D.W.M. van Dijk

On the sets of critical points and critical values

Bachelor's Thesis

Thesis Supervisor: dr. O.W. van Gaans

October 23, 2014

Mathematical Institute, Leiden University

Introduction

The set of critical points of a function f is defined as the set of points where all partial derivatives of f exist and are zero. The set of critical values of f is the image under f of the set of critical points. In particular, for functions $f : \mathbb{R} \to \mathbb{R}$, we talk about the sets of stationary points and stationary values, meaning the points where the function has zero derivative.

In this thesis we consider the sets of critical (stationary) points and values and explore what can be said about them. For example, the set of extreme values of a function is, under certain conditions, a subset of the set of stationary values. In *Chapter 1* we prove that the set of extreme values is always at most countable, but also show that the set of stationary values can be uncountable by constructing a special function.

In *Chapter 2* we wish to describe the sets of stationary points and values as intersections and unions of open and closed sets, to check if they are any form of G_δ or F_σ sets. We prove results in this direction under some additional conditions. It turns out it is not that easy to categorize the sets of stationary points and values in this way in general.

In *Chapter 3* we use a small piece of measure theory to prove that the set of stationary values is always a set of measure zero. In *Chapter 4* we extend this proof, which eventually leads to a formulation of the Morse-Sard Theorem [1], which says that, under few conditions, the set of critical values of a function always has measure zero.

Contents

1 Countability

This chapter is concerned with describing the set of local extrema and the set of stationary values in terms of countability. We are interested in what the maximum amount of elements in those sets can be.

1.1 Definitions

We assume some familiarity with the notions of countability and uncountability and state the following definition.

1.1.1 Definition. A set X is *countable* if there exists an injective function $f : X \to \mathbb{N}$ from X to the natural numbers $\mathbb{N} = \{0, 1, 2, ...\}$.

We also assume the following theorem to be familiar.

1.1.2 Theorem. *The Cartesian product of finitely many countable sets is countable.*

The following example illustrates the above.

1.1.3 Example. The sets $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ are countable. The sets $\mathbb{R}$ and $[0, 1] \subseteq \mathbb{R}$ are uncountable. The sets $\mathbb{N} \times \mathbb{N}$, $\mathbb{Z} \times \mathbb{Z}$ and $\mathbb{Q} \times \mathbb{Q}$ are countable.

Also, we will need the definition of a local extremum.

1.1.4 Definition. A *local extremum* is a point where a function reaches a local maximum or local minimum. That is, $f(a)$ is a local extremum of f if there exists an $\epsilon > 0$ such that we have either $f(x) \leq f(a)$ (local maximum) or $f(a) \leq f(x)$ (local minimum) for all $x \in (a - \epsilon, a + \epsilon)$. We note that by this definition a constant function has one local extremum, everywhere on its domain.

With these two definitions, we can start looking closer at the set of local extrema of a function.

1.2 Extrema and stationary values

Let $f : \mathbb{R} \to \mathbb{R}$ be an arbitrary function and let E be the set of points where f attains a local extremum:

$$E := \{x \in \mathbb{R} : f(x) \text{ is a local extremum of } f\}. \tag{1.2.1}$$

We have the following theorem concerning the set of local extrema $f(E)$ of f.

1.2.1 Theorem. *The set*

$$f(E) := \{f(x) : f(x) \textit{ is a local extremum of } f\}$$

is at most countable.

Proof. First, we show that the set of local maxima is at most countable. Put

$$E_+ := \{x \in \mathbb{R} : f \text{ attains a local maximum at } x\}.$$

For every $m \in E_+$, we can take $a, b \in \mathbb{Q}$ with $a < m < b$, such that

$$f(m) = \max\{f(x) : x \in (a, b)\}.$$

We note that if we pick the same a and b for different $m, n \in E_+$, then $f(m) = f(n)$, since both $f(m)$ and $f(n)$ are the maximum value f attains on the interval (a, b). Hence, we constructed an injective function $g : f(E_+) \to \mathbb{Q} \times \mathbb{Q}$ given by $f(m) \mapsto (a, b)$. By *Theorem 1.1.2*, $\mathbb{Q} \times \mathbb{Q}$ is countable, and we find that $f(E_+)$ is at most countable. In the same fashion, we can show that the set $f(E_-)$ of local minima is at most countable. Since $f(E) = f(E_+) \cup f(E_-)$, and since the union of two countable sets is countable, we conclude that $f(E)$ must be at most countable. $\square$

So, *Theorem 1.2.1* proves that the set of local extrema is at most countable for any function $f : \mathbb{R} \to \mathbb{R}$. Apparently it does not matter how irregular the function gets. Countability always follows since an extremum must be the extreme value on some interval, from where we can construct an injective function onto $\mathbb{Q} \times \mathbb{Q}$.

Next, we would like to consider the set of stationary points D of an arbitrary function $f : \mathbb{R} \to \mathbb{R}$:

$$D := \{x \in \mathbb{R} : f \text{ is differentiable at } x \text{ and } f'(x) = 0\}. \tag{1.2.2}$$

Then, the set $f(D)$ is the set of stationary values of f.

We note that we have $E \subseteq D$ and thus $f(E) \subseteq f(D)$ provided f is differentiable at x for at least all $x \in E$, since if f is differentiable at x and has an extremum at x, then $f'(x) = 0$.

