MT4004

REAL AND ABSTRACT ANALYSIS

Notes by

LARS OLSEN

"A Stroll through the Garden of Analysis"

Table of Contents

1. METRIC SPACES

In the beginning there were operations – hundreds of them – limits, derivaties, integrals sums: all of the many operations on functions, sequences, sets, vectors, matrices, and whatever else you might encounter at in calculus. The hallmark of the 20'th century mathematics is that we now view the operations as functions defined on entire collections of "abstract" objects rather than as specific actions taken on individual objects, one at a time. Maurice Frechet, in a short expository from 1950 had this to say:

> In modern times it has been recognized that it is possible to elabarate full mathematical theories dealing with elements of which the nature is not specified, that is, with abstract elements. A collection of these abstract elements will be called an abstract set. If to this set there is added some rule of association of these elements, or some relation between them, the set will be called an abstract space. A natural generalization of function consists is associating with any element x of an abstract set E a number $f(x)$. Functional analysis is the study of such functionals $f(x)$. More generally *abstract analysis* is the theory of the transformations $y = F(x)$ of an element x of an abstract set set E into an element y of another (or the same) abstract set. It is obvious that the study of general analysis should be preceded by a discussion of abstract spaces.

Early examples of this type of abstraction appeared in 1906 is Frechet's thesis lq'Sur quelques points du calcul functionnel", in which he introduced a notion of distance defined on abstyract set of points. In particular, Frechet considered the collection $C([0,1])$ consisting of all continuous real valued functions defined on the closed unit interval $[0,1]$, where he measured the distance between two functions by

$$d(f,g) \max_{x\in[0,1]} |f(x)-g(x)|$$

for $f, g \in C([0,1])$. This distance function was actually well-known in 1906 but Frechet was the first to view it as a small part of a much bigger picture. Given the notion of distance between two functions $f, g \in C([0,1])$ it makes sense to ask questions like: is integeration contunous? That is, are the numbers $\int_0^1 f(x)\,dx$ and $\int_0^1 g(x)\,dx$ "close" whenever f and g are "close"?

This new point of view proved to have immediate applications; in the same year F. Riesz used Frechet's ideas to give a new proof of a result be E Schmidt stating that any orthonormal system in $C([0,1])$ must be countable. In fact, Riesz extended this result to to another collection of of functions and in doing so introduced the so-called L^p spaces (these spaces will be studied in detail in MT5825). Reisz techniques revolutionized the the study of trigonometric series. to say that Frechet's ideas caught on would be an understatement; the study of contemporary analysis would be lost without them. We will now give an introduction to the key idea in Frechet's work, namely, the notion of a metric space.

Maurice Frechet (2 September 1878 – 4 June 1973) was a French mathematician. He made major contributions to the topology of point sets and introduced the entire concept of metric spaces. His dissertation opened the entire field of functionals on metric spaces and introduced the notion of compactness. Independently of Riesz, he discovered the representation theorem in the space of Lebesgue square integrable functions.

1.1. Metric spaces.

One of the basic ideas in analysis is that of convergence . Recall earlier courses in analysis that a sequence $(x_n)_n$ of real numbers converge to a real number x if, as n gets large, x_n gets very close to s. In order words, the distance between x and x_n gets very small. The key concept is that of distance; as long as we have a way of measuring the distance between two points, we can talk about convergence. The reader is familira with various distances. For example, if x and y are two real numbers then the distance between them is $|x-y|$, and if (x_1, x_2) and (y_1, y_2) are two points in the plane, then the Euclidean distance between then is $((x_1-y_1)^2 + (x_2-y_2)^2)^{1/2}$. What general properties must a distance satisfy in order that we can talk about convergence sensibly? It turns out that only very few are needed.

Definition. Metric space. *Let X be a non-empty set. A function $d : X \times X \to \mathbb{R}$ is called a metric if it satisfies the following properties:*

(1) *For all $x, y \in X$, we have*

$$d(x, y) \geq 0\,.$$

(2) *For all $x, y \in X$, we have*

$$d(x, y) = 0 \text{ if and only if } x = y\,.$$

(3) *For all* $x, y \in X$, *we have*

$$d(x, y) = d(y, x) .$$

(4) *For all* $x, y, z \in X$, *we have*

$$d(x, z) \le d(x, y) + d(y, z) .$$

If d *is a metric on the set* X, *then the pair* (X, d) *is called a metric space. Statement (4) is called the triangle inequality.*

Metric spaces are due to Maurice Frechet in about 1915.

Definition. Ball. *Let* (X, d) *be a metric space. For* $x \in X$ *and* $r \ge 0$, *then ball* $B(x, r)$ *with centre* x *and radius* r *is defined by*

$$B(x, r) = \{y \in X \mid d(x, y) < r\} .$$

We now list a number of important examples of metric spacesInitially, they provide exercises to test the comprehension of the definition. However, they have other equally important but enduring purposes. The definition of a metric space is an abstraction based largely on our experience with the real numbers, and so it is natural, particularly since we live in a world which is (at amy rate approximately) Euclidean, that our intuitions about metric spaces should be Euclidean in character. There is no great harm in this provided we remember that any object which satisfies the definition of a metric space is a metric space and, as the examples listed below demonstrate, that the abstraction is wide enough to include spaces which have surprisingly non-Euclidean characteristics. This means, of course, that we must take some care when making statements about metric spaces in general, and one fruitful way of testing the plausibility of such statements is to see if they are true in some more extreme examples of metric spaces. However, the variety of examples makes for a rich and useful theory.

Example. The discrete metric. Let X be any non-empty set. Define $d : X \times X \to \mathbb{R}$ by

$$d(x, y) = \begin{cases} 0 & \text{if } x = y; \\ 1 & \text{if } x \ne y. \end{cases}$$

We leave it for the reader to verify that d is a metric on X. The metric d is called the discrete metric. Observe that balls look somewhat unusual in a discrete metric space. Namely if $x \in X$, then

$$B(x, r) = \begin{cases} \{x\} & \text{if } r \le 1; \\ X & \text{if } r > 1. \end{cases}$$

Example. $\mathbb{R}$ with the usual metric. For $x, y \in \mathbb{R}$, we define

$$d(x,y) = |x - y| .$$

We leave it for the reader to verify that d is a metric on $\mathbb{R}$. The metric d is called the usual metric on $\mathbb{R}$. Observe that if $x \in X$, then

$$B(x,r) = (x - r, x + r) .$$

Example.. Let $X = [0,1) \cup \{7\}$ and equip X with the usual metric d, i.e. for $x, y \in \mathbb{R}$, we put

$$d(x,y) = |x - y| .$$

Then

$$\begin{aligned} B(7,2) &= \{7\} , \\ B(0, \tfrac{1}{2}) &= [0, \tfrac{1}{2}) , \\ B(0,2) &= [0,1) , \\ B(\tfrac{1}{2}, 9) &= [0,1) \cup \{7\} . \end{aligned}$$

Example.. Let $X = (0, \infty)$. Define $d : X \times X \to [0, \infty)$ by

$$d(x,y) = \left| \frac{1}{x} - \frac{1}{y} \right| .$$

This d is a metric. (Show it!)

For example,

$$d(1,2) = \left| \frac{1}{1} - \frac{1}{2} \right| = \frac{1}{2} ,$$

i.e. the distance between 1 and 2 is $\frac{1}{2}$. Also,

$$d(1, 1000000) = \left| \frac{1}{1} - \frac{1}{1000000} \right| = \frac{999999}{1000000} ,$$

i.e. the distance between 1 and 1000000 is $\frac{999999}{1000000}$.

We will now compute various balls in $((0,\infty), d)$. We have

$$\begin{aligned} B(1,1) &= \{ y \in (0,\infty) \,|\, d(1,y) < 1 \} \\ &= \left\{ y \in (0,\infty) \,\middle|\, \left| \frac{1}{1} - \frac{1}{y} \right| < 1 \right\} \\ &= \left\{ y \in (0,\infty) \,\middle|\, -1 < \frac{1}{1} - \frac{1}{y} < 1 \right\} \\ &= \left\{ y \in (0,\infty) \,\middle|\, -2 < -\frac{1}{y} < 0 \right\} \\ &= \left(\frac{1}{2}, \infty \right) . \end{aligned}$$

Example. $\mathbb{R}^n$**.** There are several natural metrics on $\mathbb{R}^n$. We define the metrics $d_1, d_2, d_\infty : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$
\begin{aligned}
d_1(\,(x_1, \dots, x_n), (y_1, \dots, y_n)\,) &= \sum_i |x_i - y_i|\,,\\
d_2(\,(x_1, \dots, x_n), (y_1, \dots, y_n)\,) &= \left(\sum_i |x_i - y_i|^2\right)^{1/2},\\
d_\infty(\,(x_1, \dots, x_n), (y_1, \dots, y_n)\,) &= \max_i |x_i - y_i|\,,
\end{aligned}
$$

for $(x_1, \dots, x_n), (y_1, \dots, y_n) \in \mathbb{R}^n$. The metric d_1 is known as the 1-metric, The metric d_2 is known as the 2-metric (or the Euclidean metric), and the metric d_∞ is known as the ∞-metric (or the uniform metric). The metrics were first systematically studied by David Hilbert in the beginning of the previous century in his studies of convergent of Fourier series.

David Hilbert (23 January 1862 –14 February 1943) was a German mathematician. He is recognized as one of the most influential and universal mathematicians of the 19th and early 20th centuries. Hilbert discovered and developed a broad range of fundamental ideas in many areas, including invariant theory and the axiomatization of geometry. He also formulated the theory of Hilbert spaces, one of the foundations of functional analysis. Hilbert and his students contributed significantly to establishing rigor and developed important tools used in modern mathematical physics. Hilbert is known as one of the founders of proof theory and mathematical logic.

We will now prove that d_1, d_2, d_∞ are metrics. Indeed, it is clear that d_1, d_2, d_∞ satisfy conditions (1)–(3). We will now prove that the triangle inequality is satisfied.

We first prove the triangle inequality for d_1. Indeed, for $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n), z = (z_1, \dots, z_n) \in \mathbb{R}^n$, we have

$$\begin{aligned} d_1(x,z) &= \sum_i |x_i - z_i| \\ &\le \sum_i (|x_i - y_i| + |y_i - z_i|) \\ &= \sum_i |x_i - y_i| + \sum_i |y_i - z_i| \\ &= d_1(x,y) + d_1(y,z)\,. \end{aligned}$$

This proves the triangle inequality for d_1.

Next, first prove the triangle inequality for d_2. In order to prove this we need the following inquality, known as Cauchy's inequality: for $(a_1, \dots, a_n), (b_1, \dots, b_n) \in \mathbb{R}^n$ we have

$$\sum_i a_i b_i \le \left(\sum_i a_i^2\right)^{1/2} \left(\sum_i b_i^2\right)^{1/2}$$

For $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n), z = (z_1, \dots, z_n) \in \mathbb{R}^n$, we now have

$$\begin{aligned} d_2(x,z)^2 &= \sum_i |x_i - z_i|^2 \\ &\le \sum_i (|x_i - y_i| + |y_i - z_i|)^2 \\ &= \sum_i \Big(|x_i - y_i|^2 + 2|x_i - y_i||y_i - z_i| + |y_i - z_i|^2\Big) \\ &= \sum_i |x_i - y_i|^2 + 2\sum_i |x_i - y_i||y_i - z_i| + \sum_i |y_i - z_i|^2\,. \end{aligned}$$

Using Cauchy's inequality we obtain

$$\begin{aligned} d_2(x,z)^2 &\le \sum_i |x_i - y_i|^2 + 2\sum_i |x_i - y_i||y_i - z_i| + \sum_i |y_i - z_i|^2 \\ &\le \sum_i |x_i - y_i|^2 + 2\sum_i |x_i - y_i||y_i - z_i| + \sum_i |y_i - z_i|^2 \\ &= d_2(x,y)^2 + 2d_2(x,y)d_2(y,z) + d_2(y,z)^2 \\ &= \Big(d_2(x,y) + d_2(y,z)\Big)^2\,. \end{aligned}$$

This proves the triangle inequality for d_1.

Finally we prove the triangle inequality for d_∞. Indeed, for $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n), z = (z_1, \dots, z_n) \in \mathbb{R}^n$, we have for all i,

$$\begin{aligned} |x_i - z_i| &\le |x_i - y_i| + |y_i - z_i| \\ &\le \max_j |x_j - y_j| + \max_j |y_j - z_j| \\ &= d_\infty(x,y) + d_\infty(y,z)\,. \end{aligned}$$

Since i was arbirary this implies that

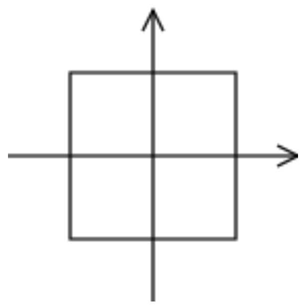
$$\max_i |x_i - z_i| \leq d_\infty(x,y) + d_\infty(y,z)\,.$$

Hence

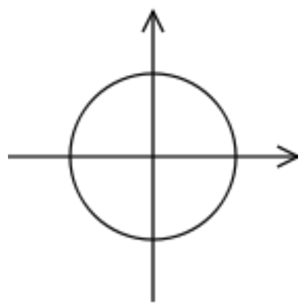
$$d_\infty(x,z) \leq d_\infty(x,y) + d_\infty(y,z)\,.$$

This proves the triangle inequality for d_∞.

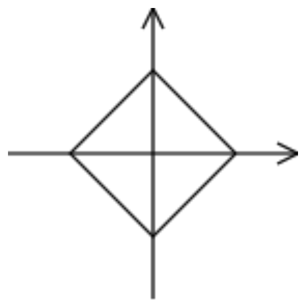
Below we sketch the unit ball $B(0,1)$ in $(\mathbb{R}^2, d_\infty)$, in $(\mathbb{R}^2, d_2)$, and in $(\mathbb{R}^2, d_1)$.



The unit circle in $(\mathbb{R}^2, d_\infty)$.



The unit circle in $(\mathbb{R}^2, d_2)$.



The unit circle in $(\mathbb{R}^2, d_1)$.

The three metrics d_1, d_2, d_∞ are related. Indeed, observe that for $(a_1, \dots, a_n) \in \mathbb{R}^n$, we have

$$\begin{aligned}
\left(\max_i |a_i|\right)^2 &= \max_i |a_i|^2 \\
&\le \sum_i |a_i|^2 \\
&\le \sum_{i,j} |a_i||a_j| \\
&= \left(\sum_i |a_i|\right)\left(\sum_j |a_j|\right) \\
&= \left(\sum_k |a_k|\right)^2 \\
&\le \left(n \max_i |a_i|\right)^2,
\end{aligned}$$

whence

$$\max_i |a_i| \le \left(\sum_i |a_i|^2\right)^{1/2} \le \sum_k |a_k| \le n \max_i |a_i| .$$

This clearly implies that for all $x, y \in \mathbb{R}^n$, we have

$$d_\infty(x,y) \le d_2(x,y) \le d_1(x,y) \le n d_\infty(x,y) .$$

Example. The space of bounded functions.. Let X be a non-empty set. A function $f : X \to \mathbb{R}$ is called bounded if there is a constant M such that

$$|f(x)| \le M \quad \text{for all } x \in M.$$

Let

$$B(X) = \left\{ f : X \to \mathbb{R} \,\middle|\, f \text{ is bounded} \right\} .$$

We now define a metric $d_\infty : B(X) \times B(X) \to \mathbb{R}$ by

$$d_\infty(f,g) = \sup_{x \in X} |f(x) - g(x)| \,.$$

The metric d_∞ is called the ∞-metric (or the uniform metric) on $B(X)$. We will now prove that d_∞ is a metrics. Indeed, it is easily seen that d_∞ satisfis conditions (1)–(3). We will now prove that the triangle inequality is satisfied. For $f, g, h \in B(X)$ we have for all $x \in X$,

$$\begin{aligned} |f(x) - h(x)| &\leq |f(x) - g(x)| + |g(x) - h(x)| \\ &\leq \sup_{y \in X} |f(y) - g(y)| + \sup_{y \in X} |g(y) - h(y)| \\ &= d_\infty(f,g) + d_\infty(g,h) \,. \end{aligned}$$

Since x was arbirary this implies that

$$\sup_{x \in X} |f(x) - h(x)| \leq d_\infty(f,g) + d_\infty(g,h) \,.$$

Hence

$$d_\infty(f,h) \leq d_\infty(f,g) + d_\infty(g,h) \,.$$

This proves the triangle inequality for d_∞.

Recalling uniform convergence from earlier courses in analysis, we see that if $(f_n)_n$ is a sequence in $B(X)$ and $f \in B(X)$ then clearly

$$f_n \to f \quad \text{uniformly on } X$$

if and only if

$$d_\infty(f_n, f) \to 0 \,.$$

Hence, uniform convergence is simply convergence of sequence in the matric space $(B(X), d_\infty)$.

Example. The space of continuous functions on $[0,1]$ and d_∞.. Let

$$C([0,1]) = \left\{ f : [0,1] \to \mathbb{R} \,\middle|\, f \text{ is continous} \right\}.$$

We now define a metric $d_\infty : C([0,1]) \times C([0,1]) \to \mathbb{R}$ by

$$d_\infty(f,g) = \sup_{x \in X} |f(x) - g(x)| \,.$$

The metric d_∞ is called the ∞-metric (or the uniform metric) on $C([0,1])$. We will now prove that d_∞ is a metrics. Indeed, it is easily seen that d_∞ satisfis conditions (1)–(3). We will now prove that the triangle inequality is satisfied. For $f, g, h \in B(X)$ we have for all $x \in X$,

$$\begin{aligned} |f(x) - h(x)| &\leq |f(x) - g(x)| + |g(x) - h(x)| \\ &\leq \sup_{y \in X} |f(y) - g(y)| + \sup_{y \in X} |g(y) - h(y)| \\ &= d_\infty(f,g) + d_\infty(g,h) \,. \end{aligned}$$

Since x was arbirary this implies that

$$\sup_{x\in X} |f(x) - h(x)| \le d_\infty(f,g) + d_\infty(g,h)\,.$$

Hence

$$d_\infty(f,h) \le d_\infty(f,g) + d_\infty(g,h)\,.$$

This proves the triangle inequality for d_∞.

For example if $f(x) = x$, then

$$d_\infty(f,0) = \sup_{x\in[0,1]} |x-0| = 1\,.$$

Example. The space of continuous functions on $[0,1]$ and d_1.. Let

$$C([0,1]) = \Big\{ f : [0,1] \to \mathbb{R} \,\Big|\, f \text{ is continous} \Big\}\,.$$

We now define a metric $d_1 : C([0,1]) \times C([0,1]) \to \mathbb{R}$ by

$$d_1(f,g) = \int_0^1 |f(x) - g(x)|\, dx\,.$$

The metric d_1 is called the 1-metric on $C([0,1])$. We will now prove that d_∞ is a metrics. Indeed, it is easily seen that d_∞ satisfis conditions (1)–(3). We will now prove that the triangle inequality is satisfied. For $f,g,h \in B(X)$ we have for all $x \in X$,

$$|f(x) - h(x)| \le |f(x) - g(x)| + |g(x) - h(x)|\,.$$

Since x was arbirary this implies that

$$\begin{aligned} d_1(f,h) &= \int_0^1 |f(x) - h(x)|\, dx \\ &\le \int_0^1 (|f(x) - g(x)| + |g(x) - h(x)|)\, dx \\ &= \int_0^1 |f(x) - g(x)|\, dx + \int_0^1 |g(x) - h(x)|)\, dx \\ &= d_1(f,g) + d_1(g,h)\,. \end{aligned}$$

This proves the triangle inequality for d_1.

For example if $f(x) = x$, then

$$d_\infty(f,0) = \sup_{x\in[0,1]} |x-0| = 1\,,$$

but

$$d_1(f,0) = \int_0^1 |x-0|\, dx = \frac{1}{2}\,,$$

Hence, in this case, we have

$$d_1(f,0) \le d_\infty(f,0)\,.$$

The above inequality is, in fact, true for all $f, g \in C([0,1])$. Indeed, for $f, g \in C([0,1])$ we have

$$\begin{aligned} d_1(f,g) &= \int_1^0 |f(x)-g(x)|\,dx \\ &\le \int_1^0 \sup_{y\in[0,1]} |f(y)-g(y)|\,dx \\ &= \int_1^0 d_\infty(f,g)\,dx \\ &= d_\infty(f,g)\,. \end{aligned}$$

Example.. Define $d : (0,\infty)\times(0,\infty) \to \mathbb{R}$ by

$$d(x,y) = \left|\frac{1}{x} - \frac{1}{y}\right|$$

Then d is a metric on $(0,\infty)$.

1.2. Sequences and their limits.

We shall first introduce the meaning of convergence of a sequence of real numbers and establish some elementary (but useful) results about convergent sequences. We present some important criteria for the convergence of sequences. It is important for the reader to learn both the theorems and how the theorems apply to special sequences. We begin with a definition.

Definition. Sequence. *Let X be a set. A sequence in X is a function*

$$f : \mathbb{N} \to X$$

on the set of natural numbers $\mathbb{N}$ whose range is contained in the set X. We will often denote the value of f at n by x_n rather than $f(n)$, and write the sequence as

$$(x_n)_n\,, \quad \textit{or} \quad (x_n)_{n\in\mathbb{N}}\,, \quad \textit{or} \quad (x_n)_{n=1}^{\infty}\,, \quad \textit{or} \quad (x_1, x_2, x_3, \dots)\,.$$

We use paratheses to indicate that the ordering induced by that in $\mathbb{N}$ is a matter of importance. Thus we disguish notationally between the sequence $(x_1, x_2, x_3, \dots)$, whose terms have an ordering, and the set $\{x_1, x_2, x_3, \dots\}$ of values of the terms of this sequence. For example, for the sequence $(x_1, x_2, x_3, \dots) = (1, -1, 1 - 1, 1, -1, 1, -1, \dots)$ we have

$$\{x_1, x_2, x_3, \dots\} = \{-1, 1\}\,.$$

Also, the order of the terms in the sequence is impotant. For example

$$(1, 2, 3, 4, 5, 6, 7, \dots) \neq (2, 1, 3, 4, 5, 6, 7, \dots)\,.$$

In defining sequences it is often convenient to list in order the terms in the sequence, stopping when the rule of formation seems evident. Thus we may write

$$(x_1, x_2, x_3, \dots) = (2, 4, 6, 8, \dots)$$

for the sequence of even natural numbers listed in increasing order, or

$$(y_1, y_2, y_3, \dots) = \left(\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right)$$

for the sequence of reciprocals of the natural numbers listen in decreasing order,

$$(z_1, z_2, z_3, \dots) = \left(\frac{1}{1^2}, \frac{1}{2^2}, \frac{1}{3^2}, \frac{1}{4^2}, \dots\right)$$

for the sequence of reciprocals of the squares of the natural numbers listen in decreasing order. A more satisfactory method is to specify a formula for the general term of the sequence such as

$$(x_n)_n = (2n)_n\,, \quad (y_n)_n = \left(\frac{1}{n}\right)_n\,, \quad (z_n)_n = \left(\frac{1}{n^2}\right)_n\,.$$

Example. If $b \in \mathbb{R}$, then the sequence

$$(x_n)_n = (b, b, b, b, b, \dots),$$

all of whose terms equal b (i.e. $x_n = b$ for all n), is called the constant sequence b. Thus the constant sequence 0 is the sequence $(0, 0, 0, 0, \dots)$, all of whose terms equal 1, and the constant sequence 1 is the sequence $(1, 1, 1, 1, 1, \dots)$.

Example. For $x_n = (-1)^n$ for $n \in \mathbb{N}$, we obtain the sequence

$$(x_n)_n = ((-1)^n)_n = (-1, 1, -1, 1, -1, \dots)\,.$$

Example. For $a_n = \frac{1}{n^3}$ for $n \in \mathbb{N}$, we obtain the sequence

$$(a_n)_n = \left(\frac{1}{n^3}\right) = \left(\frac{1}{1^3}, \frac{1}{2^3}, \frac{1}{3^3}, \dots\right)\,.$$

Definition. Limit of a sequence. *Let (X, d) be a metric space. Let $(x_n)_n$ be a sequence in X and let $x \in X$. The sequence $(x_n)_n$ is said to converge to x (as n tends to infinity) if and only if*

$$\forall \varepsilon > 0 : \exists N \in \mathbb{N} : \forall n \in \mathbb{N} : n \geq N \Rightarrow d(x, x_n) \leq \varepsilon.$$

If $(x_n)_n$ converges to x we will write

$$\lim_n x_n = x \quad \text{or} \quad \lim_{n \to \infty} x_n = x \quad \text{or} \quad x_n \to x \text{ as } n \to \infty,$$

and the element x is called a limit of the sequence.

The sequence $(x_n)_n$ is called convergent if and only if there exists $x \in X$ such that

$$\lim_n x_n = x\,.$$

The sequence $(x_n)_n$ is called divergent if and only if $(x_n)_n$ is not convergent.

Before considering various examples we list and prove two important properties of convergent sequences. The first result says that a sequence cannot converge to two distinct limits.

Theorem. Uniqueness of limits. *Let (X, d) be a metric space. If a sequence converges in (X, d), then its limit is unique.*

Proof. Let $(x_n)_n$ be a sequence and suppose that $(x_n)_n$ converges to both s and t in (X, d). We must now show that $s = t$. Let $\varepsilon > 0$. Since $x_n \to s$ there exists $N \in \mathbb{N}$ such that

$$d(x_n, s) \leq \frac{\varepsilon}{2} \quad \text{for all } n \geq N.$$

Similarly there exists $M \in \mathbb{N}$ such that

$$d(x_n, t) \leq \frac{\varepsilon}{2} \quad \text{for all } n \geq M.$$

Therefore, if $n \geq \max(N, M)$, then the triangle inequality implies that

$$d(s, t) \leq d(s, x_n) + d(x_n, t) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.$$

Since this holds for all $\varepsilon > 0$ we must have $d(s, t)$ whence $s = t$. $\square$

Definition. Bounded sequence. *A sequence $(x_n)_n$ in a metric space (X, d) is called bounded if there exists $M \geq 0$*

$$\sup_{n, m \in \mathbb{N}} d(x_n, x_m) \leq M\,.$$

The next result shows that any convergent sequence is automatically bounded.

Theorem. *Let (X, d) be a metric space. Every convergent sequence in (X, d) is bounded.*

Proof. Let $(x_n)_n$ be a convergent sequence in (X, d) with limit $x \in X$. It follows from the definition of convegence that there exists $N \in \mathbb{N}$ such that $d(x_n, x) \leq 1$ for all $n \geq N$. Hence

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) \leq 1 + 1 = 2 \quad \text{for all } n, m \geq N.$$

Writing $M_0 = \max(d(x, x_1), \dots, d(x, x_{N-1}))$, we also have

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) \leq M_0 + 1 \quad \text{for all } m \geq N \text{ and } n < N,$$

and similarly

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) \leq 1 + M_0 \quad \text{for all } n \geq N \text{ and } m < N.$$

Finally,

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) \leq M_0 + M_0 = 2M_0 \quad \text{for all } m, n < N.$$

Hence

$$d(x_n, x_m) \leq \max(2, 1 + M_0, 2M_0) \quad \text{for all } m, n < N.$$

for all $m, n \in \mathbb{N}$. $\square$

We now consider a large number of examples.

Example. The sequence $(\frac{1}{n})_n$ is convergent in $(\mathbb{R}, |\cdot|)$, in fact,

$$\frac{1}{n} \to 0 \quad \text{as } n \to \infty.$$

Observe that we had to guess the limit of the sequence before we could prove convergence.

Proof. Let $\varepsilon >$. Next choose $N \in \mathbb{N}$ such that $N \geq \frac{1}{\varepsilon}$. For $n \in \mathbb{N}$ with $n \geq N$ we have

$$\left| \frac{1}{n} - 0 \right| = \frac{1}{n} \leq \frac{1}{N} \leq \varepsilon\,.$$

This completes the proof. $\square$

Example. The sequence $(\frac{3n+7}{5n+8})_n$ is convergent in $(\mathbb{R}, |\cdot|)$, in fact,

$$\frac{3n+7}{5n+8} \to \frac{3}{5} \quad \text{as } n \to \infty.$$

Observe that we had to guess the limit of the sequence before we could prove convergence.

Proof. Let $\varepsilon >$. Next choose $N \in \mathbb{N}$ such that $N \geq \frac{25}{11\varepsilon}$. For $n \in \mathbb{N}$ with $n \geq N$ we have

$$\left|\frac{3n+7}{5n+8} - \frac{3}{5}\right| = \left|\frac{5(3n+7) - 3(5n+8)}{5(5n+8)} - \frac{3}{5}\right| = \frac{11}{25n+40} \leq \frac{11}{25n} \leq \varepsilon\,.$$

This completes the proof. □

Example. Let (X, d) be any metric space. Let $x \in X$. The sequence $(x_n)_n = (x, x, x, x, x, \dots)$ is convergent in (X, d), in fact,

$$x_n \to x\,.$$

Observe that we had to guess the limit of the sequence before we could prove convergence.

Proof. Let $\varepsilon >$. Next choose any $N \in \mathbb{N}$. For $n \in \mathbb{N}$ with $n \geq N$ we have

$$d(x_n, x) = d(x, x0 = 0 \leq \varepsilon\,.$$

This completes the proof. □

Example. Consider the metric space $(X, d) = ((0, 1], |\cdot|)$. Let $x_n = \frac{1}{n}$. Then $(x_n)_n$ is a sequence in X. We will show that $(\frac{1}{n})_n$ diverges in $(X, d) = ((0, 1], |\cdot|)$.

Proof. Assume that $(\frac{1}{n})_n$ converges in $(X, d) = ((0, 1], |\cdot|)$. Hence there is $x \in (0, 1]$ such that

$$\tfrac{1}{n} \to x \ \text{ in } ((0, 1], |\cdot|).$$

In particular, since $x > 0$, this implies that there is an N such that

$$|\tfrac{1}{n} - x| \leq \tfrac{x}{2}\,,$$

for $n \geq N$. We may also (since $x > 0$) choose M such that

$$M > \tfrac{2}{x}\,.$$

Let $K = \max(N, M)$. We now have $|\frac{1}{K} - x| \leq \frac{x}{2}$, and so

$$-\tfrac{x}{2} \leq \tfrac{1}{K} - x \leq \tfrac{x}{2}$$

Rearragning this inequality gives

$$\tfrac{x}{2} \leq \tfrac{1}{K}$$

for $n \geq N$, whence

$$K \leq \tfrac{2}{x} < M \leq K\,.$$

This gives the desired contradiction. □

Example. Define $d : (0, \infty) \times (0, \infty) \to \mathbb{R}$ by

$$d(x, y) = \left| \frac{1}{x} - \frac{1}{y} \right|$$

Then d is a metric on $(0, \infty)$. Let $x_n = 1 + \frac{1}{n}$. Then

$$x_n \to 1 \text{ in } ((0, \infty), d).$$

Proof. Let $\varepsilon > 0$. Let $N \geq \varepsilon 1 \varepsilon$. For $n \in \mathbb{N}$ with $n \geq N$, we have

$$\begin{aligned} d(x_n, 1) &= \left| \frac{1}{x_n} - \frac{1}{1} \right| \\ &= \left| \frac{1}{1 + \frac{1}{n}} - \frac{1}{1} \right| \\ &= \left| \frac{-1}{1 + n} \right| \\ &= \frac{1}{1 + n} \\ &\leq \frac{1}{n} \\ &\leq \frac{1}{N} \\ &\leq \varepsilon\,. \end{aligned}$$

This completes the proof. □

Example. Consider the metric space $(C([0, 0]), d_1)$. Define $f_n, f \in C([0, 0])$ by

and

$$f(x) = 0\,.$$

Then

$$f_n \to f \quad \text{in } (C([0,0]), d_1).$$

Proof.
Let $\varepsilon > 0$. Choose $N \geq \frac{1}{2\varepsilon}$. For $n \geq N$ we have

$$\begin{aligned} d_1(f_n, f) &= \int_0^1 |f_n)(x) - f(x)|\, dx \\ &= \int_0^1 f_n(x)\, dx \\ &= \frac{1}{2n} \\ &\leq \frac{1}{2N} \\ &\leq \varepsilon\,. \end{aligned}$$

This completes the proof. $\square$

Example. Consider the metric space $(C([0,0]), d_\infty)$. Define $f_n, f \in C([0,0])$ by

and

$$f(x) = 0\,.$$

It follows from the previous example that

$$f_n \to f \ \text{ in } (C([0,0]), d_1).$$

However, we now make the following Claim.

Claim 1. *We have*

$$f_n \not\to f \ \textit{ in } (C([0,1]), d_\infty).$$

Proof. Observe that for each positive integer n we have

$$\begin{aligned} d_\infty(f_n, f) &= \sup_{x \in [0,1]} |f_n)(x) - f(x)| \\ &= \sup_{x \in [0,1]} f_n(x) \\ &= n \\ &\geq 1\,. \end{aligned}$$

In partcular, this shows that there is no N such that for all $n \geq N$ we have

$$d_\infty(f_n, f) \leq \frac{1}{2}\,.$$

This completes the proof of Claim 1. □

It follows from Claim 1 that f_n not converge to f w.r.t. d_∞. However, perhaps there is another function $g \in C([0,1])$ such that

$$f_n \to g \ \text{ in } (C([0,1]), d_\infty).$$

That this is not the case is the content of the next claim.

Claim 2. *The sequence* $(f_n)_n$ *is divergent in* $(C([0,1]), d_\infty)$*. In particular, for all* $g \in C([0,1])$ *we have We have*

$$f_n \not\to g \quad in \ (C([0,0]), d_\infty).$$

Proof. We want the show that $(f_n)_n$ is divergent in $(C([0,1]), d_\infty)$. Since any convergent sequence is bounded, in order to show that $(f_n)_n$ is divergent, it suffices to show that $(f_n)_n$ is unbounded in $(C([0,1]), d_\infty)$. Hence, we must show that

$$\sup_{n,m} d_\infty(f_n, f_m) = \infty\,.$$

We have

$$n = d_\infty(f_n, f) \leq d_\infty(f_n, f_1) + d_\infty(f_1, f) = d_\infty(f_n, f_1) + \frac{1}{2}\,.$$

This implies that

$$d_\infty(f_n, f_1) \geq n - \frac{1}{2}\,.$$

We conclude from this that

$$\begin{aligned} \sup_{n,m} d_\infty(f_n, f_m) &\geq \sup_n d_\infty(f_n, f_1) \\ &\geq \sup_n \left(n - \frac{1}{2}\right) \\ &= \infty\,. \end{aligned}$$

This completes the proof of Claim 2. □

Example. Let $a \in \mathbb{R}$

(1) If $|a| < 1$, then the sequence $(a^n)_n$ converges to 0 in $(\mathbb{R}, |\cdot|)$, i.e.

$$a^n \to 0 \quad \text{as } n \to \infty.$$

(2) If $a > 1$, then the sequence $(a^n)_n$ converges to infinity, i.e.

$$a^n \to \infty \quad \text{as } n \to \infty.$$

Proof. We start by reminding the reader about the following inequality known as Bernoulli's inequality. For $x \geq 0$ and $n \in \mathbb{N}$ we have

$$(1 + x)^n \geq 1 + nx\,.$$

It is straight forward to prove this inequality by induction on n.

(1) Let $\varepsilon > 0$. Next choose $N \in \mathbb{N}$ such that $N \geq \frac{1}{\varepsilon(\frac{1}{|a|}-1)}$. For $n \in \mathbb{N}$ with $n \geq N$ we have (using Bernoulli's inequality)

$$\begin{aligned} |a^n - 0| = |a|^n &= \frac{1}{\left(1+\left(\frac{1}{|a|}-1\right)\right)^n} \leq \frac{1}{1+n\left(\frac{1}{|a|}-1\right)} \\ &\leq \frac{1}{n\left(\frac{1}{|a|}-1\right)} \leq \frac{1}{N\left(\frac{1}{|a|}-1\right)} \leq \varepsilon\,. \end{aligned}$$

(2) Let $A > 0$. Choose $N \in \mathbb{N}$ such that $N \geq \frac{A}{a-1}$. For $n \in \mathbb{N}$ with $n \geq N$ we now have (using Bernoulli's inequality)

$$a^n = (1+(a-1))^n \geq 1 + n(a-1) \geq n(a-1) \geq N(a-1) \geq A\,.$$

This completes the proof. □

Example. Let $a \in \mathbb{N}$ and $b > 1$. Then the sequence $(\frac{n^a}{b^n})_n$ is convergent in $(\mathbb{R}, |\cdot|)$. In fact,

$$\frac{n^a}{b^n} \to 0 \quad \text{as } n \to \infty.$$

Proof. We first notice that for $0 \leq x \leq 1$ and $a \in \mathbb{N}$ we have the following "reverse Bernoulli" inequality $(1+x)^a \leq 1+a^a x$ (this follows by an easy inductive proof). For brevity write $x_n = \frac{n^a}{b^n}$ for all n and observe that the "reverse Bernoulli" inequality implies that

$$x_n = x_{n-1}\frac{1}{b}\left(1+\frac{1}{n-1}\right)^a \leq x_{n-1}\frac{1}{b}\left(1+a^a\frac{1}{n-1}\right)\,.$$

Now, let $\varepsilon > 0$. Since $\frac{1}{b} < 1$, there exists $c \in \mathbb{R}$ such that $\frac{1}{b} < c < 1$. As $bc - 1 > 0$ we may choose $K \in \mathbb{N}$ such that

$$\frac{1}{K} < \frac{bc-1}{a^a} \quad \text{i.e.} \quad \frac{1}{b}\left(1+a^a\frac{1}{K}\right) < c\,.$$

Also, since $c^n \to 0$ as $n \to \infty$ (because $0 < c < 1$), we may choose $M \in \mathbb{N}$ such that

$$c^n \leq \varepsilon\frac{c^K}{x_K} \quad \text{for } n \geq M.$$

Finally, let $N = \max(K, M)$. For $n \in \mathbb{N}$ with $n \geq N$ we have

$$\begin{aligned} \left|\frac{n^a}{b^n} - 0\right| &= x_n \leq x_{n-1}\frac{1}{b}\left(1+a^a\frac{1}{n-1}\right) \leq x_{n-2}\frac{1}{b^2}\left(1+a^a\frac{1}{n-1}\right)\left(1+a^a\frac{1}{n-2}\right) \\ &\leq \cdots \leq x_K\frac{1}{b^{n-K}}\left(1+a^a\frac{1}{n-1}\right)\left(1+a^a\frac{1}{n-2}\right)\cdots\left(1+a^a\frac{1}{K}\right) \\ &\leq x_K\frac{1}{b^{n-K}}\left(1+a^a\frac{1}{K}\right)^{n-K} = x_K\left(\frac{1}{b}\left(1+a^a\frac{1}{K}\right)\right)^{n-K} \leq x_K c^{n-K} \leq \varepsilon\,. \end{aligned}$$

This completes the proof. $\square$

Example. The sequence $((-1)^n) = (-1, 1, -1, 1, -1, 1, \dots)$ is divergent in $(\mathbb{R}, |\cdot|)$.

Proof. Assume in order to reach a contradiction that the sequence $((-1)^n) = (-1, 1, -1, 1, -1, 1, \dots)$ is convergent. Hence there exists $x \in \mathbb{R}$ such that $(-1)^n \to x$ as $n \to \infty$. We can therefore choose $N \in \mathbb{N}$ such that $|(-1)^n - x| \leq \frac{1}{2}$ for $n \geq N$. As $2N+1, 2N \geq N$ we thus conclude that

$$|1 - x| = |(-1)^{2N} - x| \leq \frac{1}{2} \quad \text{and} \quad |-1 - x| = |(-1)^{2N+1} - x| \leq \frac{1}{2},$$

whence $2 = |1 - (-1)| \leq |1 - x| + |x - (-1)| \leq \frac{1}{2} + \frac{1}{2} = 1$. This yields the desired contradiction. $\square$

1.3. Limit theorems.

In the previous section we saw that the definition of convergence can sometimes be messy to use even for sequences given by relatively simple formulas. We will now derive som basic results that simplify our work.

