

$$\|f(x + h) - f(x)\| \leq b\|h\|^{k+1}, \tag{9}$$

for $x \in C_k \cap K$ and $x + h \in K$,where b depends only on f and K. We now pick an integer l and decompose K into l^m cubes with edge a/l. Let K_1 be one of these cubes intersecting C_k. Then for $x \in C_k \cap K_1$ and $x + h \in K_1$: $\|h\| \leq \sqrt{m}(a/l)$. It follows with (7) that $\|f(x + h) - f(x)\| \leq b(\frac{\sqrt{m}a}{l})^{k+1}$, so $f(K_1)$ is contained in a cube of edge

$$2 \cdot b \cdot (\frac{\sqrt{m}a}{l})^{k+1} = \frac{c}{l^{k+1}},$$

where c, like b, depends only on f and K. Now the volume V of the union of these l^m cubes satisfies $V \leq l^m(\frac{b}{l^{k+1}})^n = b^n l^{m-n(k+1)}$. Since $k > m/n - 1$, $m - n(k + 1) < 0$, and it follows that the volume V tends to zero as $l \to \infty$. So we have proven that $f(C_k \cap K)$ has measure zero.

proof of 2:
Since $C - C_1 = \{x \in U \mid 1 \leq \text{rank } Df(x) < n\}$, it follows that if $n = 1$ the statement is trivially true. So assuming $n \geq 2$, we will prove the claim by induction over m. For $m = 1, n \geq 2$ we know from Lemma 5 that $f(U)$ has measure zero, so let's assume that the claim holds for $m-1$, that is, 1 holds for every $f : \mathbb{R}^{m-1} \supset U \to \mathbb{R}^n$, and prove that the claim also holds for m. Around each $x \in C - C_1$, we will find a neighborhood V such that $f(V \cap C)$. Since $C - C_1$ can be covered by a countable number of such neighborhoods, this proves that $f(C - C_1)$ has measure zero. For $x \in C - C_1$ there is at least one partial derivative that does not vanish, say $D_1 f_1(x) \neq 0$. Consider the map $g : (x_1, ..., x_m) \mapsto (f_1(x), x_2, ..., x_m)$, it obviously has an invertible Jacobian at x, so by the inverse function theorem there exists a neighborhood V of x such that g is a C^∞ diffeomorphism to a neighborhood $V^{'}$ of $g(x)$. The function $h = f \circ g^{-1}$ is defined on this neighborhood $V^{'}$ and has the form: $h : (y_1, ..., y_m) \mapsto (y_1, h_2(y), ..., h_n(y))$. We see that h maps the hyperplane $\{t\} \times \mathbb{R}^{m-1}$ into $\{t\} \times \mathbb{R}^{n-1}$ and the set of critical points $C^{'}$ of h is $g(V \cap C)$. Let $h^t := g|_{\{t\} \times \mathbb{R}^{m-1}} : (\{t\} \times \mathbb{R}^{m-1}) \cap V^{'} \to \{t\} \times \mathbb{R}^{n-1}$, then the Jacobian of h is given by:

$$Dh(y) = \begin{pmatrix} 1 & 0 \\ ? & Dh^t(y) \end{pmatrix}$$

which is invertible iff Dh^t is invertible. By the induction assumption, the set of critical values of h^t has measure zero in $\{t\} \times \mathbb{R}^{m-1}$. So $h(C^{'}) \cap \mathbb{R}^{m-1}_t$ has measure zero for all $t \in \mathbb{R}$. $h(C^{'})$ is a countable union of compact sets, so Fubini's lemma applies, it follows that $h(C^{'}) = f(V \cap C)$ has measure zero.
proof of 3:
This claim is like the previous one proven by induction over m, so again we assume that the claim holds for every $f : \mathbb{R}^{m-1} \supset U \to \mathbb{R}^n$. Since $C_i - C_{i+1} = \{x \in U \mid D^j f(x) = 0 \text{ for } j = 1, .., i \text{ and } D^{i+1} f(x) \neq 0\}$, for every $x \in C_i - C_{i+1}$ there is a (i+1)-th derivative that does not vanish at x. Hence we may assume $D_1 w(x) \neq 0$, where w has the form $w(x) = D_{\sigma(1)} ... D_{\sigma(i)} f_r(x)$. Furthermore $w(x) = 0$ since $w(y) = 0$ for $\forall y \in C_i$. Now, as above $g : x \mapsto (w(x), x_2, ..., x_m)$ is a diffeomorphism on a neighborhood V of x to a neighborhood $V^{'}$ of g(x). Consider again the map $h = f \circ g^{-1} : V^{'} \to \mathbb{R}^n$ and define $h^0 := h|_{\mathbb{R}^{m-1}_0} : (\{0\} \times \mathbb{R}^{m-1}) \cap V^{'} \to \mathbb{R}^n$. By the induction assumption, the set of critical values of h^0 has measure zero. But each point of $g(C_i \cap V)$ is critical for h^0 since all partial derivatives of h, thus also of h^0, of order $\leq k$ vanish. It follows that $f(C_i \cap V) = h \circ g(C_i \cap V)$ has measure zero. The set $C_i - C_{i+1}$ is covered by a countable union of these sets $C_i \cap V$, so we conclude that $f(C_i - C_{i+1})$ has measure zero. □

More generally, Sard's theorem also holds for smooth mappings between manifolds M and N of dimensions m and n, respectively. Given $f : M \to N$, a point $p \in M$ is called regular if the differential $T_p f : T_p M \to T_{f(p)} N$ is surjective. The set of critical values is then given by those points $q \in N$ for which not every $p \in f^{-1}(q)$ is a regular value. This formulation of the result now follows from the above version for Euclidean spaces by taking a countable set of coordinate patches.

References

[1] D. M. Bressoud, *A Radical approach to Lebesgue's theory of integration*, Cambridge University Press, New York, 2008.

[2] M. Demazure, *Bifurcations and catastrophes*, Springer Verlag, 2000.

[3] J. E. Marsden, T. Ratiu, R. Abraham, *Manifolds, Tensor Analysis and applications*, Springer-Verlag, Berlin, 2003.

[4] M. Müger, *Lebesgue's characterization of Riemann integrable functions*, June 20, 2006.

[5] M. Müger, *An Introduction to Differential Topology, de Rham Theory and Morse Theory*, 2005.

[6] E. M. Stein, R. Shakarchi *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, Princeton, 2005.

[7] V. A. Zorich, *Mathematical Analysis I*, Springer-Verlag, Berlin, 2004.

[8] V. A. Zorich, *Mathematical Analysis II*, Springer-Verlag, Berlin, 2004.