Since the set $f(E)$ is at most countable, we wonder if a similar statement holds for $f(D)$. It is easy to see that our only concern is the set of points where we have $f'(x) = 0$, but where f does not attain a local extremum, since the set of points where f attains a local extremum is at most countable by *Theorem 1.2.1*. We will refer to the points x where we have $f'(x) = 0$ and $x \notin E$ as *saddle points* (see *Figure 1*).

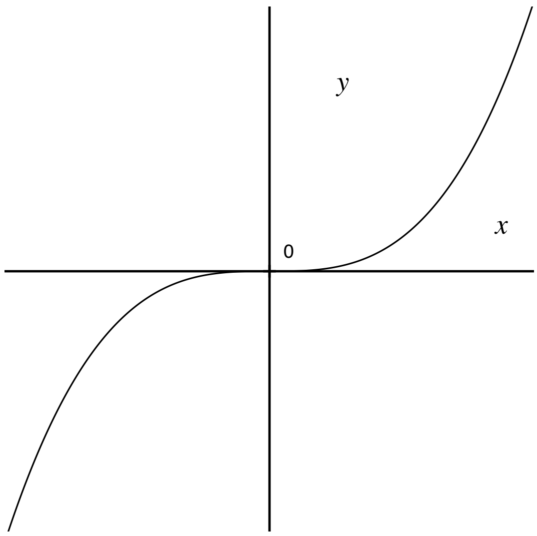

Figure 1: A graph of $y = x^3$, with a saddle point at the origin.

A similar argument as used in the proof of *Theorem 1.2.1* would be hard to formulate, since we can not specifically 'pick' a saddle point as the minimum or maximum of f on an interval. Actually, it turns out that a statement similar to *Theorem 1.2.1* will not hold for the set of stationary values.

1.3 The Cantor set

We will show that the set of stationary values can be uncountable using an idea by van Rooij and Schikhof [2, p.44]. To construct a function with an uncountable set of stationary values, we will use a set known as the Cantor set.

1.3.1 Definition. The *Cantor set* (also referred to as the *Cantor middle third set* or the *Cantor ternary set*) is the set constructed by the following steps. We take the interval $[0, 1]$ and remove the open middle third interval, i.e., $(\frac{1}{3}, \frac{2}{3})$. We now have two closed intervals, from which we again remove the open middle third intervals. We continue in this way indefinitely. What remains is the Cantor set. In *Figure 2* (found on the internet), from top to bottom, starting with the full interval $[0, 1]$, the first six steps of the described process are depicted.

Figure 2: A figure showing the first six steps in constructing the Cantor set.

An explicit formula for the Cantor set is

$$\mathbb{D} = \bigcap_{n=1}^{\infty} \bigcap_{k=0}^{3^{n-1}-1} \left(\left[0, \frac{3k+1}{3^n}\right] \cup \left[\frac{3k+2}{3^n}, 1\right] \right). \tag{1.3.1}$$

The Cantor set has several remarkable properties. The following lemmata will prove to be useful.

1.3.2 Lemma. *The Cantor set $\mathbb{D}$ is uncountable.*

There are a few different ways to prove *Lemma 1.3.2*, but we will not do so here. Most proofs use *Cantor's diagonal argument* which is outside the scope of this thesis. For the curious reader, a proof can be found in [5, p.58].

1.3.3 Lemma. *The Cantor set $\mathbb{D}$ does not contain any intervals of non-zero length.*

In *Chapter 3* we will introduce some tools which will make proving *Lemma 1.3.3* quite easy and we will prove it as a corollary of *Theorem 3.1.8*.

1.4 Constructing a function with an uncountable set of stationary values

Using the Cantor set, we can construct a function f whose set of stationary values is uncountable. First, we define a function $\phi : \mathbb{R} \to \mathbb{R}$ as

$$\phi : \ x \mapsto \min\{|x - a| : a \in \mathbb{D}\}, \qquad \text{for } x \in \mathbb{R}. \tag{1.4.1}$$

Looking more closely at our function ϕ, we can make the following remarks. We note that the minimum in (1.4.1) exists since $x \mapsto |x - a|$ is continuous and $\mathbb{D}$ is compact.

1.4.1 Claim. $\phi(x) = 0$ if and only if $x \in \mathbb{D}$. Hence, $\phi(x) = 0$ for an uncountable number of x, but never on an interval by *Lemmata 1.3.2* and *1.3.3*.

1.4.2 Remark. ϕ is continuous and non-negative anywhere on $\mathbb{R}$. This immediately follows from the definition of ϕ.

Next, we integrate (1.4.1) to get the desired function f:

$$f : \ x \mapsto \int_0^x \phi(t)\ dt, \qquad \text{for } x \in \mathbb{R}. \tag{1.4.2}$$

The remarks we made earlier now directly imply the following:

1.4.3 Remark. $f'(x) = 0$ if and only if $x \in \mathbb{D}$. Consequently, $f'(x) = 0$ holds for an uncountable number of x, but never on an interval.

1.4.4 Remark. f is a strictly increasing function on $\mathbb{R}$, since its derivative $f'(x) = \phi(x)$ is non-negative and never equal to zero on an interval. So, for any $a, b > 0$ with $a < b$ we have $\int_a^b \phi(x)dx > 0$, which implies $f(b) - f(a) > 0$.

We can now prove that a set of stationary values can be uncountable.