Theorem. *Let $k \in \mathbb{R}$ and let $(x_n)_n$ and $(y_n)_n$ be convegent sequences in $(\mathbb{R}, |\cdot|)$ with*

$$x_n \to x \quad \text{and} \quad y_n \to y\,.$$

Then

(1) $x_n + y_n \to x + y$ *in* $(\mathbb{R}, |\cdot|)$.
(2) $x_n y_n \to xy$ *in* $(\mathbb{R}, |\cdot|)$.
(3) *If* $y_n \neq 0$ *for all* n *and* $y \neq 0$, *then* $\frac{x_n}{y_n} \to \frac{x}{y}$ *in* $(\mathbb{R}, |\cdot|)$.
(4) $kx_n \to kx$ *and* $k + x_n \to k + x$ *in* $(\mathbb{R}, |\cdot|)$.

Proof.
(1) Let $\varepsilon > 0$. Since $x_n \to x$ there exists $N_1 \in \mathbb{N}$ such that

$$|x_n - x| \leq \frac{\varepsilon}{2} \quad \text{for all } n \geq N_1.$$

Similary, there exists $N_2 \in \mathbb{N}$ such that

$$|y_n - y| \leq \frac{\varepsilon}{2} \quad \text{for all } n \geq N_2.$$

Thus, if we let $N = \max(N_1, N_2)$, then we have for all $n \geq N$,

$$|(x_n + y_n) - (x + y)| = |(x_n - x) + (y_n - y)| \leq |x_n - x| + |y_n - y| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.$$

(2) Let $\varepsilon > 0$. We know from the previous Theorem that the convergent sequence $(x_n)_n$ is bounded, i.e. there exists $M \geq 0$ such that

$$|x_n| \leq M \quad \text{for all } n.$$

Since $x_n \to x$ there exists $N_1 \in \mathbb{N}$ such that

$$|x_n - x| \le \frac{\varepsilon}{2(|y|+1)} \quad \text{for all } n \ge N_1.$$

Similary, there exists $N_2 \in \mathbb{N}$ such that

$$|y_n - y| \le \frac{\varepsilon}{2(M+1)} \quad \text{for all } n \ge N_2.$$

Thus, if we let $N = \max(N_1, N_2)$, then we have for all $n \ge N$,

$$\begin{aligned} |x_n y_n - xy| &= |x_n y_n - x_n y + x_n y - xy| \le |x_n y_n - x_n y| + |x_n y - xy| \\ &= |x_n||y_n - y| + |y||x_n - x| \le M|y_n - y| + |y||x_n - x| \\ &\le M\frac{\varepsilon}{2(M+1)} + |y|\frac{\varepsilon}{2(|y|+1)} \le \varepsilon\,. \end{aligned}$$

(3) Since $\frac{x_n}{y_n} = x_n \frac{1}{y_n}$, Part (2) shows that it suffices to prove that $\frac{1}{y_n} \to \frac{1}{y}$. Since $y_n \to y$ there exists $N_1 \in \mathbb{N}$ such that

$$|y_n - y| \le \frac{|y|}{2} \quad \text{for all } n \ge N_1.$$

Observe that if $n \ge N_1$, then

$$|y_n| = |y - (y - y_n)| \ge \big||y| - |y - y_n|\big| \ge |y| - \frac{|y|}{2} = \frac{|y|}{2}\,.$$

Similary there exists $N_2 \in \mathbb{N}$ such that

$$|y_n - y| \le \frac{\varepsilon |y|^2}{2} \quad \text{for all } n \ge N_2.$$

Thus, if we let $N = \max(N_1, N_2)$, then we have for all $n \ge N$, that

$$\left|\frac{1}{y_n} - \frac{1}{y}\right| = \left|\frac{y - y_n}{y_n y}\right| = \frac{|y - y_n|}{|y_n||y|} \le \frac{2}{|y|^2}|y_n - y| \le \varepsilon\,.$$

(4) This is left as an exercise for the reader. $\square$

1.4. Open sets and closed sets.

First recall the definition of a ball.

Definition. Ball. *Let* (X, d) *be a metric space. For* $x \in X$ *and* $r \geq 0$*, then ball* $B(x, r)$ *with centre* x *and radius* r *is defined by*

$$B(x, r) = \{y \in X \mid d(x, y) < r\} .$$

We can now define the notion of an open set and of a closed set.

Definition. Open set. *Let* (X, d) *be a metric space. A subset* $G \subseteq X$ *is called open if and only if for each* $x \in G$ *there is a positive number* $r > 0$ *such that*

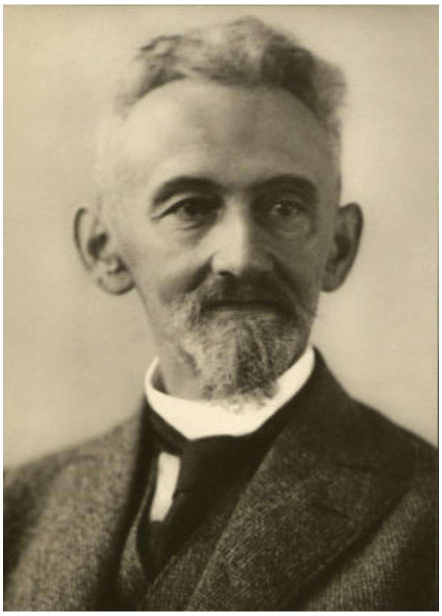
$$B(x, r) \subseteq G .$$

That is, a subset $G \subseteq X$ *is called open if and only if*

$$\forall x \in G : \exists r > 0 : B(x, r) \subseteq G .$$

Definition. Closed set. *Let* (X, d) *be a metric space. A subset* $F \subseteq X$ *is called closed if the complement* $X \setminus F$ *is open.*

The significant of open and closed sets and their importance in mathematical was first realised by Felix Hausdorff in the 1920's and is elaborated on in Hausdorff's seminal text "Grundzuge der Mengen Lehre".



Felix Hausdorff (November 8, 1868 – January 26, 1942) was a German mathematician who is considered to be one of the founders of modern topology and who contributed significantly to set theory, descriptive set theory, measure theory, function theory, and functional analysis.

We will now look at several examples.

Example. Balls are open Let (X, d) be a metric space. Then every ball

$$B(x, r) = \{y \in X \mid d(x, y) < r\}$$

is open
Proof. Let $y \in B(x, r)$. We must now find $\rho > 0$ such that

$$B(y, \rho) \subseteq B(x, r) .$$

Put

$$\rho = r - d(x, y) .$$

Observe that since $y \in B(x, r)$, we have $d(x, y) < r$, whence $\rho = r - d(x, y) > 0$. We now claim that $B(y, \rho) \subseteq B(x, r)$. Therefor fix $z \in B(y, \rho)$. We must now prove that $z \in B(x, r)$. Since $z \in B(y, \rho)$, we conclude that $d(z, y) < \rho$, whence

$$\begin{aligned} d(z, x) &\leq d(z, y) + d(y, x) \\ &< \rho + d(y, x) \\ &= r - d(x, y) + d(y, x) \\ &= r . \end{aligned}$$

This shows that $z \in B(x, r)$.

Example. X **is open** Let (X, d) be a metric space. Then X is open.
Proof. Let $x \in X$. We must now find $r > 0$ such that

$$B(x, r) \subseteq X\,.$$

However, for any $r > 0$, we have $B(x, r) \subseteq X$.

Example. $\varnothing$ **is open** Let (X, d) be a metric space. Then $\varnothing$ is open.
Proof. Let $x \in \varnothing$. We must now find $r > 0$ such that

$$B(x, r) \subseteq X\,.$$

However, this is clearly true since there is no $x \in \varnothing$.

Example. This example shows that there are metric spaces in which sets are both open and closed. Let (X, d) be the metric space given by $((0, 2] \cup \{9\}, |\cdot|)$.

We now have:

(1) The set $(0, 2]$ is open in X.
Proof. Indeed, for all $x \in (0, 2]$, it is clear that

$$B(x, 3) = \{y \in X \mid |y - x| < 3\} = (0, 2] \subseteq X\,.$$

This completes the proof. ☐

(1) The set $\{9\}$ is open in X.
Proof. Indeed, for all $x \in \{9\}$, i.e. for $x = 9$, it is clear that

$$B(x, 3) = \{y \in X \mid |y - x| < 3\} = \{9\} \subseteq X\,.$$

This completes the proof. ☐

(1) The set $(0, 2]$ is closed in X.
Proof. Indeed, since

$$X \setminus (0, 2] = \{9\}$$

and $\{9\}$ is open in X (by the above), we conclude that $(0, 2]$ is closed in X. This completes the proof. ☐

(1) The set $\{9\}$ is closed in X.
Proof. Indeed, since

$$X \setminus \{9\} = (0, 2]$$

and $(0, 2]$ is open in X (by the above), we conclude that $(0, 2]$ is closed in X. This completes the proof. ☐

The sets $(0, 2]$ and $\{9\}$ are both open and closed in X.

Example. Consider the metric spaces $(C([0,1]), d_\infty)$ and $(C([0,1]), d_1)$, and let

$$G = \left\{ f \in C([0,1]) \,\middle|\, f(0) \neq 0 \right\}.$$

We will now show that G is open in $(C([0,1]), d_\infty)$ but that G is not open in $(C([0,1]), d_1)$

Claim 1. *The set G is open in $(C([0,1]), d_\infty)$.*

Proof.
Let $f \in G$.
We must now find $r > 0$ such that

$$B(f,r) \subseteq G\,.$$

Let $r = \frac{|f(0)|}{2}$. Since $f \in G$, we conclude that $r > 0$. We will now prove that

$$B(f,r) \subseteq G\,.$$

Therefore, fix $g \in B(f,r)$. We must now show that $g \in G$, i.e. we must show that $g(0) \neq 0$.
Since $g \in B(f,r)$ we conclude that

$$\begin{aligned} |f(0)| &\leq |f(0) - g(0)| + |g(0)| \\ &\leq \sup_{x \in [0,1]} |f(x) - g(x)| + |g(0)| \\ &= d_\infty(f,g) + |g(0)| \\ &< r + |g(0)| \\ &= \frac{|f(0)|}{2} + |g(0)|\,, \end{aligned}$$

and so

$$0 < \frac{|f(0)|}{2} = |f(0)| - \frac{|f(0)|}{2} \leq |g(0)|\,.$$

This completes the proof. $\square$

Claim 2. *The set G is not open in $(C([0,1]), d_1)$.*

Proof.
We must show the following:

> there is $f \in G$ such that for all $r > 0$, we have $B(f,r) \not\subseteq G$.

In fact, we will show the significantly stronger statement (showing the G is "very very far" from being open in $(C([0,1]), d_1)$):

> for all $f \in G$ and for all $r > 0$, we have $B(f,r) \not\subseteq G$.

Therefore fix $f \in G$ and $r > 0$. We must now show that $B(f,r) \not\subseteq G$, i.e. we must find $g \in B(f,r)$ with $g \in C([0,1]) \setminus G$.

First define $h \in C([0,1])$ by

$$h(x) = \begin{cases} -f(0)\frac{|f(0)|}{r}x + f(0) & \text{for } x \in [0, \frac{r}{|f(0)|}]; \\ 0 & \text{for } x \in (\frac{r}{|f(0)|}, 1]. \end{cases}$$

Now simply put

$$g = f - h\,.$$

We then have $g \in B(f,r)$ with $g \in C([0,1]) \setminus G$.
We first prove that $g \in B(f,r)$, i.e. $d_1(f,g) < r$. Indeed, this follows from the fact that

$$d_1(f,g) = \int_0^1 |f(x) - g(x)|\,dx = \int_0^1 |h(x)|\,dx = \frac{1}{2}\,|f(0)|\,\frac{r}{|f(0)|} = \frac{r}{2} < r\,.$$

Next we prove that $g \in C([0,1]) \setminus G$, i.e. $g(0) = 0$. Indeed, this follows from the fact that

$$g(0) = f(0) - h(0) = f(0) - f(0) = 0\,.$$

This completes the proof. $\square$

After having presented a number of examples, we return to the theoretical development again. We start by presenting a useful alternative characterization of closedness.

Theorem. Sequence characterization of closedness. *Let (X,d) be a metric space and let $F \subseteq X$. Then the following two statements are equivalent:*

(1) *F is closed.*
(2) *If $(x_n)_n$ is a sequence in F and $x \in X$ and*

$$x_n \to x\,,$$

then

$$x \in F\,.$$

Proof.
(1) $\Rightarrow$ (2). Assume that F is closed. Let $(x_n)_n$ be a sequence in F and let $x \in X$. Assume that

$$x_n \to x\,.$$

We must now prove that

$$x \in F\,.$$

Assume in order to reach a contradiction that

$$x \notin F\,.$$

Hence $x \in X \setminus F$. Since F is closed, the complement $X \setminus F$ is open. Hence x is an element of the open set $X \setminus F$. It therefore follows from the definition of openess that there is a positive number $r > 0$ such that

$$B(x,r) \subseteq X \setminus F\,.$$

However, we also know that $x_n \to x$. This implies that there is an integer N such that $d(x_n,x) \leq \frac{r}{2} < r$ for all $n \geq N$. Hence

$$x_n \in B(x,r) \subseteq X \setminus F \quad \text{for all } n \geq N. \tag{1}$$

However, $(x_n)_n$ is a sequence in F and so

$$x_n \in F \quad \text{for all } n \geq N. \tag{2}$$

The desired contradiction follows from (1) and (2).
(2) $\Rightarrow$ (1). We must now prove that F is closed, i.e. we must prove that $X \setminus F$ is open. Therefore fix $x \in X \setminus F$. We must now show that there is a positive number $r > 0$ such that

$$B(x,r) \subseteq X \setminus F\,.$$

Assume, in order to reach a contradiction, that this is not satisfied, i.e. we are assuming that

$$\forall r > 0 \,:\, B(x,r) \not\subseteq X \setminus F\,.$$

In particular this implies, that for all positive integers n we have

$$B(x, \tfrac{1}{n}) \not\subseteq X \setminus F\,,$$

whence there is $x_n \in B(x, \frac{1}{n}) \cap F$. This clearly implies that $x_n \to x$, and since also $x_n \in F$ for all n, we therefore deduce from (2) that $x \in F$. However, this contradicts that fact that $x \in X \setminus F$. $\square$

Example. Uniform convergence of continuous functions. Recall the following main theorem from earlier courses in analysis.

Theorem A. Uniform convergence of continuous functions. *Let $f, f_1, f_2, \ldots \in C([0,1])$ and assume that f_n is continuous for all n. If*

$$f_n \to f \quad \textit{uniformly on } [0,1],$$

f is continuous.

The prupose of this example is to show (using the above theorem obn the sequence charaqcterization of closedness) that this result simply says that the set of continuous functions is a closed subset of the set of bounded functions w.r.t. the d_∞ metric.

Recall, that

$$B([0,1]) = \left\{ f : [0,1] \to \mathbb{R} \,\middle|\, f \text{ is bounded} \right\},$$

and

$$C([0,1]) = \left\{ f : [0,1] \to \mathbb{R} \,\middle|\, f \text{ is continuous} \right\},$$

Theorem A can now be reformulated in the following way.

Theorem. Uniform convergence of continuous functions (reformulation). *The set $C([0,1])$ is a closed subset of the metric space $(B([0,1]), d_\infty)$.*

Proof.
First it follows from a result from earlier courses in analysis that any continuous function defined on a bounded interval is bounded, and so $C([0,1]) \subseteq B([0,1])$.

We will now show that $C([0,1])$ is a closed subset of the metric space $(B([0,1]), d_\infty)$. In order to to this we will use the sequence characterization of closedness. Therefore let $(f_n)_n$ be a sequence in $C([0,1])$ and let $f \in B([0,1])$ and assume that $f_n \to f$ in $(B([0,1]), d_\infty)$. We must now prove that $f \in C([0,1])$.

However, since $f_n \to f$ in $(B([0,1]), d_\infty)$, we conclude that

$$f_n \to f \quad \text{uniformly on } [0,1],$$

and since all the f_n's are continuous, we now deduce from Theorem A that f is continuous, i.e. $f \in C([0,1])$. This completes the proof. $\square$

We now return to development of some of the properties of closed and open sets.

Theorem. Properties of open sets. *Let (X, d) be a metric space*

(1) *If $G_1, \dots, G_n$ are finitely many open subsets of X, then*

$$\bigcap_{i=1}^{n} G_i$$

is also open.

(2) *If $(G_i)_{i \in I}$ is an arbitrary (and possibly uncountable) family of open subsets of X, then*

$$\bigcup_{i \in I} G_i$$

is also open.

Proof.
(1). Let $x \in \cap_{i=1}^{n} G_i$. This implies that $x \in G_i$ for all $i = 1, \dots, n$. Since G_i is open and $x \in G_i$, it follows from the definition of openess that there is a positive number $r_i > 0$ such that

$$B(x, r_i) \subseteq G_i \,.$$

Now put $r = \min_{i=1,\dots,n} r_i$. Observe that $r > 0$ (since r is the minumum of *finitely* many strictly positive numbers). It is clear that

$$B(x, r) \subseteq B(x, r_i) \subseteq G_i$$

for all i, whence $B(x, r) \subseteq \cap_{i=1}^{n} G_i$. This proves that $\cap_{i=1}^{n} G_i$ is open.
(1). Let $x \in \cup_{i \in I} G_i$. This implies that there an $i_0 \in I$ such that $x \in G_{i_0}$. Since G_{i_0} is open and $x \in G_{i_0}$, it follows from the definition of openess that there is a positive number $r > 0$ such that

$$B(x, r) \subseteq G_{i_0} \,.$$

In particular,

$$B(x, r) \subseteq G_{i_0} \subseteq \bigcup_{i \in I} G_i \,.$$

This proves that $\cup_{i \in I} G_i$ is open. □

Theorem. Properties of closed sets. *Let (X, d) be a metric space*

(1) *If $F_1, \dots, F_n$ are finitely many closed subsets of X, then*

$$\bigcup_{i=1}^{n} G_i$$

is also open.

(2) *If $(F_i)_{i \in I}$ is an arbitrary (and possibly uncountable) family of closed subsets of X, then*

$$\bigcap_{i \in I} F_i$$

is also closed.

Proof. This is left as an exercise to the reader. □

2. COMPLETENESS AND COMPACTNESS

The two main notions in this chapter, namely, completness and compactness are properties not enjoyed by metric spaces in general. However, it is hard to overestimate the significance and importance of these notions.

2.1. Cauchy sequences and complete metric spaces.

It is important for us to have a condition implying the convergence of a sequence that does not require the knowledge of the value of the limit.

Definition. Cauchy sequence. *Let (X, d) e a metric space. A sequence $(x_n)_n$ in X is called a Cauchy sequence if and only if*

$$\forall \varepsilon > 0 : \exists N \in \mathbb{N} : \forall n, m \in \mathbb{N} : n, m \geq N \Rightarrow d(x_n, x_m) \leq \varepsilon .$$

The reader should compare this definition closely with the definition of a convergent sequence.

Example. Define $f_n \in C([0,1])$ by

$$f_n(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} .$$

Then $(f_n)_n$ is a Cauchy sequence in $(C([0,1]), d_\infty)$.
Proof. Let $\varepsilon > 0$.
Choose $N \geq \frac{1}{\varepsilon}$.
For $n \geq m \geq N$, we have

$$\begin{aligned}
d_\infty(f_n, f_m) &= \sup_{x\in[0,1]} |f_n(x) - f_m(x)| \\
&= \sup_{x\in[0,1]} \left| \left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^m}{m!} + \cdots + \frac{x^n}{n!}\right) - \left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^m}{m!}\right) \right| \\
&= \sup_{x\in[0,1]} \left| \frac{x^{m+1}}{(m+1)!} + \cdots + \frac{x^n}{n!} \right| \\
&\leq \frac{1}{(m+1)!} + \cdots + \frac{1}{n!} \\
&\leq \frac{1}{m(m+1)} + \cdots + \frac{1}{(n-1)n} \\
&= \left(\frac{1}{m} - \frac{1}{m+1}\right) + \left(\frac{1}{m+1} - \frac{1}{m+2}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) \\
&= \frac{1}{m} - \frac{1}{n} \\
&\leq \frac{1}{m} \\
&\leq \frac{1}{N} \\
&\leq \varepsilon .
\end{aligned}$$

This completes the proof and the example. □

Do we know any other Cauchy sequences? Yes, in fact, the next result shows that any convergent sequence is Cauchy.

Proposition. *Let (X,d) e a metric space. Every convergent sequence in (X,d) is a Cauchy sequence.*

Proof. Let $(x_n)_n$ be a convergent sequence with limit x. Let $\varepsilon > 0$. Since $x_n \to x$ there exists $N \in \mathbb{N}$ such that $d(x_n, x) \leq \frac{\varepsilon}{2}$ for all $n \geq N$. Hence for $n, m \in \mathbb{N}$ with $n, m \geq N$ we have

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.$$

This completes the proof. □

The natural question now is:

Is the converse true, i.e. is any Cauchy sequence convergent?

It is known from earlier courses in analysis that in the metric space $(\mathbb{R}, |\cdot|)$, the converse to the above proposition is true, i.e. in in the metric space $(\mathbb{R}, |\cdot|)$ any Cauchy sequence is convergent. However, the next example shows that is need necessarily the case in an arbitrary metric space.

Example. We have

(1) The sequence $(x_n)_n = (\frac{1}{n})_n$ is Cauchy in $(\mathbb{R} \setminus \{0\}, |\cdot|)$.
(2) The sequence $(x_n)_n = (\frac{1}{n})_n$ is not convergent in $(\mathbb{R} \setminus \{0\}, |\cdot|)$.

Proof
(1) Let $\varepsilon > 0$. Choose $N \geq \frac{2}{\varepsilon}$. For $n, m \geq N$, we have

$$\begin{aligned} |x_n - x_m| &= \left|\frac{1}{n} - \frac{1}{m}\right| \\ &\leq \frac{1}{n} + \frac{1}{m} \\ &\leq \frac{1}{N} + \frac{1}{N} \\ &= \frac{2}{N} \\ &\leq \varepsilon\,. \end{aligned}$$

This shows that $(x_n)_n$ is Cauchy in $(\mathbb{R} \setminus \{0\}, |\cdot|)$.
(2) Assume in order to reach a contradiction that there is $x \in \mathbb{R} \setminus \{0\}$ such that $x_n \to x$ w.r.t. $|\cdot|$.

Recall that the following follws from the tri-angle inequality: if $u, v \in \mathbb{R}$, then

$$\big|\,|u| - |v|\,\big| \leq |u - v|\,. \tag{1}$$

(Prove this inequality.)

Since $\frac{|x|}{2} > 0$ (because $x \neq 0$), we can find N such that if $n \geq N$, then (using (1))

$$\begin{aligned} \frac{1}{n} - |x| &= \left|\frac{1}{n}\right| - |x| \\ &\leq \left|\frac{1}{n} - x\right| \\ &\leq \frac{|x|}{2}\,. \end{aligned}$$

It follows from this that, if $n \geq N$, then

$$\frac{|x|}{2} \leq \frac{1}{n}\,.$$

This in turn implies that, if $n \geq N$, then

$$n \leq \frac{2}{|x|}\,,$$

and so

$$\infty = \sup_{n \geq N} n \leq \frac{2}{|x|} < \infty\,.$$

This gives the desired contradiction. □

The previous example shows that there are metric spaces in which Cauchy sequences need not be convergent. It is therefore sensible to isolate those metric spaces for which this is not the case. This is the content of the next definition.

Definition. Completeness. *A metric (X, d) is called complete if and only if every Cauchy sequence is convergent.*

While some notion resembling the Cauchy condition goes back to Augustin Cauchy, the real significance of this idea and the related notion of completeness is due to Stefan Banach.

Augustin Cauchy (21 August 1789 – 23 May 1857) was a French mathematician reputed as a pioneer of analysis. He was one of the first to state and prove theorems of calculus rigorously, rejecting the heuristic principle of the generality of algebra of earlier authors. He almost singlehandedly founded complex analysis and the study of permutation groups in abstract algebra. His writings range widely in mathematics and mathematical physics.

Stefan Banach (30 March 1892 – 31 August 1945) was a Polish mathematician. He was one of the founders of modern functional analysis, and an original member of the Lwow School of Mathematics. His major work was the 1932 book, Theorie des operations lineaires, the first monograph on the general theory of functional analysis

Theorem. *Let (X, d) be a complete metric space and let F be a subset of X. Then the following two statements are equivalent:*

(1) *The metric space (F, d) is complete.*
(2) *The set F is a closed subset of X.*

Proof.
(1) $\Rightarrow$ (2). Assume that (F, d) is a complete metric space. We must now prove that F is a closed subset of X. Therefore let $(x_n)_n$ be a sequence from F and let $x \in X$, and assume that

$$x_n \to x \,.$$

We mus now prove that

$$x \in F \,.$$

Since $(x_n)_n$ converges to x in (X, d) it is easily seen that $(x_n)_n$ is Cauchy in (F, d). (Indeed, let $\varepsilon > 0$. Since $x_n \to x$ in X there exists $N \in \mathbb{N}$ such that $d(x_n, x) \le \frac{\varepsilon}{2}$ for all $n \ge N$. Hence for $n, m \in \mathbb{N}$ with $n, m \ge N$ we have

$$d(x_n, x_m) \le d(x_n, x) + d(x, x_m) \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \,.$$

This shows that $(x_n)_n$ is Cauchy in (F, d).) Since $(x_n)_n$ is Cauchy in (F, d) and (F, d) is complete there is $y \in F$ such that

$$x_n \to y \,.$$

As $x_n \to x$ and $x_n \to y$, we conclude that $x = y$, and so $x = y \in F$.
(2) $\Rightarrow$ (1). Assume that F is a cosed subset of X. We must now prove that (F, d) is a complete metric space. Therefore let $(x_n)_n$ be a Cauchy sequence in (F, d). It is clear that $(x_n)_n$ is also a Cauchy sequence in (X, d) and since (X, d) is complete, we conclude that there is $x \in X$ such that

$$x_n \to x \,.$$

However, as F is a closed subset of X and $x_n \in F$ for all n, we deduce that $x \in F$. Hence $x \in F$ and $x_n \to x$. This shows that (F, d) is complete. $\square$

Theorem. *Let F be a subset of $\mathbb{R}$. Then the following two statements are equivalent:*

(1) *The metric space $(F, |\cdot|)$ is complete.*
(2) *The set F is a closed subset of $\mathbb{R}$.*

Proof. This follows from the previous Theorem since it is known from earlier courses in analysis that $(\mathbb{R}, |\cdot|)$ is complete. $\square$

Example. The completeness of $(C(X), d_\infty)$.
Let X be a subset of $\mathbb{R}$ and recall that

$$C(X) = \left\{ f : X \to \mathbb{R} \,\middle|\, f \text{ is continuous} \right\}.$$

Then the metric space $(C(X), d_\infty)$ is complete.
Proof. Let $(f_n)_n$ be a Cauchy sequence in $(C(X), d_\infty)$. We first prove the following claim.

Claim 1. For each $x \in X$, the sequence $(f_n(x))_n$ is Cauchy in $\mathbb{R}, |\cdot|)$. In particular, For each $x \in X$, the sequence $(f_n(x))_n$ is convergent in $\mathbb{R}, |\cdot|)$.
Proof of Claim 1. Let $x \in X$ and let $\varepsilon > 0$. Since $(f_n)_n$ is Cauchy in $(C(X), d_\infty)$, there is an N such that if $n, m \geq N$, then

$$d_\infty(f_n, f_m) \leq \varepsilon\,.$$

Hence, for $n, m \geq N$, we have

$$|f_n(x) - f_m(x)| \leq \sup_{y \in X} |f_n(y) - f_m(y)| \leq d_\infty(f_n, f_m) \leq \varepsilon\,.$$

This proves Claim 1.

It follows from Claim 1, that for each $x \in X$, the sequence $(f_n(x))_n$ is convergent in $\mathbb{R}, |\cdot|)$. Hence, we can define $f : X \to \mathbb{R}$ by

$$f(x) = \lim_n f_n(x)\,.$$

We must now prove that $f \in C(X)$ and that $f_n \to f$ w.r.t. d_∞; this is done in Claim 2 and Claim 3 below.

Claim 2. $f_n \to f$ w.r.t. d_∞.
Proof of Claim 2. Let $\varepsilon > 0$.
Since $(f_n)_n$ is Cauchy in $(C(X), d_\infty)$ we can find N such that if $n, m \geq N$, then

$$d_\infty(f_n, f_m) \leq \varepsilon\,. \tag{1}$$

We now claim that if $n \geq N$, then

$$d_\infty(f_n, f) \leq \varepsilon\,. \tag{2}$$

Therefore fix $n \geq N$. Also, fix $x \in X$.
Next, observe that

$$\text{for all } m \geq N, \text{ we have } |f_n(x) - f_m(x)| \leq d_\infty(f_n, f_m) \leq \varepsilon \tag{3}$$

$$|f_n(x) - f_m(x)| \to |f_n(x) - f(x)| \quad \text{as } m \to \infty. \tag{4}$$

It follows from (3) and (4) that

$$|f_n(x) - f(x)| \leq \varepsilon\,.$$

Since $x \in X$ was arbitrary this shows that

$$d_\infty(f_n, f) \leq \sup_{x \in X} |f_n(x) - f(x)| \leq \varepsilon \,.$$

This completes the proof of Claim 2.

Claim 3. $f \in C(X)$.
Proof of Claim 3. Let $\varepsilon > 0$.
Since $f_n \to f$ w.r.t. d_∞, we conclude that

$$f_n \to f \quad \text{uniformaly on } X.$$

Since each f_n is continuous, it therefore follows that $f \in C(X)$. This completes the proof of Claim 2.

Combining Claim 2 and Claim 3 proves that $(C(X), d_\infty)$ is complete. This completes the Exanmple. $\square$

As an application of the previous example and various of the theorems in this section, we provide the following example showing that a certain subspace of $C([0,1])$ is complete.

Example. Let

$$M = \left\{ f \in C([0,1]) \,\middle|\, f(0) = 0 \right\}.$$

Then (M, d_∞) is a complete metric space.
Proof. It follows from the previous Example that $C([0,1]), d_\infty)$ is complete, and since $M \subseteq C([0,1])$, it suffuces (see Theorem ???) to show that M is a closed subset of $C([0,1]), d_\infty)$. To prove this observe that it follows from Example ??? that

$$G = C([0,1]) \setminus M = \left\{ f \in C([0,1]) \,\middle|\, f(0) \neq 0 \right\}$$

is an open subset of $C([0,1]), d_\infty)$. Hence, M is a closed subset of $C([0,1]), d_\infty)$. $\square$

Example. The metric space $([0,\infty),|\cdot|)$ is clearly complete (why?). We will now change the metric in $[0,\infty)$ in such that way that the set $[0,\infty)$ with ther new metric is no longer complete.

Defihne the metric d in $[0,\infty)$ by

$$d(x,y)=\left|\frac{x}{1+x}-\frac{y}{1+y}\right|.$$

It is now difficult to see that d is a metric in $[0,\infty)$ (write down a complete proof of this).

We now have

(1) The sequence $(x_n)_n=(n)_n$ is Cauchy in $([0,\infty),d)$.
(2) The sequence $(x_n)_n=(n)_n$ is not convergent in $([0,\infty),d)$.

It follows from (1) and (2) that $(x_n)_n=(n)_n$ is a divergent sequence in $([0,\infty),d)$, and the metric space $([0,\infty),d)$ is therefore not complete.

Proof.

(1) Let $\varepsilon>0$. Choose $N\geq\frac{21}{\varepsilon}$. For $n,m\geq N$, we have

$$\begin{aligned}
d(x_n,x_m)&=d(n,m)\\
&=\left|\frac{n}{1+n}-\frac{m}{1+m}\right|\\
&=\left|\frac{n-m}{nm}\right|\\
&=\left|\frac{1}{m}-\frac{1}{n}\right|\\
&\leq\frac{1}{m}+\frac{1}{n}\\
&\leq\frac{1}{N}+\frac{1}{N}\\
&=\frac{2}{N}\\
&\leq\varepsilon\,.
\end{aligned}$$

This completes the proof of (1).

(2) Assume in order to reach a contradiction that the sequence $(x_n)_n=(n)_n$ is convergent in $([0,\infty),d)$ and let $a=\lim_n x_n\in[0,\infty)$. We can thus find N such that, if $n\geq N$, then

$$d(n,a)=d(x_n,a)\leq\frac{1}{2(1+a)}\,,$$

i.e. if $n\geq N$, then

$$\frac{n}{1+n}-\frac{a}{1+a}\leq\frac{1}{2(1+a)}\,.$$

Rearranging this inequality shows that if $n\geq N$, then

$$\frac{n}{1+n}\leq\frac{a}{1+a}+\frac{1}{2(1+a)}=\frac{2a+1}{2a+2}\,,$$

and so

$$n(2a+2) \leq (1+n)(2a+1)\,,$$

whence

$$2an + 2n \leq 2a + 1 + 2an + n\,.$$

This implies that if $n \geq N$, then

$$n \leq 2a+1\,.$$

Hence

$$\infty = \sup_{n \geq N} n \leq 2a + 1 < \infty\,,$$

which gives the desired contradiction. $\square$

Example Let $(x_n)_n$ be a sequence of real numbers and assume that there exists $A > 0$ such that

$$|x_{n+1} - x_n| \leq \frac{A}{n^2} \quad \text{for all } n.$$

Then $(x_n)_n$ is convergent.

Proof. It suffices to show that $(x_n)_n$ is Cauchy. Let $\varepsilon > 0$. Choose $k \in \mathbb{N}$ such that $\frac{1}{N-1} \leq \frac{\varepsilon}{A}$. For $m \geq n \geq N$ we have

$$\begin{aligned}
|x_n - x_m| &\leq |x_n - x_{n+1}| + |x_{n+1} - x_{n+1}| + \cdots + |x_{m-1} - x_m| \\
&\leq A\left(\frac{1}{n^2} + \frac{1}{(n+1)^2} + \cdots + \frac{1}{m^2}\right) \\
&\leq A\int_{n-1}^{m} \frac{1}{x^2}\,dx \\
&\leq A\left(\frac{1}{n-1} - \frac{1}{m}\right) \\
&\leq A\frac{1}{n-1} \\
&\leq \varepsilon\,.
\end{aligned}$$

Example The sequence $(x_n)_n$ where

$$x_n = 1 + \frac{1}{2} + \frac{1}{3}\cdots + \frac{1}{n} - \log n$$

is convergent. The number $\gamma = \lim_n(1 + \frac{1}{2} + \frac{1}{3}\cdots + \frac{1}{n} - \log n)$ is known as Euler's constant. It is not know whether γ is rational or irrational.

Proof. By the previous Example it suffices to show that

$$|x_{n+1} - x_n| \leq \frac{2}{n^2} \quad \text{for all } n.$$

Indeed, this inequality follows easily

$$\begin{aligned}
|x_{n+1} - x_n| &= \left|\frac{1}{n+1} - \log\left(1 + \frac{1}{n}\right)\right| \leq \left|\frac{1}{n+1} - \frac{1}{n}\right| + \left|\frac{1}{n} - \log\left(1 + \frac{1}{n}\right)\right| \\
&= \frac{1}{n(n+1)} + \left|\int_1^{1+\frac{1}{n}} dt - \int_1^{1+\frac{1}{n}} \frac{1}{t}\,dt\right| \leq \frac{1}{n^2} + \left|\int_1^{1+\frac{1}{n}} \frac{t-1}{t}\,dt\right| \\
&= \frac{1}{n^2} + \int_1^{1+\frac{1}{n}} \frac{t-1}{t}\,dt \leq \frac{1}{n^2} + \int_1^{1+\frac{1}{n}} (t-1)\,dt \\
&\leq \frac{1}{n^2} + \int_1^{1+\frac{1}{n}} \frac{1}{n}\,dt = \frac{2}{n^2}\,.
\end{aligned}$$

2.2. Compactness.

We start with the key definition

Definition. Compactness. *A metric space (X, d) is called compact, if every sequence has a subsequence that converges in X. More precisely, a metric space (X, d) is called compact is the folowing is satisfied: if $(x_n)_n$ is a sequence in X, then there exist a subsequeence $(x_{n_k})_k$ and an element $x \in X$ such that*

$$x_{n_k} \to x\,.$$

We will now consider variouis examples of compact and non-compact spaces. However, we first recall the Bolzano-Weierstrass Theorem known from earlier course in analysis.

Bolzano-Weierstrass Theorem. *If $(x_n)_n$ is a bounded sequence in $\mathbb{R}$, then there is a subsequence $(x_{n_k})_k$ and a real number x such that*

$$x_{n_k} \to x\,.$$

The fisrt example shows that the Bolzona-Weierstrass is simply saying the any closed and bounded interval is compact.

Example. Closed and bounded intervals are compact. Let $a, b \in \mathbb{R}$. Then the metric space $([a, b], |\cdot|)$ is compact
Proof. Let $(x_n)_n$ be a sequence in $[a, b]$. Since $(x_n)_n$ is boundfed, it follows from the Bolzan o-Weierstrass theorem that there is a subseeqnce $x_{n_k})_k$ and a real number x such that

$$x_{n_k} \to x\,.$$

However, since $x_{n_k} \in [a, b]$ for all k, we conclude that $x = \lim x_{n_k} \in [a, b]$. hence there is $x \in [a, b]$ such that

$$x_{n_k} \to x\,.$$

This completes the proof. $\square$

We will now give an example of a non-compact space.

Example. The space $(C([0,1]), d_\infty)$ is not compact.
Proof. We must find a sequence $(f_n)_n$ in $C([0,1])$ such that all subsequences of $(f_n)_n$ are divergent in $(C([0,1]), d_\infty)$.

We define $f_n \in C(([0,1])$ by

We now claim that all subsequences of $(f_n)_n$ are divergent in $(C([0,1]), d_\infty)$. Indeed, otherwise there is a subsequence $(f_{n_k})_k$ and an $f \in C([0,1])$ such that

$$f_{n_k} \to f \quad \text{w.r.t. } d_\infty.$$

In particular, there is a K such that

$$d_\infty(f_{n_k}, f) \leq \frac{1}{4} \quad \text{for all } k \geq K. \tag{1}$$

Also note that

$$d_\infty(f_n, f_m) = 1 \quad \text{for all } n \text{ and } m \text{ with } n \neq m. \tag{2}$$

It follows from (1) and (2) that

$$\begin{aligned} 1 &= d_\infty(f_{n_K}, f_{n_{K+1}}) \\ &\leq d_\infty(f_{n_K}, f) + d_\infty(f, f_{n_{K+1}}) \\ &\leq \frac{1}{4} + \frac{1}{4} \\ &= \frac{1}{2}, \end{aligned}$$

giving the desired contradiction. This completes the proof. $\square$

Definition. Bounded. *Let (X, d) be a metric space. A subset M of X is called bounded if and only if*

$$\sup_{x,y\in M} d(x, y) < \infty \,.$$

Theorem. *Let (X, d) be a metric space and let C be a subset of X. If (C, d) is compact, then C is closed and bounded.*

Proof. We first prove that C is closed. Therefore let $(x_n)_n$ be a sequence in C and let $x \in X$, and assume that

$$x_n \to x \quad \text{in } X.$$

We must now show that $x \in C$. Since $(x_n)_n$ is a squence in C, the compactness of C shows that there is subsequence $(x_{n_k})_k$ a point $y \in C$ such that

$$x_{n_k} \to y \quad \text{in } C.$$

In particular, this implies that

$$x_{n_k} \to y \quad \text{in } X.$$

Since both $x_{n_k} \to x$ and $x_{n_k} \to y$ in X, we conclude that $x = y \in C$. This shows that C is closed.