1.4.5 Theorem. *Consider the function f defined by (1.4.2), (1.4.1) and (1.3.1). The set of stationary values $f(D)$ is uncountable.*

Proof. From *Remark 1.4.3* we know the set of stationary points D of f is uncountable. From *Remark 1.4.4* it follows that the function f is injective, since it is strictly increasing. Thus, distinct points of D get mapped to distinct points in $f(D)$, which allows us to conclude that $f(D)$ has as many elements as D, which has as many elements as $\mathbb{D}$, which is uncountable. □

Since the set of extreme values of a function is at most countable by *Theorem 1.2.1*, we also conclude that the set of saddle points of a function f can be uncountable; and that the uncountability of $f(D)$ is due to these saddle points.

2 Topological structure

2.1 The set of stationary points

Recall that D is defined as the set of stationary points of a function $f : \mathbb{R} \to \mathbb{R}$ as in (1.2.2):

$$D := \{x \in \mathbb{R} : f \text{ is differentiable at } x \text{ and } f'(x) = 0\}.$$

Then, the set $f(D)$ is the set of stationary values of f. In this chapter, we require that f is continuous.
We wonder what we can say about the sets D and $f(D)$. In particular, we are interested in whether they are any form of G_δ or F_σ sets.

2.1.1 Definition[1]. A G_δ set is a countable intersection of open sets. So, a set X is G_δ if and only if it can be written as

$$X = \bigcap_{n=1}^{\infty} A_n,$$

where A_n is open for all $n \in \mathbb{N}$.
Similarly, an F_σ set is a countable union of closed sets. So, a set X is F_σ if and only if it can be written as

$$X = \bigcup_{n=1}^{\infty} A_n,$$

where A_n is closed for all $n \in \mathbb{N}$.

We note that, trivially, every open set is G_δ and every closed set is F_σ. However, the converse of this statement is not true as the following example shows.

2.1.2 Example. The G_δ set X_1 defined by

$$X_1 = \bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, 1 + \frac{1}{n}\right)$$

is closed, since we have $X_1 = [0, 1]$. Likewise, the F_σ set X_2 defined by

$$X_2 = \bigcup_{n=1}^{\infty} \left[\frac{1}{n+1}, 1 - \frac{1}{n+1}\right]$$

is open, since we have $X_2 = (0, 1)$.

We would also like to give a name to countable unions of G_δ sets, which we will call $G_{\delta\sigma}$ sets. Continuing in this way we can also define, say, a $G_{\delta\sigma\delta\sigma\delta}$ set which is the countable intersection of the countable union of the countable intersection of the countable union of G_δ sets, or: $\bigcap_k \bigcup_l \bigcap_m \bigcup_n X_{k,l,m,n}$, where every $X_{k,l,m,n}$ is a G_δ set for $k, l, m, n \in \mathbb{N}$. In the same way we define a countable intersection of F_σ sets as an $F_{\sigma\delta}$ set.

We can now state the following result about the set of stationary points D.

2.1.3 Theorem. *The set D of stationary points of a continuous function $f : \mathbb{R} \to \mathbb{R}$ is a $G_{\delta\sigma\delta}$ set.*

Proof. We take the sequence $(g_n)_{n\in\mathbb{N}}$, where g_n is defined by

$$g_n(x) := \frac{f\left(x + \frac{1}{n}\right) - f(x)}{\frac{1}{n}}, \quad x \in \mathbb{R}.$$

By definition of the derivative we see that $\lim_{n\to\infty} g_n(x_0)$ converges to the derivative of f at x_0 if and only if f is differentiable at x_0. We can now write

$$D := \{x \in \mathbb{R} : \lim_{n\to\infty} g_n(x) = 0\}.$$

By definition of the limit, we can rewrite this to

$$D := \left\{x \in \mathbb{R} : \forall m \in \mathbb{N}\ \exists N \in \mathbb{N} \text{ such that } \forall n > N : |g_n(x)| < \frac{1}{m}\right\}. \tag{2.1.1}$$

[1] The term G_δ originates in Germany, where the G stands for *Gebiet*, area, and the δ for *Durchsnitt*, which means intersection. The term F_σ however originates in France, with the F standing for *fermé*, closed, and the σ for *sommé*, which means union.

Now we use some formal logic to rewrite D as a set defined by intersections and unions:

$$D := \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N+1}^{\infty} D_n^m, \tag{2.1.2}$$

where

$$D_n^m := \Big\{ x \in \mathbb{R} : |g_n(x)| < \frac{1}{m} \Big\}. \tag{2.1.3}$$

Since we required f to be continuous, we know that $|g_n(x)|$ is continuous for every $n \in \mathbb{N}$. It follows that D_n^m is open for every $n, m \in \mathbb{N}$. Thus, since D can be written as in (2.1.2), we conclude that D is a $G_{\delta\sigma\delta}$ set. □

2.1.4 Remark. We remark that if we change the $<$ sign to $\leq$ in (2.1.1), we get

$$D := \Big\{ x \in \mathbb{R} : \forall m \in \mathbb{N}\ \exists N \in \mathbb{N} \text{ such that } \forall n > N : |g_n(x)| \leq \frac{1}{m} \Big\}, \tag{2.1.4}$$

and the same rewriting as above will then lead to D being an $F_{\sigma\delta}$ set:

$$D := \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N+1}^{\infty} D_n^m,$$

where

$$D_n^m := \Big\{ x \in \mathbb{R} : |g_n(x)| \leq \frac{1}{m} \Big\}.$$

Every D_n^m is closed, and since any intersection of closed sets is closed, we indeed get the intersection of the union of closed sets, hence D is also an $F_{\sigma\delta}$ set.

2.2 The set of stationary values

Next, we are interested in the topological properties of the set $f(D)$ of stationary values. We will use the following lemmata.

2.2.1 Lemma. *For any sequence of sets $A_n \subseteq \mathbb{R}$ $(n \in \mathbb{N})$ and a function $f : \mathbb{R} \to \mathbb{R}$ we have*

$$f\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} f(A_n), \qquad \text{and} \qquad f\left(\bigcap_{n=1}^{\infty} A_n\right) \subseteq \bigcap_{n=1}^{\infty} f(A_n).$$

Furthermore, the equality

$$f\left(\bigcap_{n=1}^{\infty} A_n\right) = \bigcap_{n=1}^{\infty} f(A_n)$$

holds if and only if f is injective, in particular, if f is strictly increasing or strictly decreasing.

Proof. We prove the statements for any two sets $A, B \subseteq \mathbb{R}$, since the general cases follow by repetition.

For all $y \in f(A \cup B)$ there exists $x \in A \cup B$ with $y = f(x)$. Without loss of generality, assume $x \in A$. Then $f(x) \in f(A)$, hence $f(x) \in f(A) \cup f(B)$. So $f(A \cup B) \subseteq f(A) \cup f(B)$.
Conversely, for all $y \in f(A) \cup f(B)$ we assume without loss of generality $y \in f(A)$. This implies existence of an $x \in A$ with $y = f(x)$, thus $x \in A \cup B$, and $f(x) \in f(A \cup B)$ follows immediately. So $f(A) \cup f(B) \subseteq f(A \cup B)$, and the first equality follows.

For $y \in f(A \cap B)$ we have an $x \in A \cap B$ with $y = f(x)$. So both $x \in A$ and $x \in B$ hold. Hence, we have $f(x) \in f(A)$ and $f(x) \in f(B)$, thus $f(x) \in f(A) \cap f(B)$. We conclude $f(A \cap B) \subseteq f(A) \cap f(B)$.
For the converse, assume f is injective. Then, $f(x) \in f(A) \cap f(B)$ implies $f(x) \in f(A)$ and $f(x) \in f(B)$. By injectivity, both $x \in A$ and $x \in B$ hold. Thus, $f(x) \in f(A \cap B)$ and $f(A) \cap f(B) \subseteq f(A \cap B)$ follows, and the last equality in our lemma holds.
We note that f must be injective, since otherwise $f(x) \in f(A) \cap f(B)$ could imply $f(x) = f(x_1) \in f(A)$ and $f(x) = f(x_2) \in f(B)$, with $x_1 \neq x_2$. Now, $x_1 \in A$ and $x_2 \in B$ do not imply that either x_1 or x_2 is in $A \cap B$. □

2.2.2 Lemma. *Every interval of the real line is a G_δ set.*

Proof. Take $a, b \in \mathbb{R}$. Then every open interval (a, b) is G_δ since we can write $(a, b) = \bigcap_{n=1}^{\infty} (a, b)$.
Every closed interval is also G_δ by writing $[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right)$.
The case for half-open intervals follows by adding or subtracting $\frac{1}{n}$ to the closed side and intersecting the open intervals. With obvious adaptations, one can also handle unbounded intervals. So, we conclude every interval of the real line is indeed a G_δ set. □

2.2.3 Theorem. *The set $f(D)$ of stationary values of an injective continuous function $f : \mathbb{R} \to \mathbb{R}$ is a $G_{\delta\sigma\delta\sigma\delta}$ set.*

Proof. Repeatedly applying *Lemma 2.2.1* to (2.1.2) yields

$$f(D) \subseteq \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N+1}^{\infty} f(D_n^m),$$

with equality if f is injective. From (2.1.3) we know that D_n^m consists of open intervals, and that there are at most countably many open intervals in D_n^m. Since f is continuous, $f(D_n^m)$ contains at most countably many intervals. By *Lemma 2.2.2* it then follows that $f(D_n^m)$ is a countable union of G_δ sets. We conclude $\bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N+1}^{\infty} f(D_n^m)$ is a $G_{\delta\sigma\delta\sigma\delta}$ set, and the theorem follows. □

Of course, not nearly every continuous function is injective. However, some functions can be divided into intervals on which they are injective, for example between two successive local extrema. This gives rise to the idea that if we 'cut' a function into pieces, right at the points where it attains its local extrema, we get a set of intervals on which the function is indeed injective. And since a function only has countably many local extrema (*Theorem 1.2.1*), we can even create a countable union of injective pieces of the function that together make up the entire function. This however does require that the function is really injective between every two successive local extrema.
Expanding our theorem above to even more continuous functions is now not so hard; especially when we note that on an interval where a function is constant it attains only one stationary value.