Next, we prove that C is bounded. Assume in order to reach a contradiction that C is not bounded, i.e.

$$\sup_{x,y\in C} d(x, y) = \infty \,.$$

Next fix $n \in \mathbb{N}$. Since

$$n < \infty = \sup_{x,y\in C} d(x, y) \,,$$

we can find $x_n, y_n \in C$ such that

$$n < d(x_n, y_n) \,.$$

Now consider the sequence $(x_n)_n$. Since $(x_n)_n$ is a squence in C, the compactness of C shows that there is subsequence $(x_{n_k})_k$ a point $x_0 \in C$ such that

$$x_{n_k} \to x_0 \quad \text{in } C.$$

Next consider the sequence $(y_{n_k})_k$. Since $(y_{n_k})_k$ is a squence in C, the compactness of C shows that there is subsequence $(y_{n_{k_l}})_l$ a point $y_0 \in C$ such that

$$y_{n_{k_l}} \to y_0 \quad \text{in } C.$$

For brevity write $x_{n_{k_l}} = u_l$ and $y_{n_{k_l}} = v_l$. Then we have

$$u_l \to x_0 \,, \quad v_l \to y_0 \,,$$

where $x_0, y_0 \in C$. In paricular, there is an L such that for $l \geq L$,

$$d(u_l, x_0) \leq 1 \,, \quad d(y_0, v_l) \leq 1 \,.$$

Hence, for $l \geq L$, we have

$$d(u_l, v_l) \leq d(u_l, x_0) + d(x_0, y_0) + d(y_0, v_l) \leq 2 + d(x_0, y_0)\,,$$

and so

$$sup_{l\geq L} d(u_l, v_l) \leq 2 + d(x_0, y_0)\,.$$

But this contradicts the fact that

$$\sup_{l\geq L} d(u_l, v_l) = \sup_{l\geq L} d(x_{n_{k_l}}, y_{n_{k_l}}) \geq \sup_{l\geq L} n_{k_l} = \infty\,.$$

This completes the proof. □

Theorem. *Let C be a subset of $\mathbb{R}$. Then the following two statements are equivalent:*

(1) *The metric space $(C, |\cdot|)$ is compact.*
(2) *The set C is a closed and bounded subset of $\mathbb{R}$.*

Proof.
(1) $\Rightarrow$ (2). This follows from the previous Theorem.

(2) $\Rightarrow$ (1). In order to prove this result, recall the following result known from earlier courses in analysis.

Bolzano-Weierstrass Theorem. If $(x_n)_n$ is a bounded sequence in $\mathbb{R}$, then there is an $x \in \mathbb{R}$ and a subsequence $(x_{n_k})_k$ such that $x_{n_k} \to x$.

Let C be a closed an bounded subset of $\mathbb{R}$. We will now prove that C is compact. Therefore let $(x_n)_n$ be a sequence in C. Since C is bounded, we conclude that the sequence $(x_n)_n$ is bounded, and it therefore follows from Bolzano-Weierstrass Theorem's that there is an $x \in \mathbb{R}$ and a subsequence $(x_{n_k})_k$ such that

$$x_{n_k} \to x\,.$$

Since $x_{n_k} \in C$ and C is closed, we deduce from the sequence characterization of closedness that $x = \lim_l x_{n_l} \in C$. □

Theorem. *Let (X, d) be a compact metric space and let M be a subset of X.*

(1) *The metric space (M, d) is compact.*
(2) *The set M is a closed and bounded subset of X.*

Proof.
(1) $\Rightarrow$ (2). This follows from Theorem ????.

(2) $\Rightarrow$ (1). Let M be a closed an bounded subset of X. We will now prove that M is compact. Therefore let $(x_n)_n$ be a sequence in M. Since $(x_n)_n$ is a sequence in X and X is compact, we conclude that the is an $x \in X$ and a subsequence $(x_{n_k})_k$ such that

$$x_{n_k} \to x \;\text{ in } X.$$

Since $x_{n_k} \in M$ and M is closed, we deduce from the sequence characterization of closedness that $x = \lim_l x_{n_l} \in M$. Hence $x \in M$ and we therefore deduce that

$$x_{n_k} \to x \;\text{ in } M.$$

This completes the proof. □

3. CONTINUOUS FUNCTIONS

3.1. Continuous Functions.

Definition. Continuous function. *Let $(X, d_X)$and (Y, d_Y) be metric spaces and let*

$$f : X \to Y$$

be a function on X.

The function f is called continuous at a point $x_0 \in X$ if

$$\forall \varepsilon > 0 \,:\, \exists \delta > 0 \,:\, \forall x \in X \,:\, d_X(x, x_0) \leq \delta \;\Rightarrow\; d_Y(f(x), f(x_0)) \leq \varepsilon\,.$$

We will say the f is discontinuous at a point $x_0 \in X$ if f is not continuous at x_0.

We will say that f is continuous if f is continuous at all points $x_0 \in X$.

Finally, we will say that f is discontinuous if f is not continuous.

Theorem. Sequence characterization of continuity. *Let $(X, d_X)$and (Y, d_Y) be metric spaces and let*

$$f : X \to \mathbb{Y}$$

be a function on X. Fix $x_0 \in X$. Then the follwing statements are equivalent.

(1) *f is continuous at x_0.*
(2) *If $(x_n)_n$ is any sequence in X such that*

$$x_n \to x_0\,,$$

then

$$f(x_n) \to f(x_0)\,.$$

Proof.
(1)⇒(2) Let $(x_n)_n$ be a sequence in X such that $x_n \to x_0$. We must now prove that $f(x_n) \to f(x_0)$. Therefore let $\varepsilon > 0$. Since f is continuous at x_0 there exists $\delta > 0$ such that:

$$\text{if } x \in X \text{ and } d_X(x, x_0) \leq \delta, \text{ then } d_Y(f(x), f(x_0)) \leq \varepsilon.$$

Also, since $x_n \to x_0$ there exists $N \in \mathbb{N}$ such that:

$$\text{if } n \in \mathbb{N} \text{ and } n \geq N, \text{ then } d_X(x_n, x_0) \leq \delta.$$

Hence if $n \in \mathbb{N}$ and $n \geq N$, we have $d_X(x_n, x_0) \leq \delta$, whence $d_Y(f(x_n), f(x_0)) \leq \varepsilon$.
(2)⇒(1) Assume in order to reach a contradiction that f is discontinuous at x_0, i.e.

$$\text{non}(\forall \varepsilon > 0 \,:\, \exists \delta > 0 \,:\, \forall x \in X \,:\, d_X(x, x_0) \leq \delta \;\Rightarrow\; d_Y(f(x), f(x_0)) \leq \varepsilon)\,.$$

Hence

$$\exists\varepsilon > 0 \,:\, \forall\delta > 0 \,:\, \exists x \in X \,:\, d_X(x, x_0) \leq \delta \text{ and } d_Y(f(x), f(x_0)) > \varepsilon\,. \qquad (1)$$

Let $\varepsilon > 0$ be chosen such that (1) holds, i.e.

$$\forall\delta > 0 \,:\, \exists x \in X \,:\, d_X x, x_0) \leq \delta \text{ and } d_Y(f(x), f(x_0)) > \varepsilon\,.$$

In particular for each $n \in \mathbb{N}$ it follows that there exists $x_n \in X$ such that

$$d_X(x_n, x_0) \leq \frac{1}{n} \text{ and } d_Y(f(x_n), f(x_0)) > \varepsilon\,.$$

Hence $(x_n)_n$ is a sequence in X such that $x_n \to x_0$, but $(f(x_n))_n$ does not converge to $f(x_n)$. This contradicts (2). $\square$

Example. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = 2x^2 + 1\,.$$

Then f is continuous.
Proof using the definition of continuity. Let $x_0 \in \mathbb{R}$. Fix $\varepsilon > 0$. Next choose $\delta > 0$ such that $\delta = \min(1, \frac{2\varepsilon}{2(2|x_0|+1)})$. For $x \in \mathbb{R}$ with $|x - x_0| \leq \delta$ we have

$$\begin{aligned} |f(x) - f(x_0)| &= |2x^2 - 2x_0^2| = 2|x - x_0||x + x_0| \\ &\leq 2|x - x_0|(|x - x_0| + 2|x_0|) \leq 2\delta(\delta + 2|x_0|) \leq 2\delta(1 + 2|x_0|) \leq \varepsilon\,. \end{aligned}$$

Proof using the sequence characterization of continuity. Let $x_0 \in \mathbb{R}$. Fix a sequence $(x_n)_n$ such that $x_n \to x_0$. It now follows from Theorem ????? in Section 2.2 that

$$f(x_n) = 2x_n + 1 \to 2x_0 + 1\,.$$

This shows that f is continuous. $\square$

Example. Let

$$B([0,1]) = \left\{ f : [0,1] \to \mathbb{R} \,\middle|\, f \text{ is bounded} \right\},$$

and equip $B([0,1])$ with the uniform metric $d_\infty : B([0,1]) \times B([0,1]) \to \mathbb{R}$ defined by

$$d_\infty(f, g) = \sup_{x \in X} |f(x) - g(x)|\,.$$

Define $F : B([0,1]) \to \mathbb{R}$ by

$$F(f) = f(0)\,.$$

Then $F : (B([0,1]), d_\infty) \to (\mathbb{R}, |\cdot|)$ is continuous.

Proof. Let $f_0 \in B([0,1])$ and let $\varepsilon > 0$. Choose $\delta = \varepsilon$. For $f \in B([0,1])$ with $d_\infty(f, f_0) \leq \varepsilon$, we have

$$|F(f_n) - F(f)| = |f_n(0) - f(0)| \leq \sup_{x \in [0,1]} |f_n(x) - f(x)| = d_\infty(f_n, f) \leq \delta = \varepsilon\,.$$

This completes the proof $\square$

Example. Let $\varphi : [0,1] \to [0,1]$ be a continuous function. Define $F : C([0,1]) \to C([0,1])$ by

$$F(f) = f \circ \varphi\,.$$

Then $F : (C([0,1]), d_\infty) \to (C([0,1]), d_\infty)$ is continuous.
Proof. Let $f_0 \in C([0,1])$ and let $\varepsilon > 0$. Choose $\delta = \varepsilon$. For $f \in C([0,1])$ with $d_\infty(f_0, f) \le \delta$, we have

$$\begin{aligned}
d_\infty(F(f_0), F(f)) &= d_\infty(f_0 \circ \varphi, f \circ \varphi)\\
&= \sup_{x\in[0,1]} |f_0(\varphi(x)) - f(\varphi(x))|\\
&\le \sup_{y\in[0,1]} |f_0(y) - f(y)|\\
&= d_\infty(f_0, f)\\
&\le \delta\\
&= \varepsilon\,.
\end{aligned}$$

This completes the example. $\square$

Example. Define $F : C([0,1]) \to C([0,1])$ by

$$F(f) = f^2\,.$$

Then $F : (C([0,1]), d_\infty) \to (C([0,1]), d_\infty)$ is continuous.
Proof. Let $f_0 \in C([0,1])$ and let $\varepsilon > 0$. Choose $\delta \le \min(1, \frac{\varepsilon}{1+2d_\infty(0,f_0)})$. For $f \in C([0,1])$ with $d_\infty(f_0, f) \le \delta$, we have

$$\begin{aligned}
d_\infty(F(f_0), F(f)) &= d_\infty(f_0^2, f^2)\\
&= \sup_{x\in[0,1]} |f_0^2(x) - f^2(x)|\\
&= \sup_{x\in[0,1]} |f_0(x) - f(x)|\,|f_0(x) + f(x)|\\
&= \sup_{x\in[0,1]} |f_0(x) - f(x)|\,|f_0(x) - f(x) + 2f(x)|\\
&\le \sup_{x\in[0,1]} |f_0(x) - f(x)|\,\big(\,|f_0(x) - f(x)| + 2|f(x)|\,\big)\\
&= \sup_{x\in[0,1]} |f_0(x) - f(x)|\,\big(\,|f_0(x) - f(x)| + 2|f(x) - 0|\,\big)\\
&\le \sup_{x\in[0,1]} d_\infty(f_0, f)\,\big(\,d_\infty(f_0, f) + 2d_\infty(f, 0)\,\big)\\
&= d_\infty(f_0, f)\,\big(\,d_\infty(f_0, f) + 2d_\infty(f, 0)\,\big)\\
&\le d_\infty(f_0, f)\,\big(\,1 + 2d_\infty(f, 0)\,\big)\\
&\le \frac{\varepsilon}{1 + 2d_\infty(0, f_0)}\,\big(\,1 + 2d_\infty(f, 0)\,\big)\\
&\le \varepsilon\,.
\end{aligned}$$

Example. Write

$$C^1 = \left\{ f : [0,1] \to \mathbb{R} \,\middle|\, f \text{ is differentiable and } f' \text{ is continuous} \right\}.$$

Define $D : C^1([0,1]) \to C([0,1])$ by

$$D(f) = f'.$$

Then $F : (C^1([0,1]), d_\infty) \to (C([0,1]), d_\infty)$ is discontinuous at all points $f \in C^1([0,1])$.

Proof. Let $f \in C^1([0,1])$. We must now show that D is discontinuous at f. We will prove this using the sequence charaterization of continuity. We thus have to find a sequence $(f_n)_n$ in $C^1([0,1])$ such that

$$f_n \to f \quad \text{w.r.t. } d_\infty,$$

but

$$D(f_n) \not\to D(f) \quad \text{w.r.t. } d_\infty.$$

First define $h_n : [0,1] \to \mathbb{R}$ by

$$h_n(x) = \frac{1}{n}\sin(nx),$$

and finally define $f_n \in C^1([0,1])$ by

$$f_n = f + h_n.$$

We first show that

$$f_n \to f \quad \text{w.r.t. } d_\infty.$$

Indeed, we clearly have

$$d_\infty(f, f_n) = \sup_{x\in[0,1]} |f(x) - f_n(x)| = \sup_{x\in[0,1]} |h_n(x)| \le \frac{1}{n} \to 0.$$

Next, we show that

$$D(f_n) \not\to D(f) \quad \text{w.r.t. } d_\infty.$$

Indeed, for all positive integers n we have

$$\begin{aligned} d_\infty(D(f), D(f_n)) &= \sup_{x\in[0,1]} |f'(x) - f_n'(x)| \\ &= \sup_{x\in[0,1]} |f'(x) - f'(x) - h_n'(x)| \\ &\le \sup_{x\in[0,1]} |h_n'(x)| \\ &= \sup_{x\in[0,1]} |\cos(nx)| \\ &\ge |\cos(n0)| \\ &= 1. \end{aligned}$$

If follows from this that $D(f_n) \not\to D(f)$ w.r.t. d_∞. $\square$

Example. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 0 & \text{for } x < 0; \\ 1 & \text{for } 0 \le x. \end{cases}$$

Then f is discontinuous at 0 and continuous at all $x \in \mathbb{R} \setminus \{0\}$.

Proof. First we show that f is dicontinuous at 0. Indeed, define the sequence $(t_n)_n$ by $t_n = -\frac{1}{n}$ for $n \in \mathbb{N}$. Then $t_n \to 0$, but the sequence $(f(t_n))_n = (0, 0, 0, 0, 0, 0, \dots)$ does not converge to $f(0) = 1$.

Next we prove that f is continuous at $x_0 \in \mathbb{R} \setminus \{0\}$. Indeed, let $\varepsilon > 0$. Since $x_0 \neq 0$, we can choose $\delta > 0$ such that $\delta < \frac{|x_0|}{2}$. For $x \in \mathbb{R}$ with $|x - x_0| \le \delta$ we have $x > 0$ if $x_0 > 0$ (since $x \ge x_0 - \frac{|x_0|}{2} = \frac{x_0}{2} > 0$), and $x < 0$ if $x_0 < 0$ (since $x \le x_0 + \frac{|x_0|}{2} = \frac{x_0}{2} < 0$). Hence, for $x \in \mathbb{R}$ with $|x - x_0| \le \delta$ we have $f(x) = f(x_0)$, and so $|f(x) - f(x_0)| = 0 \le \varepsilon$. $\square$

Example. A function that is continuous at exactly one point. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x & \text{for } x \in \mathbb{Q}; \\ 0 & \text{for } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

Then f is continuous exactly at the 0.

Proof. We first prove that f is continous at 0. Let $\varepsilon > 0$. Now choose $\delta = \varepsilon$. For $x \in \mathbb{R}$ with $|x - 0| \le \delta$ we have

$$|f(x) - f(0)| = |f(x)| \le |x| \le \delta = \varepsilon\,.$$

Next, we prove that f is discontinuous at all $x \in \mathbb{R} \setminus \{0\}$. Therefore fix $x \in \mathbb{R} \setminus \{0\}$. We divide the proof into two cases.

Case 1: $x \in \mathbb{Q}$. For $n \in \mathbb{N}$, let $x_n = x + \frac{\sqrt{2}}{n}$. The $x_n \to x$. However, since $x_n \in \mathbb{R} \setminus \mathbb{Q}$, we have that the sequence $(f(x_n))_n = (0, 0, 0, 0, \dots)$ does not converge to $f(x) = x \neq 0$.

Case 2: $x \in \mathbb{R} \setminus \mathbb{Q}$. For $n \in \mathbb{N}$, choose $x_n \in \mathbb{Q}$ such that $x_n \to x$. Hence (by construction) $x_n \to x$. However, $f(x_n) = x_n \to x \neq 0 = f(x)$. $\square$

Example. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & \text{for } x \neq 0; \\ 0 & \text{for } x = 0. \end{cases}$$

Then f is continuous at the 0.

Proof. Let $\varepsilon > 0$. Choose $\delta = \varepsilon > 0$. For $x \in \mathbb{R}$ with $|x - 0| \le \delta$ we have

$$\begin{aligned} |f(x) - f(0)| = |f(x)| &= \begin{cases} \left|x \sin\left(\frac{1}{x}\right)\right| & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases} \\ &\le |x| \le \delta = \varepsilon\,. \end{aligned}$$

This completes the proof. $\square$

Theorem. Construction of continuous functions from continuous function. *Let (X,d) be a metric space and let $x_0 \in X$. Let $f,g : X \to \mathbb{R}$ be continous functions and let $\lambda \in \mathbb{R}$. Define*

$$f+g, fg, \lambda f, \min(f,g), \max(f,g), |f| : A \to \mathbb{R}$$

by

$$\begin{aligned}
&(f+g)(x) = f(x)+g(x)\,,\\
&(fg)(x) = f(x)g(x)\,,\\
&(\lambda f)(x) = \lambda f(x)\,,\\
&\min(f,g)(x) = \min(f(x),g(x))\,,\\
&\max(f,g)(x) = \max(f(x),g(x))\,,\\
&|f|(x) = |f(x)|\,.
\end{aligned}$$

Also, if $g(x) \neq 0$ for all $x \in A$, define

$$\frac{f}{g} : A \to \mathbb{R}$$

by

$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}\,.$$

Then the following statements hold:

(1) *$f+g$ is continuous at x_0*
(2) *fg is continuous at x_0*
(3) *λf is continuous at x_0*
(4) *$\min(f,g)$ is continuous at x_0*
(5) *$\max(f,g)$ is continuous at x_0*
(6) *$|f|$ is continuous at x_0*
(7) *If $g(x) \neq 0$ for all $x \in A$, then $\frac{f}{g}$ is continuous at x_0.*

Proof.
(1) To be written.
(5) Observe that

$$\max(f,g) = \frac{1}{2}(f-g) + \frac{1}{2}|f-g|\,.$$

The result now follows from (1), (3) and (6).
(4) Observe that

$$\max(f,g) = \frac{1}{2}(f-g) - \frac{1}{2}|f-g|\,.$$

The result now follows from (1), (3) and (6). □

Theorem. Continuity of composite functions. *Let (X,d_X), (Y,d_Y) and (Z,d_Z) be metric spaces. Let $f : X \to Y$ and $g : Y \to Z$ be functions. If f is continuous at $x_0 \in X$ and g is continuous at $f(x_0) \in Y$, then $g \circ f$ is continuous at $x_0 \in X$.*

Proof. To be written. □

3.2. Properties of continuous functions.

Thereom. The abstract extremum theorem for continuous functions. *Let* (X, d_X) *and* (Y, d_Y) *be metric spaces. Let*

$$f : X \to Y$$

be continuous. Assume that (X, d_X) *is compact. Then*

$$f(X) = \left\{ f(x) \,\middle|\, x \in X \right\}$$

is a compact subset of Y*.*

Proof. Let $(y_n)_n$ be sequence in $f(X)$. We mus now prove that there is a subsequence $(y_{n_k})_k$ and a point $y \in f(X)$ such that $y_{n_k} \to y$. First, since $y_n \in f(X)$ there is an element $x_n \in X$ such that $y_n = f(x_n)$. hence $(x_n)_n$ is a sequence in the compact space X, and it therefore follows that there is a subsequence $(x_{n_k})_k$ and a point $x \in X$ such that $x_{n_k} \to x$. Now put $y = f(x)$. Then clearly $y = f(x) \inf(X)$ and the sequence characterization of continuity f at x, now implies that

$$y_{n_k} = f(x_{n_k}) \to f(x) = y\,.$$

This completes the proof. □

Corollary. The extremum theorem for continuous functions. *Let* (X, d) *be a metric space. Let*

$$f : X \to \mathbb{R}$$

be continuous. Assume that (X, d) *is compact. Then the following statements hold:*

(1) f *is bounded.*
(2) f *attains its maximum, i.e. there exists* $x_0 \in [a, b]$ *such that*

$$f(x) \le f(x_0) \quad \textit{for all } x \in X.$$

(3) f *attains its minimum, i.e. there exists* $y_0 \in [a, b]$ *such that*

$$f(y_0) \le f(x) \quad \textit{for all } x \in X.$$

Proof.
It follows from the previous theorem that $f(X)$ is compact, and we therefore conclude from Theorem ??? that $f(X)$ is closed and bounded. This clearly implies the statements in the Theorem. □

Remark. Part (1) (and hence also Parts (2) and (3)) of the Extremum Theorem for Continuous Functions is in general false if f is defined on a non-compact space as the example below shows.

An example of a continuous and unbounded function defined a bounded and non-closed interval

Define $f : (0,1) \to \mathbb{R}$ by
$$f(x) = \frac{1}{x}\,.$$
Then f is continuous but unbounded.

Part (2) of the Extremum Theorem for Continuous Functions is in general false even if f is a bounded and continuous function defined on a bounded and non-closed interval as the example below shows.

An example of a bounded and continuous function whose range does not have a maximum defined a bounded and non-closed interval
Define $f : (0,1) \to \mathbb{R}$ by
$$f(x) = x\,.$$
Then f is bounded and continuous but f does not attain its maximum since $\{f(x) \mid x \in (0,1)\} = (0,1)$.

3.3. Uniform Continuity.

Let $A \subseteq \mathbb{R}$ and let $f : A \to \mathbb{R}$ be a function. Recall that the definition of continuity tells us that f is continuous if and only if:

$$\forall x_0 \in A \,:\, \forall \varepsilon > 0 \,:\, \exists \delta > 0 \,:\, \forall x \in A \,:\, |x - x_0| \Rightarrow |f(x) - f(x_0)| \le \varepsilon\,. \qquad (*)$$

In $(*)$ the choice of δ depends on $\varepsilon > 0$ as well as on the point $x_0 \in A$.

As an example consider the function $f : A = (0, \infty) \to \mathbb{R}$ defined by

$$f(x) = \frac{1}{x^2} \quad \text{for } x > 0.$$

We will now show that $(*)$ is satisfied. Let $x_0 \in A = (0, \infty)$ and $\varepsilon > 0$. Choose

$$\delta = \min\left(\frac{x_0}{2}, \frac{x_0^3 \varepsilon}{10}\right).$$

For all $x \in A$ with $|x - x_0| \le \delta$ we have:

$$x \ge x_0 - \delta \ge x_0 - \frac{x_0}{2} = \frac{x_0}{2}\,,$$

and:

$$x \le x_0 + \delta \le x_0 + \frac{x_0}{2} = \frac{3x_0}{2}\,.$$

Hence, for all $x \in A$ with $|x - x_0| \le \delta$ we have

$$\begin{aligned} |f(x_0) - f(x)| &= \left|\frac{1}{x_0^2} - \frac{1}{x^2}\right| = \frac{|x_0^2 - x^2|}{x_0^2 x^2} = \frac{|x_0 - x|(x_0 + x)}{x_0^2 x^2} \\ &\le \frac{|x_0 - x|(\frac{3x_0}{2} + x_0)}{x_0^2 (\frac{x_0}{2})^2} = \frac{10}{x_0^3}|x_0 - x| \le \frac{10}{x_0^3}\frac{x_0^3 \varepsilon}{10} = \varepsilon\,. \end{aligned}$$

This proves $(*)$ for f on $A = (0, \infty)$. Note that δ depends on both ε and x_0. Even if ε is fixed, δ gets small when x_0 is small. This shows that *our* choice of δ depends on x_0 as well as ε, though this may of course be because *we* obtained δ via sloppy estimates. In fact, in this case δ *must* depend on x_0 as well as ε; see Example ???? below.

It turns out to be extremely useful to know when the δ appearing in condition $(*)$ can be chosen to depend only on ε and not on the particular point x_0. Functions for which the δ appearing in condition $(*)$ can be chosen to depend only on ε are said to be uniformly continuous on A. Aften this informal discussion we provide the reader with the formal definition of uniform continuity.

Definition. Uniform continuity. *Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is called uniformly continuous on X if and only if*

$$\forall \varepsilon > 0 \,:\, \exists \delta > 0 \,:\, \forall x, y \in M \,:\, d_X(x, y) \le \delta \Rightarrow d_Y(f(x), f(y)) \le \varepsilon\,.$$

Note that if a function $f : X \to Y$ is uniformly continuous on X, then f is continuous on X. (This should be obvious; if it isn't the definition of continuity and the definition of uniform continuity should be very carefully scrutinized.) Note also that uniform continuity is a property concerning a function and a set (on which it is defined).

Example. Define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = 5x + 2 .$$

Then f is uniformly continuous on $\mathbb{R}$.

Proof. Let $\varepsilon > 0$. Choose $\delta = \frac{\varepsilon}{5} > 0$. For $x, y \in \mathbb{R}$ with $|x - y| \leq \delta$, we have

$$|f(x) - f(y)| = 5|x - y| \leq 5\delta = \varepsilon .$$

This completes the proof. □

Example. Define the function $f : (0, \infty) \to \mathbb{R}$ by

$$f(x) = \frac{1}{x^2}$$

Then f is uniformly continuous on $[a, \infty)$ for any $a > 0$.

Proof. Let $\varepsilon > 0$. Choose $\delta = \frac{a^3}{2}\varepsilon > 0$. For $x, y \in [a, \infty)$ with $|x - y| \leq \delta$, we have

$$\begin{aligned} |f(x) - f(y)| = \left| \frac{1}{x^2} - \frac{1}{y^2} \right| &= \frac{|x^2 - y^2|}{x^2 y^2} = \frac{|x - y|(x + y)}{x^2 y^2} = |x - y| \left(\frac{1}{xy^2} + \frac{1}{x^2 y} \right) \\ &\leq |x - y| \left(\frac{1}{a^3} + \frac{1}{a^3} \right) = \frac{2}{a^3}|x - y| \leq \varepsilon . \end{aligned}$$

This completes the proof. □

Example. Define the function $f : (0, \infty) \to \mathbb{R}$ by

$$f(x) = \frac{1}{x^2}$$

Then f is not uniformly continuous on $(0, \infty)$.

Proof. We must show that

$$\text{non}(\forall \varepsilon > 0 \,:\, \exists \delta > 0 \,:\, \forall x, y \in (0, \infty) \,:\, |x - y| \leq \delta \,\Rightarrow\, |f(x) - f(y)| \leq \varepsilon) ,$$

i.e. we must show that

$$\exists \varepsilon > 0 \,:\, \forall \delta > 0 \,:\, \exists x, y \in (0, \infty) \,:\, |x - y| \leq \delta \text{ and } |f(x) - f(y)| \leq \varepsilon . \qquad (1)$$

We claim that $\varepsilon = 1$ satifies (1). Therefore let $\delta > 0$. Put $\eta = \min(\delta, \frac{1}{2}) > 0$. Choose

$$x = \eta > 0 \quad \text{and} \quad y = \eta + \frac{1}{2}\eta > 0 .$$

Then clearly

$$|x-y| = \frac{1}{2}\eta \leq \delta\,,$$

and

$$|f(x)-f(y)| = \left|\frac{1}{\eta^2} - \frac{1}{(\eta+\frac{1}{2}\eta)^2}\right| = \frac{5}{9\eta^2} \geq \frac{5}{9(\frac{1}{2})^2} = \frac{20}{9} > 1\,.$$

This shows that f is not uniformly continuous on $(0,\infty)$. $\square$

Example. Define the function $f : \to\mathbb{R}$ by

$$f(x) = x^2$$

Then f is uniformly continuous on any closed and bounded interval $[a,b]$.

Proof. Let $\varepsilon > 0$. Choose $\delta = \frac{\varepsilon}{2\max(|a|,|b|)+1} > 0$. For $x,y \in \mathbb{R}$ with $|x-y| \leq \delta$, we have

$$|f(x)-f(y)| = |x^2-y^2| = |x-y||x+y| \leq |x-y|(|x|+|y|) \leq |x-y|2\max(|a|,|b|) \leq \varepsilon\,.$$

This completes the proof. $\square$

The previous example shows that the continuous function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is uniformly continuous on any closed and bounded interval $[a,b]$. This is not an accident as the next important Theorem shows.

Theorem. *Let (X,d_X) and (Y,d_Y) be metric spaces. Let*

$$f : X \to Y$$

be a continuous function. If X is compact, then f is uniformly continuous on X.

Proof. Assume in order to reach a contradiction that f is not uniformly continuous on X, i.e.

$$\text{non}(\forall\varepsilon > 0 \,:\, \exists\delta > 0 \,:\, \forall x,y \in [a,b] \,:\, |x-y| \leq \delta \,\Rightarrow\, |f(x)-f(y)| \leq \varepsilon)\,,$$

Hence

$$\exists\varepsilon > 0 \,:\, \forall\delta > 0 \,:\, \exists x,y \in X \,:\, d_X(x,y) \leq \delta \text{ and } d_Y(f(x),f(y)) > \varepsilon\,. \tag{2}$$

Choose ε such that (2) is satisfied. Then for each $n \in \mathbb{N}$ there exist $x_n, y_n \in X$ such that

$$d_X(x_n,y_n) \leq \frac{1}{n}\,,$$

and

$$d_Y(f(x_n),f(y_n)) \geq \varepsilon\,.$$

It follows from the compactness of X that there exist a subsequence $(x_{m_k})_k$ and a point $x_0 \in X$ such that

$$x_{m_k} \to x_0\,.$$

Also, it it easily seen that

$$y_{m_k} \to x_0 \,.$$

(Indeed, let $\eta > 0$. Since $x_{m_k} \to x_0$ there exists $K_1 \in \mathbb{N}$ such that $d_X(x_{m_k}, x_0) \le \frac{\eta}{2}$ for $k \ge K_1$. Also, since $(m_k)_k$ is a strictly increasing sequence of positive integers there exists $K_2 \in \mathbb{N}$ such that $\frac{1}{m_k} \le \frac{\eta}{2}$ for $k \ge K_2$. Let $K = \max(K_1, K_2)$. For $k \ge K$ we have $d_X(y_{m_k}, x_0) \le d_X(y_{m_k}, x_{m_k}) + d_X(x_{m_k}, x_0) \le \frac{1}{m_k} + \frac{\eta}{2} \le \eta$.) Since $x_0 \in X$ and f is continuous we now conclude that

$$f(x_{m_k}) \to f(x_0) \text{ and } f(y_{m_k}) \to f(x_0) \,.$$

We can thus choose $K \in \mathbb{N}$ such that

$$d_Y(f(x_{m_k}), f(y_{m_k})) \le d_Y(f(x_{m_k}), f(x_0)) + d_Y(f(x_0), f(y_{m_k})) < \varepsilon \quad \text{for all } k \ge K.$$

However this contradicts the fact that

$$d_Y(f(x_{m_k}), f(y_{m_k})) \ge \varepsilon$$

for all k. This completes the proof. □

Remark. The previous theorem is in general false if f is defined on a non-compact space as the example below shows.

An example of a continuous but not uniformly continuous function defined on open and bounded interval

Define $f : (0,1) \to \mathbb{R}$ by

$$f(x) = \frac{1}{x^2} \,.$$

Then f is continuous but not uniformly continuous on $(0,1)$. Observe that f is unbounded. The conclusion in the previous theorem is in general false even if f is a bounded and continuous function defined on an open and bounded interval.

An example of a bounded and continuous but not uniformly continuous function defined on open and bounded interval

Define $f : (0,1) \to \mathbb{R}$ by

$$f(x) = \sin\left(\frac{1}{x}\right) \,.$$

Then f is bounded and continuous but not uniformly continuous on $(0,1)$

Proof. It is obvious that f is bounded and continuous. We will now prove that f is not uniformly continuous on $(0,1)$, i.e. we must prove that

$$\text{non}(\forall \varepsilon > 0 \,:\, \exists \delta > 0 \,:\, \forall x, y \in (0,1) \,:\, |x-y| \le \delta \;\Rightarrow\; |f(x) - f(y)| \le \varepsilon) \,,$$

Hence, we must prove that

$$\exists \varepsilon > 0 \,:\, \forall \delta > 0 \,:\, \exists x, y \in (0,1) \,:\, |x-y| \le \delta \text{ and } |f(x) - f(y)| > \varepsilon \,. \tag{3}$$

We claim that $\varepsilon = 1$ satifies (3). Therefore let $\delta > 0$. Pick an integer n such that $n \geq \frac{1}{\sqrt{\delta}}$. Now, choose

$$x = \frac{2}{(4n+1)\pi} \quad \text{and} \quad y = \frac{2}{(4n+3)\pi}\,.$$

Then clearly

$$|x-y| = \frac{2}{\pi}\,\frac{2}{(4n+1)(4n+3)} \leq \frac{4}{\pi}\,\frac{1}{16n^2} \leq \frac{1}{n^2} \leq \delta\,,$$

and

$$|f(x)-f(y)| = |1-(-1)| = 2 > 1\,.$$

This shows that f is not uniformly continuous on $(0,1)$. $\square$

4. APPLICATIONS OF COMPLETE METRIC SPACES

4.1. The Contraction Mapping Theorem.

We begin we a definition.

Definition. Fixed point. *Let X be a set and let $f : X \to X$ be a map. A point $x \in X$ is called a fixed point of f if*

$$f(x) = x\,.$$

In this section, we will give a simple condition on a map $f : X \to X$ defined on a complete metric space X that guarantees the existence of a fixed point x of f. In the next section, we will give some applications of that condition to differential equations. We now present the key definition.

Definition. Contraction. *Let (X, d) be a metric space. A map $f : X \to X$ is called a contraction if there is a constant $c \in [0, 1)$ such that*

$$d(f(x), f(y)) \le c\, d(x, y)$$

for all $x, y \in X$.

Example. Let (X, d) be a metric space and let $f : X \to X$ be a contraction. Then f is continuous.

Proof. Since f is a contraction there is a constant $c \in [0, 1)$ such that

$$d(f(x), f(y)) \le c\, d(x, y)$$

for all $x, y \in X$.

Let $x_0 \in X$ and $\varepsilon > 0$.

Choose $\delta = \varepsilon > 0$.

For $x \in X$ with $d(x, x_0) \le delta$, we have

$$\begin{aligned} d(f(x), f(x_0)) &\le c\, d(x, x_0) \\ &\le c\delta \\ &\le \delta \\ &= \varepsilon\,. \end{aligned}$$

This completes the proof.

Example. Let $a \in [0, 1]$ and define $f : 0, 1] \to [0, 1]$ by

$$f(x) = ax\,;$$

note that since $a \in [0, 1]$, we conclude that $f([0, 1]) \subseteq [0, 1]$.

(1) If $a \in [0, 1)$, then f is a contraction.

(2) If $a = 1$, then f is not a contraction.

Proof. (1) For $x, y \in [0,1]$, we have

$$\begin{aligned}|f(x) - f(y)| &= |ax - ay| \\ &= a|x - y|\end{aligned}$$

where $a \in [0,1)$. This completes the proof.
(2) We must show that there is no constant $c \in [0,1)$ such that

$$|f(x) - f(y)| \leq c|x - y|$$

for all $x, y \in [0,1]$. We do this by contradiction. We therefore assume that there is constant $c \in [0,1)$ such that

$$|f(x) - f(y)| \leq c|x - y|$$

for all $x, y \in [0,1]$. In particular, it follows that

$$\begin{aligned}|f(1) - f(0)| &\leq c|1 - 0| \\ &= c\,. \end{aligned} \tag{1}$$

However, we also have

$$\begin{aligned}|f(1) - f(0)| &= |1 - 0| \\ &= 1\,. \end{aligned} \tag{2}$$

Comparing (1) and (2), we conclude that $1 \leq c$, which contradicts the fact that $c < 1$. This completes the proof.

Example. Define $T : (C([0, \frac{1}{2}]), d_\infty) \to (C([0, \frac{1}{2}]), d_\infty)$ by

$$(Tf)(x) = \int_0^x f(t)\sin(t)\,dt + \exp(x)$$

for $x \in [0, \frac{1}{2}]$. Then T is a contraction.

Proof. (1) We must show that there is a constant $c \in [0,1)$ such that

$$d_\infty(Tf, Tg) \leq c\, d_\infty(f, g)$$

for all $f, g \in C([0, \frac{1}{2}])$. We claim that

$$d_\infty(Tf, Tg) \leq \frac{1}{2}\, d_\infty(f, g)$$

for all $f, g \in C([0, \frac{1}{2}])$. Indeed, for $f, g \in C([0, \frac{1}{2}])$ we have

$$\begin{aligned}
d_\infty(Tf, Tg) &= \sup_{x\in[0,\frac{1}{2}]} |(Tf)(x) - (Tg)(x)| \\
&= \sup_{x\in[0,\frac{1}{2}]} \left| \int_0^x f(t)\sin(t)\,dt + \exp(x) - \int_0^x g(t)\sin(t)\,dt - \exp(x) \right| \\
&= \sup_{x\in[0,\frac{1}{2}]} \left| \int_0^x f(t)\sin(t)\,dt - \int_0^x g(t)\sin(t)\,dt \right| \\
&\le \sup_{x\in[0,\frac{1}{2}]} \left| \int_0^x (f(t)\sin(t) - g(t)\sin(t))\,dt \right| \\
&= \sup_{x\in[0,\frac{1}{2}]} \int_0^x |f(t)\sin(t) - g(t)\sin(t)|\,dt \\
&= \sup_{x\in[0,\frac{1}{2}]} \int_0^x |f(t) - g(t)|\,|\sin(t)|\,dt \\
&\le \sup_{x\in[0,\frac{1}{2}]} \int_0^x d_\infty(f,g)\,dt \\
&= \sup_{x\in[0,\frac{1}{2}]} d_\infty(f,g)\,x \\
&\le \frac{1}{2} d_\infty(f,g)\,.
\end{aligned}$$

This completes the proof.

We will now state and prove the main result in this section.

The Contraction Mapping Theorem. *Let (X,d) be a complete metric space and let $f:X\to X$ be a contraction.*

(1) *The map f has a unique fixed point, i.e. there is a unique point $p\in X$ such that*
$$f(p)=p\,.$$
(2) *If $x\in X$, then*
$$f^n(x)\to p\,.$$

Proof. Since f is a contraction there is a constant $c\in[0,1)$ such that

$$d(f(x),f(y))\leq c\,d(x,y)$$

for all $x,y\in X$. We now prove the following claim.

Claim. If $x\in X$, then the sequence $(f^n(x)_n$ converges and if we write

$$p=\lim_n f^n(x)\,,$$

then
$$f(p)=p\,.$$

Proof of Claim. Since X is complete it suffices to show that the sequence $(f^n(x)_n$ is Cauchy.

Let $\varepsilon>0$.

Choose $N\in\mathbb{N}$ such that if $n\geq N$, then $d(f(x),x)\,c^n\,\frac{1}{1-c}\leq\varepsilon$

For $m\geq n\geq N$, we have

$$d(f^m(x),f^n(x))\leq\sum_{i=1}^{m-n}d(f^{n+i}(x),f^{n+i-1}(x))\,,$$

and

$$\begin{aligned}d(f^{n+i}(x),f^{n+i-1}(x))&\leq c\,d(f^{n+i-1}(x),f^{n+i-2}(x))\\&\leq c^2\,d(f^{n+i-2}(x),f^{n+i-3}(x))\\&\quad\cdots\\&\leq c^{n+i-1}\,d(f(x),x)\end{aligned}$$

Combining the above two inequalities we see that for $m \geq n \geq N$, we have

$$\begin{aligned} d(f^m(x), f^n(x)) &\leq \sum_{i=1}^{m-n} d(f^{n+i}(x), f^{n+i-1}(x)) \\ &\leq \sum_{i=1}^{m-n} c^{n+i-1}\, d(f(x), x) \\ &\leq d(f(x), x) \sum_{i=1}^{m-n} c^{n+i-1} \\ &\leq d(f(x), x)\, c^n \sum_{k=0}^{\infty} c^k \\ &\leq d(f(x), x)\, c^n \, \frac{1}{1-c} \\ &\leq \varepsilon\,. \end{aligned}$$

This shows that the sequence $(f^n(x)_n$ is Cauchy and therefore convergent.
Write

$$p = \lim_n f^n(x)\,.$$

Using the continuity of f at p and the sequence characterisation of continuity applied to the sequence $(x_n)_n$ defined by $x_n = f^n(x)$, we now conclude that

$$f(p) = \lim_n f(x_n) = \lim_n f(f^n(x)) = \lim_n f^{n+1}(x) = p\,.$$

This completes the proof of the Claim

(1) *Existence of a fixed point for f.* We first prove that f has a fixed point. Indeed, this follows immediately from the Claim.