2.2.4 Theorem. *Assume $f : \mathbb{R} \to \mathbb{R}$ is continuous and there exist intervals I_k for some $k \in \mathbb{N}$ such that $\mathbb{R} = \bigcup_k I_k$ and f is injective on I_k for every k. Then the set $f(D)$ of stationary values of f can be written as a $G_{\delta\sigma\delta\sigma\delta\sigma}$ set.*

At this point it is time to consider the results we are achieving. Of course, it is always nice for us mathematicians to be able to describe sets in a way that shows how they are constructed, but once we are talking about unions of intersections of unions of ... etcetera, we are extensively widening our choice of possibilities. It would almost be possible to say that any set can be such a set, which would make all the above rather useless. For further research, it would be interesting to see if there is another way to categorize the set of stationary values, or, for example, to check what conditions need to be met for a $G_{\delta\sigma\delta}$ set to be a G_δ set.
Also, we could reverse the question and check if every $G_{\delta\sigma\delta\sigma\delta\sigma}$ set of measure zero is the set of stationary values of some function.

3 Lebesgue Measure

This section describes the Lebesgue measure and how it can be used to measure a set, specifically the set of stationary values. It will lead us to a formulation of the Morse-Sard Theorem for functions $f : \mathbb{R} \to \mathbb{R}$.

3.1 Definitions

The *Lebesgue measure* is a standard way in mathematics to assign a measure to a subset of an n-dimensional Euclidean space. Just as *countability* describes the *amount* of elements in a set, the *measure* of a set describes the *size* of the elements in the set. In a 1-dimensional space this measure corresponds with the length of an interval, in a 2-dimensional space the measure corresponds with the area of a surface, and in 3-dimensional space the measure corresponds with the volume of a shape. Since we are only interested in functions from $\mathbb{R}$ to $\mathbb{R}$ at this point, we will restrict our definition of the Lebesgue measure to the length of an interval. However, we do note that measure theory is a very extensive and interesting field in mathematics and that we are not doing it justice here. For more on the subject we refer the reader to one of the many books written about it, for example [3].

3.1.1 Definition. The *Lebesgue measure* λ of an interval of $\mathbb{R}$ is equal to the length of that interval. More precisely, the measure of the interval $[a, b]$ is given by:

$$\lambda\big([a, b]\big) = |b - a| .$$

We have the following corollaries.

3.1.2 Corollary. The measure of a point is equal to zero.

This follows readily since we can write a point $a \in \mathbb{R}$ as the closed interval $[a, a] \in \mathbb{R}$ with measure $\lambda\big([a, a]\big) = |a - a| = 0$.

3.1.3 Corollary. The measure of an open interval (a, b) is equal to the measure of the closed interval $[a, b]$. In fact, we have $\lambda\big([a, b]\big) = \lambda\big((a, b]\big) = \lambda\big([a, b)\big) = \lambda\big((a, b)\big)$.

It is easy to see that since we are removing a point, of zero length, from the endpoints of a closed interval, we are left with the same length.

We introduce the following definition to determine the measure of a set of intervals.

3.1.4 Definition. The measure of a set consisting of a union of disjoint intervals is equal to the sum of the measures of each of the intervals. Furthermore, if some of the intervals are not disjoint, then we can combine overlapping intervals into one interval.

To illustrate the statements above we will give a few examples.

3.1.5 Example. The measure of the set $\{0\} \cup \{1\} \cup [2, 3] \cup (4, 5)$ is $0 + 0 + 1 + 1 = 2$.

3.1.6 Example. The measure of the set $[0, 2] \cup (1, 3) \cup [5, 6]$ is equal to $\lambda\big([0, 3) \cup [5, 6]\big) = 3 + 1 = 4$.

Sets of measure zero will be given a special name.

3.1.7 Definition. A set of measure zero is called a *null set*.

We will conclude this section by considering the Cantor set $\mathbb{D}$ as defined in *Definition 1.3.1*.

3.1.8 Theorem. *The Cantor set $\mathbb{D}$ is a null set.*

Proof. We defined the Cantor set as the interval $[0, 1]$ from which we repeatedly remove middle third open intervals. Hence, we start out with a set of measure $\lambda\big([0, 1]\big) = 1$ from which we remove a set of measure $\lambda\big((\frac{1}{3}, \frac{2}{3})\big) = \frac{1}{3}$. We are left with a set of measure $1 - \frac{1}{3} = \frac{2}{3}$, from which we remove two times a set of measure $\lambda\big((\frac{1}{9}, \frac{2}{9})\big) = \lambda\big((\frac{7}{9}, \frac{8}{9})\big) = \frac{1}{9}$. This process continues indefinitely, allowing us to give the following formula for the measure of $\mathbb{D}$:

$$\lambda\big(\mathbb{D}\big) = 1 - 1 \cdot \frac{1}{3} - 2 \cdot \frac{1}{9} - 4 \cdot \frac{1}{3^3} - ... = 1 - \sum_{n=0}^{\infty} 2^n \cdot \frac{1}{3^{n+1}} = 1 - \frac{1}{2} \sum_{n=1}^{\infty} \left(\frac{2}{3}\right)^n = 1 - \frac{1}{2} \cdot 2 = 0.$$

So, since $\lambda\big(\mathbb{D}\big) = 0$, we may conclude that the Cantor set is a null set. □

As a corollary we can now prove *Lemma 1.3.3* which we left unproven before.

3.1.9 Corollary. The Cantor set $\mathbb{D}$ does not contain any intervals of non-zero length.

Proof. Since the Cantor set $\mathbb{D}$ has measure zero and a set can not have negative measure, the Cantor set must be a union of sets of measure zero. Thus, the Cantor set does not contain any intervals of non-zero length. $\square$

3.2 A weak version of the Morse-Sard Theorem

We are interested in the Lebesgue measure of the sets of stationary points and stationary values. Recall that the set of stationary points D is defined as (1.2.2):

$$D := \{x \in \mathbb{R} : f \text{ is differentiable at } x \text{ and } f'(x) = 0\}.$$

In general there is not much we can say about the measure of the set of stationary points. We give two examples to show this.

3.2.1 Example. The constant function $f : \mathbb{R} \to \mathbb{R}$ given by $x \mapsto 2$ is constant everywhere on $\mathbb{R}$. Hence, its derivative is zero everywhere and the measure of the set of stationary points is $\lambda\big(D\big) = \lambda\big(\mathbb{R}\big) = \infty$.

3.2.2 Example. The function $f : \mathbb{R} \to \mathbb{R}$ given by $x \mapsto x$ has derivative $f'(x) = 1$ everywhere. Hence, the set of stationary points is empty, and we have $\lambda\big(D\big) = \lambda\big(\{\emptyset\}\big) = 0$.

It turns out that the measure of the set of stationary values is much more interesting.

3.2.3 A weak version of the Morse-Sard Theorem. *(mentioned in [2, p.154])*
The set of stationary values $f(D)$ of an arbitrary function $f : \mathbb{R} \to \mathbb{R}$ is a null set.

To prove *Theorem 3.2.3*, we need some tools to work with. We introduce the *Vitali Covering Lemma.*

3.2.4 Definition. A *Vitali cover* of a set $X \subseteq \mathbb{R}$ is a set $\mathcal{V}$ of closed intervals with positive length such that, for every $\alpha > 0$ and every $x \in X$, there is some $I \in \mathcal{V}$ with $\lambda(I) < \alpha$ and $x \in I$.

3.2.5 Vitali Covering Lemma. *Given a set $X \subseteq \mathbb{R}$, with $\lambda^*(X) < \infty$, a Vitali cover $\mathcal{V}$ of X, and some $\epsilon > 0$, there are disjoint $I_1, ..., I_m \in \mathcal{V}$ with $\lambda^*\left(X \setminus \bigcup_{k=1}^m I_k\right) < \epsilon$. It follows that $\lambda^*(X) < \lambda^*(\bigcup_{k=1}^m I_k) + \epsilon$.*

The definition of a *Vitali cover* states that any subset X of $\mathbb{R}$ can be covered by a number of closed intervals of arbitrarily small length. The lemma then states that, if the subset has finite measure, it is possible to choose a countable number of disjoint intervals from the Vitali cover $\mathcal{V}$ that covers almost all of X. We will not prove this lemma here, as its proof is outside the scope of this thesis. A proof can be found in [4].
Note that the Vitali Covering Lemma uses the *outer measure* λ^* rather than the measure λ we defined. These two 'measures' are in general not the same, but for our purposes we can regard the outer measure λ^* as to be the same as the measure λ, namely in defining the length of an interval.

We can now state our proof for *Theorem 3.2.3*, the weak Morse-Sard Theorem, based on an idea of the course *Real Functions*, given at Leiden University in the spring of 2008.

Proof. We prove the theorem for an arbitrary function $f : [a, b] \mapsto \mathbb{R}$. We would like to apply the Vitali Covering Lemma to the set $f(D)$. Therefore, we will need to make sure that $\lambda^*(f(D)) < \infty$, i.e., that $f(D)$ has finite (outer) measure. To make sure this is the case, we pick $E = f(D) \cap [-M, M]$ for arbitrary $M \in \mathbb{N}$. Then $\lambda^*(E) \leq 2M < \infty$. Next, we will need to find a suitable Vitali cover $\mathcal{V}$ for $f(D)$. We take, for fixed $\epsilon > 0$,

$$\begin{aligned}\mathcal{V} = \{[f(x) - \delta\epsilon, f(x) + \delta\epsilon] : x \in D, \delta > 0 \text{ such that}& \\ &(x - \delta, x + \delta) \subset [a, b], \\ &u \in (x - \delta, x + \delta) \implies f(u) \in [f(x) - \delta\epsilon, f(x) + \delta\epsilon]\}.\end{aligned}$$

We need to check whether this is indeed a Vitali cover for $f(D)$. Indeed, $\mathcal{V}$ consists of closed intervals with positive length. We also need to make sure that for every $\alpha > 0$ and every $y \in f(D)$, there is some $I \in \mathcal{V}$ with $\lambda(I) < \alpha$ and $y \in I$.
We let $y \in f(D)$, $\epsilon > 0$, and $\alpha > 0$. Since $y \in f(D)$, there exists $x \in D$ with $f(x) = y$. By definition of D, there exists a $\delta_0 > 0$ such that

$$0 < |u - x| < \delta_0 \implies \left|\frac{f(u) - f(x)}{u - x} - f'(x)\right| < \epsilon,$$

and $f'(x) = 0$. We now take $\delta = \min\{\delta_0, \alpha/(3\epsilon), x - a, b - x\}$. Then $(x - \delta, x + \delta) \subset [a, b]$ and

$$\begin{aligned}u \in (x - \delta, x + \delta) &\implies |u - x| < \delta_0 \\ &\implies |f(u) - f(x)| < \epsilon\,|u - x| < \epsilon\delta \\ &\implies f(u) \in [y - \delta\epsilon, y + \delta\epsilon].\end{aligned}$$

So, for $y \in f(D)$ there indeed exists $x \in D$, $\delta > 0$ such that $y = f(x)$, and $[y - \delta\epsilon, y + \delta\epsilon] \in \mathcal{V}$. Moreover, the length of this interval is $\lambda([y-\delta\epsilon, y+\delta\epsilon]) = 2\delta\epsilon \leq 2\alpha\epsilon/(3\epsilon) = 2\alpha/3 < \alpha$. So $\mathcal{V}$ is indeed a Vitali cover for $f(D)$.