Uniqueness of the fixed point for f. Next first prove that f has at most one fixed point. Indeed, t if p and q are fixed points for f, then

$$d(p,q) = d(f(p), f(q)) \leq c\, d(p,q)\,,$$

and since $c < 1$ we therefore conclude that

$$d(p,q) = 0\,,$$

when $p = q$.

(2) The statement in (2) follows immediately from the Claim. $\square$

4.2. Applications of the Contraction Mapping Theorem: Solutions to Differential Equations.

We will illustrate the contraction mapping theorem by using it to obtain an existence and uniqueness result for a certain (very general) class of differential equations. We first start by defining what a differential equation is and what a solution to a differential solution is.

Definition. Differential equation. *A (1'st order) differential equation is a pair*

$$(F, I)$$

where

(1) *I is an open and non-empty subinterval of $\mathbb{R}$.*
(2) *F is a continuous function from $I \times \mathbb{R}$ to $\mathbb{R}$, i.e. F is a function*

$$F : I \times \mathbb{R} \to \mathbb{R}$$

which is continuous.

Definition. Solution to a differential equation. *Let (F, I) be a differential equation. A solution to (F, I) is function*

$$f : I \to \mathbb{R}$$

such that

(1) *f is differentaible.*
(2) *for all $x \in I$, we have*

$$f'(x) = F(x, f(x)) \,.$$

Using the contraction mapping theorem we will now show that certain differential equation have unique solutions locally. This result is due to Picard and is known as Picard's theorem. Any function satisfying condition $(*)$ in Picard's theorem is called a Lipschitz function after Lipschitz who was the first to realise the condition's importance in the theory of differential equations.

Rudolf Lipschitz (14 May 1832 – 7 October 1903) was a German mathematician who made contributions to mathematical analysis (where he gave his name to the Lipschitz continuity condition) and differential geometry, as well as number theory, algebras with involution and classical mechanics.

Charles Emile Picard (24 July 1856 11 December 1941) was a French mathematician. He made important contributions to complex analysis and in the theory of differential equations,

Picard's Theorem. *Let (F, I) be a differential equation and assume that there is a constant $M > 0$ such that*

$$|F(x, y_1) - F(x, y_2)| \leq M|y_1 - y_y| \tag{$*$}$$

for all $x \in I$ and all $y_1, y_2 \in \mathbb{R}$.

For each $(x_0, y_0) \in I \times \mathbb{R}$ there is an open and non-empty subinterval I_0 of I with $x_0 \in I_0$ such that the following holds:

If F_0 denotes the restriction of F to $I_0 \times \mathbb{R}$, i.e.

$$F_0 : I_0 \times \mathbb{R} \to \mathbb{R}$$

is the map defined by

$$F_0(x, y) = F(x, y)$$

for $(x, y) \in I_0 \times \mathbb{R}$, then the differential equation (F_0, I_0) has a unique solution f with

$$f(x_0) = y_0 \,.$$

in order to prove Picard's theorem we first reformulate the problem in terms of "solution to integral equations".

Lemma. *Let (F, I) be a differential equation and $(x_0, y_0) \in I \times \mathbb{R}$. Fix $f \in C(I)$. Then the following two statements are equivalent.*

(1) *f is a solution to (F, I) with*

$$f(x_0) = y_0 \,.$$

(2) *We have*

$$f(x) = y_0 + \int_{x_0}^{x} F(t, f(t))\, dt$$

for all $x \in I$.

Proof.
The proof of this lemma follows easily from the fundamental theorem of calculus and is therefore omitted. ☐

We can now prove Picard's theorem.

Proof of Picard's theorem.
Since F is bounded (by assumption), there is a constant $K > 0$ such that

$$\forall (x, y) \in I \times \mathbb{R} \ : \ |F(x, y)| \leq K \,.$$

Next, observe that since I is an open interval and $x_0 \in I$ we can choose $\delta > 0$ such that

$$[x_0 - \delta, x_0 + \delta] \subseteq I \,,$$

and

$$M\delta < 1\,.$$

Finally put

$$I_0 = (x_0 - \delta, x_0 + \delta)\,.$$

We must now show that (F_0, I_0) has a unique solution f with

$$f(x_0) = y_0\,.$$

It follows from the previous lemma (applied to the differential equation (F_0, I_0)) that this is equivalent to proving that:

there is $f \in C(I_0)$ such that

$$f(x) = y_0 + \int_{x_0}^{x} F_0(t, f(t))\,dt$$

for all $x \in I_0$.

We will now prove this statement. For brevity write

$$X = [x_0 - \delta, x_0 + \delta]$$

and note that $X \subseteq I$. Next, define

$$T : \big(\, C(X)\,,\, d_\infty\,\big) \to \big(\, C(X)\,,\, d_\infty\,\big)$$

by

$$(Tg)(x) = y_0 + \int_{x_0}^{x} F_0(t, g(t))\,dt$$

for $x \in X$. Next, observe that:

Claim 1. T maps $C(X)$ into $C(X)$.

Proof of Claim 1. We fix $g \in X$ and must now prove that $Tg \in tC(I_0)$. This is not difficult to prove and is therefore left as an exercise for the reader. This completes the proof of Claim 1

Claim 2. The space $\big(\, C(X)\,,\, d_\infty\,\big)$ is a complete.

Proof of Claim 1. Since X is a closed interval we conclude from ????? that $\big(\, C(X)\,,\, d_\infty\,\big)$ is complete. This completes the proof of Claim 2

Claim 2. The map $T : \big(\, C(X)\,,\, d_\infty\,\big) \to \big(\, C(X)\,,\, d_\infty\,\big)$ is a contraction.

Proof of Claim 3. For $g, h \in C(X)$ we have

$$
\begin{aligned}
d_\infty(Tg, Th) &= \sup_{x\in X} |(Tg)(x) - (Th)(x)| \\
&= \sup_{x\in X} \left| \left(y_0 + \int_{x_0}^{x} F_0(t, g(t))\, dt \right) - \left(y_0 + \int_{x_0}^{x} F_0(t, h(t))\, dt \right) \right| \\
&= \sup_{x\in X} \left| \int_{x_0}^{x} F_0(t, g(t))\, dt - \int_{x_0}^{x} F_0(t, h(t))\, dt \right| \\
&= \sup_{x\in X} \left| \int_{x_0}^{x} (F_0(t, g(t)) - F_0(t, h(t)))\, dt \right| \\
&\le \sup_{x\in X} \int_{x_0}^{x} |F_0(t, g(t)) - F_0(t, h(t))|\, dt \\
&\le \sup_{x\in X} \int_{x_0}^{x} |F(t, g(t)) - F(t, h(t))|\, dt \\
&\le \sup_{x\in X} \int_{x_0}^{x} M\, |g(t) - h(t)|\, dt \\
&\le \sup_{x\in X} \int_{x_0}^{x} M\, d_\infty(g, h)\, dt \\
&= M\, |x - x_0|\, d_\infty(g, h) \\
&= M\, \delta\, d_\infty(g, h)\,.
\end{aligned}
$$

Since $M\delta < 1$, it follows from the previous inequality that T is a contraction with respect to d_∞. This completes the proof of Claim 3.

Finally, it follows from Claim 1. Claim 2, Claim 3 and the contraction mapping theorem that there is a unique function $\varphi \in C(X)$ such that

$$
T\varphi = \varphi\,,
$$

i.e. there is a unique function $\varphi \in C(X)$ such that

$$
y_0 + \int_{x_0}^{x} F_0(t, \varphi(t))\, dt = \varphi(x)
$$

for all $x \in X$. In particular, if we let f denote the restricton of φ to I_0, then we conclude that for all $x \in I_0$ we have

$$
y_0 + \int_{x_0}^{x} F_0(t, f(t))\, dt = y_0 + \int_{x_0}^{x} F_0(t, \varphi(t))\, dt = \varphi(x) = f(x)\,.
$$

This proves the theorem. $\square$

4.3. Baire's Category Theorem.

Suppose we think of a subset A metric space as being "small" if the smallest closed set containing A does not have any interior points. The Baire Category Theorem assets that a complete metric space cannot be expressed as the union of countably many "small" sets. Rather than using the terminology "small" we adopt the following classical terminology.

1. Nowhere dense and meagre sets.

We start with the basic definitions.

Definition. Nowhere dense.. *A subset N of a metric space X is called nowhere dense if for all $x \in X$ and all $r > 0$, there is a $y \in X$ and $\rho > 0$ such that*

$$B(y, \rho) \subseteq B(x, r)$$

and

$$B(y, \rho) \subseteq X \setminus N\,.$$

Hence, a subset N of X is called nowhere dense if every ball in X contains a sub-ball that lies completely in the complement of N.

Example. Every finite subset of a metric space is nowehere dense. (Why?)

Example. The set $\mathbb{N}$ is a nowhere dense subset of $\mathbb{R}$. (Why?)

Example. The set $\mathbb{Z}$ is a nowhere dense subset of $\mathbb{R}$. (Why?)

Example. The set

$$N = \left\{ f \in C([0,1]) \,\middle|\, f(0) = 0 \right\}$$

is a nowhere dense subset of $(C([0,1]), d_\infty)$.

Proof. Let $f \in C([0,1])$ and $r > 0$. Let $g \in C([0,1])$ denote the continuous function defined by

$$g(x) = \tfrac{r}{2}$$

for $x \in [0,1]$, and put $\delta = \frac{r}{4}$. We claim that

$$B(g, \delta) \subseteq B(f, r)\,.$$

Indeed, if $h \in B(g, \delta)$, then

$$|h(0) - \tfrac{r}{2}| = |h(0) - g(0)| \leq d_\infty(h, g) < \tfrac{r}{4}\,,$$

whence

$$-\tfrac{r}{4} < h(0) - \tfrac{r}{2} < \tfrac{r}{4}\,,$$

and so

$$\tfrac{r}{4} < h(0) < \tfrac{3r}{4}\,.$$

In particular, we conclude that $h(0) \neq 0$, and so $h \notin N$. □

The next example show that if we change the metric, then the set N is not necessarily nowhere dense.

Example. The set

$$N = \left\{ f \in C([0,1]) \,\middle|\, f(0) = 0 \right\}$$

is not a nowhere dense subset of $(C([0,1]), d_1)$.
Proof. See Problem ??? on Tutorial sheet no. ????. □

The previous examples should convince the reader that nowhere dense sets are "very small".

Definition. Meagre. *A subset M of a metric space (X, d) is called meagre (or of the 1'st category) if there exist countable many nowhere dense subsets $N_1, N_2, \ldots$ of X such that such that*

$$N = \bigcup_n N_n \, .$$

If a set is not meagre (i.e. of the 1'st category), then it is said to be of the 2'nd category.

Hence, if nowhere dense sets should be thought of as being "very small", then meagre sets should be thought of as being "small".

Example. Every nowehere dense set is meagre. (Why?)

Example. Every countable subset of a metric space is nowehere dense. (Why?)

Example. The set $\mathbb{Q}$ is not nowhere dense in $\mathbb{R}$, but the set $\mathbb{Q}$ is a meagre subset of $\mathbb{R}$.

Example. The set $\mathbb{Q}$ is not nowhere dense in $\mathbb{R}$, but the set $\mathbb{Q}$ is a meagre subset of $\mathbb{R}$.

Example. The set

$$M = \left\{ f \in C([0,1]) \,\middle|\, f(0) \in \mathbb{Q} \right\}$$

is a meagre subset of $(C([0,1]), d_\infty)$.
Proof. See Problem ??? on Tutorial sheet no. ????. □

Here are some natural questions:

Question. Is the "big" set $\mathbb{R}$ of real numbers "small" after all?
More precisely: Is $\mathbb{R}$ meagre? I.e. can we find nowhere dense subsets $N_1, N_2, \ldots$ of $\mathbb{R}$ such that

$$\mathbb{R} = \bigcup_n N_n \, ?$$

Question. Is the "big" set $\mathbb{R} \setminus \mathbb{Q}$ of irrational numbers "small" after all?

More precisely: Is $\mathbb{R} \setminus \mathbb{Q}$ meagre? I.e. can we find nowhere dense subsets $N_1, N_2, \dots$ of $\mathbb{R}$ such that

$$\mathbb{R} \setminus \mathbb{Q} = \bigcup_n N_n \, ?$$

I.e. can we find nowhere dense subsets $N_1, N_2, \dots$ of $\mathbb{R}$ such that

$$\mathbb{Q} = \mathbb{R} \setminus \bigcup_n N_n = \bigcap_n \mathbb{R} \setminus N_n \, ?$$

This seems plausible. Simply, let $\mathbb{Q} = \{q_1, q_2, \dots\}$ be an arbitrary enumeration of $\mathbb{Q}$ and put

$$N_n = \mathbb{R} \setminus \bigcup_m \left(q_m - \tfrac{1}{2^{n+m}}, q_m - \tfrac{1}{2^{n+m}} \right).$$

Then N_n is nowhere dense and

$$\mathbb{R} \setminus N_n = \bigcup_m \left(q_m - \tfrac{1}{2^{n+m}}, q_m - \tfrac{1}{2^{n+m}} \right).$$

Is now seems very plausible that

$$\mathbb{Q} = \bigcap_n \mathbb{R} \setminus N_n \, .$$

(Why?)

To answer the above questions (and many more) we need one of the most powerful results in analysis, namely, Baire's Category Theorem.

Rene Baire (21 January 1874 – 5 July 1932) was a French mathematician most famous for his Baire category theorem. His theory was published originally in his dissertation Sur les fonctions de variable reelles (”On the Functions of Real Variables”) in 1899.

Baire's Category Theorem. *Let (X, d) be a complete metric space. Then X is not meagre.*

In fact, we even have the following: If $N_1, N_2, \ldots$ are nowhere dense subsets of X, then the following holds,

(1) *the set*
$$X \setminus \bigcup_n N_n$$
is dense.

(2) *the set*
$$X \setminus \bigcup_n N_n$$
is uncountable.

Proof.
(1) Let $N_1, N_2, \ldots$ be nowehere dense subsets of X. We first prove that the set

$$X \setminus \bigcup_n N_n$$

is dense. Therefore fix $x \in X$ and $r > 0$. We must now find

$$y \in B(x, r) \ \cap \ \left(X \setminus \bigcup_n N_n \right)$$

Write $x = x_0$ and $r = r_0$ We first show that we can find a sequence $B(x_1, r_1), B(x_2, r_2), \ldots$ of balls in X such that

$$0 < r_n < \tfrac{1}{2^n}$$

and

$$B(x_{n+1}, r_{n+1}) \subseteq B(x_n, \tfrac{r_n}{2}) \setminus N_{n+1}\,.$$

Indeed, this follows immediately from the fact that each N_n is nowhere dense.

We now show that the sequence $(x_n)_n$ is Cauchy, Indeed, let $\varepsilon > 0$, and choose N such hat

$$\tfrac{1}{2^{n-1}} < \varepsilon$$

for $n \geq N$. Observe that if $k \geq N$, then $x_k \in B(x_k, r_k) \subseteq B(x_{k-1}, \frac{r_{k-1}}{2})$, whence

$$d(x_k, x_{k-1}) \leq \tfrac{r_{k-1}}{2} \leq \tfrac{1}{2}\ \tfrac{1}{2^{k-1}} = \tfrac{1}{2^k}\,.$$

For $m \geq n \geq N$, we therefore have

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \ldots + d(x_{m-1}, x_m) \\ &\leq \tfrac{1}{2^n} + \tfrac{1}{2^n} + \ldots + \tfrac{1}{2^n} \\ &\leq \tfrac{1}{2^{n-1}} \\ &\leq \varepsilon\,. \end{aligned}$$

This shows that $(x_n)_n$ is Cauchy.

Since X is complete and $(x_n)_n$ is Cauchy we conclude that there is a point y such that $x_n \to y$.

We now claim that

$$y \in B(x, r)\ \cap\ \left(X \setminus \bigcup_n N_n\right).$$

We first show that

$$y \in B(x, r)\,.$$

Indeed, for all $k > 0$, we have

$$\begin{aligned} x_k \in B(x_k, r_k) &\subseteq B(x_{k-1}, \tfrac{r_{k-1}}{2}) \\ &\subseteq B(x_{k-1}, r_{k-1}) \subseteq B(x_{k-2}, \tfrac{r_{k-2}}{2}) \\ &\vdots \\ &\subseteq B(x_2, r_2) \subseteq B(x_1, \tfrac{r_1}{2}) \\ &\subseteq B(x_1, r_1) \subseteq B(x_0, \tfrac{r_0}{2}) \\ &= B(x, r) \subseteq \left\{ z \in X \,\middle|\, d(z, x) \leq \tfrac{r}{2} \right\} = F\,. \end{aligned}$$

Hence, the sequence $(x_k)_k$ is a sequence in F, and since F is closed we therefore conclude that

$$y = \lim_k x_k \in F \subseteq B(x, r)\,.$$

This proves that $y \in B(x, r)$.

Next we prove that for each positive integer n, we have

$$y \in X \setminus N_n\,.$$

Indeed, for all $k > n$ we have

$$\begin{aligned}
x_k \in B(x_k, r_k) &\subseteq B(x_{k-1}, \tfrac{r_{k-1}}{2}) \\
&\subseteq B(x_{k-1}, r_{k-1}) \subseteq B(x_{k-2}, \tfrac{r_{k-2}}{2}) \\
&\vdots \\
&\subseteq B(x_{n+2}, r_{n+2}) \subseteq B(x_{n+1}, \tfrac{r_{n+1}}{2}) \\
&\subseteq B(x_{n+1}, r_{n+1}) \subseteq B(x_n, \tfrac{r_n}{2}) \\
&\subseteq B(x_n, \tfrac{r_n}{2}) \\
&\subseteq \left\{ z \in X \,\middle|\, d(z, x_n) \leq \tfrac{r_n}{2} \right\} = F_n \\
&\subseteq B(x_n, r_n) \subseteq B(x_{n-1}, \tfrac{r_{n-1}}{2}) \setminus N_n \subseteq X \setminus N_n\,.
\end{aligned}$$

Hence, the sequence $(x_k)_{k>n}$ is a sequence in F_n, and since F_n is closed we therefore conclude that

$$y = \lim_{k>n} x_k \in F_n \subseteq X \setminus N_n\,.$$

This proves that $y \in X \setminus N_n$.

(2) Let $N_1, N_2, \ldots$ be nowehere dense subsets of X. Assume, ion order to reach a contradiction, that the set

$$N = X \setminus \bigcup_n N_n$$

is countable. In particular, thre are countable many points $x_1, x_2, \ldots \in X$ such that

$$N = \bigcup_n \{x_n\}\,.$$

Then we have

$$X = N \;\cup\; \bigcup_n N_n = \bigcup_n \{x_n\} \;\cup\; \bigcup_n N_n\,,$$

and all the sets $\{x_1\}, \{x_2\}, \ldots, N_1, N_2, \ldots$ are nowhere dense. This shows that X is meagre contradicting part (1). $\square$

4.4. Applications of Baire's Category Theorem: $\mathbb{Q}$ is not $\mathcal{G}_\delta$

We now present several applications of Baire's Category Theorem.

Application 1. Our first application shows that $\mathbb{R}$ is not the union of countably many nowhere dense sets.

Theorem. *The set $\mathbb{R}$ of real numbers is not meagre, i.e. we cannot find nowhere dense subsets $N_1, N_2, \dots$ of $\mathbb{R}$ such that*

$$\mathbb{R} = \bigcup_n N_n \,.$$

In fact, if $N_1, N_2, \dots$ are nowhere dense subsets of $\mathbb{R}$, then

$$\mathbb{R} \setminus \bigcup_n N_n$$

is both dense and uncountable.

Proof. This statement follows immediately ffrom Baire's Category Theorem since $\mathbb{R}$ is complete. $\square$

Application 2. Our second application shows that the set $\mathbb{R} \setminus \mathbb{Q}$ of irrational numbers cannot be expressed as a intersection of open sets. Recall that if X is a metric space, then $\mathcal{G} = \mathcal{G}(X)$ denotes the family of open subsets of X, i.e.

$$\mathcal{G} = \mathcal{G}(X) = \left\{ G \subseteq X \,\middle|\, G \text{ is open} \right\}.$$

Also, if $\mathcal{A}$ is a family of subsets of X, the we will write $\mathcal{A}_\delta$ for the family of all sets that can be written as a countable intersections of sets from $\mathcal{A}$. i.e.

$$\mathcal{A}_\delta = \left\{ \bigcap_{n \in \mathbb{N}} A_n \,\middle|\, A_n \in \mathcal{A} \text{ for all } n \in \mathbb{N} \right\}.$$

Using this notation, our second application of Baire's Category Theorem says that

$$\mathbb{Q} \notin \mathcal{G}_\delta \,.$$

Theorem. *We have*

$$\mathbb{Q} \notin \mathcal{G}_\delta \,,$$

i.e. the set $\mathbb{Q}$ of rational numbers cannot be expressed as a countable intersection of open sets, i.e. we cannot find open subsets $G_1, G_2, \dots$ of $\mathbb{R}$ such that

$$\mathbb{Q} = \bigcap_n G_n \,.$$

Proof. Indeed, assume that we can find open subsets $G_1, G_2, \dots$ of $\mathbb{R}$ such that

$$\mathbb{Q} = \bigcap_n G_n \,.$$

Also, let $\mathbb{Q} = \{q_1, q_2, \dots\}$. Then

$$\begin{aligned}
\mathbb{R} &= \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q}) \\
&= \bigcup_m \{q_m\} \cup \left(\mathbb{R} \setminus \bigcap_n G_n \right) \\
&= \bigcup_m \{q_m\} \cup \bigcup_n (\mathbb{R} \setminus G_n) \,. \qquad (1)
\end{aligned}$$

Next note that

$$\mathbb{R} \setminus G_n \text{ is nowhere dense for all } n \qquad (2)$$

(since G_n is open and $\mathbb{Q} \subseteq G_n$; you are strongly encouraged to complete the details showing that $\mathbb{R} \setminus G_n$ is nowhere dense), and

$$\{q_m\} \text{ is nowhere dense for all } m. \qquad (3)$$

Combining (1), (2) and (3), we conclude that

$$\mathbb{R} = \bigcup_m \{q_m\} \cup \bigcup_n (\mathbb{R} \setminus G_n)$$

is meagre. However, this contradicts the fact that $\mathbb{R}$ is not meagre by Application no. 1. $\square$

In particular, if we let $\mathbb{Q} = \{q_1, q_2, \dots\}$ and

$$G_n = \bigcup_m \left(q_m - \tfrac{1}{2^{n+m}}, q_m - \tfrac{1}{2^{n+m}} \right)$$

then G_n is open and dense, and clearly

$$\mathbb{Q} \subseteq \bigcap_n G_n \,.$$

Also,

$$\operatorname{length}\left(\left(q_m - \tfrac{1}{2^{n+m}}, q_m - \tfrac{1}{2^{n+m}} \right) \right) = \tfrac{2}{2^{n+m}}$$

whence

$$\begin{aligned}
\operatorname{length}(G_n) &\leq \sum_m \operatorname{length}\left(\left(q_m - \tfrac{1}{2^{n+m}}, q_m - \tfrac{1}{2^{n+m}} \right) \right) \\
&\leq \sum_m \tfrac{2}{2^{n+m}} \\
&= \sum_m \tfrac{2}{2^n} \,.
\end{aligned}$$

Hence each set G_n contains $\mathbb{Q}$ and has a length less than $\frac{2}{2^n}$. It is therefore very plausible that

$$\mathbb{Q} = \bigcap_n G_n \,.$$

However, it follows from Application 2 that (very suprpringsly)

$$\mathbb{Q} \neq \bigcap_n G_n \,.$$

4.5. Applications of Baire's Category Theorem: Typical

We now present several applications of Baire's Category Theorem.

Recall that if X is a metric space, then Baire's cartegory Theorem says that X is n ot meagre. I.e. if M is a meagre subset of X, then

$$X \setminus M$$

is not empty. In fact,

$$X \setminus M$$

is both dense and uncountable, i.e. $X \setminus M$ is is "very big".

Hence if X is a metric space and for each $x \in X$ let $P(x)$ be a statement about x (e.g. if $X = \mathbb{R}$, then $P(x)$ might be the statement "$x \geq 42$"). Assume that we want to mshow that there is an $x \in X$ such that $P(x)$ is a true statement. Then we might proceed as follows: let

$$M = \left\{ x \in X \,\middle|\, P(x) \text{ is not a true statement} \right\}.$$

If we could show that

"M is meagre"

the it would follow from the above that

$$X \setminus M = \left\{ x \in X \,\middle|\, P(x) \text{ is a true statement} \right\}.$$

is not empty. In fact,

$$X \setminus M = \left\{ x \in X \,\middle|\, P(x) \text{ is a true statement} \right\}.$$

is both dense and uncountable, i.e. $X \setminus M$ is is "very big". Hence, not only is there an element x is X such that $P(x)$ is true, but, in fact, the set of x for which $P(x)$ is true is "huge".

This motivates the next definition.

Definition. Typical. *Let X be a complete metric space and for each $x \in X$ let $P(x)$ be a statement about x. We will say that a typical element has property P (or that property P is typical) if the set*

$$M = \left\{ x \in X \,\middle|\, P(x) \textit{ is not a true statement} \right\}.$$

is meagre.

Example. A typical real number is different from $\sqrt{2}$. (Why?)

Example. A typical real number is not rational. (Why?)

Example. Consider the metric space $(C([0,1]), d_\infty)$. A typical continuous function $f \in C([0,1])$ has the property that $f(0) \neq 0$. (Why?)

Example. Consider the metric space $(C([0,1]), d_\infty)$. A typical continuous function $f \in C([0,1])$ has the property that $f(0) \notin \mathbb{Q}$. (Why?)

Example. Consider the metric space $(C([0,1]), d_\infty)$. A typical continuous function $f \in C([0,1])$ is not differentiable at $\frac{1}{2}$. (Why?)

4.6. Weierstrass Approximation Theorem

In this section we prove the following theorem due to Weierstrass.

Weierstrass' approximation theorem. *Let $a, b \in \mathbb{R}$ with $a < b$. Write*

$$P([a,b]) = \left\{ p : [a,b] \to \mathbb{R} \,\middle|\, p \text{ is a polynomial} \right\}.$$

Then $P([a,b])$ is dense in $\big(C([a,b])\,,\, d_\infty\big)$, i.e. for each $f \in C([0,1])$ and each $r > 0$, there is a polynomial $p \in P([0,1])$ such that

$$d_\infty(f,p) \le r\,,$$

or, alternatively, $f \in C([0,1])$ there is a sequence $(p_n)_n$ with $p_n \in P([0,1])$ for all n such that

$$d_\infty(f,p_n) \to 0\,.$$

Before proving Weiestrass Approximation Theorem, we state and prove a more fundamental result that will also have applications later. Prior to doing so, we need the following definitions.

Definition. *Let $a > 0$ and $Q : [-a,a] \to \mathbb{R}$ be a Riemann integrable function. For $\delta > 0$, we will write*

$$\int_{\{\delta \le |t|\}} Q(t)\,dt = \int_{-a}^{-\delta} Q(t)\,dt + \int_{\delta}^{a} Q(t)\,dt\,.$$

Definition. Approximate identity. *Let $a > 0$. A sequence $(Q_n)_n$ of functions $Q_n : [-a,a] \to \mathbb{R}$ is called an approximate identity on $[-a,a]$ (or a Dirac sequence on $[-a,a]$) if*

1. *The function Q_n is Riemann integrable on $[-a,a]$ for all n.*
2. *$Q_n \ge 0$ for all n.*
3. *$\int_{-a}^{a} Q(t)\,dt = 1$ for all n.*
4. *For all $\delta > 0$, we have*

$$\int_{\{\delta \le |t|\}} Q_n(t)\,dt \to 0\,.$$

Approximate identities play very important roles in analysis. An example of an approximate identity $(Q_n)_n$ is given by

$$Q_n(t) = \begin{cases} \frac{n}{2} & \text{for } t \in [-\frac{1}{n}, \frac{1}{n}]\ : \\ 0 & \text{for } t \in \mathbb{R} \setminus [-\frac{1}{n}, \frac{1}{n}]\,. \end{cases}$$

A further example of an approximate identity will be encountered in the proof of Weierstrtass approximation theorem, and still others when be introduced when we study Fourier series. The fact that the integral over the set $\{t \in \mathbb{R} \mid |t| \geq \delta\}$ become small as $n \to \infty$ seems to suggest the in some sense the functions themselves become small as n becomes large. On the other hand, since the integral over $[-a, a]$ are always equal to 1, we have

$$\lim_n \int_{-\delta}^{\delta} Q_n(t)\, dt = 0$$

for all $\delta > 0$. This seems to indicate that the functions are concentrated near 0 and must become very large near 0; this statement will be made precise in Question ??? on Tutorial Sheet no. ???. The graphs of the first few function of a typical approximate identity is show in the figure below.

We can now state and prove our main result on approximate identities.

Theorem. *Let $a > 0$ and let $(Q_n)_n$ be an approximate identity on $[-a, a]$. Let $f : \mathbb{R} \to \mathbb{R}$ be a function with period equal to $2a$ and assume that f is Riemann integrable on $[-a, a]$. Then the following statements hold.*

(1) *For each $x \in \mathbb{R}$ and $n \in \mathbb{R}$, the function*

$$t \to f(x+t)Q_n(t) \quad \textit{where} \quad t \in [-a, a]\,,$$

is Riermann integrable. Hence we can define $S_n : [-a, a] \to \mathbb{R}$ by

$$S_n(x) = \int_{-a}^{a} f(x+t)Q_n(t)\,dt$$

(2) *If $x \in \mathbb{R}$ and f is continuous at x, then*

$$S_n(x) \to f(x)\,.$$

(3) *If f is continuous on $[-a, a]$, then*

$$S_n \to f \quad \textit{with respect to } d_\infty \textit{ on } [-a, a].$$

Proof.

First note that f is bounded, i.e. there is a constant M such that $|f(t)| \leq M$ for all $t \in \mathbb{R}$ (you are inviter to think about why this statement is true).

(1) This statement is clear.

(2) Now let $x \in \mathbb{R}$ and fix $\varepsilon > 0$.

Since f is continuous at x, there is a $\delta > 0$ such that

$$|t| \leq \delta \quad \Rightarrow \quad |f(x+t) - f(x)| \leq \frac{\varepsilon}{2}\,. \tag{1}$$

Also, since $(Q_n)_n$ is an approximate identity we conclude that

$$\int_{\{\delta \leq |t|\}} Q_n(t)\,dt \to 0\,,$$

and we therefore find a positive integer N such that

$$n \geq N \quad \Rightarrow \quad \int_{\{\delta \leq |t|\}} Q_n(t)\,dt \ \leq \frac{\varepsilon}{4M}\,. \tag{2}$$

For $n \geq N$, we now have

$$\begin{aligned}
|S_n(x) - f(x)| &= \left| \int_{-a}^{a} f(x+t) Q_n(t)\, dt - f(x) \right| \\
&= \left| \int_{-a}^{a} f(x+t) Q_n(t)\, dt - f(x) \int_{-a}^{a} Q_n(t)\, dt \right| \\
&= \left| \int_{-a}^{a} (f(x+t) - f(x)) Q_n(t)\, dt \right| \\
&= \int_{-a}^{a} |f(x+t) - f(x)| Q_n(t)\, dt \\
&= \int_{-\delta}^{\delta} |f(x+t) - f(x)| Q_n(t)\, dt \\
&\qquad + \int_{\{\delta \leq |t|\}} |f(x+t) - f(x)| Q_n(t)\, dt \\
&\leq \int_{-\delta}^{\delta} |f(x+t) - f(x)| Q_n(t)\, dt \\
&\qquad + \int_{\{\delta \leq |t|\}} \big(|f(x+t)| + |f(x)| \big) Q_n(t)\, dt \\
&\leq \int_{-\delta}^{\delta} \frac{\varepsilon}{2} Q_n(t)\, dt \\
&\qquad + \int_{\{\delta \leq |t|\}} 2M Q_n(t)\, dt \\
&\leq \frac{\varepsilon}{2} \int_{-a}^{a} Q_n(t)\, dt \\
&\qquad + 2M \int_{\{\delta \leq |t|\}} Q_n(t)\, dt \\
&\leq \frac{\varepsilon}{2} + 2M \frac{\varepsilon}{4M} \\
&= \varepsilon\,.
\end{aligned}$$

(3) First note that since f is continuous on $[-a, a]$, then f is uniformly continuous on $[-a, a]$, and the periodicity of f therefore implies that f is uniformly continuous on $\mathbb{R}$ (you are inviter to think about why this statement is true).

Now fix $\varepsilon > 0$.

Since f is uniformly continuous on $\mathbb{R}$, there is a $\delta > 0$ such that for all $x, t \in \mathbb{R}$, we have

$$|t| \leq \delta \quad \Rightarrow \quad |f(x+t) - f(x)| \leq \frac{\varepsilon}{2}\,. \tag{1}$$

Also, since $(Q_n)_n$ is an approximate identity we conclude that

$$\int_{\{\delta \leq |t|\}} Q_n(t)\, dt \to 0\,,$$

and we therefore find a positive integer N such that

$$n \geq N \quad \Rightarrow \quad \int_{\{\delta \leq |t|\}} Q_n(t)\, dt \;\leq \frac{\varepsilon}{4M}\,. \tag{2}$$

For $n \geq N$, we now have

$$
\begin{aligned}
\sup_{x\in[-a,a]} |S_n(x) - f(x)| &= \sup_{x\in[-a,a]} \left| \int_{-a}^{a} f(x+t)Q_n(t)\,dt - f(x) \right| \\
&= \sup_{x\in[-a,a]} \left| \int_{-a}^{a} f(x+t)Q_n(t)\,dt - f(x) \int_{-a}^{a} Q_n(t)\,dt \right| \\
&= \sup_{x\in[-a,a]} \left| \int_{-a}^{a} (f(x+t) - f(x))Q_n(t)\,dt \right| \\
&= \sup_{x\in[-a,a]} \int_{-a}^{a} |f(x+t) - f(x)| Q_n(t)\,dt \\
&= \sup_{x\in[-a,a]} \Bigg(\int_{-\delta}^{\delta} |f(x+t) - f(x)| Q_n(t)\,dt \\
&\qquad\qquad + \int_{\{\delta \leq |t|\}} |f(x+t) - f(x)| Q_n(t)\,dt \Bigg) \\
&\leq \sup_{x\in[-a,a]} \Bigg(\int_{-\delta}^{\delta} |f(x+t) - f(x)| Q_n(t)\,dt \\
&\qquad\qquad + \int_{\{\delta \leq |t|\}} \big(|f(x+t)| + |f(x)| \big) Q_n(t)\,dt \Bigg) \\
&\leq \int_{-\delta}^{\delta} \frac{\varepsilon}{2} Q_n(t)\,dt \\
&\qquad + \int_{\{\delta \leq |t|\}} 2M Q_n(t)\,dt \\
&\leq \frac{\varepsilon}{2} \int_{-a}^{a} Q_n(t)\,dt \\
&\qquad + 2M \int_{\{\delta \leq |t|\}} Q_n(t)\,dt \\
&\leq \frac{\varepsilon}{2} + 2M \frac{\varepsilon}{4M} \\
&= \varepsilon\,.
\end{aligned}
$$

This completes the proof. $\square$

We can now prove Weierstrass approximation theorem.

Proof of Weierstrass approximation theorem.
We divide the proof into three cases.

Case 1: $a = 0$, $b = 1$ *and* $f(0) = f(1) = 0$.
First we extend f to $[-1, 1]$, by putting

$$f(x) = 0 \quad \text{for } x \in [-1, 0).$$

as follows.
Next, we extend f to $\mathbb{R}$ as follows. For $x \in \mathbb{R}$, let k_x be the unique integer such that $x - 2k_x \in [-1, 1]$, and put

$$f(x) = f(x - k_x) \quad \text{for } x\mathbb{R}.$$

The f is continuous on $\mathbb{R}$ and f has period equal to 2.
For a positive integer n, let

$$c_n = \frac{1}{\int_{-1}^{1} (1-t^2)^n \, dt} \, ,$$

and define $Q_n : [-1, 1] \to \mathbb{R}$ by

$$Q_n(t) = c_n (1 - t^2)^n \, .$$

We willow prove that $(Q_n)_n$ is an approximate identity on $[-1, 1]$.

Claim 1: Q_n *is Riemann integrable on* $[-1, 1]$.
Proof of Claim 1. This is clear since Q_n is continuous. This proves Claim 1.

Claim 2: $Q_n \geq 0$.
Proof of Claim 2. This is clear. This proves Claim 2.

Claim 3: $\int_{-1}^{-1} Q_n(t) \, dt = 1$.
Proof of Claim 3. This follows from the definition of c_n. This proves Claim 3.

Claim 4: For all $\delta > 0$, *we have* $\int_{\{\delta \leq |t|\}} Q_n(t) \, dt \to 0$ . .
Proof of Claim 4. We have

$$\begin{aligned}
\frac{1}{c_n} &= \int_{-1}^{1} (1-t^2)^n \, dt \\
&= 2 \int_{0}^{1} (1-t^2)^n \, dt \\
&\geq 2 \int_{0}^{\frac{1}{\sqrt{n}}} (1-t^2)^n \, dt \\
&\geq 2 \int_{0}^{\frac{1}{\sqrt{n}}} (1-nt^2) \, dt \\
&\qquad \text{[since } (1-x)^n \geq 1 - nx \text{ for all } x \geq 0 \text{ and all } n \in \mathbb{N}\text{]} \\
&= 2 \left(\frac{1}{\sqrt{n}} - \frac{1}{3\sqrt{n}} \right) \\
&= \frac{4}{3\sqrt{n}} \\
&\geq \frac{1}{\sqrt{n}} \, ,
\end{aligned}$$

whence

$$c_n \leq \sqrt{n}\,.$$

Hence

$$\begin{aligned}\int_{\{\delta\leq|t|\}} Q_n(t)\,dt &\leq c_n \int_{\{\delta\leq|t|\}} (1-t^2)^n\,dt\\ &\leq \sqrt{n}\int_{\{\delta\leq|t|\}} (1-\delta^2)^n\,dt\\ &\leq \sqrt{n}\int_{-1}^{1} (1-\delta^2)^n\,dt\\ &\leq 2\sqrt{n}(1-\delta^2)^n\\ &\to 0\,.\end{aligned}$$

This proves Claim 4.

Combining Claim 1–Claim 4, we deduce that $(Q_n)_n$ is an approximate identity on $[-1,1]$.

Next, since $(Q_n)_n$ is an approximate identity on $[-1,1]$ and since f is continuous on $\mathbb{R}$ and f has period equal to 2, we conclude from the previous theorem that if we define $S_n : [-1,1] \to \mathbb{R}$ by

$$S_n(x) = \int_{-1}^{1} f(x+t)Q_n(t)\,dt\,,$$

then

$$S_n \to f \quad \text{with respect to } d_\infty \text{ on } [-1,1]. \tag{1}$$

Finally define $p_n : [0,1] \to \mathbb{R}$ by

$$p_n(x) = S_n(x)\,.$$

We now have:

Claim 5: $p_n \to f$ with respect to d_∞ on $[0,1]$.
Proof of Claim 5. This follows from (1). This proves Claim 5.

Claim 6: p_n is a polynomial.
Proof of Claim 6. Since $f = 0$ on $[-1,0]\cup[1,2]$, we deduce that for all $x \in [0,1]$ we have

$$\begin{aligned}p_n(x) &= S_n(x)\\ &= \int_{-1}^{1} f(x+t)Q_n(t)\,dt\\ &= \int_{-x}^{1-x} f(x+t)Q_n(t)\,dt\\ &= \int_{0}^{1} f(s)Q_n(s-x)\,ds\\ &= c_n\int_{0}^{1} f(s)\,(1-(s-x)^2)^n\,ds\,,\end{aligned}$$

which proves Claim 6 since $c_n \int_0^1 f(s)\,(1-(s-x)^2)^n\,ds$ is clearly a polynomial in x. This proves Claim 6.