We can now use the Vitali covering lemma. We want to show $\lambda^*(E) = 0$. Let us suppose $\lambda^*(E) > 0$. First, we pick

$$\epsilon = \frac{\frac{1}{2}\lambda^*(E)}{1+b-a}, \tag{3.2.1}$$

and consider $\mathcal{V}$ for this ϵ. The lemma states that there are disjoint $I_1, ..., I_m \in \mathcal{V}$ such that

$$\lambda^*(E) < \lambda^*\left(\bigcup_{k=1}^{m} I_k\right) + \epsilon.$$

For each k we have an interval $I_k = [y_k - \delta_k\epsilon, y_k + \delta_k\epsilon]$, where $y_k = f(x_k)$ with $x_k \in D$. Hence, the length of each interval in the union is $\lambda^*([y_k - \delta_k\epsilon, y_k + \delta_k\epsilon]) = 2\delta_k\epsilon$. So, we get

$$\lambda^*(E) < \lambda^*\left(\bigcup_{k=1}^{m} I_k\right) + \epsilon \leq \sum_{k=1}^{m}(2\delta_k\epsilon) + \epsilon. \tag{3.2.2}$$

We will need to show that $\sum_{k=1}^{m} 2\delta_k$ is small. Since the I_k are mutually disjoint, the intervals $(x_k - \delta_k, x_k + \delta_k)$ have to be mutually disjoint. Otherwise, there exists

$$u \in (x_k - \delta_k, x_k + \delta_k) \cap (x_l - \delta_l, x_l + \delta_l),$$

which then implies

$$f(u) \in [f(x_k) - \delta_k\epsilon, f(x_k) + \delta_k\epsilon] \cap [f(x_l) - \delta_l\epsilon, f(x_l) + \delta_l\epsilon] = I_k \cap I_l,$$

which is impossible for $k \neq l$ since the I_k are mutually disjoint (and thus $I_k \cap I_l = \emptyset$). Therefore, we have

$$\lambda^*\left(\bigcup_{k=1}^{m}(x_k - \delta_k, x_k + \delta_k)\right) = \sum_{k=1}^{m} 2\delta_k,$$

which is at most $b - a$ since $\bigcup_{k=1}^{m}(x_k - \delta_k, x_k + \delta_k) \subset [a, b]$. Rewriting (3.2.2) gives us

$$\lambda^*(E) < \epsilon\sum_{k=1}^{m}(2\delta_k) + \epsilon \leq \epsilon(b-a) + \epsilon = (1+b-a)\epsilon,$$

and filling in the ϵ we picked in (3.2.1) gives us

$$\lambda^*(E) < \frac{1}{2}\lambda^*(E),$$

which contradicts our assumption $\lambda^*(E) > 0$. Thus, we infer $\lambda^*(E) = 0$ and conclude $E = f(D) \cap [-M, M]$ is a null set. Since $f(D)$ is the union of countably many null sets:

$$f(D) = \bigcup_{M=1}^{\infty} f(D) \cap [-M, M],$$

we can conclude $f(D)$ is indeed a null set. □

The weak Morse-Sard Theorem tells us that for any function its set of stationary values is a null set. For simple functions, this is not too hard to comprehend: when the derivative of a function is zero, then it is constant and thus yields only one stationary value. Values close to this stationary value will not be reached as long as the derivative stays zero, hence when they are reached they are not stationary values since the derivative is not zero anymore. There is always a value between two stationary values where the derivative is not zero, so the set of stationary values can never contain an interval.

For more irregular functions the theorem still holds. We will analyze the function we constructed in *Section 1.4* using the Cantor set. This function is more complicated since we showed in *Theorem 1.4.5* that its set of stationary values is uncountable. Would it not be likely that this set contains at least one interval? The answer of course lies in the property of the Cantor set $\mathbb{D}$ that it does not contain any intervals of non-zero length (*Lemma 1.3.3*). Since the function we constructed is injective and continuous, this property carries over to the set of stationary values, indeed making the set of stationary values a null set.

4 Higher dimensions

This chapter discusses the sets of critical points and critical values of functions mapping from n-dimensional to m-dimensional space. Ultimately we will state the full Morse-Sard Theorem for these functions.

4.1 The case $\mathbb{R} \to \mathbb{R}^m$

In the previous chapters we considered functions $f : \mathbb{R} \to \mathbb{R}$. We will now see that extending our knowledge to functions $f : \mathbb{R} \to \mathbb{R}^m$ is intuitively not so hard.