The desired result now follows from Claim 5 and Claim 6.

Case 2: $a = 0$ *and* $b = 1$.
Define the function $g : [0,1] \to \mathbb{R}$ by

$$g(x) = f(x) - f(0) - (f(1) - f(0))x\,.$$

Then g is continuous with $g(0) = g(1) = 0$. It therefore follows from Case 1 that there is a sequence of polynomials $(p_n)_n$ such that

$$p_n \to g \quad \text{with respect to } d_\infty \text{ on } [0,1]. \tag{2}$$

Now define the sequence of polynomials $(q_n)_n$ by $q_n(x) = p_n(x) + f(0) + (f(1) - f(0))x$, and note that it follows from (2) that

$$q_n \to f \quad \text{with respect to } d_\infty \text{ on } [0,1].$$

Case 3: $a, b \in \mathbb{R}$.
This case follows easily from Case 2. □

5. DIFFERENTIABLE FUNCTIONS

5.1. The Derivative.

Having developed our skills at working with limits, we now apply this understanding to the important process of differentiation. Most of the topics covered here will be at least somewhat familiar to the reader from the standard calculus course in school. In school a good deal of time was spend on the applications of derivatives in physics, geometry, and the like. By way of contrast, the focus of this chapter will be on the theoretical aspects of differentiation that are ofter (i.e. always) treated very superficially in school.

Definition. The derivative. *Let $I \subseteq \mathbb{R}$ be an open interval and let*

$$f : I \to \mathbb{R}$$

be a function on I.

The function f is called differentiable at a point $x \in I$ with derivative $f'(x) \in \mathbb{R}$ if

$$\forall \varepsilon > 0 \,:\, \exists \delta > 0 \,:\, \forall 0 < |h| < \delta \,:\, x + h \in I \Rightarrow \left| \frac{f(x+h) - f(x)}{h} - f'(x) \right| \leq \varepsilon \,.$$

We will say that f is differentiable (on I) if f is differentiable at all points $x \in I$.

If f is differentiable on I, then the function

$$f' = f^{(1)} : x \to f'(x)$$

for $x \in I$ is called the derivative of f.

Definition. Higher order derivatives. *Let $I \subseteq \mathbb{R}$ be an open interval and let*

$$f : I \to \mathbb{R}$$

be a differentiable function on I. If f' is differentiable at $x \in I$, the derivative of f' at x is called the second derivative of f at x and is denoted by

$$f''(x) = f^{(2)}(x)\,.$$

Similarly we define

$$f^{(n)}(x)$$

for all $n \in \mathbb{N}$.

Theorem. Sequence characterization of differenbtiablity. *Let $I \subseteq \mathbb{R}$ be an open interval and let*

$$f : I \to \mathbb{R}$$

be a function on I. Fix $x \in I$. Then the follwing statements are equivalent.

(1) *f is differentiable at x with derivative $f'(x)$.*
(2) *If $(h_n)_n$ is any sequence with $h_n \neq 0$ for all n such that*

$$h_n \to 0\,,$$

and $x + h_n \in I$ for all n, then

$$\frac{f(x+h_n) - f(x)}{h} \to f'(x)\,.$$

Proof. The proof is very similar to the proof of the Sequence Characterization of Continuity and is therefore omitted. □

Example. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = x^2\,.$$

Then f is differentiable and $f'(x) = x^2$ for all $x \in \mathbb{R}$.

Proof. Let $x \in \mathbb{R}$. Fix a sequence $h_n \in \mathbb{R}$ such that $h_n \to 0$ and $h_n \neq 0$ for all n. We now have

$$\frac{f(x+h_n) - f(x)}{h_n} = \frac{(x+h_n)^2 - x^2}{h_n} = 2x + h_n \to 2x\,.$$

This completes the proof. □

Example. Let $m \in \mathbb{R}$ and define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = x^m$$

Then f is differentiable and $f'(x) = mx^{m-1}$ for all $x \in \mathbb{R}$.

Proof. Let $x \in \mathbb{R}$. Fix a sequence $h_n \in \mathbb{R}$ such that $h_n \to 0$ and $h_n \neq 0$ for all n. We now have

$$\begin{aligned}
\frac{f(x+h_n) - f(x)}{h_n} &= \frac{\sum_{k=0}^{m} \binom{m}{k} x^{m-k} h_n^k}{h_n} \\
&= \frac{x^m + \binom{m}{1} x^{m-1} h_n^1 + \binom{m}{2} x^{m-2} h_n^2 + \cdots \binom{m}{m} x^{m-m} h_n^m - x^m}{h_n} \\
&= \binom{m}{1} x^{m-1} + \binom{m}{2} x^{m-2} h_n^1 + \cdots + \binom{m}{m} x^{m-m} h_n^{m-1} \\
&\to \binom{m}{1} x^{m-1} = mx^{m-1}\,.
\end{aligned}$$

This completes the proof. □

Example. A differentiable function f such that f' is discontinuous. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x^2 \cos\left(\frac{1}{x}\right) & \text{for } x \neq 0; \\ 0 & \text{for } x = 0. \end{cases}$$

Then we have:

(1) f is differentiable.
(2) f' is not continuous at 0.

Proof.
(1) The function f is clearly differentiable at all $x \neq 0$ with

$$f'(x) = 2x \cos\left(\frac{1}{x}\right) + \sin\left(\frac{1}{x}\right). \tag{1}$$

We now claim that f is differentiable at 0 with

$$f'(0) = 0. \tag{2}$$

Indeed, let h_n be a sequence such that $h_n \to 0$ and $h_n \neq 0$ for all n. Let $\varepsilon > 0$. Choose N such that $|h_n| \leq \varepsilon$ for $n \geq N$. For $n \geq N$ we have

$$\left|\frac{f(h_n) - f(0)}{h_n}\right| = \left|\frac{f(h_n)}{h_n}\right| = \frac{h_n^2 |\cos(\frac{1}{h_n})|}{|h_n|} \leq |h_n| \leq \varepsilon.$$

(2) We must prove that the function

$$f'(x) = \begin{cases} 2x \cos\left(\frac{1}{x}\right) + \sin\left(\frac{1}{x}\right) & \text{for } x \neq 0; \\ 0 & \text{for } x = 0, \end{cases}$$

is discontinuous at 0. Indeed, define the sequence $(x_n)_n$ by $x_n = \frac{2}{(4n+1)\pi}$. Then clearly $x_n \to 0$, but since

$$f'(x_n) = 2x_n \cos\left(\frac{1}{x_n}\right) + \sin\left(\frac{1}{x_n}\right) = 1 \quad \text{for all } n,$$

the sequence $(f(x_n)_n = (1, 1, 1, 1, \dots)$ does not converge to $f(0) = 0$. ☐

Theorem. *Let $I \subseteq \mathbb{R}$ be an open interval and let*

$$f : I \to \mathbb{R}$$

be a function on I. If f is differentiable at $x \in I$, then f is continuous at x.

Proof. Let $\varepsilon > 0$. Since f is differentiable at x there exists $\eta > 0$ such that

$$\forall 0 < |h| < \eta \, : \, x + h \in I \Rightarrow \left|\frac{f(x+h) - f(x)}{h} - f'(x)\right| \leq 1.$$

Now let $\delta = \min(\eta, \frac{\varepsilon}{1+|f'(x)|})$. For $y \in I$ with $0 < |x-y| < \delta$ we have

$$\begin{aligned}
|f(x) - f(y)| &= \left| \frac{f(x)-f(y)}{x-y} - f'(x) + f'(x) \right| \; |x-y| \\
&\leq \left(\left| \frac{f(x)-f(y)}{x-y} - f'(x) \right| + |f'(x)| \right) \; |x-y| \\
&\leq (1 + |f'(x)|)|x-y| \\
&\leq (1 + |f'(x)|)\delta \\
&\leq \varepsilon \, .
\end{aligned}$$

Also, for $y \in I$ with $y = x$ we have

$$|f(x) - f(y)| = 0 \leq \varepsilon \, .$$

Hence, for all $y \in I$ with $|x-y| < \delta$ we have

$$|f(x) - f(y)| \leq \varepsilon \, .$$

This completes the proof. □

We now present the useful (and familiar) rules for taking the derivative of sums, products, and quotients of function.

Theorem. *Let $I \subseteq \mathbb{R}$ be an open interval and let*

$$f, g : I \to \mathbb{R}$$

be functions on I. Let $k \in \mathbb{R}$. Assume that f and g are differentiable at $x \in I$. Then the following statements hold.

(1) *The function kf is differentiable at x, with*

$$(kf)'(x) = kf'(x) \, .$$

(2) *The function $f+g$ is differentiable at x, with*

$$(f+g)'(x) = f'(x) + g'(x) \, .$$

(3) *The function fg is differentiable at x, with*

$$(fg)'(x) = f(x)g'(x) + f'(x)g(x) \, .$$

(4) *If $g(y) \neq 0$ for all $y \in I$, then function $\frac{f}{g}$ is differentiable at x, with*

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2} \, .$$

Proof.
(1) This is left as an exercise for the reader.
(2) Let $(h_n)_n$ be a sequence with $h_n \neq 0$ and $x + h_n \in I$ for all n. Since f and g are differentiable at x we conclude that

$$\begin{aligned} \frac{f(x+h_n)+f(x)}{h_n} &\to f'(x)\,, \\ \frac{g(x+h_n)+g(x)}{h_n} &\to g'(x)\,. \end{aligned}$$

Hence

$$\begin{aligned} \frac{(f+g)(x+h_n)-(f+g)(x)}{h_n} &= \frac{(f(x+h_n)+g(x+h_n))-(f(x)+g(x))}{h_n} \\ &= \frac{f(x+h_n)-f(x)+g(x+h_n)-g(x)}{h_n} \\ &= \frac{f(x+h_n)-f(x)}{h_n} + \frac{g(x+h_n)-g(x)}{h_n} + \\ &\to f'(x)+g'(x)\,. \end{aligned}$$

(3) Let $(h_n)_n$ be a sequence with $h_n \neq 0$ and $x + h_n \in I$ for all n. Since f and g are differentiable at x we conclude that

$$\begin{aligned} \frac{f(x+h_n)+f(x)}{h_n} &\to f'(x)\,, \\ \frac{g(x+h_n)+g(x)}{h_n} &\to g'(x)\,. \end{aligned}$$

Also, since $x + h_n \to x$ and f is differentiable at x and hence continuous at x, we conclude that

$$f(x+h_n) \to f(x)\,.$$

Hence

$$\begin{aligned} \frac{(fg)(x+h_n)-(fg)(x)}{h_n} &= \frac{f(x+h_n)g(x+h_n)-f(x)g(x)}{h_n} \\ &= \frac{f(x+h_n)g(x+h_n)-f(x+h_n)g(x)+f(x+h_n)g(x)-f(x)g(x)}{h_n} \\ &= f(x+h_n)\frac{g(x+h_n)-g(x)}{h_n} + g(x)\frac{f(x+h_n)-f(x)}{h_n} \\ &\to f(x)g'(x)+g(x)f'(x)\,. \end{aligned}$$

(4) Let $(h_n)_n$ be a sequence with $h_n \neq 0$ and $x + h_n \in I$ for all n. Since f and g are differentiable at x we conclude that

$$\begin{aligned} \frac{f(x+h_n)+f(x)}{h_n} &\to f'(x)\,, \\ \frac{g(x+h_n)+g(x)}{h_n} &\to g'(x)\,. \end{aligned}$$

Also, since $x + h_n \to x$ and g is differentiable at x and hence continuous at x, we conclude that

$$\frac{1}{g(x+h_n)} \to \frac{1}{g(x)}\,.$$

Hence

$$\begin{aligned}
\frac{\frac{f}{g}(x+h_n) - \frac{f}{g}(x)}{h_n} &= \frac{\frac{f(x+h_n)}{g(x+h_n)} - \frac{f(x)}{g(x)}}{h_n} \\
&= \frac{g(x)f(x+h_n) - f(x)g(x+h_n)}{g(x+h_n)g(x)}\,\frac{1}{h_n} \\
&= \frac{g(x)f(x+h_n) - g(x)f(x) + g(x)f(x) - f(x)g(x+h_n)}{g(x+h_n)g(x)}\,\frac{1}{h_n} \\
&= \left(g(x)\frac{f(x+h_n) - f(x)}{h_n} - f(x)\frac{g(x+h_n) - g(x)}{h_n}\right)\frac{1}{g(x+h_n)g(x)} \\
&\to \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}\,.
\end{aligned}$$

This completes the proof. $\square$

In Section 2 we proved that the composition of two continuous functions is continuous. A similar result holds for the composition of differentiable functions, and is known as the chain rule.

Theorem. The Chain Rule. *Let $I, J \subseteq \mathbb{R}$ be open intervals and let*

$$g : I \to J$$

and

$$f : J \to \mathbb{R}$$

be functions. If g is differentable at $x \in I$ and f is differentiable at $g(x) \in J$, the the composite function $f \circ g$ is differentiable at $x \in I$ and

$$(f \circ g)'(x) = f'(g(x))g'(x)\,.$$

Discussion. Let $(h_n)_n$ be a sequence with $h_n \neq 0$ and $x + h_n \in I$ for all n. We must now prove that

$$\frac{(f \circ g)(x+h_n) - (f \circ g)(x)}{h_n} \to f'(g(x))g'(x)\,.$$

Naively we would simply rewrite the quotient $\frac{(f\circ g)(x+h_n)-(f\circ g)(x)}{h_n}$ as follows

$$\begin{aligned}
\frac{(f \circ g)(x+h_n) - (f \circ g)(x)}{h_n} &= \frac{f(g(x+h_n)) - f(g(x))}{h_n} \\
&= \frac{f(g(x+h_n)) - f(g(x))}{g(x+h_n) - g(x)}\,\frac{g(x+h_n) - g(x)}{h_n} \\
&= \frac{f(g(x)+k_n) - f(g(x))}{k_n}\,\frac{g(x+h_n) - g(x)}{h_n}\,,
\end{aligned} \tag{1}$$

where $k_n = g(x+h_n) - g(x)$. Since g is differentiable at x and thus continuous at x we conclude that $k_n \to 0$. Hence (by taking limits in (1)),

$$\begin{aligned}\frac{(f\circ g)(x+h_n)-(f\circ g)(x)}{h_n} &= \frac{f(g(x)+k_n)-f(g(x))}{k_n}\;\frac{g(x+h_n)-g(x)}{h_n}\\ &\to f'(g(x))g'(x)\,.\end{aligned}$$

The only problem with this approach is that the quantity $k_n = g(x+h_n) - g(x)$ may be zero for many or possible infinitely many values of n. Thus the first factor in the right-hand side of (1) may have a zero denominator.

Proof of the Chain Rule. Let $(h_n)_n$ be a sequence with $h_n \neq 0$ and $x + h_n \in I$ for all n. We must now prove that

$$\frac{(f\circ g)(x+h_n)-(f\circ g)(x)}{h_n} \to f'(g(x))g'(x)\,.$$

Define the function $h : J \to \mathbb{R}$ by

$$h(y) = \begin{cases} \dfrac{f(y)-f(g(x))}{y-g(x)} & \text{for } y \in J \setminus \{g(x)\};\\ f'(g(x)) & \text{for } y = g(x). \end{cases}$$

(Observe that h is well-defined since f is differentiable at $g(x)$.) Note that

$$f(y) - f(g(x)) = h(y)(y - g(x)) \quad \text{for all } y \in J. \tag{2}$$

Since f is differentiable at $g(x)$ we conclude that h is continuous at $g(x)$. Also, g is differentiable at x and hence continuous at x. We therefore deduce that the composite $h \circ g : I \to \mathbb{R}$ is continuous at x, whence

$$(h\circ g)(x+h_n) \to (h\circ g)(x) = h(g(x)) = f'(g(x))\,.$$

Hence (using (2))

$$\begin{aligned}\frac{(f\circ g)(x+h_n)-(f\circ g)(x)}{h_n} &= \frac{f(g(x+h_n))-f(g(x))}{h_n}\\ &= \frac{h(g(x+h_n))\big(g(x+h_n)-g(x)\big)}{h_n}\\ &= h(g(x+h_n))\frac{g(x+h_n)-g(x)}{h_n}\\ &\to f'(g(x))g'(x)\,.\end{aligned}$$

This completes the proof. $\square$

5.2. The Mean Value Theorem.

The mean value theorem is one of the most important results in differential calculus. Its proof depends on the fact that a continuous function defined on a closed interval attains its maximum and its minimum values. In this section we establish the theorem and several of its corollaries. We begin with a preliminary result about maxima and minima.

The Interior Extremum Theorem. *Let $I \subseteq \mathbb{R}$ be an open interval and let*

$$f : I \to \mathbb{R}$$

be a differentiable function on I.

(1) *If f assumes it maximum at $x \in I$, then $f'(x) = 0$.*
(2) *If f assumes it minimum at $x \in I$, then $f'(x) = 0$.*

Proof.
(1) We first prove that $f'(x) \geq 0$. Let $(h_n)_n$ be a sequence with $h_n \to 0$ and $h_n < 0$ for all n such that $x + h_n \in I$ for all n. (Since I is open and $x \in I$ there exists $N \in \mathbb{N}$ such that $x - \frac{1}{n} \in I$ for $n \geq N$, and we may take $h_n = -\frac{1}{n}$ for $n \geq N$.) Then since f is differentiable at x we have

$$f'(x) = \lim_n \frac{f(x + h_n) - f(x)}{h_n} \,. \tag{3}$$

However, since f has a maximum at x we conclude that $f(x + h_n) \leq f(x)$ for all n, whence

$$\frac{f(x + h_n) - f(x)}{h_n} \geq 0 \tag{4}$$

for all n. It follows from (3) and (4) that $f'(x) \geq 0$.

Next we prove that $f'(x) \leq 0$. Let $(k_n)_n$ be a sequence with $k_n \to 0$ and $k_n > 0$ for all n such that $x + h_n \in I$ for all n. (Since I is open and $x \in I$ there exists $N \in \mathbb{N}$ such that $x + \frac{1}{n} \in I$ for $n \geq N$, and we may take $k_n = x + \frac{1}{n}$ for $n \geq N$.) Then since f is differentiable at x we have

$$f'(x) = \lim_n \frac{f(x + k_n) - f(x)}{k_n} \,. \tag{5}$$

However, since f has a maximum at x we conclude that $f(x + k_n) \leq f(x)$ for all n, whence

$$\frac{f(x + k_n) - f(x)}{k_n} \leq 0 \tag{6}$$

for all n. It follows from (5) and (6) that $f'(x) \leq 0$. $\square$

In beginning calculus when we wished to find the maximum (or minimum) of a continuous function $f : [a, b] \to \mathbb{R}$ on a closed interval, we considered the following three type of points

(1) The points x at which f is differentiable and $f'(x) = 0$.
(2) The endpoints a and b.
(3) The points where f is not differentiable.

In many (but not all) applications the extreme value will actually occur at a point of the first type where the derivative is zero. It is the Interior Extremum Theorem that justufies this approach.

For our present puposes, we shall use the Interior Extremum Theorem to prove Rolle's Theorem which is a special case of the mean value theorem.

Rolle's Theorem. *Let $a, b \in \mathbb{R}$ with $a < b$. Let $f : [a, b] \to \mathbb{R}$ be a continuous function and assume that f is differentiable on (a, b) and that $f(a) = f(b) = 0$. Then then there exists $\xi \in (a, b)$ such that*

$$f'(\xi) = 0\,.$$

Proof. We divide the proof into three cases.

Case 1: $f(x) = 0$ for all $x \in [a, b]$. In the case the result is trivial since $f'(\xi) = 0$ for all $\xi \in (a, b)$ (why?).

Case 2: $f(x) > 0$ for some $x \in [a, b]$. Since f is a continuous function defined on a closed and bounded interval the Extremum Theorem for Continuous Functions shows that there exists $\xi \in [a, b]$ such that

$$f(\xi) \geq f(y)$$

for all $y \in [a, b]$. In particular, we conclude that $f(\xi) \geq f(x) > 0$, whence $\xi \neq a$ and $\xi \neq b$, i.e. $\xi \in (a, b)$. Hence $\xi \in (a, b)$ and f has a maximum at ξ. The previous theorem therefore implies that $f'(\xi) = 0$.

Case 3: $f(x) < 0$ for some $x \in [a, b]$. The proof in Case 3 is very similar to the proof in Case 2 and is therefor omitted. ☐

The geometric interpretation of Rolle's Theorem is that, if the graph of a differentiable function touches the x-axis at a and b, then for some ξ between a and b there is a horizontal tangent. If we allow the function to have different values at the endpoints , then we cannot be assured of a horizontal tangent, but there will be a point ξ in (a, b) such that the tangent to the graph at ξ will be parallel to the line joining the endpoints of the graph. This is the essence of the mean value theorem.

The Mean Value Theorem. *Let $a, b \in \mathbb{R}$ with $a < b$. Let $f : [a, b] \to \mathbb{R}$ be a continuous function and assume that f is differentiable on (a, b). Then then there exists $\xi \in (a, b)$ such that*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}\,.$$

Proof. Define $\varphi : [a, b] \to \mathbb{R}$ by

$$\varphi(x) = f(x) - \left(\frac{f(b) - f(a)}{b - a}(x - a) + f(a)\right)\,.$$

Then φ is continuous on $[a, b]$ and differentiable on (a, b) with $\varphi(a) = \varphi(b) = 0$. Rolle's Theorem therefore implies that that there exists $\xi \in (a, b)$ such that

$$0 = \varphi'(\xi) = f'(\xi) - \frac{f(b) - f(a)}{b - a}\,.$$

This completes the proof. ☐

5.3. Applications of the Mean Value Theorem.

In this section we will give several applications of how the mean value theorem can be used to relate the properties of a differentiable function and its derivative.

Theorem. *Let $a, b \in \mathbb{R}$ with $a < b$. Let $f : [a, b] \to \mathbb{R}$ be a continuous function and assume that f is differentiable on (a, b) with $f'(x) = 0$ for all $x \in (a, b)$. Then there exists $c \in \mathbb{R}$ such that*

$$f(x) = c$$

for all $x \in [a, b]$.

Proof. Suppose that f were not constant on $[a, b]$. Then there exist x_1 and x_2 such that $a \le x_1 < x_2 \le b$ and $f(x_1) \ne f(x_2)$. Since f is continuous on $[x_1, x_2]$ and differentiable on (x_1, x_2), the mean value theorem implies that there exsits $\xi \in (x_1, x_2) \subset (a, b)$ such that

$$f'(\xi) = \frac{f(x_2) - f(x_1)}{x_2 - x_1} \ne 0\,,$$

contradicting the fact that $f'(x) = 0$ for all $x \in (a, b)$. □

Corollary. *Let $a, b \in \mathbb{R}$ with $a < b$. Let $f, g : [a, b] \to \mathbb{R}$ be continuous functions and assume that f and g are differentiable on (a, b) with $f'(x) = g'(x)$ for all $x \in (a, b)$. Then there exists $c \in \mathbb{R}$ such that*

$$f(x) = g(x) + c$$

for all $x \in [a, b]$.

Proof. This follows by applying the previous theorem to the function $f - g : [a, b] \to \mathbb{R}$. □

Definition. Increasing and decreasing function. *Let $I \subseteq \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be a function on I.*

The function is called increasing if

$$f(x) \le f(y)$$

for all $x, y \in I$ with $x < y$.

The function is called decreasing if

$$f(x) \ge f(y)$$

for all $x, y \in I$ with $x < y$.

Definition. Strictly increasing and strictly decreasing function. *Let $I \subseteq \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be a function on I.*

The function is called strictly increasing if

$$f(x) < f(y)$$

for all $x, y \in I$ with $x < y$.

The function is called decreasing if

$$f(x) > f(y)$$

for all $x, y \in I$ with $x < y$.

Theorem. *Let $I \subseteq \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be a differentiable function on I.*

(1) *If $f'(x) > 0$ for all $x \in I$, then f is strictly increasing on I.*
(2) *If $f'(x) < 0$ for all $x \in I$, then f is strictly decreasing on I.*

Proof. (1) Let $x, y \in I$ with $x < y$. It now follows from the mean value theorem that there exists $\xi \in (x, y)$ such that $f(y) - f(x) = f'(\xi)(y - x) > 0$, whence $f(y) > f(x)$.
(2) The proof of (2) is similar to the proof of (1). $\square$

Observe that the implication in the last theorem cannot be reversed. Indeed, define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = x^3 .$$

Then f is differentiable and strictly increasing, but

$$f'(0) = 0 .$$

Theorem. *Let $I \subseteq \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be a differentiable function on I.*

(1) *The following statements are equivalent.*
  (a) $f'(x) \geq 0$ for all $x \in I$.
  (b) f is increasing on I.
(2) *The following statements are equivalent.*
  (a) $f'(x) \leq 0$ for all $x \in I$.
  (b) f is decreasing on I.

Proof.
(1) We first prove that (a) implies (b). Let $x, y \in I$ with $x < y$. It now follows from the mean value theorem that there exists $\xi \in (x, y)$ such that $f(y) - f(x) = f'(\xi)(y - x) \geq 0$, whence $f(y) \geq f(x)$.

Next we prove that (b) implies (a). Let $x \in I$. We must now prove that $f'(x) \geq 0$. Let $(h_n)_n$ be a sequence with $h_n \to 0$ and $h_n > 0$ for all n such that $x + h_n \in I$ for all n. (Since I is open and $x \in I$ there exists $N \in \mathbb{N}$ such that $x + \frac{1}{n} \in I$ for $n \geq N$, and we may take $h_n = \frac{1}{n}$ for $n \geq N$.) Then since f is differentiable at x we have

$$f'(x) = \lim_n \frac{f(x + h_n) - f(x)}{h_n} . \tag{7}$$

However, since f increasing we conclude that $f(x) \leq f(x+h_n)$ for all n, whence

$$\frac{f(x+h_n)-f(x)}{h_n} \geq 0 \tag{8}$$

for all n. It follws from (7) and (8) that $f'(x) \geq 0$.
(2) The proof of (2) is similar to the proof of (1). $\square$

6. INTEGRATION

In this chapter we present the theory of Riemann integration. While the process of integration had been developed much earlier in the seventeenth century by Isaac Newton (1642–1727) and Gottfried Leibniz (1646–1716) it was Berhard Riemann (1826–1866) who formulated the modern definition of the definite integral that is commonly used today. Subsequent to Riemann's work, other more satisfactory approaches to integration were developed, most notably by Henri Lebesgue (1875–1941). The theory and applications (to ergodic theory) of Lebesgue's theory of integration will be covered subsequent courses.

Since the reader has already in school seen many of the important applications of integration in the sciences, we shall concentrate on a rigorous development of the theory. We begin by defining the Riemann integral in terms of upper and lower sums. In section 6.2 we identify an important class of functions that are integrable and then derive several related algebraic properties. The Fundamental Theorem of Calculus is discussed in Section 6.3.

Georg Friedrich Bernhard Riemann (September 17, 1826 – July 20, 1866) was an influential German mathematician who made lasting and revolutionary contributions to analysis, number theory, and differential geometry. In the field of real analysis, he is mostly known for the first rigorous formulation of the integral, the Riemann integral, and his work on Fourier series. His contributions to complex analysis include most notably the introduction of Riemann surfaces, breaking new ground in a natural, geometric treatment of complex analysis. His famous 1859 paper on the prime-counting function, containing the original statement of the Riemann hypothesis, is regarded, although it is his only paper in the field, as one of the most influential papers in analytic number theory. Through his pioneering contributions to differential geometry, Riemann laid the foundations of the mathematics of general relativity..

Henri Lebesgue (June 28, 1875 – July 26, 1941) was a French mathematician most famous for his theory of integration, which was a generalization of the 17th century concept of integration. His theory was published originally in his dissertation Integrale, longueur, aire ("Integral, length, area") at the University of Nancy during 1902.

Jean-Gaston Darboux (14 August 1842 – 23 February 1917) was a French mathematician. Darboux made several important contributions to geometry and mathematical analysis, including, the approach to the Riemann integral based on upper and lower sums.

6.1. The Riemann Integral.

We begin with some definitions.

Definition. Partitions. *Let $a, b \in \mathbb{R}$ with $a < b$. A partition P of $[a, b]$ is a finite set of points*

$$\{x_0, x_1, \ldots, x_{n-1}, x_n\}$$

in $[a, b]$ such that

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\,.$$

The partition P will be written as

$$P = \{x_0 < x_1 < \cdots < x_{n-1} < x_n\}\,.$$

If P and Q are two partitions of $[a, b]$ with $P \subseteq Q$, then Q is called a refinement of P.

Definition. Upper sum and lower sum. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$ be a bounded function. For $S \subseteq [a, b]$ we write*

$$M(f, S) = \sup\{f(x) \mid x \in S\}\,,$$

and

$$m(f, S) = \inf\{f(x) \mid x \in S\}\,.$$

Let

$$P = \{x_0 < x_1 < \cdots < x_{n-1} < x_n\}$$

be a partition of $[a, b]$. We define the upper sum of f over $[a, b]$ with respect to P by

$$U(f, P) = U_a^b(f, P) = \sum_{k=1}^{n} M(f, [x_{k-1}, x_k])(x_k - x_{k-1})\,,$$

and we define the lower sum of f over $[a, b]$ with respect to P by

$$L(f, P) = L_a^b(f, P) = \sum_{k=1}^{n} m(f, [x_{k-1}, x_k])(x_k - x_{k-1})\,.$$

Sometimes $U(f, P)$ and $L(f, p)$ are called the upper and lower Darboux sums in honor of Gaston Darboux (1842–1917), who first developed this approach to the Riemann integral.

Since we are assuming that f is a bounded function on $[a, b]$, there exist real numbers m and M such that $m \leq f(x) \leq M$ for all $x \in [a, b]$. Thus for any partition P of $[a, b]$ we have

$$m(b - a) \leq L(f, P) \leq U(f, p) \leq M(b - a)\,.$$

This implies that the sets of the upper and lower sums for f form bounded sets, and it guarantees the existence of the following upper and lower integrals

Definition. Upper integral, lower integral and the integral. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a,b] \to \mathbb{R}$ be a bounded function. We define the upper (Riemann) integral of f over $[a,b]$ by*

$$U(f) = U_a^b(f) = \inf\{U(f,P) \mid P \text{ is a partition of } [a,b]\}\,,$$

and we define the lower (Riemann) integral of f over $[a,b]$ by

$$L(f) = L_a^b(f) = \sup\{L(f,P) \mid P \text{ is a partition of } [a,b]\}\,.$$

If the upper integral $U(f)$ and the lower integral $L(f)$ coincide, then we say that f is (Riemann) integrable on $[a,b]$, and we denote their common value by $\int_a^b f = \int_a^b f(t)\,dt$, i.e. if $U(f) = L(f)$, then the (Riemann) integral $\int_a^b f = \int_a^b f(t)\,dt$ is defined by

$$\int_a^b f = \int_a^b f(t)\,dt = L(f) = U(f)\,.$$

Theorem. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a,b] \to \mathbb{R}$ be a bounded function.*

(1) *If P and Q are partitions of $[a,b]$ and Q is a refinement of P, then*

$$L(f,P) \le L(f,Q) \le U(f,Q) \le U(f,P)\,.$$

(2) *If P and Q are partitions of $[a,b]$, then*

$$L(f,P) \le U(f,Q)\,.$$

(3) *We have*

$$L(f) \le U(f)\,.$$

Proof.
(1) The middle inequality follows directly from the definitions of $L(f,Q)$ and $U(f,Q)$.

To prove $L(f,P) \le U(f,Q)$ we suppose that $P = \{x_0 < x_1 < \cdots < x_{n-1} < x_n\}$ and consider the partition

$$P_* = \{x_0 < x_1 < \ldots x_{k-1} < x_* < x_k \cdots < x_{n-1} < x_n\}$$

former by joining some point x_* to P where $x_{k-1} < x_* < x_k$ for some $k = 1, \ldots, n$.

Now the terms in $L(f,P_*)$ and $L(f,P)$ are all the same except those over the subinterval $[x_{k-1}, x_k]$. Thus we have

$$\begin{aligned}
L(f,P_*) - L(f,P) &= \Big(m(f,[x_{k-1},x_*])(x_* - x_{k-1}) + m(f,[x_*,x_k])(x_k - x_*)\Big) \\
&\quad - m(f,[x_{k-1},x_k])(x_k - x_{k-1}) \\
&= \Big(m(f,[x_{k-1},x_*])(x_* - x_{k-1}) + m(f,[x_*,x_k])(x_k - x_*)\Big) \\
&\quad - \Big(m(f,[x_{k-1},x_k])(x_* - x_{k-1}) + m(f,[x_{k-1},x_k])(x_k - x_*)\Big) \\
&= \Big(m(f,[x_{k-1},x_*]) - m(f,[x_{k-1},x_k])\Big)(x_* - x_{k-1}) \\
&\quad + \Big(m(f,[x_*,x_k]) - m(f,[x_{k-1},x_k])\Big)(x_k - x_*)\,. \qquad (1.1)
\end{aligned}$$

Since, clearly

$$m(f,[x_{k-1},x_*]) \geq m(f,[x_{k-1},x_k])$$

and

$$m(f,[x_*,x_k]) \geq m(f,[x_{k-1},x_k])\,,$$

it follows from (1.1) that

$$\begin{aligned} L(f,P_*) - L(f,P) &= \Big(m(f,[x_{k-1},x_*]) - m(f,[x_{k-1},x_k])\Big)(x_* - x_{k-1}) \\ &\quad + \Big(m(f,[x_*,x_k]) - m(f,[x_{k-1},x_k])\Big)(x_k - x_*) \\ &\geq 0\,. \end{aligned}$$

Finally, if the partition Q contains m more points than P, we apply the argument above m times to conclude to obtain $L(f,P) \leq L(f,Q)$.

The proof that $U(f,Q) \leq U(f,P)$ is similar.
(2) Since $P \cup Q$ is a refinement of both P and Q we obtain by (1)

$$L(f,P) \leq L(f,P\cup Q) \leq U(f,P\cup Q) \leq U(f,Q)\,.$$

(3) Let P and Q be partitions of $[a,b]$. It follows from (2) that

$$L(f,P) \leq U(f,Q)\,.$$

Hence

$$L(f) = \sup\{L(f,P) \mid P \text{ is a partition of } [a,b]\} \leq U(f,Q)\,.$$

This now implies that

$$L(f) \leq \inf\{U(f,Q) \mid Q \text{ is a partition of } [a,b]\} = U(f)\,.$$

This completes the proof. $\square$

Example. Define $f:[0,1]\to\mathbb{R}$ by

$$f(x) = x^2\,.$$

Then f is integrable and

$$\int_0^1 f = \frac{1}{3}\,.$$

Proof. For each $n \in \mathbb{N}$ consider the partition

$$P_n = \left\{0, \frac{1}{n}, \frac{12}{n}, \ldots, \frac{n-1}{n}, 1\right\}\,.$$

Clearly (since f is increasing)

$$m\left(f, \left[\frac{k-1}{n}, \frac{k}{n}\right]\right) = f\left(\frac{k-1}{n}\right) = \left(\frac{k-1}{n}\right)^2$$

and

$$M\left(f,\left[\frac{k-1}{n},\frac{k}{n}\right]\right)=f\left(\frac{k}{n}\right)=\left(\frac{k}{n}\right)^2.$$

Hence (since $\sum_{k=1}^n k^2=\frac{n(n+1)(2n+1)}{6}$ for all $n\in\mathbb{N}$)

$$\begin{aligned}L(f,P_n)&=\sum_{k=1}^n m\left(f,\left[\frac{k-1}{n},\frac{k}{n}\right]\right)\left(\frac{k}{n}-\frac{k-1}{n}\right)\\&=\sum_{k=1}^n\left(\frac{k-1}{n}\right)^2\frac{1}{n}\\&=\frac{1}{n^3}\sum_{k=1}^{n-1}k^2=\frac{1}{n^3}\ \frac{(n-1)n(2(n-1)+1)}{6}=\frac{1}{3}\ \left(1-\frac{1}{n}\right)\ \left(1-\frac{1}{2n}\right),\end{aligned}$$

and so

$$L(f)\geq\sup_n L(f,P_n)=\sup_n\frac{1}{3}\ \left(1-\frac{1}{n}\right)\ \left(1-\frac{1}{2n}\right)=\frac{1}{3}.$$

Similarly, we obtain

$$\begin{aligned}U(f,P_n)&=\sum_{k=1}^n M\left(f,\left[\frac{k-1}{n},\frac{k}{n}\right]\right)\left(\frac{k}{n}-\frac{k-1}{n}\right)\\&=\sum_{k=1}^n\left(\frac{k}{n}\right)^2\frac{1}{n}\\&=\frac{1}{n^3}\sum_{k=1}^{n}k^2=\frac{1}{n^3}\ \frac{n(n+1)(2n+1)}{6}=\frac{1}{3}\ \left(1+\frac{1}{n}\right)\ \left(1+\frac{1}{2n}\right),\end{aligned}$$

and so

$$U(f)\leq\inf_n U(f,P_n)=\inf_n\frac{1}{3}\ \left(1+\frac{1}{n}\right)\ \left(1+\frac{1}{2n}\right)=\frac{1}{3}.$$

Hence $\frac{1}{3}\leq L(f)\leq U(f)\leq\frac{1}{3}$, and so $\int_0^1 f=L(f)=U(f)=\frac{1}{3}$. $\square$

Example. Define $f:[0,1]\to\mathbb{R}$ by

$$f(x)=\begin{cases}0 & \text{for } x\in 0,1\setminus\mathbb{Q};\\ 1 & \text{for } x\in 0,1\cap\mathbb{Q}.\end{cases}$$

Then

$$L(f)=0\ ,\ U(f)=1.$$

In particular, f is not integrable.

Proof. Let $P=\{0=x_0<x_1<\cdots<x_{n-1}<x_n=1\}$ be a partition of $[0,1]$. Since each subinterval $[x_{k-1},x_k]$ contains irrational numbers we have

$$m(f,[x_{k-1},x_k])=0.$$

Similary, since each subinterval $[x_{k-1}, x_k]$ contains rational numbers we have

$$M(f, [x_{k-1}, x_k]) = 1 .$$

Hence

$$L(f, P) = \sum_{k=1}^{n} m(f, [x_{k-1}, x_k])(x_k - x_{k-1}) = \sum_{k=1}^{n} 0(x_k - x_{k-1}) = 0 ,$$

and

$$U(f, P) = \sum_{k=1}^{n} M(f, [x_{k-1}, x_k])(x_k - x_{k-1}) = \sum_{k=1}^{n} 1(x_k - x_{k-1}) = x_n - x_0 = 1 - 0 = 1 .$$

This implies that $L(f) = 0$ and $U(f) = 1$. □

6.2. Conditions for integrability.

Since the previous example shows that not every function is integrable, we are faced with the problem of determine when a function is integrable. Our next result will be very useful to us in that task.

Theorem. Riemann's integrability criterium. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then the folowing statements are equivalent.*

1. *f is integrable.*
2. *For every $\varepsilon > 0$ there exists a partition P of $[a, b]$ such that*

$$U(f, P) - L(f, P) \le \varepsilon .$$

Proof.
(a) ⇒ (b): Let $\varepsilon > 0$. Since f is integrable, we have $L(f) = U(f)$. We may choose a partition P_1 of $[a, b]$ such that

$$L(f, P_1) \ge L(f) - \frac{\varepsilon}{2} .$$

(This follows from the definition of $L(f)$ as a supremum.) Similarly, there exists a partition P_2 of $[a, b]$ such that

$$U(f, P_2) \le U(f) + \frac{\varepsilon}{2} .$$

Now let $P = P_1 \cup P_2$. As P is a refinement of P_1 and P_2 it follows from Theorem ????? that

$$L(f, P_1) \le L(f, P) \ , \ U(f, P) \le U(f, P_2) .$$

Hence

$$\begin{aligned} U(f,P) - L(f,P) &\leq U(f,P_2) - L(f,P_1) \\ &\leq \Big(U(f) + \frac{\varepsilon}{2}\Big) - \Big(L(f) - \frac{\varepsilon}{2}\Big) \\ &= U(f) - L(f) + \varepsilon = \varepsilon\,. \end{aligned}$$

(b) $\Rightarrow$ (a): Let $\varepsilon > 0$. It follows from (b) that there exists a partition P of $[a,b]$ such that

$$U(f,P) - L(f,P) \leq \varepsilon\,.$$

Hence

$$U(f) \leq U(f,P) \leq L(f,P) + \varepsilon \leq L(f) + \varepsilon\,.$$

Since $\varepsilon > 0$ was arbitrary, this implies that $U(f) \leq L(f)$. But then Theorem ????? implies that $L(f) = U(f)$. $\square$

We will now show that all continuous functions defined on closed and bounded intervals are integrable. The proof of this result relies on the fact that a continuous function on a closed and bounded interval is, in fact, uniformly continuous (cf. Theorem ????).

Theorem. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then f is integrable.*

Proof. First observe that since f is a continuous function defined on a closed and bounded interval, Theorem ???? implies that f is bounded. We will show that condition (b) in Riemann's Integrability Criterion Theorem is satisfied. Therefore let $\varepsilon > 0$. Since f is continuous on the closed and bounded interval $[a,b]$, we deduce from Theorem ???? that f is uniformly continuous on $[a,b]$. Hence, we can choose $\delta > 0$ such that

$$|f(x) - f(y)| \leq \frac{\varepsilon}{b-a} \quad \text{for all } x, y \in [a,b] \text{ with } |x-y| \leq \delta.$$

Next let $P = \{x_0 < x_1 < \cdots < x_{n-1} < x_n\}$ be any partition of $[a,b]$ such that

$$|x_{k-1} - x_k| \leq \delta \quad \text{for all } k = 1, 2, 3, \ldots, n.$$

(Such a partition clearly exists.) Now, since f is continuous the Extremum value Theorem for Continuous Functions, shows that f assumes its maximum and minimum on each closed and bounded interval $[x_{k-1}, x_k]$ for $k = 1, 2, 3, \ldots, n$. That is, for each $k = 1, 2, 3, \ldots, n$ there exist $s_k, t_k \in [x_{k-1}, x_k]$ such that

$$f(s_k) = m(f, [x_{k-1}, x_k]) = m_k$$

and

$$f(t_k) = M(f, [x_{k-1}, x_k]) = M_k\,.$$

Since $s_k, t_k \in [x_{k-1}, x_k]$ we have $|s_k - t_k| \leq |x_{k-1} - x_k| \leq \delta$, and so

$$M_k - m_k = f(t_k) - f(s_k) \leq \frac{\varepsilon}{b-a}$$

for all k. It now follows that

$$\begin{aligned}
U(f,P) - L(f,P) &= \sum_{k=1}^{n} M_k(x_k - x_{k-1}) - \sum_{k=1}^{n} m_k(x_k - x_{k-1}) \\
&= \sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1}) \\
&\leq \sum_{k=1}^{n} \frac{\varepsilon}{b-a}(x_k - x_{k-1}) = \frac{\varepsilon}{b-a}\sum_{k=1}^{n}(x_k - x_{k-1}) \\
&= \frac{\varepsilon}{b-a}(x_n - x_0) = \frac{\varepsilon}{b-a}(b-a) = \varepsilon\,.
\end{aligned}$$

This completes the proof. □

6.3. Algebraic properties of the Riemann integral.

We now turn our attention to proving several algebraic properties of the Riemann integral.

Theorem. *Let $a, b \in \mathbb{R}$ with $a < b$. Let $f, g : [a,b] \to \mathbb{R}$ be bounded functions and $\lambda \in \mathbb{R}$.*

(1) *The upper integral is sub-additive: we have*

$$U(f+g) \leq U(f) + U(g)\,.$$

(2) *The lower integral is super-additive: we have*

$$L(f) + L(g) \leq L(f+g)\,.$$

(3) *If f and g are integrable, then $f+g$ is integrable and*

$$\int_a^b (f+g) = \int_a^b f + \int_a^b g\,.$$

(4) *We have*

$$U(\lambda f) \leq \begin{cases} \lambda U(f) & \text{for } 0 \leq \lambda; \\ \lambda L(f) & \text{for } \lambda \leq 0. \end{cases}$$

(5) *We have*

$$L(\lambda f) \leq \begin{cases} \lambda L(f) & \text{for } 0 \leq \lambda; \\ \lambda U(f) & \text{for } \lambda \leq 0. \end{cases}$$

(6) *If f is integrable, then λf is integrable and*

$$\int_a^b \lambda f = \lambda \int_a^b f\,.$$

Proof. We first make the following observations. For all $S \subseteq [a,b]$ we have

$$\begin{aligned} M(f+g,S) &= \sup_{x\in S}(f(x)+g(x)) \\ &\le \sup_{x\in S} f(x) + \sup_{x\in S} g(x) = M(f,S)+M(g,S)\,. \end{aligned} \tag{I}$$

$$\begin{aligned} m(f+g,S) &= \inf_{x\in S}(f(x)+g(x)) \\ &\ge \inf_{x\in S} f(x) + \inf_{x\in S} g(x) = m(f,S)+m(g,S)\,. \end{aligned} \tag{II}$$

$$M(-f,S) = \sup_{x\in S} -f(x) = -\inf_{x\in S} f(x) = -m(f,S) \tag{III}$$

(1) Let P, Q be partitions of $[a,b]$. Let

$$R = P\cup Q = \{x_0 < x_1 < x_2 < \ldots x_{n-1} < x_n\}\,.$$

Since R is a refinement of P and Q we obtain (by (I)),

$$\begin{aligned} U(f+g) \le U(f+g,R) &= \sum_{k=1}^{n} M(f+g,[x_{k-1},x_k])(x_k - x_{k-1}) \\ &\le \sum_{k=1}^{n} \big(M(f,[x_{k-1},x_k]) + M(g,[x_{k-1},x_k])\big)(x_k - x_{k-1}) \\ &= U(f,R)+U(g,R) \\ &\le U(f,P)+U(g,Q)\,. \end{aligned}$$

Since P was arbitrary this implies that

$$U(f+g) \le U(f) + U(g,Q)\,.$$

Finally, since Q was arbitrary this shows that

$$U(f+g) \le U(f)+U(g)\,.$$

(2) The proof of (2) is similar to the proof of (1) and is therefore omitted.
(3) We must prove that $L(f+g) = U(f+g) = \int_a^b f + \int_a^b g$. It follows immediately from (1) and (2) that

$$\int_a^b f + \int_a^b g = L(f)+L(g) \le L(f+g) \le U(f+g) \le U(f)+U(g) = \int_a^b f + \int_a^b g\,.$$

(4) Let $\mathcal{P}$ denote the family of partitions of $[a,b]$. We divide the proof inte two cases.

Case 1: $\lambda \geq 0$. Let $P = \{x_0 < x_1 < x_2 < \dots x_{n-1} < x_n\}$ be a partition of $[a, b]$. It now follows that

$$\begin{aligned} U(\lambda f, P) &= \sum_{k=1}^{n} M(\lambda f, [x_{k-1}, x_k])(x_k - x_{k-1}) \\ &= \sum_{k=1}^{n} \lambda M(f, [x_{k-1}, x_k])(x_k - x_{k-1}) \\ &= \lambda U(f, P)\,. \end{aligned}$$

Hence,

$$U(\lambda f) = \inf_{P \in \mathcal{P}} U(\lambda f, P) = \inf_{P \in \mathcal{P}} \lambda U(f, P) = \lambda \inf_{P \in \mathcal{P}} U(f, P) = \lambda U(f)\,.$$

Case 2: $\lambda \leq 0$. Let $P = \{x_0 < x_1 < x_2 < \dots x_{n-1} < x_n\}$ be a partition of $[a, b]$. It now follows from (III) that

$$\begin{aligned} U(\lambda f, P) &= \sum_{k=1}^{n} M(-|\lambda| f, [x_{k-1}, x_k])(x_k - x_{k-1}) \\ &= \sum_{k=1}^{n} -|\lambda| m(f, [x_{k-1}, x_k])(x_k - x_{k-1}) \\ &= -|\lambda| L(f, P)\,. \end{aligned}$$

Hence,

$$U(\lambda f) = \inf_{P \in \mathcal{P}} U(\lambda f, P) = \inf_{P \in \mathcal{P}} -|\lambda| L(f, P) = -|\lambda| \sup_{P \in \mathcal{P}} L(f, P) = -|\lambda| L(f) = \lambda L(f)\,.$$

(5) The proof of (5) is similar to the proof of (4) and is therefore omitted.
(6) We must prove that $L(\lambda f) = U(\lambda f) = \lambda \int_a^b f$. This follows immediately from (4) and (6). $\square$

Corollary. *Let $a, b \in \mathbb{R}$ with $a < b$. Let $f, g : [a, b] \to \mathbb{R}$ be integrable functions with $f \leq g$. Then*

$$\int_a^b f \leq \int_a^b g\,.$$

Proof. Since f and g are integrable, the previous theorem shows that $g - f$ is integrable and

$$\int_a^b g - \int_a^b f = \int_a^b (g - f)\,.$$

Since by assumption $g - f \geq 0$, it is easily seen that $\int_a^b (g - f) \geq 0$, and we therefore conclude that $\int_a^b g - \int_a^b f = \int_a^b (g - f) \geq 0$. $\square$

Theorem. *Let $a, b \in \mathbb{R}$ with $a < b$. Let $f : [a, b] \to \mathbb{R}$ be an integrable function. Then $|f|$ is integrable and*

$$\left| \int_a^b f \right| \leq \int_a^b |f| \,.$$

Proof.
We first prove that $|f|$ is integrable, i.e. we must prove that

$$U(|f|) \leq L(|f|) \,.$$

Therefore let $\varepsilon \geq 0$. Since f is integrable there exists (by Theorem ???) a partition $P = \{x_0 < x_1 < x_2 < \dots x_{n-1} < x_n\}$ of $[a, b]$ such that

$$U(f, P) - L(f, P) \leq \varepsilon \,.$$

We now have for all $k = 1, 2, \dots, n$,

$$\begin{aligned} \forall x, y \in [x_{k-1}, x_k] \,:\, |f(x)| - |f(y)| &\leq |f(x) - f(y)| \\ &\leq M(f, [x_{k-1}, x_k]) - m(f, [x_{k-1}, x_k]) \,. \end{aligned}$$

Hence

$$M(|f|, [x_{k-1}, x_k]) - m(|f|, [x_{k-1}, x_k]) \leq M(f, [x_{k-1}, x_k]) - m(f, [x_{k-1}, x_k]) \,.$$

Summing over all k now yields

$$U(|f|, P) - L(|f|, P) \leq U(f, P) - L(f, P) \leq \varepsilon \,.$$

This shows (cf. Theorem ???) that $|f|$ is integrable.

Next we show that $|\int_a^b f| \leq \int_a^b |f|$. Since clearly $-|f| \leq f \leq |f|$ we have by Corollary ???,

$$-\int_a^b |f| \leq \int_a^b -|f| \leq \int_a^b f \leq \int_a^b |f| \,.$$

This clearly implies that $|\int_a^b f| \leq \int_a^b |f|$. $\square$

Theorem. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$ be a bounded function. Let $u_0, u_1, u_2, \dots, u_{n-1}, u_n \in \mathbb{R}$ with $a = u_0 < u_1 < u_2 < \dots u_{n-1} < u_n = b$. Assume that f is integrable over $[u_{k-1}, u_k]$ for all k. Then f is integrable over $[a, b]$ and*

$$\int_a^b f = \sum_{k=1}^n \int_{u_{k-1}}^{u_k} f \,.$$

Proof. We must prove that

$$L(f) = U(f) = \sum_{k=1}^n \int_{u_{k-1}}^{u_k} f \,.$$

Let P_k be a partition of $[u_{k-1}, u_k]$ and put $P = P_1 \cup \cdots \cup P_n$. Since P is a partition of $[a, b]$ we clearly have

$$U(f) \le U(f, P) = \sum_{k=1}^{n} U_{u_{k-1}}^{u_k}(f, P_k)\,.$$

This implies that

$$U(f) \le \sum_{k=1}^{n} U_{u_{k-1}}^{u_k} f = \sum_{k=1}^{n} \int_{u_{k-1}}^{u_k} (f)\,.$$

Similarly we prove that

$$L(f) \ge \sum_{k=1}^{n} L_{u_{k-1}}^{u_k} f = \sum_{k=1}^{n} \int_{u_{k-1}}^{u_k} (f)\,.$$

This completes the proof. □

Theorem. The Mean value Theorem for Integrals. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$ be a continuous (and thus integrable) function. Then there exists $\xi \in [a, b]$ such that*

$$\frac{1}{b-a} \int_a^b f = f(\xi)\,.$$

Proof.
Since f is continuous function defined on a closed and bounded interval, the Extremum Value Theorem for Continuous Functions (Theorem ?????) shows that there exist $u, v \in [a, b]$ such that

$$f(u) \le f(x) \le f(v) \quad \text{for all } x \in [a, b].$$

It therefore follows from Corollary ???? that

$$\frac{1}{b-a} \int_a^b f(x)\,dx \le \frac{1}{b-a} \int_a^b f(v)\,dx = \frac{1}{b-a} f(v) \int_a^b dx = \frac{1}{b-a} f(v)(b-a) = f(v)$$

and similarly

$$\frac{1}{b-a} \int_a^b f(x)\,dx \ge f(u)\,.$$

Hence

$$f(u) \le \frac{1}{b-a} \int_a^b f(x)\,dx \le f(v)\,.$$

The intermediate value theorem for continuous therefore shows that there exists ξ between u and v such that

$$f(\xi) = \frac{1}{b-a} \int_a^b f\,.$$

This completes the proof. □

6.4. The Fundamental Theorem of calculus.

The Fundamental Theorem of Calculus is really two theorems, each expressing that differentiation and integration are inverse operations. The first result says in essence that "the derivative of the integral of a function is the original function", and the second result establishes the reverse: "the integral of the derivative of a function is again the same original function".

Historically, the operations of integration and differentiation were developed to solve seemingly unrelated problems. These problems may be described geometrically as finding the area under a curve and finding the slope of a curve at a point. The discovery of their inverse relationship was one of the important theoretical contributions of Newton and Leibniz in the seventeenth century.

In our discussion of integrals we have defined $\int_a^b f$ when $a \leq b$. It will be convenient now to extend this definition to the case $b \leq a$.

Definition. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$ be an integrable function. We define $\int_b^a f$ by*

$$\int_b^a f = -\int_a^b f\,.$$

The Fundamental Theorem of Calculus. Part I. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$ be a bounded integrable function. Define the function $F : [a, b] \to \mathbb{R}$ by*

$$F(x) = \int_a^x f\,.$$

(1) *There exists $M \geq 0$ such that*

$$|F(x) - F(y)| \leq M|x - y|$$

for all $x, y \in [a, b]$. In particular, it follows that F is uniformly continuous.

(2) *Let $x \in (a, b)$. If f is continuous at x, then F is differentiable at x with*

$$F'(x) = f(x)\,.$$

Proof.
(1) Since f is bounded there exists $M \geq 0$ such that $|f(x)| \leq M$ for all $x \in [a, b]$. Next, let $x, y \in [a, b]$. We may clearly assume that $x \leq y$. We have

$$\begin{aligned} |F(x) - F(y)| &= |F(y) - F(x)| \\ &= \left|\int_a^y f - \int_a^x f\right| = \left|\int_x^y f\right| \\ &\leq \int_x^y |f| \ \leq \int_x^y M = M|y - x|\,. \end{aligned}$$

(2) Let $\varepsilon > 0$. Since f is continuous at x there exists $\delta > 0$ such that

$$\forall t \in [a, b] \,:\, |t - x| \leq \delta \,\Rightarrow\, |f(t) - f(x)| \leq \varepsilon\,.$$

In particular, it follows that if $h \in \mathbb{R}$ with $0 < |h| \le \delta$ and $x + h \in (a, b)$, then

$$|f(t) - f(x)| \le \varepsilon$$

for all t between x and $x + h$.

Next, let $h \in \mathbb{R}$ with $0 < |h| \le \delta$ and $x + h \in (a, b)$. For $h > 0$, we obtain

$$\begin{aligned}
\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| &= \left| \frac{1}{h} \left(\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt \right) - f(x) \right| \\
&= \left| \frac{1}{h} \int_x^{x+h} f(t)\,dt - f(x) \right| \\
&= \left| \frac{1}{h} \int_x^{x+h} f(t)\,dt - \frac{1}{h} \int_x^{x+h} f(x)\,dt \right| \\
&= \left| \frac{1}{h} \int_x^{x+h} (f(t) - f(x))\,dt \right| \\
&\le \frac{1}{|h|} \int_x^{x+h} |f(t) - f(x)|\,dt \\
&\le \frac{1}{|h|} \int_x^{x+h} \varepsilon\,dt \\
&= \varepsilon\,.
\end{aligned}$$

Similarly, for $h < 0$ we obtain

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| \le \varepsilon\,.$$

This completes the proof. □

The Fundamental Theorem of Calculus. Part II. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$ be a function satisfying*

(1) *f is continuous on $[a, b]$;*
(2) *f is differentiable on (a, b); extend f' to $[a, b]$ by defining*

$$f'(a) = f'(b) = 0\,;$$

(3) *f' is bounded on (a, b).*

Then

$$L(f') \le f(b) - f(a) \le U(f')\,.$$

Assume, in addition, that f satisfies

(4) *f' is integrable on $[a, b]$.*

Then

$$\int_a^b f' = f(b) - f(a)\,.$$

Proof.
Let P, Q be partitions of $[a, b]$. Now write

$$R = P \cup Q = \{x_0 < x_1 < \ldots < x_{n-1} < x_n\}.$$

By assumption f is continuous on $[a, b]$ and differentiable on (a, b). Hence, for each k there exists by the Mean Value Theorem a point $\xi_k \in (x_{k-1}, x_k)$ such that

$$f(x_k) - f(x_{k-1}) = (x_k - x_{k-1})f'(\xi_k).$$

Hence

$$\begin{aligned} m(f', [x_{k-1}, x_k])(x_k - x_{k-1}) &\leq f'(\xi_k)(x_k - x_{k-1}) \\ &= f(x_k) - f(x_{k-1}) \\ &= f'(\xi_k)(x_k - x_{k-1}) \\ &\leq M(f', [x_{k-1}, x_k])(x_k - x_{k-1}). \end{aligned}$$

If we add all these inequalities over all subintervals in the partition and note that the middle term "telescopes", we obtain,

$$L(f', R) \leq f(b) - f(a) \leq U(f', R).$$

But since R is a refinement of P and Q we obtain,

$$L(f', P) \leq L(f', R) \leq f(b) - f(a) \leq U(f', R) \leq U(f', Q).$$

Taking supremum over all partitions P and infimum over all partitions Q gives

$$L(f') \leq f(b) - f(a) \leq U(f')$$

In, in addition, f' is assumed to be integrable, this inequality implies that

$$\int_a^b f' = L(f') \leq f(b) - f(a) \leq U(f') = \int_a^b f',$$

whence $\int_a^b f' = f(b) - f(a)$. This completes the proof. $\square$

Example. Define $f : [1, 4] \to \mathbb{R}$ by

$$f(x) = x^2$$

Then f is continuous and thus integrable, and we have

$$\int_1^4 x^2 \, dx = 21$$

Proof. Indeed, since $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) = \frac{1}{3}x^3$ is differentiable with a continuous derivative $F' = f$, it follows from the Fundamental Theorem of Calculus, Part II, that

$$\int_1^4 x^2 \, dx = F(4) - F(1) = \frac{1}{3}(4^3 - 1^3) = 21.$$

This completes the proof. $\square$

Example. Let $a, b \in \mathbb{R}$ with $a < b$ and let $r > 0$. Suppose $f, g : [a, b] \to \mathbb{R}$ are continuous functions and that f and g are differentiable on (a, b). Extend f' and g' to $[a, b]$ by defining $f'(a) = f'(b) = g'(a) = g'(b) = 0$ and assume that f' and g' are integrable on $[a, b]$. Then

$$\int_a^b fg' = f(b)g(b) - f(a)g(a) - \int_a^b f'g\,.$$

Proof. Let $h = fg$. Note, that h is differentiable on (a, b) and the derivative $h' = f'g + gf'$ is integrable on $[a, b]$. (Indeed, f and g are continuous, and thus integrable, on $[a, b]$, and f' and g' are integrable on $[a, b]$, by assumption. Problem 7 on Tut. Sheet no. 8 therefore implies that the products $f'g$ and fg' are integrable on $[a, b]$. This shows that the sum $h' = f'g + gf'$ is integrable on $[a, b]$.) The Fundamental Theorem of Calculus, Part II, therefore implies that

$$\begin{aligned}\int_a^b f'g + \int_a^b fg' &= \int_a^b h' \\ &= h(b) - h(a) \\ &= f(b)g(b) - f(a)g(a)\,.\end{aligned}$$

This completes the proof. $\square$

6.5. Uniform convergence and integration.

Theorem. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $(f_n)_n$ be a sequence of integrable functions defined on $[a,b]$. Let $f : [a,b] \to \mathbb{R}$ be a function on A Assume that*

$$f_n \to f \quad \text{uniformly on } A.$$

Then

(1) *f is integrable.*
(2) *We have*

$$\int_a^b f_n \to \int_a^b f\,.$$

Proof.
(1) Let $\varepsilon > 0$. Since $f_n \to f$ uniformly, there exists N such that

$$\|f_N - f\|_\infty \leq \frac{\varepsilon}{2(b-a)}\,,$$

i.e.

$$f_N(x) - \frac{\varepsilon}{2(b-a)} \leq f(x) \leq f_N(x) + \frac{\varepsilon}{2(b-a)} \quad \text{for all } x \in [a,b]. \tag{1}$$

Let P and Q be partitions of $[a,b]$, and write $R = P \cup Q = \{x_0 < x_1 < \ldots < x_n\}$. It follows from (1) that

$$\begin{aligned} m(f_N, [x_{k-1}, x_k]) - \frac{\varepsilon}{2(b-a)} &\leq m(f, [x_{k-1}, x_k]) \\ &\leq M(f, [x_{k-1}, x_k]) \\ &\leq M(f_N, [x_{k-1}, x_k]) + \frac{\varepsilon}{2(b-a)}\,. \end{aligned}$$

Summing this inequality over k gives

$$\begin{aligned} L(f_N, R) - \frac{\varepsilon}{2} &\leq L(f, R) \\ &\leq L(f) \\ &\leq U(f) \\ &\leq U(f, R) \\ &\leq U(f_N, R) + \frac{\varepsilon}{2}\,. \end{aligned}$$

Finally, since R is a refinement of both P and Q this implies that

$$\begin{aligned} L(f_N, P) - \frac{\varepsilon}{2} &\leq L(f_N, R) - \frac{\varepsilon}{2} \\ &\leq L(f) \\ &\leq U(f) \\ &\leq U(f_N, R) + \frac{\varepsilon}{2} \\ &\leq U(f_N, Q) + \frac{\varepsilon}{2}\,. \end{aligned}$$

Taking infimum over Q and taking supremum over P yields

$$\begin{aligned} L(f_N) - \frac{\varepsilon}{2} &\le L(f) \\ &\le U(f) \\ &\le U(f_N) + \frac{\varepsilon}{2}\,. \end{aligned}$$

Hence
$$U(f) - L(f) \le U(f_N) - L(f_N) + \varepsilon = \varepsilon\,,$$

since f_N is integrable. Finally, since this holds for all $\varepsilon > 0$, we now conclude that $L(f) = U(f)$.

(2) It follows from ?????? that

$$\begin{aligned} \left|\int_a^b f_n - \int_a^b f\right| &= \left|\int_a^b (f_n - f)\right| \\ &\le \int_a^b |f_n - f| \\ &\le \int_a^b \|f_n - f\|_\infty \\ &= (b-a)\|f_n - f\|_\infty \to 0\,. \end{aligned}$$

This completes the proof. □

7.1. Lebesgue Outer Measure and Almost Everewhere

We now define the Lebesgue outer measure and and prove a few of its simplest properties Among other things, we will see that this measure is a function that assigns to each subset of $\mathbb{R}$ a number in $[0, \infty]$ in such a way that each interval is assigned its length. Next, we use the Lebesgue measure to define the notion of "almost everywhere".

We begin with the fundamental definition.

Definition. Lebesgue Measure. *For an interval I we define the length $\ell(I)$ of I by*

$$\ell(I) = \sup(I) - \inf(I)$$

for $I \neq \varnothing$, and

$$\ell(I) = 0$$

for $I = \varnothing$. For a subset E of $\mathbb{R}$, we define the (outer) Lebesgue measure of E by

$$\lambda(E) = \inf \left\{ \sum_i \ell(I_i) \;\middle|\; (I_i)_{i \in \mathbb{N}} \text{ is a family of open intervals with } E \subseteq \bigcup_i I_i \right\}.$$

Henri Lebesgue (June 28, 1875 – July 26, 1941) was a French mathematician most famous for his theory of integration, which was a generalization of the 17th century concept of integration. His theory was published originally in his dissertation Integrale, longueur, aire ("Integral, length, area") at the University of Nancy during 1902.

In the following theorems we give a number of important properties of the Lebesgue measure.

Theorem.

(1) *For* $E \subseteq \mathbb{R}$, *we have* $\lambda(E) \in [0, \infty]$.
(2) $\lambda(\varnothing) = 0$.
(3) *For* $E \subseteq F \subseteq \mathbb{R}$, *we have* $\lambda(E) \leq \lambda(F)$.
(4) *For* $E_1, E_2, \ldots \subseteq \mathbb{R}$, *we have*

$$\lambda\left(\bigcup_n E_n\right) \leq \sum_n \lambda(E_n)\,.$$

(5) *If* $C \subseteq \mathbb{R}$ *is countable, then* $\lambda(C) = 0$.
(6) *If* $I \subseteq \mathbb{R}$ *is an interval, then* $\lambda(I) = \ell(I)$.

Proof.
(1) This statement is clear.
(2) This statement is clear.
(3) This statement is clear.
(4) Let $\varepsilon > 0$. For each n, we can choose a family $(I_{n,i})_{i \in \mathbb{N}}$ such that

$$\sum_i \ell(I_{n,i}) \leq \lambda(E_n) + \frac{\varepsilon}{2^n}\,.$$

Since $\cup_n E_n \subseteq \cup_n \cup_i I_{n,i}$, it now follows from the definition of $\lambda(\cup_n E_n)$ that

$$\begin{aligned}
\lambda\left(\bigcup_n E_n\right) &\leq \sum_n \sum_i \ell(I_{n,i}) \\
&\leq \sum_n \left(\lambda(E_n) + \frac{\varepsilon}{2^n}\right) \\
&\leq \sum_n \lambda(E_n) + \sum_n \frac{\varepsilon}{2^n} \\
&\leq \sum_n \lambda(E_n) + \varepsilon\,.
\end{aligned}$$

Letting $\varepsilon \searrow 0$, now gives the desired result.
(5) Let $C = \{x_1, x_n, \ldots\}$ be a countable subset of $\mathbb{R}$. Since it is easily seen that $\lambda(\{x_n\}) = 0$ for all n, it now follows from Part (4) that $\lambda(C) \leq \sum_n \lambda(\{x_n\}) = \sum_n 0 = 0$.
(6) "$\leq$" Let $\varepsilon > 0$ Write $a =\in (I)$ and $b = \sup(I)$, and put $J = (a - \varepsilon, b + \varepsilon)$, Then J is an open interval and $I \subseteq J$, Hence

$$\lambda(I) \leq \ell(J) + \ell(\varnothing) + \ell(\varnothing) + \ell(\varnothing) + \cdots = (b + \varepsilon) - (a - \varepsilon) = \ell(I) + 2\varepsilon\,.$$

Letting $\varepsilon \searrow 0$, now gives the desired result.
"$\geq$" The proof of this inequality use Heine-Borel's theorem on compact sets and is therefore omitted. $\square$

Theorem. *Let $E \subseteq \mathbb{R}$ and $\varepsilon > 0$. The there is an open set U such that $E \subseteq U$ and*

$$\lambda(U) \leq \lambda(E) + \varepsilon\,.$$

Proof.
It follows from the definition of the Lebesgue measure $\lambda(E)$ that there is a countable familyy $(I_i)_i$ of open intervals I_i such that $E \subseteq \cup_i I_i$ and

$$\sum_i \ell(I_i) \leq \lambda(E) + \varepsilon\,.$$

Now put $U = \cup_i I_i$. Then U is open with $E \subseteq \cup_i I_i = U$ and Parts (4) and (6) of the previous theorem implies that

$$\begin{aligned} \lambda(U) &= \lambda\left(\bigcup_i I_i\right) \\ &\leq \sum_i \lambda(I_i) \\ &= \sum_i \ell(I_i) \\ &\leq \lambda(E) + \varepsilon\,. \end{aligned}$$

This completes the proof. $\square$

Remark. Is it possible to construct disjoint subsets A and B of $\mathbb{R}$ such that

$$\lambda(A \cup B) < \lambda(A) + \lambda(B)\,.$$

This can be proven using the Axiom of Choice . However, it is true that there is a "very" large family $\mathcal{B}$ of subsets of $\mathbb{R}$ (namely, the family of Borel sets) having the property that if $(E_n)_n$ is a family of pairwise disjoint sets with $E_n \in \mathcal{B}$ for all n, then

$$\lambda\left(\bigcup_n E_n\right) = \sum_n \lambda(E_n)\,.$$

This matter is examined in detail in MT5826.

Using the Lebesgue measure we can now define the notion of "almost everywhere".

Definition. Almost Everywhere. *Let E be a subset of $\mathbb{R}$. For each $x \in E$ let $P(x)$ be a statement about x. We will say that a almost all x in E satisfies the property P if the set*

$$\left\{x \in E \,\middle|\, P(x) \textit{ is not a true statement}\right\}.$$

is has zero Lebesgue measure, i.e. if

$$\lambda\Big(\left\{x \in E \,\middle|\, P(x) \textit{ is not a true statement}\right\}\Big) = 0\,.$$

Example. Almost all real numbers are different from $\sqrt{2}$. (Why?)

Example. Almost all real numbers are not rational. (Why?)

7.2. Dini Derivatives

A great many elementary facts about derivatives was proven in Chapter 4 without the notion of Dini derivatives. However, more sophisticated facts, e.g. that every monotone function is differentiable almost everywhere (proved in the next Section), require some such notion. Thus is is convenient to begin with these derivatives in order to avoid needless repetition in definition. First, another definition.

Definition. Limit inferior and limit superior.. *Let $\varphi : (a,b) \to \mathbb{R}$ be given and $x \in (a,b)$. We define the right limit inferior and the right limit superior of φ at x by*

$$\liminf_{h \searrow x} \varphi(h) = \sup_{t>x} \inf_{x<h<t} \varphi(h)$$

and

$$\limsup_{h \searrow x} \varphi(h) = \inf_{t>x} \sup_{x<h<t} \varphi(h)\,.$$

Similarly, we define the left limit inferior and the left limit superior of φ at x by

$$\liminf_{h \nearrow x} \varphi(h) = \sup_{t<x} \inf_{t<h<x} \varphi(h)$$

and

$$\limsup_{h \nearrow x} \varphi(h) = \inf_{t<x} \sup_{t<h<x} \varphi(h)\,.$$

We have the following result justifying the above terminology.

Theorem. *Let $\varphi : (a,b) \to \mathbb{R}$ be given and $x \in (a,b)$. Also, let $y \in \mathbb{R}$. Then the following statements are equivalent.*

(1) *We have*

$$\varphi(t) \to y \quad \text{as } t \to x.$$

(2) *We have*

$$\liminf_{h \searrow x} \varphi(h) = \limsup_{h \searrow x} \varphi(h) = \liminf_{h \nearrow x} \varphi(h) = \limsup_{h \nearrow x} \varphi(h) = y\,.$$

Proof.
"(1) $\Rightarrow$ (2)" Let $\varepsilon > 0$.

Since $y - \varepsilon < y = \liminf_{h \searrow x} \varphi(h) = \sup_{t>x} \inf_{x<h<t} \varphi(h)$, we can find $t_+ > x$ such that

$$y - \varepsilon < \inf_{x<h<t_+} \varphi(h)\,,$$

whence

$$\forall h \in \mathbb{R} \,:\, x < h < t_+ \;\Rightarrow\; y - \varepsilon \le \varphi(h)\,. \tag{1}$$

Since $y + \varepsilon > y = \limsup_{h \searrow x} \varphi(h) = \inf_{t>x} \sup_{x<h<t} \varphi(h)$, we can find $t^+ > x$ such that

$$y + \varepsilon > \sup_{x<h<t^+} \varphi(h)\,,$$

whence

$$\forall h \in \mathbb{R} \,:\, x < h < t^+ \;\Rightarrow\; y + \varepsilon \ge \varphi(h)\,. \tag{2}$$

Since $y - \varepsilon < y = \liminf_{h \nearrow x} \varphi(h) = \sup_{t<x} \inf_{t<h<x} \varphi(h)$, we can find $t_- < x$ such that

$$y - \varepsilon < \inf_{t_- < h < x} \varphi(h)\,,$$

whence

$$\forall h \in \mathbb{R} \;:\; t_- < h < x \;\Rightarrow\; y - \varepsilon \leq \varphi(h)\,. \tag{3}$$

Since $y + \varepsilon > y = \limsup_{h \nearrow x} \varphi(h) = \inf_{t<x} \sup_{t<h<x} \varphi(h)$, we can find $t^- < x$ such that

$$y + \varepsilon > \sup_{t^- < h < x} \varphi(h)\,,$$

whence

$$\forall h \in \mathbb{R} \;:\; t^- < h < x \;\Rightarrow\; y + \varepsilon \geq \varphi(h)\,. \tag{4}$$

Now put $\delta = \min(t_+ - x, t^+ - x, x - t_-, x - t^-) > 0$. It now follows immediately from (1), (2),. (3) and (4) that if $h \in (x - \delta, x + \delta)$, then

$$\varphi(h) \in (y - \varepsilon, y + \varepsilon)\,.$$

This proves the statement.
"(2) $\Rightarrow$ (1)" The proof of this implication is left as an exercise for the reader. $\square$

We can now define the Dini derivatives.

Dini derivatives.. *Let $f : (a, b) \to \mathbb{R}$ be given and $x \in (a, b)$. We define the right lower derivative and the right upper derivative of f at x by*

$$(D_+ f)(x) = \liminf_{h \searrow x} \frac{f(x+h) - f(x))}{h}$$

and

$$(D^+ f)(x) = \limsup_{h \searrow x} \frac{f(x+h) - f(x))}{h} .$$

Similarly, we define the left lower derivative and the left upper derivative of f at x by

$$(D_- f)(x) = \liminf_{h \nearrow x} \frac{f(x+h) - f(x))}{h}$$

and

$$(D^- f)(x) = \limsup_{h \nearrow x} \frac{f(x+h) - f(x))}{h} .$$

These four derivatives are known as the Dini derivatives of f at x.

Ulisse Dini (14 November 1845 – 28 October 1918) was an Italian mathematician and politician, born in Pisa. He is known for his contribution to real analysis, partly collected in his book "Fondamenti per la teorica delle funzioni di variabili reali".

Using the previous theorem we can now characterise differentiability in terms of the Dini derivatives.

Theorem. *Let $\varphi : (a,b) \to \mathbb{R}$ be given and $x \in (a,b)$. Also, let $y \in \mathbb{R}$. Then the following statements are equivalent.*

(1) *The function f is differentiable at x with*

$$f'(x) = y$$

(2) *We have*

$$(D_+ f)(x) = (D^+ f)(x) = (D_- f)(x) = (D^- f)(x) = y\,.$$

Proof.
This follows immediately from the definitions and the previous theorem. □

7.3. Differentiation of Monotone Functions.

The main purpose of this section is to prove that a monotone function is differentiable almost everywhere. However, we start with a definition and a technical covering lemma.

Definition. Vitali Cover. *Let $E \subseteq \mathbb{R}$. A family $\mathcal{V}$ of closed, bounded and non-trivial intervals is called a Vitaly cover of E if for all $x \in E$ and all $\varepsilon > 0$, there is $I \in \mathcal{V}$ such that*

$$x \in I \, ,$$

and

$$\ell(I) \leq \varepsilon \, .$$

Vitali's Covering Theorem. *Let $E \subseteq \mathbb{R}$ be a bounded set and let $\mathcal{V}$ be a Vitali cover of E.*

Then there is a subfamily $\mathcal{I}$ of $\mathcal{V}$ such that the following conditions are satisfied.

(1) *The family $\mathcal{I}$ is finite.*
(2) *The intervals in $\mathcal{I}$ are pairwise disjoint.*
(3) *We have*

$$\lambda\left(E \setminus \bigcup_{I \in \mathcal{I}} I \right) \leq \varepsilon \, .$$

In particular, this implies that

$$\lambda(E) - \varepsilon \leq \lambda\left(E \cap \bigcup_{I \in \mathcal{I}} I \right) .$$

Proof.
Recall, that if I is an interval, then we write $\ell(I) = \sup(I) - \inf(I)$. Choose $R > 0$ such that $E \subseteq (-R, R)$. We now choose a sequence $(I_n)_n$ of intervals from $\mathcal{V}$ with $I_n \subseteq (-R, R)$ inductively as follows.

The start of the induction. Let I_1 be any interval from $\mathcal{V}$ with $I_1 \subseteq (-R, R)$; this choice of I_1 is possible since $\mathcal{V}$ is a Vitali cover.

The inductive step. Assume that the intervals $I_1, \ldots, I_n$ have been chosen. Let

$$l_n = \sup\left\{ \ell(I) \,\middle|\, I \in \mathcal{V} \, , \, I \subseteq (-R, R) \, , \;\; I \cap I_k = \varnothing \text{ for all } k \leq n \right\} .$$

We can now choose $I_{n+1} \in \mathcal{V}$ such that $I_{n+1} \subseteq (-R, R)$ with $I_{n+1} \cap I_k = \varnothing$ for all $k \leq n$ and

$$\ell(I_{n+1}) > \frac{1}{2} l_n \, .$$

Then the sequence $(I_n)_n$ consists of pairwise disjoint intervals from $\mathcal{V}$ with $I_n \subseteq (-R, R)$. In particular $\sum_n \ell(I_n) \leq \ell((-R,R)) = 2R < \infty$, and we can therefore choose an integer N such that

$$\sum_{n>N} \ell(I_n) \leq \frac{\varepsilon}{5}\,.$$

Now put $\mathcal{I} = \{I_1, \ldots, I_N\}$. Then $\mathcal{I}$ is a finite subfamily of $\mathcal{V}$ consisting of pairwise disjoint intervals. Next, we show that

$$\lambda\left(E \setminus \bigcup_{I \in \mathcal{I}} I\right) \leq \varepsilon\,.$$

For each n, let I_n be the closed interval with the same midpoint as I_n and $\ell(I_n) = 5\ell(I_n)$. We now prove that

$$E \setminus \bigcup_{I \in \mathcal{I}} I = E \setminus \bigcup_{n \leq N} I_n \subseteq \bigcup_{n>N} I_n\,. \tag{1}$$

Indeed, let $x \in E \setminus \cup_{n \leq N} I_n$. Then $x \in E$ and $x \notin \cup_{n \leq N} I_n$, and since $\mathcal{V}$ is a Vitali cover of E, we can therefore choose an interval $I \in \mathcal{V}$ with $I \subseteq (-R, R)$ such that $x \in I$ and $I \cap I_n = \varnothing$ for $n \leq N$. Next, observe that there is an integer $n_x > N$ such that

$$I \cap I_{n_x} \neq \varnothing\,. \tag{2}$$

(Otherwise $I \cap I_n = \varnothing$ for all $n > N$, whence $I \cap I_n = \varnothing$ for all n, and so

$$\ell(I_n) \leq l_n \quad \text{for all } n.$$

In particular, since $l_n \leq 2\ell(I_{n+1})$, this implies that

$$\ell(I) \leq \inf_n l_n \leq 2 \inf_n \ell(I_{n+1})\,.$$

However, since $\sum_n \ell(I_n) < \infty$, we conclude that $\inf_n \ell(I_{n+1}) = 0$, and the previous inequality therefore implies that $\ell(I) \leq \inf_n l_n \leq \inf_n \ell(I_{n+1}) = 0$, contradicting the fact that I is a non-trivial interval. This completes the proof of (2).) In particular, since $I \cap I_{n_x} \neq \varnothing$, we deduce that $x \in I_{n_x}$. This completes the proof of (1).