First, we define an arbitrary function

$$f : \mathbb{R} \to \mathbb{R}^m : \; x \mapsto \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix}.$$

We see that the variable $x \in \mathbb{R}$ gets mapped to a vector consisting of m functions, each of which may depend on x. The set of critical points D is now defined by

$$D := \{x \in \mathbb{R} : \forall i \in \{1, ..., m\} : f_i \text{ is differentiable at } x \text{ and } f_i'(x) = 0\}. \tag{4.1.1}$$

Hence, the set of critical points is the set of points $x \in \mathbb{R}$ for which every component f_i of f is differentiable at x and has zero derivative. We can also rewrite this as

$$D = \bigcap_{i=1}^{m} \{x \in \mathbb{R} : f_i \text{ is differentiable at } x \text{ and } f_i'(x) = 0\} = \bigcap_{i=1}^{m} D_i, \tag{4.1.2}$$

where D_i is exactly D for a function $f = f_i$ as defined in (1.2.2). This means that, considering the set of critical points, we can look at functions $f : \mathbb{R} \to \mathbb{R}^m$ as m distinct one-dimensional cases.

With the above observations, we can generalize the results from the previous chapters for functions $f : \mathbb{R} \to \mathbb{R}^m$. Of course, the set of critical values can still be uncountable. To show this, we can simply take every component f_i of f to be the function constructed in *Section 1.4*, and the resulting set of critical values will definitely be uncountable.

More interesting is that we can find a somewhat stronger version of the weak Morse-Sard Theorem (*Theorem 3.2.3*). However, we will need to extend our definition of the Lebesgue measure in order to be able to measure the set $f(D) \subseteq \mathbb{R}^m$.

4.1.1 Definition. The *m-dimensional Lebesgue measure* λ^m of a Cartesian product of intervals is equal to the product of the corresponding one-dimensional measures.

For example, the measure of the Cartesian product of intervals $[a, b]$ and $[c, d]$ is $\lambda^2\big([a, b] \times [c, d]\big) = \lambda\big([a, b]\big) \cdot \lambda\big([c, d]\big) = |b - a| \cdot |d - c|$. The following lemma follows immediately.

4.1.2 Lemma. *If any of the sets in an m-dimensional Cartesian product of sets is a null set, then the measure of the Cartesian product is zero.*

Proof. By *Definition 4.1.1* we have a product of m one-dimensional measures. Since at least one of these is zero, the entire product is zero. □

We can now prove a stronger version of the Morse-Sard Theorem than we had before.

4.1.3 Theorem. *The set of critical values $f(D)$ of an arbitrary function $f : \mathbb{R} \to \mathbb{R}^m$ is a null set.*

Proof. From (4.1.2) we know that for all $i \in \{1, ..., m\}$ we have $D \subseteq D_i$, since D is an intersection. Also, for all $i \in \{1, ..., m\}$ we have

$$\lambda^m\big(f(D_i)\big) = \lambda\big(f_1(D_i)\big) \cdot \lambda\big(f_2(D_i)\big) \cdot ... \cdot \lambda\big(f_m(D_i)\big) = 0,$$

since $\lambda(f_i(D_i)) = 0$ by the weak Morse-Sard Theorem (*Theorem 3.2.3*). In other words, the i'th component is a null set, thus the set of stationary values of f is a null set by *Lemma 4.1.2.* □

So, extending the theorem to be able to handle more functions was intuitively not so hard. However, we would like to extend the theorem to all functions $f : \mathbb{R}^n \to \mathbb{R}^m$, which leads us to the next section.

4.2 The case $\mathbb{R}^n \to \mathbb{R}^m$

We will now give the full Morse-Sard Theorem, as proven by A. Sard in 1942 in [1].

4.2.1 The Morse-Sard Theorem.
Consider an arbitrary function $f : \mathbb{R}^n \to \mathbb{R}^m$. We distinguish two cases.
(i) If $n \leq m$, then the set of critical values of f is a null set if all first-order partial derivatives exist and are continuous.
(ii) If $n > m$, then the set of critical values of f is a null set if each of the components f_j ($j \in \{1, ..., m\}$) is of class C^q for $q \geq n - m + 1$, that is, if for every function f_j all partial derivatives of order q exist and are continuous.

Let us look at this theorem more closely. It seems that the theorem holds without restriction as long as the amount of variables the functions depend on is less than or equal to the amount of component functions defining f, and the first-order derivatives of all components of f exist and are continuous. This is a stronger condition than we assumed in *Theorem 3.2.3* and *Theorem 4.1.3*, since these theorems applied to arbitrary functions. The first case in *Theorem 4.2.1*, for $n = 1$, is thus a special case of *Theorems 3.2.3* and *4.1.3*.
On the contrary, if $n > m$, i.e., if we have more variables x_i ($i \in \{1, ..., n\}$) than functions f_j ($j \in \{1, ..., m\}$), the theorem can not always be applied. It turns out that this then depends on the *smoothness* of the functions f_j. As Sard shows in [1], it is possible to construct a function not smooth enough that yields a set of critical values of positive measure. This also proves that the condition $q \geq n - m + 1$ can not be weakened.

References

[1] Arthur Sard, The measure of the critical values of differentiable maps, *Bulletin of the American Mathematical Society* **48**, p.883–890, 1942.

[2] A.C.M. van Rooij and W.H. Schikhof, *A second course on real functions*, Cambridge University Press, 1982.

[3] Heinz Bauer, *Measure and Integration Theory; Translated from the German by Robert B. Burckel*, de Gruyter, 2001.

[4] H.L. Royden, *Real Analysis*, Macmillan and Prentice Hall, 1963-2010.

[5] Frank Morgan, *Real Analysis*, American Mathematical Society, 2012.