It now follows from (1) that

$$\lambda\left(E \setminus \bigcup_{I \in \mathcal{I}} I\right) \leq \lambda\left(\bigcup_{n>N} I_n\right) \leq \sum_{n>N} \lambda(I_n) = 5 \sum_{n>N} \ell(I_n) \leq \varepsilon\,.$$

This concludes the proof of Vitali's covering theorem. $\square$

Giuseppe Vitali (26 August 1875 – 29 February 1932) was an Italian mathematician who worked in several branches of mathematical analysis. He gives his name to several entities in mathematics, most notably the Vitali set with which he was the first to give an example of a non-measurable subset of real numbers.

Before showing the a monotone function is differentiable almost everywhere, we start by proving the following technical proposition.

Proposition. *Let $f : [a, b] \to \mathbb{R}$ be monotone.*

(1) *For $u, v \in \mathbb{R}$ with $u < v$, we have*

$$\lambda\left(\left\{x \in [a,b] \,\middle|\, (D_- f)(x) < u < v < (D^+ f)(x)\right\}\right) = 0\,.$$

(2) *For $u, v \in \mathbb{R}$ with $u < v$, we have*

$$\lambda\left(\left\{x \in [a,b] \,\middle|\, (D_+ f)(x) < u < v < (D^- f)(x)\right\}\right) = 0\,.$$

Proof.
Without loss of generality we may assume that f is increasing.

(1) For brevity write

$$E = \left\{x \in [a,b] \,\middle|\, (D_- f)(x) < u < v < (D^+ f)(x)\right\}.$$

Let $\varepsilon > 0$.

First note that we can find an open set U with $E \subseteq U$ such that

$$\lambda(U) \leq \lambda(E) + \varepsilon\,. \tag{1}$$

Fix $x \in E$. Since $x \in E$, we have $(D_- f)(x) < u$, and we can therefore find a sequence $(\delta_{x,n})_n$ with $\delta_{x,n} > 0$ such that

$$\delta_{x,n} \searrow 0\,,$$

$$\frac{f(x) - f(x - \delta_{x,n})}{\delta_{x,n}} < u\,, \tag{2}$$

$$[x - \delta_{x,n}, x] \subseteq U\,. \tag{3}$$

Now put

$$\mathcal{V} = \Big\{ [x - \delta_{x,n}, x] \,\Big|\, x \in E\,, n \in \mathbb{N} \Big\}\,.$$

It is not difficult to see that $\mathcal{V}$ is a Vitali cover of E, and it therefore follows from Vitali's covering theorem that there is a finite subfamily $\mathcal{I} = \big(I_i = [x_i - \delta_{x_i,n_i}, x_i] \big)_{i=1}^N$ of $\mathcal{V}$ consisting of pairwise disjoint intervals such that

$$\lambda(E) - \varepsilon \le \lambda\left(E \cap \bigcup_i I_i \right). \tag{4}$$

Next, note that

$$\begin{aligned}
\sum_i \big(f(x_i) - f(x_i - \delta_{x_i,n_i}) \big) &\le u \sum_i \delta_{x_i n_i} && \text{[by (2)]} \\
&= u\,\lambda\left(\bigcup_i I_i \right) \\
&\le u\,\lambda(U) && \text{[by (3)]} \\
&\le u(\lambda(E) + \varepsilon) && \text{[by (1)]}\,.
\end{aligned} \tag{5}$$

For i, write

$$E_i = E \cap \mathring{I}_i$$

Fix $x \in E_i$. Since $v < (D^+ f)(x)$ and $x \in \mathring{I}_i$, we can find a sequence and we can therefore find a sequence $(\gamma_{x,n})_n$ with $\gamma_{x,n} > 0$ such that

$$\gamma_{x,n} \searrow 0\,,$$

$$v < \frac{f(x+\gamma_{x,n}) - f(x)}{\gamma_{x,n}}\,, \tag{6}$$

$$[x, x+\gamma_{x,n}] \subseteq I_i\,. \tag{7}$$

Now put

$$\mathcal{V}_i = \Big\{ [x, x+\gamma_{x,n}] \,\Big|\, x \in E_i\,, n \in \mathbb{N} \Big\}\,.$$

It is not difficult to see that $\mathcal{V}_i$ is a Vitali cover of E_i, and it therefore follows from Vitali's covering theorem that there is a finite subfamily $\mathcal{I}_i = \big(I_{i,j} = [x_{i,j}, x_{i,j} + \delta_{x_{i,j},n_{i,j}}] \big)_{j=1}^{N_i}$ of $\mathcal{V}_i$ consisting of pairwise disjoint intervals such that

$$\lambda(E_i) - \tfrac{\varepsilon}{N} \le \lambda\left(E_i \cap \bigcup_j I_{i,j} \right). \tag{8}$$

Next, note that since f is increasing and $I_{i,j} = [x_{i,j}, x_{i,j} + \delta_{x_{i,j},n_{i,j}}] \subseteq I_i$ (by (7)) and the intervals $(I_{i,j})_j$ are pairwise disjoint, we have

$$f(x_i) - f(x_i - \delta_{x_i,n_i}) \ge \sum_j \big(f(x_{i,j} + \gamma_{x_{i,j},n_{i,j}}) - f(x_{i,j}) \big)$$

for all i. Hence

$$
\begin{aligned}
\sum_i \big(f(x_i) - f(x_i - \delta_{x_i,n_i}) \big) &= \sum_i \sum_j \big(f(x_{i,j} + \gamma_{x_{i,j},n_{i,j}}) - f(x_{i,j}) \big) \\
&\geq v \sum_i \sum_j \gamma_{x_{i,j} n_{i,j}} && \text{[by (6)]} \\
&= v \sum_i \lambda\left(\bigcup_j I_{i,j} \right) \\
&\geq v \sum_i \lambda\left(E_i \cap \bigcup_j I_{i,j} \right) \\
&\geq v \sum_i (\lambda(E_i) - \tfrac{\varepsilon}{N}) && \text{[by (8)]} \\
&= v \sum_i (\lambda(E \cap \mathring{I}_i) - \tfrac{\varepsilon}{N}) \\
&= v \sum_i (\lambda(E \cap I_i) - \tfrac{\varepsilon}{N}) \\
&= v \left(\sum_i \lambda(E \cap I_i) - N \tfrac{\varepsilon}{N} \right) \\
&\geq v \left(\lambda\left(\bigcup_i (E \cap I_i) \right) - \varepsilon \right) \\
&= v \left(\lambda\left(E \cap \bigcup_i I_i \right) - \varepsilon \right) \\
&\geq v\, (\lambda(E) - \varepsilon - \varepsilon) && \text{[by (4)]} \\
&= v\, (\lambda(E) - 2\varepsilon)\,. && (9)
\end{aligned}
$$

Combining (5) and (9) we now conclude that

$$v\,(\lambda(E) - 2\varepsilon) \leq \sum_i \big(f(x_i) - f(x_i - \delta_{x_i,n_i}) \big) \leq u\,(\lambda(E) + \varepsilon)$$

for all $\varepsilon > 0$, i.e.

$$v\,(\lambda(E) - 2\varepsilon) \leq u\,(\lambda(E) + \varepsilon)$$

for all $\varepsilon > 0$. Letting $\varepsilon \searrow 0$, now gives

$$v\,\lambda(E) \leq u\,\lambda(E)\,,$$

and since $u < v$, we therefore deduce that $\lambda(E) = 0$.
(2) The proof of this statement is similar to the proof of the statement in (1) and is therefore omitted. $\square$

Lebesgue's Differentiation Theorem for Monotone Functions. *Let $f : [a,b] \to \mathbb{R}$ be monotone. Then f is differentiable almost everywhere.*

Proof.
We must prove that

$$(D_- f)(x) = (D^- f)(x) = (D_+ f)(x) = (D^+ f)(x) \quad \text{for almost all } x.$$

It follows from the previous proposition that

$$\begin{aligned}
&\lambda\Big(\Big\{x \in [a,b] \,\Big|\, (D_- f)(x) < (D^+ f)(x)\Big\}\Big) \\
&\qquad = \lambda\left(\bigcup_{\substack{u,v\in\mathbb{Q}\\ u<v}} \Big\{x \in [a,b] \,\Big|\, (D_- f)(x) < u < v < (D^+ f)(x)\Big\}\right) \\
&\qquad \le \sum_{\substack{u,v\in\mathbb{Q}\\ u<v}} \lambda\Big(\Big\{x \in [a,b] \,\Big|\, (D_- f)(x) < u < v < (D^+ f)(x)\Big\}\Big) \\
&\qquad \le \sum_{\substack{u,v\in\mathbb{Q}\\ u<v}} 0 \\
&\qquad = 0\,,
\end{aligned}$$

and

$$\begin{aligned}
&\lambda\Big(\Big\{x \in [a,b] \,\Big|\, (D_+ f)(x) < (D^- f)(x)\Big\}\Big) \\
&\qquad = \lambda\left(\bigcup_{\substack{u,v\in\mathbb{Q}\\ u<v}} \Big\{x \in [a,b] \,\Big|\, (D_+ f)(x) < u < v < (D^- f)(x)\Big\}\right) \\
&\qquad \le \sum_{\substack{u,v\in\mathbb{Q}\\ u<v}} \lambda\Big(\Big\{x \in [a,b] \,\Big|\, (D_+ f)(x) < u < v < (D^- f)(x)\Big\}\Big) \\
&\qquad \le \sum_{\substack{u,v\in\mathbb{Q}\\ u<v}} 0 \\
&\qquad = 0\,.
\end{aligned}$$

It follows from this that

$$(D^+ f)(x) \le (D_- f)(x) \quad \text{for almost all } x, \tag{1}$$

$$(D^- f)(x) \le (D_+ f)(x) \quad \text{for almost all } x. \tag{2}$$

Hence

$$\begin{aligned}
(D^+ f)(x) &\le (D_- f)(x) && \text{[by (1)]} \\
&\le (D^- f)(x) \\
&\le (D_+ f)(x) && \text{[by (2)]} \\
&\le (D^+ f)(x) \quad \text{for almost all } x,
\end{aligned}$$

and so

$$(D_- f)(x) = (D^- f)(x) = (D_+ f)(x) = (D^+ f)(x) \quad \text{for almost all } x.$$

This completes the proof. $\square$

7.4. Functions of Bounded Variation.

We will now introduce a more general class of function, namely, functions of bounded variation, and extend the results from the previous section to this broader class of function.

Definition. Bounded Variation.. *Let $a, b \subseteq \mathbb{R}$ with $a < b$. A partition P of $[a, b]$is a finite ordered set*

$$P = \Big\{ a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b \Big\}.$$

If $f : [a, b] \to \mathbb{R}$ is a function and $P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ is a partition of $[a, b]$, the we definition the variation of f on $[a, b]$ with respect to the partition P by

$$V_a^b(f, P) = \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|.$$

We define the total variation of f on on $[a, b]$ by

$$V_a^b(f) = \sup \Big\{ V_a^b(f, P) \Big| P \text{ is a partition of } [a, b] \Big\}.$$

If $V_a^b(f) < \infty$, then we will say that f has bounded variation ohm $[a, b]$, and we will write

$$BV([a, b]) = \Big\{ f : [a, b] \to \mathbb{R} \Big| V_a^b(f) < \infty \Big\}$$

for the family of function on $[a, b]$ with bounded variation.

We now consider some examples.

Example. If $f : [a, b] \to \mathbb{R}$ is monotone, then

$$V_a^b(f) = |f(b) - f(ba)|.$$

In particular, $f \in BV([a, b])$.
Proof. We assume that f is increasing. If $P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ is a partition of $[a, b]$, then

$$V_a^b(f, P) = \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| = \sum_{k=1}^{n} (f(x_k) - f(x_{k-1})) = f(b) - f(a).$$

Finally, taking supremum over all partitions P of $[a, b]$ gives

$$V_a^b(f) \leq f(b) - f(a).$$

The proof for decreasing functions is similar and is left as an exercise for the reader.

Example. If $f : [a, b] \to \mathbb{R}$ is Lipschitz, i.e. if there is a constant M such that

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

for all $x, y \in [a, b]$, then

$$V_a^b(f) = M(b - a)\,.$$

In particular, $f \in BV([a, b])$.
Proof. If $P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ is a partition of $[a, b]$, then

$$V_a^b(f, P) = \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \leq \sum_{k=1}^{n} M(x_k - x_{k-1}) = M(b - a)\,.$$

Finally, taking supremum over all partitions P of $[a, b]$ gives

$$V_a^b(f) \leq M(b - a)\,.$$

This completes the proof.

Example. Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} x\sin(\frac{1}{x}) & \text{for } x \in (0, 1]; \\ 0 \quad \text{for } x = 0. \end{cases}$$

Then f is continuous and

$$V_0^1(f) = \infty\,.$$

Proof. The reader is inviter to prove this himself/herself. This completes the proof.

We now list some of the basic properties of the total variation.

Proposition. *Let $a, b, u \subseteq \mathbb{R}$ with $a \leq u \leq b$ and let $\lambda \in \mathbb{R}$. Let $f, g : [a, b] \to \mathbb{R}$ be functions.*

(1) *We have*

$$V_a^b(f+g) \leq V_a^b(f) + V_a^b(g) .$$

(2) *We have*

$$V_a^b(f) = V_a^u(f) + V_u^b(f) .$$

(3) *We have*

$$V_a^b(\lambda f) = |\lambda| \, V_a^u(f) .$$

Proof.
(1) Let $P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ be a partition of $[a, b]$. Then we have

$$\begin{aligned}
V_a^b(f+g, P) &= \sum_{k=1}^{n} |(f(x_k) + g(x_k)) - (f(x_{k-1}) + g(x_{k-1}))| \\
&\leq \sum_{k=1}^{n} \Big(|f(x_k) - f(x_{k-1})| + |g(x_k) - g(x_{k-1})| \Big) \\
&= \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| + \sum_{k=1}^{n} |g(x_k) - g(x_{k-1})| \\
&= V_a^b(f, P) + V_a^b(g, P) \\
&\leq V_a^b(f) + V_a^b(g) .
\end{aligned}$$

Finally, taking supremum over all partitions P of $[a, b]$ gives

$$V_a^b(f+g) \leq V_a^b(f) + V_a^b(g) .$$

(2) "$\leq$" Let $P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ be a partition of $[a, b]$. Next choose m such that $x_{m-1} \leq u \leq x_m$ and put

$$\begin{aligned}
Q &= \Big\{ a = x_0 < x_1 < \ldots < x_{m-1} < x_m = u \Big\} , \\
R &= \Big\{ u = x_m < x_{m+1} < \ldots < x_{n-1} < x_n = b \Big\} .
\end{aligned}$$

Then Q is a partition of $[a,u]$ and R is a partition of $[u,b]$. Hence

$$\begin{aligned}
V_a^b(f,P) &= \sum_{k=1}^{n} |f(x_k)-f(x_{k-1})| \\
&= \sum_{k=1}^{m-1} |f(x_k)-f(x_{k-1})| + |f(x_m)-f(x_{m-1})| + \sum_{k=m+1}^{n} |f(x_k)-f(x_{k-1})| \\
&\le \sum_{k=1}^{m-1} |f(x_k)-f(x_{k-1})| + |f(u)-f(x_{m-1})| \\
&\qquad\qquad + |f(x_m)-f(u)| + \sum_{k=m+1}^{n} |f(x_k)-f(x_{k-1})| \\
&= V_a^u(f,Q) + V_u^b(f,R) \\
&= V_a^u(f) + V_u^b(f)\,.
\end{aligned}$$

Finally, taking supremum over all partitions P of $[a,b]$ gives

$$V_a^b(f) \le V_a^u(f) + V_u^b(f)\,.$$

"$\geq$" Let $\varepsilon > 0$. We can now choose a partition Q of $[a,u]$ such that

$$V_a^u(f) - \varepsilon \le V_a^u(f,Q)\,,$$

and we can now choose a partition R of $[u,b]$ such that

$$V_u^b(f) - \varepsilon \le V_u^b(f,R)\,,$$

Hence

$$\begin{aligned}
V_a^b(f) &\ge V_a^b(f, Q\cup R) \\
&= V_a^u(f,Q) + V_u^b(f,R) \\
&\ge V_a^u(f) - \varepsilon + V_u^b(f) - \varepsilon\,.
\end{aligned}$$

Finally, letting $\varepsilon \searrow 0$ gives the desired result.
(3) The proof of this statement is left as an exercise for the reader. $\square$

Functions of bounded variation has a remarkably simple representation in terms of monotone functions as the next result shows.

Jordan's Decomposition Theorem. *Let $f : [a, b] \to \mathbb{R}$ be a function. Then the following two statements are equivalent.*

(1) $f \in BV([a, b])$,
(2) *There are increasing functions $g, h; [a, b] \to \mathbb{R}$ such that*

$$f = g - h\,.$$

Proof.
(1) $\Rightarrow$ (2) Since $V_a^b(f) < \infty$, we can define $g, h; [a, b] \to \mathbb{R}$ by

$$\begin{aligned} g(x) &= V_a^x(f)\,, \\ h(x) &= V_a^x(f) - f(x)\,, \end{aligned}$$

for $x \in [a, b]$.

It is clear that $f = g - h$.

We now prove that h is increasing. Indeed, for $x, y \in [a, b]$ with $x < y$ we have (using the previous proposition)

$$g(x) = V_a^x(f) \le V_a^x(f) + V_x^y(f) = V_a^y(f) = g(y)\,.$$

This proves that g is increasing.

We now prove that h is increasing. Indeed, for $x, y \in [a, b]$ with $x < y$ we have (using the previous proposition)

$$\begin{aligned} h(y) - h(x) &= (V_a^y(f) - f(y)) - (V_a^x(f) - f(x)) \\ &= (V_a^y(f) - V_a^x(f)) - (f(y) - f(x)) \\ &= V_x^y(f) - (f(y) - f(x)) \\ &\ge V_x^y(f) - |f(y) - f(x)| \\ &\ge 0\,. \end{aligned}$$

This proves that h is increasing.

(2) $\Rightarrow$ (1) Using the previous proposition and the first example following the definition of bounded variation, we conclude that

$$\begin{aligned} V_a^b(f) &= V_a^b(g - h) \\ &\le V_a^b(g) + V_a^b(-h) \\ &= V_a^b(g) + V_a^b(-h) \\ &= V_a^b(g) + V_a^b(h) \\ &< \infty\,. \end{aligned}$$

This completes the proof. $\square$

Using the results in Section 7.3 and Jordan's Decomposition Theorem, we can now discuss the differentiability properties of functions of bounded variation.

Lebesgue's Differentiation Theorem for Functions of Bounded Variation. *Let $f : [a, b] \to \mathbb{R}$ be of bounded variation. Then f is differentiable almost everywhere.*

Proof.
The follows immediately from Jordan's Decomposition Theorem and Lebesgue's Differentiation Theorem for Monotone Functions. □

We now consider several applications of Lebesgue's differentiation theorem for functions of bounded variation. We first prove Lebesgue's Differentiation Theorem for Integrals.

Lebesgue's Differentiation Theorem for Integrals. *Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable.. Define $F : (a, b) \to \mathbb{R}$ by*

$$F(x) = \int_a^x f(t)\,dt\,.$$

Then F is differentiable almost everywhere.

Proof.
Using Lebesgue's differentiation theorem for functions of bounded variation it suffices to show that F is of bounded variation. i.e. it suffices to show that

$$V_a^b(F) < \infty\,.$$

Below we show that

$$V_a^b(F) < \infty\,.$$

Since f is Riemann integrable, the function f is bounded. Hence we can find a constant M such that

$$|f| \le M\,.$$

Let $P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ be a partition of $[a, b]$. Then

$$\begin{aligned}
V_a^b(F) &= \sum_{k=1}^{n} |F(x_k) - F(x_{k-1})| \\
&= \sum_{k=1}^{n} \left| \int_a^{x_k} f(t)\,dt - \int_a^{x_{k-1}} f(t)\,dt \right| \\
&= \sum_{k=1}^{n} \left| \int_{x_{k-1}}^{x_k} f(t)\,dt \right| \\
&\le \sum_{k=1}^{n} \int_{x_{k-1}}^{x_k} |f(t)|\,dt \\
&= \int_a^b |f(t)|\,dt \\
&\le \int_a^b M\,dt \\
&= M(b-a)\,.
\end{aligned}$$

Finally, taking supremum over all partitions P of $[a,b]$ now gives $V_a^b(F) \leq M(b-a) < \infty$. $\square$

Next, we prove Lebesgue's theorem on Lebesgue points. We start by defining the notion of Lebesgue points.

Definition. Lebesgue point. *Let $f : [a,b] \to \mathbb{R}$ be Riemann integrable with $f \geq 0$. A point $x \in (a,b)$ is called a Lebesgue point of f, if the following limit exists, namely*

$$\lim_{r \searrow 0} \frac{1}{2r} \int_{x-r}^{x+r} f(t)\,dt\,.$$

We now prove Lebesgue's theorem on Lebesgue points.

Lebesgue's Theorem on Lebesgue points. *Let $f : [a,b] \to \mathbb{R}$ be Riemann integrable. Then almost all points in (a,b) are Lebesgue points of f.*

Proof.
Define $F : (a,b) \to \mathbb{R}$ by

$$F(x) = \int_a^x f(t)\,dt\,,$$

and note that

$$\begin{aligned}\frac{1}{2r}\int_{x-r}^{x+r} f(t)\,dt &= \frac{1}{2r}(F(x+r) - F(x-r)) \\ &= \frac{1}{2}\left(\frac{F(x+r)-F(x)}{r} + \frac{F(x)-F(x-r)}{r}\right). \qquad (1)\end{aligned}$$

Since the previous theorem guarantees that F is differentiable almost everywhere, the desired statement now follows immediately from (1). $\square$

8. FOURIER SERIES

Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π periodic. We now define that n'th Fourier coefficient of f by

$$c_n(f) = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx}\, dx\,.$$

Mathematicians of the 18'th century (including Bernoulli, Euler, Lagrange and others) knew "experimentally" that

$$\sum_{k=-n}^{n} c_k(f)\, e^{ikx} \to f(x) \tag{1}$$

for some simple functions f and all x. Indeed, the left hand side of (1) is known as the (n'th partial sum of the) Fourier series of f, and Fourier claimed that (1) was always true and in a book of outstanding importance in the history of physics show how formulae of the kind $\sum_{k=-\infty}^{\infty} c_k(f)\, e^{ikt}$ could be used to solve partial differential equations of the kind which dominated 19'th century physics.

Jospeh Fourier (21 March 1768 – 16 May 1830) was a French mathematician and physicist born in Auxerre and best known for initiating the investigation of Fourier series and their applications to problems of heat transfer and vibrations.

After several mathematics (including Cauchy) had produced more or less fallacious proofs of the convergence stated in (1), Dirichlet took up the problem. In a paper which set up new and previously undreamed of standards of rigour and clarity in analysis, he was able to prove convergence under quite general conditions.

Fourier series still have widespread applications in science and engineering, and for this reason Fourier series are usually introduced to science students as a technique for solving certain ordinary and partial differential equations. In such courses the emphasis is on getting answers and questions about convergence are completely ignored. (In fact, scientists often view the introduction of mathematical rigour in the study of Fourier series with hostility seeming believing that imposing technical conditions on the functions involved somehow makes the theory more suspect; after all, many scientists will argue, Fourier did not impose any conditions on the functions involved, expect perhaps piecewise continuity, so why should we? To those scientists there is only one thing to say: read Section ????.2.)

8.1. Definitions.

For $n \in \mathbb{Z}$, we define the function $e_n : \mathbb{R} \to \mathbb{C}$ by

$$e_n(x) = e^{inx} .$$

We immediately note the following fact

$$\int_0^{2\pi} e_n \overline{e_m} = \int_0^{2\pi} e^{i(n-m)x} \, dx$$

$$= \begin{cases} \displaystyle\int_0^{2\pi} dx & \text{for } n = m; \\ \left[\dfrac{1}{i(n-m)} e^{i(n-m)x} \right]_{x=0}^{x=2\pi} & \text{for } n \neq m. \end{cases}$$

$$= \begin{cases} 2\pi & \text{for } n = m; \\ 0 & \text{for } n \neq m. \end{cases}$$

The next proposition shows that if a sequence of linear combinations of the functions e_n converge uniformly, then the coefficients are completely determined by the limit function.

Proposition. *Let $(\lambda_n)_{n \in \mathbb{Z}}$ be a family of complex numbers and define $f_n : \mathbb{R} \to \mathbb{C}$ by*

$$f_n = \sum_{k=-n}^{n} \lambda_k e_k .$$

Let $f : \mathbb{R} \to \mathbb{C}$ be a function and assume that

$$f_n \to f \quad uniformly.$$

Then we have:

(1) *The function f is continuous. In particular $f\overline{e_n}$ is integrable over $[0, 2\pi]$.*
(2) *For all $n \in \mathbb{Z}$, we have*

$$\begin{aligned} \lambda_n &= \frac{1}{2\pi} \int_0^{2\pi} f \overline{e_n} \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} \, dx . \end{aligned}$$

Proof.
Fix $n \in \mathbb{Z}$. Since

$$f_m \to f \quad \text{uniformly on } [0, 2\pi] \text{ as } m \to \infty,$$

we conclude that

$$f_m\overline{e_n} \to f\overline{e_n} \quad \text{uniformly on } [0,2\pi] \text{ as } m \to \infty,$$

It follows from this and Theorem ??? that $f\overline{e_n}$ is integrable over $[0,2\pi]$ and that

$$\int_0^{2\pi} f_m\overline{e_n} \to \int_0^{2\pi} f\overline{e_n} \quad \text{as } m \to \infty, \tag{1}$$

However, for $m \geq n$, we clearly have (using the linearity of the Riemann integral)

$$\begin{aligned}
\int_0^{2\pi} f_m\overline{e_n} &= \int_0^{2\pi} \left(\sum_{k=-m}^{m} \lambda_k e_k\right)\overline{e_n} \\
&= \sum_{k=-m}^{m} \lambda_k \int_0^{2\pi} e_k\overline{e_n} \\
&= 2\pi\lambda_n \,,
\end{aligned}$$

whence

$$\int_0^{2\pi} f_m\overline{e_n} \to 2\pi\lambda_n \quad \text{as } m \to \infty, \tag{2}$$

Finally, comparing (1) and (2), we deduce that

$$\int_0^{2\pi} f\overline{e_n} = \lim_m \int_0^{2\pi} f_m\overline{e_n} = 2\pi\lambda_n \,.$$

This completes the proof. □

Motivated by the previous result we make the following definition.

Defintion. . *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0,2\pi]$ and 2π periodic. For $n \in \mathbb{Z}$, this implies that $f\overline{e_n}$ is Riemann integrable. We now define that n'th Fourier coefficient of f by*

$$\begin{aligned}
c_n(f) &= \frac{1}{2\pi}\int_0^{2\pi} f\overline{e_n} \\
&= \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}\,dx \,.
\end{aligned}$$

For $n \in \mathbb{N}$, we define the n'th partial Fourier sum $s_n(f) : [0,2\pi] \to \mathbb{C}$ by

$$s_n(f) = \sum_{k=-n}^{n} c_k(f)\, e_k \,.$$

We immediately make the following observation showing that the Fourier coefficient and the partial Fourier sums of f depend linearly on the function f.

Observation. *We immediately note that if $f, g : \mathbb{R} \to \mathbb{R}$ are Riemann integrable on $[0, 2\pi]$ and 2π periodic and $\lambda \in \mathbb{R}$, then $f + g$ and λf are Riemann integrable on $[0, 2\pi]$ and 2π periodic with*

$$\begin{aligned} c_n(f+g) &= c_n(f) + c_n(g)\,, \\ c_n(\lambda f) &= \lambda\, c_n(f) \end{aligned}$$

for all $n \in \mathbb{Z}$, and

$$\begin{aligned} s_n(f+g) &= s_n(f) + s_n(g)\,, \\ s_n(\lambda f) &= \lambda\, s_n(f) \end{aligned}$$

for all $n \in \mathbb{Z}$.

The main questions is now: find conditions on f guaranteeing that:

$s_n(f)$ converges to f in some suitable sense.

We begin by analysing the Fourier coefficients in more detail.

Proposition. Bessel's Inequality. *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π periodic.*

(1) *For $n \in \mathbb{N}$, we have*

$$\frac{1}{2\pi} \int_0^{2\pi} |s_n(f)|^2 = \sum_{k=-n}^{n} |c_k(f)|^2 \,.$$

(2) *For $n \in \mathbb{N}$, we have*

$$\sum_{k=-n}^{n} |c_k(f)|^2 \leq \frac{1}{2\pi} \int_0^{2\pi} |f|^2 \,.$$

(3) *For $n \in \mathbb{N}$, we have $\|s_n(f)\|_2 \leq \|f\|_2$.*

Proof.
(1) We have

$$\begin{aligned}
\int_0^{2\pi} |s_n(f)|^2 &= \int_0^{2\pi} s_n(f)\, \overline{s_n(f)} \\
&= \int_0^{2\pi} \left(\sum_{k=-n}^{n} c_k(f)\, e_k \right) \overline{\left(\sum_{l=-n}^{n} c_l(f)\, e_l \right)} \\
&= \int_0^{2\pi} \left(\sum_{k=-n}^{n} c_k(f)\, e_k \right) \left(\sum_{l=-n}^{n} \overline{c_l(f)}\, \overline{e_l} \right) \\
&= \sum_{k,l=-n}^{n} c_k(f) \overline{c_l(f)} \int_0^{2\pi} e_k \overline{e_l} \\
&= 2\pi \sum_{k=-n}^{n} c_k(f) \overline{c_k(f)} \\
&= 2\pi \sum_{k=-n}^{n} |c_k(f)|^2 \,.
\end{aligned}$$

(2) We have

$$
\begin{aligned}
0 &\le \int_0^{2\pi} |f - s_n(f)|^2 \\
&= \int_0^{2\pi} (f - s_n(f))(\overline{f - s_n(f)}) \\
&= \int_0^{2\pi} \left(f - \sum_{k=-n}^{n} c_k(f)\, e_k \right) \left(\overline{f - \sum_{k=-n}^{n} c_k(f)\, e_k} \right) \\
&= \int_0^{2\pi} \left(f - \sum_{k=-n}^{n} c_k(f)\, e_k \right) \left(\overline{f} - \sum_{k=-n}^{n} \overline{c_k(f)}\, \overline{e_k} \right) \\
&= \int_0^{2\pi} f\overline{f} - \sum_{k=-n}^{n} c_k(f) \int_0^{2\pi} e_k\overline{f} - \sum_{k=-n}^{n} \overline{c_k(f)} \int_0^{2\pi} \overline{e_k} f + \sum_{k,l=-n}^{n} c_k(f)\overline{c_l(f)} \int_0^{2\pi} e_k\overline{e_l} \\
&= \int_0^{2\pi} f\overline{f} - \sum_{k=-n}^{n} c_k(f) \overline{\int_0^{2\pi} \overline{e_k} f} - \sum_{k=-n}^{n} \overline{c_k(f)} \int_0^{2\pi} \overline{e_k} f + \sum_{k,l=-n}^{n} c_k(f)\overline{c_l(f)} \int_0^{2\pi} e_k\overline{e_l} \\
&= \int_0^{2\pi} |f|^2 - 2\pi \sum_{k=-n}^{n} c_k(f)\overline{c_k(f)} - 2\pi \sum_{k=-n}^{n} \overline{c_k(f)} c_k(f) + \sum_{k,l=-n}^{n} c_k(f)\overline{c_l(f)} \int_0^{2\pi} e_k\overline{e_l} \\
&= \int_0^{2\pi} |f|^2 - 2\pi \sum_{k=-n}^{n} |c_k(f)|^2 - 2\pi \sum_{k=-n}^{n} |c_k(f)|^2 + 2\pi \sum_{k=-n}^{n} c_k(f)\overline{c_k(f)} \\
&= \int_0^{2\pi} |f|^2 - 2\pi \sum_{k=-n}^{n} |c_k(f)|^2 - 2\pi \sum_{k=-n}^{n} |c_k(f)|^2 + 2\pi \sum_{k=-n}^{n} |c_k(f)|^2 \\
&= \int_0^{2\pi} |f|^2 - 2\pi \sum_{k=-n}^{n} |c_k(f)|^2 \,.
\end{aligned}
$$

The desired result follows immediately from the above inequality.
(3) This follows from (1) and (2). $\square$

Corollary. *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π periodic. Then*

$$c_{\pm n}(f) \to 0 \quad \text{as } n \to \infty.$$

Proof.
Since f is Riemann integrable on $[0, 2\pi]$, it follows from ????? that $|f|^2$ is also Riemann integrable on $[0, 2\pi]$. We conclude from this that $\int_0^{2\pi} |f|^2 < \infty$. It therefore follows from the previous Proposition that $\sup_{n\in\mathbb{Z}} \sum_{k=-n}^{n} |c_n(f)|^2 \leq \frac{1}{2\pi} \int_0^{2\pi} |f|^2 < \infty$, whence $c_n(f) \to 0$ as $n \to \pm\infty$. □

The previous corollary to Bessel's Inequality shows that if $f : \mathbb{R} \to \mathbb{R}$ is Riemann integrable on $[0, 2\pi]$ and 2π periodic, then the Fourier coefficients $c_n(f)$ tend to 0 as $n \to \pm\infty$. It is therefore at least plausible that the n'th partial sum $s_n(f) = \sum_{k=-n}^{n} c_n(f)e_n$ converges in some sense. In the remaining sections we will investigate this question in more detail. Below is description of this:

Section 8.2. Many 18'th century physicist and mathematicians and many contemporary scientist believe that if $f : \mathbb{R} \to \mathbb{R}$ is continuous and 2π periodic, then

$$s_n(f) \to f \quad \text{pointwise.}$$

In Section ???.2 we show that this is false. More precisely, we will construct a 2π-perodic function $\Phi : \mathbb{R} \to \mathbb{R}$ such that

(1) Φ is continuous;
(2) $s_n(\Phi)(0)$ diverges as $n \to \infty$.

Consequently, continuity of f is not enough to guarantee pointwise convergence of $s_n(f)$ to f.

Section 8.3. In Section 8.3 we will study conditions on f guaranteeing that

$$s_n(f) \to f \quad \text{pointwise.}$$

Section 8.4. In Section 8.4 we will study conditions on f guaranteeing that

$$s_n(f) \to f \quad \text{uniformly.}$$

Section 8.5. In Section 8.5 we will study conditions on f guaranteeing that summability, i.e. we will study conditions on f guaranteeing that

$$\frac{s_0(f) + s_1(f) + \cdots + s_n(f)}{n} \to f \quad \text{uniformly.}$$

Section 8.6. In Section 8.6 we will study conditions on f guaranteeing that summability, i.e. we will study conditions on f guaranteeing that

$$s_n(f) \to f \quad \text{with respect to } \|\cdot\|_2.$$

Remark. Write

$$\mathbf{R} = \left\{ f : \mathbb{R} \to \mathbb{R} \;\middle|\; f \text{ is Riemann integrable on } [0, 2\pi] \text{ and } 2\pi \text{ periodic } \right\}$$

and

$$\ell^2(\mathbb{Z}) = \left\{ (x_n)_{n\in\mathbb{Z}} \;\middle|\; \sup_n \sum_{k=-n}^{n} |x_k|^2 < \infty \right\}.$$

Then it follows from Bessel's Inequality that we can define a map

$$\mathcal{F} : \mathbf{R} \to \ell^2(\mathbb{Z})$$

by

$$\mathcal{F}(f) = \big(c_n(f) \big)_{n\in\mathbb{Z}}.$$

The main question addressed in this section (namely, does $s_n(f)$ converge to f) can now be recast as follows: is there a subspace U of $\ell^2(\mathbb{Z})$ such that the map

$$\mathcal{E} : U \to \mathbf{R}$$

defined by

$$\mathcal{E}\big((\lambda_n)_{n\in\mathbb{Z}} \;\big) = \text{“some suitable limit of } \sum_{k=-n}^{n} \lambda_k e_k \text{ as } n \to \infty\text{”}$$

is well-defined and

$$\mathcal{E}\mathcal{F}(f) = f$$

for all $f \in \mathbf{R}$. That is, we are discussing the inversion of $\mathcal{F}$. This suggests (correctly) that powerful methods from functional analysis can be used. However, this direction will not be pursued further in these notes since readers are not assumed to be familiar with the theory of Lebesgue integration needed in order to discuss the convergence issues involved.

8.2. An example of a Continuous Function whose Fourier Series Diverges.

We will show that there is a continuous 2π periodic function whose Fourier series is not pointwise convergent at all points. More precisely, we will construct a 2π-perodic function $\Phi : \mathbb{R} \to \mathbb{R}$ such that

(1) Φ is continuous;
(2) $s_n(\Phi)(0)$ diverges as $n \to \infty$.

For a positive integer m, define $Q_m : \mathbb{R} \to \mathbb{R}$ by

$$\begin{aligned} Q_m = &\frac{\cos mx}{m} + \frac{\cos(m+1)x}{m-1} + \cdots + \frac{\cos(2m-2)x}{2} + \frac{\cos(2m-1)x}{1} \\ &- \frac{\cos(2m+1)x}{1} - \frac{\cos(2m+2)x}{2} - \cdots - \frac{\cos(3m-1)x}{m-1} - \frac{\cos(3m)x}{m} \,. \end{aligned}$$

For a positive integer k, we write

$$m_k = 2^{(k^2)} \,.$$

Also, for a positive integer n, define $\Phi_n : \mathbb{R} \to \mathbb{R}$ by

$$\Phi_n = \sum_{k=1}^{n} \frac{1}{k^2} Q_{m_k} \,.$$

With considerably amount of work it can now be shown that there is a function $\Phi : \mathbb{R} \to \mathbb{R}$ such that

$$\Phi_n \to \Phi \quad \text{unifornmly.}$$

The function Φ can now be shown to satisfy the following:

(1) Φ is continuous;
(2) $s_n(\Phi)(0)$ diverges as $n \to \infty$.

However, the proof is fairly lengthy and is therefore omitted.

8.3. Pointwise Convergence.

The main result in this section is the following.

Dini's Theorem. *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π periodic.*
Let $x, y \in \mathbb{R}$.
Assume that there is a $\delta > 0$ such that

$$\sup_{0<t\leq\delta} \left| \frac{f(x+t) + f(x-t) - 2y}{t} \right| < \infty \,.$$

Then

$$s_n(f)(x) \to y.$$

Ulisse Dini (14 November 1845 – 28 October 1918) was an Italian mathematician and politician, born in Pisa. He is known for his contribution to real analysis, partly collected in his book "Fondamenti per la teorica delle funzioni di variabili reali".

In order to prove Dini's Theorem we need the following results.

Definition. Dirichlet kernel. *For $n \in \mathbb{N}$, we define the n'th Dirichlet kernel $D_n : \mathbb{R} \to \mathbb{C}$ by*

$$D_n = \sum_{k=-n}^{n} e_k \,.$$

Proposition. Properties of the Dirichlet kernel.

(1) *We have*

$$D_n(x) = \begin{cases} \dfrac{\sin(n+\frac{1}{2})x}{\sin\frac{1}{2}x} & \text{for } x \notin 2\pi\mathbb{Z}; \\ 2n+1 & \text{for } x \in 2\pi\mathbb{Z}; \end{cases}$$

(2) *We have*

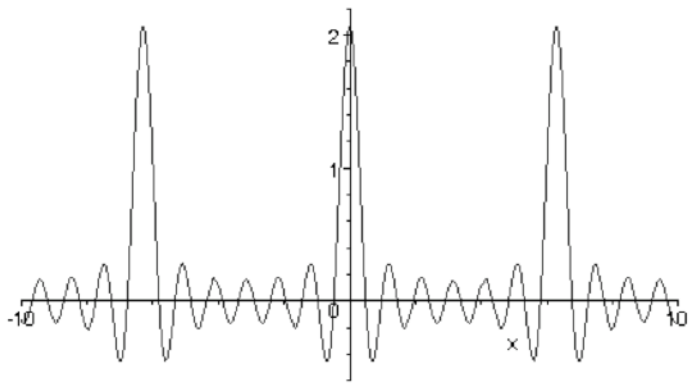
$$\int_0^{\pi} D_n = \pi \,,$$

$$\int_0^{2\pi} D_n = \int_{-\pi}^{\pi} D_n = 2\pi \,,$$

(3) *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π periodic. Then*

$$\begin{aligned} s_n(f)(x) &= \frac{1}{2\pi}\int_0^{2\pi} D_n(x-t)f(t)\,dt \\ &= \frac{1}{2\pi}\int_0^{2\pi} D_n(t)f(x+t)\,dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(t)f(x+t)\,dt \end{aligned}$$

for all x.



The graph of the Dirichlet kernel D_n for $n = 6$.

Proof.
(1) For $x \notin 2\pi\mathbb{Z}$, we have

$$
\begin{aligned}
D_n(x) &= \sum_{k=-n}^{n} e_k(x) \\
&= \sum_{k=-n}^{n} e^{ikx} \\
&= e^{-inx} \sum_{k=0}^{2n} e^{ikx} \\
&= e^{-inx} \sum_{k=0}^{2n} (e^{ix})^k \\
&= e^{-inx} \frac{(e^{ix})^{2n+1} - 1}{e^{ix} - 1} \\
&= \frac{e^{i(n+1)x} - e^{-inx}}{e^{ix} - 1} \\
&= \frac{e^{-i\frac{1}{2}x}\left(e^{i(n+1)x} - e^{-inx}\right)}{e^{-i\frac{1}{2}x}\left(e^{ix} - 1\right)} \\
&= \frac{e^{i(n+\frac{1}{2})x} - e^{-i(n+\frac{1}{2})x}}{e^{i\frac{1}{2}x} - e^{-\frac{1}{2}x}} \,.
\end{aligned}
$$

Next recalling de Moivre's formula, namely, $sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$ for all $\theta \in \mathbb{R}$, we therefore deduce that

$$
\begin{aligned}
D_n(x) &= \frac{e^{i(n+\frac{1}{2})x} - e^{-i(n+\frac{1}{2})x}}{e^{i\frac{1}{2}x} - e^{-\frac{1}{2}x}} \\
&= \frac{\sin(n+\frac{1}{2})x}{\sin\frac{1}{2}x} \,.
\end{aligned}
$$

On the other hand, for $x \in 2\pi\mathbb{Z}$, we can choose $m \in \mathbb{Z}$ such that $x = 2\pi m$, whence

$$
\begin{aligned}
D_n(x) &= \sum_{k=-n}^{n} e_k(x) \\
&= \sum_{k=-n}^{n} e^{ikx} \\
&= \sum_{k=-n}^{n} e^{2\pi ikm} \\
&= \sum_{k=-n}^{n} 1 \\
&= 2n+1 \,.
\end{aligned}
$$

(2) Since $e_0 = 1$, we have
(3) For all x, we have

$$\begin{aligned} s_n(f)(x) &= \sum_{k=-n}^{n} c_k(f)\, e_k(x) \\ &= \sum_{k=-n}^{n} \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int}\, dt\, e^{ikx} \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(t) \sum_{k=-n}^{n} e^{in(x-t)}\, dt \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(t) D_n(x-t)\, dt\,. \end{aligned}$$

This completes the proof. □

Lemma. *Let $a, b \in \mathbb{R}$ with $a \leq b$. Let $f : [a, b] \to \mathbb{C}$ be a function. Assme that*

(i) *f is bounded.*
(ii) *for each $\epsilon > 0$, the function f is Riemann integrable on $[a + \varepsilon, b]$.*

Then f is Riemann integrable on $[a, b]$.

Proof.
The proof is left to the reader. □

We can now prove Dini's Theorem.

Proof of Dini's Theorem.
First note that it follows from Proposition ??? and a change of variable that

$$\begin{aligned} s_n(f)(x) &= \frac{1}{2\pi}\int_0^{2\pi} f(t)D_n(x-t)\,dt \\ &= \frac{1}{2\pi}\int_{x-2\pi}^{x} f(x-u)D_n(u)\,du \end{aligned}$$

However, since both f and D_n are 2π periodic, we conclude that $\int_{x-2\pi}^{x} f(x-u)D_n(u)\,du = \int_0^{2\pi} f(x-u)D_n(u)\,du$, whence

$$\begin{aligned} s_n(f)(x) &= \frac{1}{2\pi}\int_{x-2\pi}^{x} f(x-u)D_n(u)\,du \\ &= \frac{1}{2\pi}\int_0^{2\pi} f(x-u)D_n(u)\,du \\ &= \frac{1}{2\pi}\int_0^{\pi} f(x-u)D_n(u)\,du + \frac{1}{2\pi}\int_{\pi}^{2\pi} f(x-u)D_n(u)\,du \end{aligned}$$

Once more using the fact that both f and D_n are 2π periodic, we deduce that $\int_{\pi}^{2\pi} f(x-u)D_n(u)\,du = \int_{-\pi}^{0} f(x-u)D_n(u)\,du$, and so

$$\begin{aligned} s_n(f)(x) &= \frac{1}{2\pi}\int_0^{\pi} f(x-u)D_n(u)\,du + \frac{1}{2\pi}\int_{\pi}^{2\pi} f(x-u)D_n(u)\,du \\ &= \frac{1}{2\pi}\int_0^{\pi} f(x-u)D_n(u)\,du + \frac{1}{2\pi}\int_{-\pi}^{0} f(x-u)D_n(u)\,du \end{aligned}$$

Finally changing the variable in $\int_{-\pi}^{0} f(x-u)D_n(u)\,du$ and using the fact that D_n is even gives

$$\begin{aligned} s_n(f)(x) &= \frac{1}{2\pi}\int_0^{\pi} f(x-u)D_n(u)\,du + \frac{1}{2\pi}\int_{-\pi}^{0} f(x-u)D_n(u)\,du \\ &= \frac{1}{2\pi}\int_0^{\pi} f(x-u)D_n(u)\,du + \frac{1}{2\pi}\int_0^{\pi} f(x+v)D_n(-v)\,dv \\ &= \frac{1}{2\pi}\int_0^{\pi} f(x-u)D_n(u)\,du + \frac{1}{2\pi}\int_0^{\pi} f(x+v)D_n(v)\,dv \\ &= \frac{1}{2\pi}\int_0^{\pi} \Big(f(x-t)+f(x+t)\Big)D_n(t)\,dt\,. \end{aligned} \tag{1}$$

Next, note that Proposition ??? implies that

$$y = \frac{1}{\pi}\int_0^{\pi} yD_n(t)\,dt\,. \tag{2}$$

Finally combining (1) and (2) yields

$$|s_n(f)(x) - y| = \left|\frac{1}{2\pi}\int_0^{\pi} \Big(f(x-t)+f(x+t)-2y\Big)D_n(t)\,dt\right|. \tag{3}$$

Define 2π periodic functions $\Phi, S : \mathbb{R} \to \mathbb{R}$ by

$$\Phi(t) = \begin{cases} 0 & \text{for } t = 0; \\ \dfrac{f(x+t) + f(x-t) - 2y}{t} & \text{for } t \in (0, \pi]; \\ 0 & \text{for } t \in (\pi, 2\pi), \end{cases}$$

$$S(t) = \begin{cases} 0 & \text{for } t = 0; \\ \dfrac{\frac{t}{2}}{\sin(\frac{t}{2})} & \text{for } t \in (0, \pi]; \\ 0 & \text{for } t \in (\pi, 2\pi), \end{cases}$$

Define 2π periodic functions $h_-, h_+ : \mathbb{R} \to \mathbb{R}$ by

$$h_-(t) = \frac{1}{i}\Phi(t)S(t)e^{-\frac{1}{2}it} \quad \text{for } t \in [0, 2\pi),$$

$$h_+(t) = \frac{1}{i}\Phi(t)S(t)e^{\frac{1}{2}it} \qquad \text{for } t \in [0, 2\pi).$$

With this notation it follows that if $t \in (0, \pi)$, then

$$\begin{aligned} \Big(f(x-t) + f(x+t) - 2y\Big)D_n(t) &= \frac{f(x-t) + f(x+t) - 2y}{t}\; t\; \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t} \\ &= 2\frac{f(x-t) + f(x+t) - 2y}{t}\; \frac{\frac{1}{2}t}{\sin\frac{1}{2}t}\; \sin(n+\tfrac{1}{2})t \\ &= 2\Phi(t)\; S(t)\; \sin(n+\tfrac{1}{2})t \\ &= 2\Phi(t)\; S(t)\; \frac{e^{i(n+\frac{1}{2})t} - e^{-i(n+\frac{1}{2})t}}{2i} \\ &= h_+(t)e^{int} - h_-(t)e^{-int}. \end{aligned}$$

We therefore deduce from (3) that

$$|s_n(f)(x) - y| = \left|\frac{1}{2\pi}\int_0^{\pi}\Big(h_+(t)e^{int} - h_-(t)e^{-int}\Big)\,dt\right|. \tag{4}$$

Next, since $h_-(t) = h_+(t) = 0$ for all $t \in (\pi, 2\pi)$, we conclude from (4) that

$$\begin{aligned} |s_n(f)(x) - y| &= \left|\frac{1}{2\pi}\int_0^{\pi}\Big(h_+(t)e^{int} - h_-(t)e^{-int}\Big)\,dt\right| \\ &= \left|\frac{1}{2\pi}\int_0^{2\pi}\Big(h_+(t)e^{int} - h_-(t)e^{-int}\Big)\,dt\right|. \end{aligned} \tag{5}$$

Next, we prove the following claim.

Claim 1. The functions h_- and h_+ are Riemann integerable on $[0, 2\pi]$.
Proof of Claim 1. We first observe that h_- and h_+ are bounded on $[0, 2\pi]$. Indeed, if $* \in \{-.+\}$, then for all $t \in [0, 2\pi]$, we have

$$|h_*(t)| = |\Phi(t)|\,|S(t)|\,.$$

By assumtion Φ is bounded on $[0, \delta]$, and since f is bunded (because f is Riemann integrable), we conclude that Φ is also bounded on $[\delta, 2\pi]$. Next, since $\frac{\sin t}{t} \to 1$ as $t \searrow$, we deduce that S is bounded on $[0, 2\pi]$. Consequently, h_* is bounded on $[0, 2\pi]$.

Next, we observe that since f is Riemann integrable on $[0, 2\pi]$, the functions h_- and h_+ are Riemann integrable on $[\varepsilon, 2\pi]$ for all $\varepsilon > 0$.

Finally, since the h_- and h_+ are bounded on $[0, 2\pi]$ and Riemann integrable on $[\varepsilon, 2\pi]$ for all $\varepsilon > 0$, we conclude from Lemma ???? that the functions h_- and h_+ are Riemann integrable on $[0, 2\pi]$. This completes the proof of Claim 1.

Since (by Claim 1) the functions h_- and h_+ are Riemmann integerable on $[0, 2\pi]$, we conclude that the integrals $\frac{1}{2\pi}\int_0^{2\pi} h_-(t)e^{int}\,dt$ and $\frac{1}{2\pi}\int_0^{2\pi} h_+(t)e^{int}\,dt$ are well-defined and that $\frac{1}{2\pi}\int_0^{2\pi} (h_+(t)e^{int} - h_-(t)e^{-int})\,dt = \frac{1}{2\pi}\int_0^{2\pi} h_+(t)e^{int}\,dt - \frac{1}{2\pi}\int_0^{2\pi} h_-(t)e^{-int}\,dt$. It therefore follows from (5) that

$$\begin{aligned}
|s_n(f)(x) - y| &= \left|\frac{1}{2\pi}\int_0^{2\pi} \Big(h_+(t)e^{int} - h_-(t)e^{-int}\Big)\,dt\right| \\
&= \left|\frac{1}{2\pi}\int_0^{2\pi} h_+(t)e^{int}\,dt - \frac{1}{2\pi}\int_0^{2\pi} h_-(t)e^{-int}\Big)\,dt\right| \\
&= |c_n(h_+) - c_{-n}(h_-)|\,. \qquad (6)
\end{aligned}$$

Finally, since the functions h_- and h_+ are Riemmann integerable on $[0, 2\pi]$, it follows from Corollary ??? that $c_n(h_+) \to 0$ as $n \to \infty$ and that $c_{-n}(h_-) \to 0$ as $n \to \infty$. Combining this with (6), now shows that $|s_n(f)(x) - y| = |c_n(h_+) - c_{-n}(h_-)| \to 0$ as $n \to \infty$. □

Corollary. *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π periodic. Let $x \in [0, 2\pi]$ and assume that*

(i) *The following limits exist, namely,*

$$\lim_{t \searrow 0} f(x - t)\,,$$
$$\lim_{t \searrow 0} f(x + t)\,.$$

Define $f_- : [a, x] \to \mathbb{R}$ and $f_+ : [x, b]$ by

$$f_-(u) = \begin{cases} f(u) & \text{for } u < x; \\ \lim_{t \searrow 0} f(x - t) & \text{for } u = x, \end{cases}$$
$$f_+(u) = \begin{cases} f(u) & \text{for } x < u; \\ \lim_{t \searrow 0} f(x + t) & \text{for } u = x. \end{cases}$$

(ii) *The functions f_- and f_+ are differentiable at x.*

Then

$$s_n(f)(x) \to \frac{\lim_{t \searrow 0} f(x + t) + \lim_{t \searrow 0} f(x - t)}{2}\,.$$

Proof.
Since f_- is differentiable at x, there is a $\delta_- > 0$ such that if $0 < t \le \delta_-$, then

$$\left| \frac{f_-(x - t) - f_-(x)}{t} - f_-'(x) \right| \le 1\,,$$

i.e. if $0 < t \le \delta_-$, then

$$\left| \frac{f(x - t) - f_-(x)}{t} - f_-'(x) \right| \le 1\,.$$

Similarly, since f_+ is differentiable at x, there is a $\delta_+ > 0$ such that if $0 < t \le \delta_-$, then

$$\left| \frac{f_+(x + t) - f_+(x)}{t} - f_+'(x) \right| \le 1\,,$$

i.e. if $0 < t \le \delta_+$, then

$$\left| \frac{f(x + t) - f_+(x)}{t} - f_+'(x) \right| \le 1\,.$$

Let $\delta = \min(\delta_-, \delta_+)$. For $0 < t \le \delta$, we now have

$$\begin{aligned} \left| \frac{f(x + t) + f(x - t) - 2\frac{f_+(x) + f_-(x)}{2}}{t} \right| &\le \left| \frac{f(x + t) - f_+(x)}{t} \right| + \left| \frac{f(x - t) - f_-(x)}{t} \right| \\ &\le |f_+'(x)| + 1 + |f_-'(x)| + 1\,. \qquad (1) \end{aligned}$$

The desired result now follows from (1) and Dini's Theorem. $\square$

Corollary. *Let $f : \mathbb{R} \to \mathbb{R}$ be piecewise differentiable and 2π periodic.*
Then

$$s_n(f)(x) \to \frac{\lim_{t \searrow 0} f(x+t) + \lim_{t \searrow 0} f(x-t)}{2}$$

for all x.

Proof.
This follows immediately from the previous lemma. □

Corollary. *Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable and 2π periodic.*
Then

$$s_n(f) \to f \quad \textit{pointwise.}$$

Proof.
This follows immediately from the previous lemma. □

8.4. Uniform Convergence.

The main result in this section is the following.

Theorem. *Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable with continuous derivative f' and 2π periodic.*

Then

$$s_n(f) \to f \quad \textit{uniformly.}$$

Proof.
We first prove the following two claims.

Claim 1. For $n \in \mathbb{Z}$, we have $c_n(f') = inc_n(f)$.
Proof of Claim 1. Since f' is continuous, it follows from partial integration that

$$\begin{aligned}
c_n(f') &= \frac{1}{2\pi}\int_0^{2\pi} f'(x)e^{-inx}\,dx \\
&= \frac{1}{2\pi}\left[f(x)e^{inx}\right]_{x=0}^{x=2\pi} - \frac{1}{2\pi}\int_0^{2\pi} f(x)(-in)e^{-inx}\,dx \\
&= in\frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}\,dx \\
&= inc_n(f)\,.
\end{aligned}$$

This completes the proof of Claim 1.

Claim 2. For $n \in \mathbb{Z}$, we have $\sum_{k=-n}^{n} \|c_k(f)e_k\|_\infty \le \frac{5}{2\pi}\int_0^{2\pi} |f'|^2$.
Proof of Claim 2. Using Claim 1, for positive integers n, we have

$$\begin{aligned}
\sum_{k=-n}^{n} \|c_k(f)e_k\|_\infty &\le \sum_{k=-n}^{n} |c_k(f)| \\
&= \sum_{k=-n}^{n} |kc_k(f)|\,\frac{1}{|k|} \\
&\le \left(\sum_{k=-n}^{n} |kc_k(f)|^2\right)^{\frac{1}{2}} \left(\sum_{k=-n}^{n} \frac{1}{k^2}\right)^{\frac{1}{2}} \qquad \text{[by Cauchy-Schwatz inequality]} \\
&= \left(\sum_{k=-n}^{n} |c_k(f')|^2\right)^{\frac{1}{2}} \left(\sum_{k=-n}^{n} \frac{1}{k^2}\right)^{\frac{1}{2}}. \qquad (1)
\end{aligned}$$

Next, since it follows from Lemma ??? that $\sum_{k=-n}^{n} \frac{1}{k^2} \le \frac{1}{2\pi}\int_0^{2\pi} |f'|^2$ and since $\sum_{k=-n}^{n} \frac{1}{k^2} = 3 + 2\sum_{k=2}^{n} \frac{1}{k^2} \le 3 + 2\sum_{k=2}^{n} \frac{1}{k(k-1)} = 3 + 2\sum_{k=2}^{n}(\frac{1}{k-1} - \frac{1}{k}) = 3 + 2(\frac{1}{1} - \frac{1}{n}) \le 5$, we conclude from (1) that

$$\begin{aligned}
\sum_{k=-n}^{n} \|c_k(f)e_k\|_\infty &\le \left(\sum_{k=-n}^{n} |c_k(f')|^2\right)^{\frac{1}{2}} \left(\sum_{k=-n}^{n} \frac{1}{k^2}\right)^{\frac{1}{2}} \\
&\le \frac{5}{2\pi}\int_0^{2\pi} |f'|^2\,. \qquad (2)
\end{aligned}$$

This completes the proof of Claim 2.

It now follows from Claim 2 and Theorem ???? that there is continuous function $\varphi : \mathbb{R} \to \mathbb{C}$ such that

$$s_n(f) = \sum_{k=-n}^{n} c_k(f) e_k \to \varphi \quad \text{uniformly.} \tag{3}$$

Since f is differentiable, it also follows from Corollary ??? that

$$s_n(f) \to f \quad \text{pointwise.} \tag{4}$$

Comparing (3) and (4), we deduce that $\varphi = f$, and so (using (3))

$$s_n(f) \to \varphi = f \quad \text{uniformly.}$$

This completes the proof. $\square$

8.5. Summability.

Defintion. . *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π periodic. For $n \in \mathbb{N}$, we define $\sigma_n(f) : [0, 2\pi] \to \mathbb{C}$ by*

$$\sigma_n(f) = \frac{s_0(f) + \cdots + s_{n-1}(f)}{n}.$$

The main result in this section is the following.

Fejer's Theorem. *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and 2π periodic. Then*

$$\sigma_n(f) \to f \quad \textit{uniformly.}$$

Leopold Fejer (9 February 1880 15 October 1959) was a Hungarian mathematician. In 1911 Fejer was appointed to the chair of mathematics at the University of Budapest and he held that post until his death. During his period in the chair at Budapest Fejer led a highly successful Hungarian school of analysis. He was the thesis advisor of mathematicians such as John von Neumann, Paul Erdös, George Polya and Pal Turan. Fejer's research concentrated on harmonic analysis and, in particular, Fourier series.

In order to prove Fejer's Theorem we need the following results.

Fejer kernel. *For $n \in \mathbb{N}$, we define the n'th Fejer kernel $K_n : \mathbb{R} \to \mathbb{C}$ by*

$$K_n = \frac{D_0 + D_1 + \cdots + D_{n-1}}{n}\,.$$

Proposition. Properties of the Fejer kernel.

(1) *We have*

$$K_n(x) = \begin{cases} \dfrac{1}{n}\left(\dfrac{\sin \frac{n}{2}x}{\sin \frac{1}{2}x}\right)^2 & \text{for } x \notin 2\pi\mathbb{Z}; \\ n & \text{for } x \in 2\pi\mathbb{Z}. \end{cases}$$

(2) *We have*

$$\int_{-\pi}^{\pi} K_n = 2\pi\,.$$

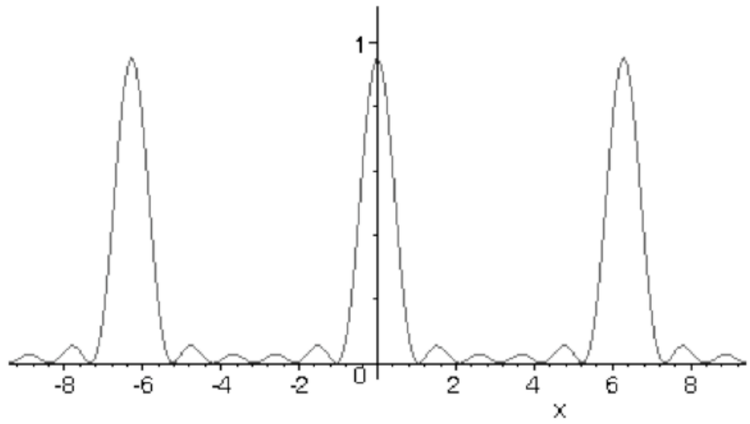
(3) *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π-periodic. Then*

$$\sigma_n(f)(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(t) f(x+t)\,dt$$

for all x.

(4) *For each $\delta > 0$, we have*

$$K_n \to 0 \quad \textit{uniformly on } [-\pi, -\delta] \cup [\delta, \pi].$$



The graph of the Fejer kernel D_n for $n = 6$.

Proof.
(1) For $x \notin 2\pi\mathbb{Z}$, we have

$$
\begin{aligned}
K_n(x) &= \frac{1}{n}\sum_{k=0}^{n-1} D_k(x) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} \frac{\sin(k+\frac{1}{2})x}{\sin\frac{1}{2}x} \\
&= \frac{1}{n}\frac{1}{(\sin\frac{1}{2}x)^2}\sum_{k=0}^{n-1} \sin(k+\tfrac{1}{2})x\,\sin\tfrac{1}{2}x \\
&= \frac{1}{2n}\frac{1}{(\sin\frac{1}{2}x)^2}\sum_{k=0}^{n-1}\Big(\cos kx - \cos(k+1)x\Big) \qquad [\text{since } \sin a\sin b = \tfrac{1}{2}(\cos(a-b)-\cos(a+b)) \text{ for } a,b\in\mathbb{R}] \\
&= \frac{1}{2n}\frac{1}{(\sin\frac{1}{2}x)^2}(\cos 0x - \cos nx) \\
&= \frac{1}{2n}\frac{1}{(\sin\frac{1}{2}x)^2}(1-\cos nx) \\
&= \frac{1}{n}\frac{1}{(\sin\frac{1}{2}x)^2}(\sin\tfrac{n}{2}x)^2\,. \qquad [\text{since } 1-\cos a = 2(\sin a)^2 \text{ for } a\in\mathbb{R}]
\end{aligned}
$$

Also, for $x \in 2\pi\mathbb{Z}$, we have

$$
K_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} D_k(x) = \frac{1}{n}\sum_{k=0}^{n-1}(2k+1) = \frac{1}{n}\left(2\sum_{k=0}^{n-1}k+n\right) = \frac{1}{n}\left(2\frac{(n-1)n}{2}+n\right) = n\,.
$$

(2) Since it follows from Proposition ?? that $\int_{-\pi}^{\pi} D_k = 2\pi$ for all k, we deduce that

$$
\begin{aligned}
\int_{-\pi}^{\pi} K_n &= \frac{1}{n}\sum_{k=0}^{n-1}\int_{-\pi}^{\pi} D_k \\
&= \frac{1}{n}\sum_{k=0}^{n-1} 2\pi \\
&= 2\pi\,.
\end{aligned}
$$

(3) Since it follows from Proposition ?? that

$$
s_n(f)(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} D_k(t) f(x+t)\,dt
$$

for all k and all x, we deduce that

$$\begin{aligned}
\sigma_n(f)(x) &= \frac{1}{n}\sum_{k=0}^{n-1} s_k(f) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} \frac{1}{2\pi}\int_{-\pi}^{\pi} D_k(t) f(x+t)\,dt \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{1}{n}\sum_{k=0}^{n-1} D_k(t)\right) f(x+t)\,dt \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(t) f(x+t)\,dt\,.
\end{aligned}$$

(4) We have (using Part (1))

$$\begin{aligned}
\sup_{x\in[-\pi,-\delta]\cup[\delta,\pi]} |K_n(x)| &= \sup_{x\in[-\pi,-\delta]\cup[\delta,\pi]} \frac{1}{n}\left(\frac{\sin\frac{n}{2}x}{\sin\frac{1}{2}x}\right)^2 \\
&\le \frac{1}{n}\frac{1}{(\sin\frac{1}{2}\delta)^2} \\
&\to 0\,.
\end{aligned}$$

This completes the proof. $\square$

We can now prove Fejer's Theorem.

Proof of Fejer's Theorem.
Let $\varepsilon > 0$

Since f is continuous and 2π-periodic, we deduce that f is bounded, i.e. there is a constant M such that

$$|f(x)| \leq M$$

for all x.

Since f is continuous and 2π-periodic, we conclude that f is uniformly continuous. We can there choose $\delta > 0$ such that for all $x, y \in \mathbb{R}$, we have

$$|x - y| \leq \delta \quad \Rightarrow \quad |f(x) - f(y)| \leq \frac{\varepsilon}{1 + 2M} \,.$$

Next, since (by the previous proposition) $K_n \to 0$ uniformly on $[-\pi, -\delta] \cup [\delta, \pi]$, we can find a positive integer N such that

$$n \geq N \quad \Rightarrow \quad \sup_{x \in [-\pi, -\delta] \cup [\delta, \pi]} |K_n(x)| \leq \frac{\varepsilon}{1 + 2M} \,.$$

For $n \geq N$, we now have (again using the previous proposition)

$$
\begin{aligned}
\sup_x |\sigma_n(f)(x) - f(x)| &= \sup_x \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} K_n(t) f(x+t)\, dt - f(x) \right| \\
&= \sup_x \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} K_n(t) f(x+t)\, dt - \frac{1}{2\pi} \int_{-\pi}^{\pi} K_n(t) f(x)\, dt \right| \\
&\leq \sup_x \frac{1}{2\pi} \int_{-\pi}^{\pi} |K_n(t)|\, |f(x+t) - f(x)|\, dt \\
&\leq \sup_x \frac{1}{2\pi} \Bigg(\int_{-\delta}^{\delta} |K_n(t)|\, |f(x+t) - f(x)|\, dt \\
&\qquad + \int_{-\pi}^{-\delta} |K_n(t)|\, |f(x+t) - f(x)|\, dt \, + \int_{\delta}^{\pi} |K_n(t)|\, |f(x+t) - f(x)|\, dt \Bigg) \\
&\leq \sup_x \frac{1}{2\pi} \Bigg(\int_{-\delta}^{\delta} |K_n(t)| \, \frac{\varepsilon}{1+2M}\, dt \\
&\qquad + \int_{-\pi}^{-\delta} |K_n(t)| \Big(|f(x+t)| + |f(x)| \Big)\, dt \, + \int_{\delta}^{\pi} |K_n(t)| \Big(|f(x+t)| + |f(x)| \Big)\, dt \Bigg) \\
&\leq \sup_x \frac{1}{2\pi} \Bigg(\int_{-\delta}^{\delta} |K_n(t)| \, \frac{\varepsilon}{1+2M}\, dt \\
&\qquad + 2M \int_{-\pi}^{-\delta} |K_n(t)|\, dt \, + 2M \int_{\delta}^{\pi} |K_n(t)|\, dt \Bigg) \\
&\leq \sup_x \frac{1}{2\pi} \Bigg(\frac{\varepsilon}{1+2M} \int_{-\pi}^{\pi} K_n(t)\, dt \\
&\qquad + 2M \int_{-\pi}^{-\delta} \frac{\varepsilon}{1+2M}\, dt \, + 2M \int_{\delta}^{\pi} \frac{\varepsilon}{1+2M}\, dt \Bigg) \\
&\leq \sup_x \frac{1}{2\pi} (2\pi + 4M\pi) \frac{\varepsilon}{1+2M} \\
&= \varepsilon
\end{aligned}
$$

This completes the proof. $\square$

8.6. Convergence with respect to the 2-norm.

The main result in this section is the following.

Theorem. Riesz-Fischer's Theorem for Riemann Integrable Functions. *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π periodic. Then*

$$\|s_n(f) - f\|_2 \to 0 .$$

Frigyes Riesz (22 January 1880 – 28 February 1956) was a Hungarian mathematician who made fundamental contributions to functional analysis.

Ernst Sigismund Fischer (12 July 1875 – 14 November 1954) was a mathematician born in Vienna, Austria. He worked as a professor at the University of Erlangen. His main area of research was mathematical analysis, specifically orthonormal sequences of functions which laid groundwork for the emergence of the concept of a Hilbert space.

However, before we can prove Riesz-Fischer's theorem for Riemann integrable functions, we first prove a version of Riesz-Fischer's theorem for continuous functions. The Riesz-Fischer's theorem for continuous functions will then be used to prove the more general Riesz-Fischer's theorem for Riemann integrable functions.

Theorem. *Let $f:\mathbb{R}\to\mathbb{R}$ be continuous and 2π periodic.*
Then

$$\|s_n(f)-f\|_2 \to 0\,.$$

Proof.
Let $\varepsilon>0$.
It follows from Fejer's Theorem that there is a positive integer N such that

$$\|\sigma_N(f)-f\|_\infty \le \frac{\varepsilon}{2\sqrt{2\pi}}\,. \tag{1}$$

Next, since clearly the function $\sigma_N(f)$ is a linear combination of of the exponential functions e_k for $k\in\{-N,\dots,0,\dots,N\}$, there exist complex numbers λ_k for $k\in\{-N,\dots,0,\dots,N\}$ such that

$$\sigma_{n_0}(f)=\sum_{k=-N}^{N}\lambda_k e_k\,.$$

Hence, for $n\ge N$, we have

$$\begin{aligned}
s_n\big(\,\sigma_N(f)\,\big) &= s_n\left(\sum_{k=-N}^{N}\lambda_k e_k\right)\\
&=\sum_{k=-N}^{N}\lambda_k s_n(e_k)\\
&=\sum_{k=-N}^{N}\lambda_k\sum_{l=-n}^{n}c_l(e_k)e_l\\
&=\sum_{k=-N}^{N}\lambda_k\sum_{l=-n}^{n}\left(\frac{1}{2\pi}\int_0^{2\pi}e_k\overline{e_l}\right)e_l\\
&=\sum_{k=-N}^{N}\lambda_k\sum_{l=-n}^{n}\begin{pmatrix}1 \text{ for } l=k\\ 0 \text{ for } l\neq k\end{pmatrix}e_l\\
&=\sum_{k=-N}^{N}\lambda_k e_k && [\text{since } |k|\le N\le n]\\
&=\sigma_N(f)\,. && (2)
\end{aligned}$$

For brevity write $g=\sigma_N(f)$ and note that (1) implies that $\|g-f\|_\infty\le\frac{\varepsilon}{2\sqrt{2\pi}}$ for $n\ge N$ and that (2) implies that $s_n(g)=g$ for $n\ge N$.

For $n \geq N$ we now have

$$\|s_n(f) - f\|_2 \leq \|s_n(f) - s_n(g)\|_2 + \|s_n(g) - g\|_2 + \|g - f\|_2 \,. \tag{3}$$

Next, using the fact that $s_n(g) = g$ (see (2)), we conclude from (3) that

$$\begin{aligned} \|s_n(f) - f\|_2 &\leq \|s_n(f) - s_n(g)\|_2 + \|s_n(g) - g\|_2 + \|g - f\|_2 \\ &= \|s_n(f) - s_n(g)\|_2 + \|g - f\|_2 \\ &= \|s_n(f - g)\|_2 + \|g - f\|_2 \,. \end{aligned} \tag{4}$$

Finally, using the fact that $\|s_n(f - g)\|_2 \leq \|f - g\|_2$ (see ????), we conclude from (4) and (1) that

$$\begin{aligned} \|s_n(f) - f\|_2 &\leq \|f - g\|_2 + \|g - f\|_2 \\ &= 2\left(\int_0^{2\pi} |g - f|^2 \right)^{\frac{1}{2}} \\ &\leq 2\left(\int_0^{2\pi} \|g - f\|_\infty^2 \right)^{\frac{1}{2}} \\ &\leq 2\sqrt{2\pi}\, \|g - f\|_\infty \\ &\leq \varepsilon \,. \end{aligned}$$

This completes the proof. $\square$

Below we use the following notation. Namely, if $E \subseteq \mathbb{R}$, then we define the function $\mathbf{1}_E \mathbb{R} \to \mathbb{R}$ by

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

Proposition. *Let $a, b \in \mathbb{R}$ with $a \leq b$. Let $f : [a,b] \to \mathbb{R}$ be Riemann integrable and $\varepsilon > 0$*

(1) *There is a partition $\{a = x_0 < x_1 < \cdots < x_N = b\}$ of $[a,b]$ and real numbers $\lambda_1, \ldots, \lambda_N$ such that if we put*

$$s = \sum_{i=1}^{N} \lambda_i \, \mathbf{1}_{(x_{i-1}, x_i]} \,,$$

then

$$\|f - s\|_2 \leq \varepsilon \,.$$

(2) *There is a continuous function $g : [a,b] \to \mathbb{R}$ such that*

$$\|f - g\|_2 \leq \varepsilon \,.$$

Proof.
(1) First npote that since f is Riemann integrable, the function f is bounded, i.e. there is a constat K such that $|f(x)| \leq K$ for all $x \in [a,b]$.

Also, since f is Riemann integrable there is a partition $P = \{a = x_0 < x_1 < \cdots < x_N = b\}$ of $[a,b]$ such that

$$U(f,P) - L(f,P) \leq \frac{\varepsilon^2}{2K} \,.$$

Now put

$$s = \sum_{i=1}^{N} f(x_i) \, \mathbf{1}_{(x_{i-1}, x_i]} \,,$$

and note that it follows from the definition of s that if $x \in [a,b]$, then $|s(x)| \leq \max_i |f(x_i)| \leq \sup_{y \in [a,b]} |f(y)| \leq K$.

We now have

$$
\begin{aligned}
\|f-s\|_2^2 &= \int_a^b |f-s|^2 \\
&= \int_a^b |f-s|\,|f-s| \\
&\leq \int_a^b |f-s|\,(|f|+|s|) \\
&\leq 2K\int_a^b |f-s| \qquad\qquad [\text{since } |f|\leq K \text{ and } |s|\leq K] \\
&= 2K\sum_{i=1}^N \int_{x_{i-1}}^{x_i} |f-f(x_i)| \\
&\leq 2K\sum_{i=1}^N \int_{x_{i-1}}^{x_i} \Big(M(f,[x_{i-1},x_i]) - m(f,[x_{i-1},x_i])\Big) \\
&= 2K\sum_{i=1}^N \Big(M(f,[x_{i-1},x_i]) - m(f,[x_{i-1},x_i])\Big)\,(x_i-x_{i-1}) \\
&= 2K\left(\sum_{i=1}^N M(f,[x_{i-1},x_i])\,(x_i-x_{i-1}) - \sum_{i=1}^N m(f,[x_{i-1},x_i])\,(x_i-x_{i-1})\right) \\
&= 2K(U(f,P)-L(f,P)) \\
&\leq \varepsilon^2\,.
\end{aligned}
$$

(2) It follows from Part (1) that there is a partition $\{a=x_0<x_1<\cdots<x_N=b\}$ of $[a,b]$ and real numbers $\lambda_1,\dots,\lambda_N$ such that if we put

$$
s=\sum_{i=1}^N \lambda_i\,\mathbf{1}_{(x_{i-1},x_i]}\,,
$$

then

$$
\|f-s\|_2\leq\frac{\varepsilon}{2}\,.
$$

Write $M=\max_{i=1,\dots,N}|\lambda_i|$ and put

$$
\delta=\frac{(\varepsilon/2)^2}{2(2M)^2(N-1)}\,.
$$

Next define the function $g:[a,b]\to\mathbb{R}$ by

$$
\begin{aligned}
g(x) &= \lambda_1 && \text{for } x\in[x_0,x_1-\delta]\,, \\
g(x) &= \lambda_2 && \text{for } x\in[x_1+\delta,x_2-\delta]\,, \\
g(x) &= \lambda_3 && \text{for } x\in[x_2+\delta,x_3-\delta]\,, \\
&\;\vdots \\
g(x) &= \lambda_{N-2} && \text{for } x\in[x_{N-3}+\delta,x_{N-2}-\delta]\,, \\
g(x) &= \lambda_{N-1} && \text{for } x\in[x_{N-2}+\delta,x_{N-1}-\delta]\,, \\
g(x) &= \lambda_N && \text{for } x\in[x_{N-1}+\delta,x_N]\,,
\end{aligned}
$$

and extend g (uniquely) to a piecewise linear continuous function on $[a, b]$.

Then

$$\begin{aligned}
\|s-g\|_2^2 &= \int_a^b |s-g|^2 \\
&= \sum_{i=1}^{N-1} \int_{x_i-\delta}^{x_i+\delta} |s-g|^2 \\
&\leq \sum_{i=1}^{N-1} \int_{x_i-\delta}^{x_i+\delta} (|s|+|g|)^2 \\
&= (2M)^2 \sum_{i=1}^{N-1} \int_{x_i-\delta}^{x_i+\delta} \\
&= (2M)^2 (N-1) 2\delta \\
&\leq \left(\frac{\varepsilon}{2}\right)^2 ,
\end{aligned}$$

whence

$$\begin{aligned}
\|f-g\|_2 &\leq \|f-s\|_2 + \|s-g\|_2 \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon .
\end{aligned}$$

This completes the proof. □

We can now prove Riesz-Fischer's Theorem for Riemann Integrable Functions.

Proof of Riesz-Fischer's Theorem for Riemann Integrable Functions.
Let $\varepsilon > 0$.

It follows from the previous proposition that there is a continuous function $g : \mathbb{R} \to \mathbb{R}$ that is 2π periodic. such that

$$\|f - g\|_2 \leq \frac{\varepsilon}{2}\,.$$

Next, since $g : \mathbb{R} \to \mathbb{R}$ is continuous and 2π periodic, we conclude from Theorem ??? $\|s_n(g) - g\|_2 \to n$ as $n \to \infty$, and we can therefore find a positive integer N such that if $n \geq N$, then

$$\|s_n(g) - g\|_2 \leq \frac{\varepsilon}{2}\,.$$

For $n \geq N$, we now have

$$\begin{aligned}
\|s_n(f) - f\|_2 &\leq \|s_n(f) - s_n(g)\|_2 + \|s_n(g) - g\|_2 + \|g - f\|_2 \\
&= \|s_n(f - g)\|_2 + \|s_n(g) - g\|_2 + \|g - f\|_2\,. \qquad (1)
\end{aligned}$$

Using the fact that $\|s_n(f - g)\|_2 \leq \|f - g\|_2$ (see ????), we now conclude from (1) that

$$\begin{aligned}
\|s_n(f) - f\|_2 &\leq \|f - g\|_2 + \|s_n(g) - g\|_2 + \|g - f\|_2 \\
&= 2\|f - g\|_2 + \|s_n(g) - g\|_2 \\
&\leq 2\frac{\varepsilon}{4} + \frac{\varepsilon}{2} \\
&= \varepsilon\,.
\end{aligned}$$

This completes the proof. $\square$

The following corollaries, showing that integrable functions are uniquely determined by their Fourier coefficients, follow immediately from the previous theorem.

Corollary. Uniqueness Theorem for Fourier Series of Riemann Integrable Functions. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be Riemann integrable on $[0, 2\pi]$ and 2π periodic.*

Then the following statements are equivalent:

(a) *For all $n \in \mathbb{Z}$, we have*

$$c_n(f) = c_n(g)\,.$$

(b) *We have*

$$\int_0^{2\pi} |f - g| = 0\,.$$

Proof.
(a) $\Rightarrow$ (b): Since $c_n(f) = c_n(g)$ for all $n \in \mathbb{Z}$, we conclude that $\sigma_n(f) = \sigma_n(g)$ for all $n \in \mathbb{Z}$. It therefore follows from the previous theorem that

$$\begin{aligned}
\|f - g\|_2 &\le \|f - \sigma_n(f)\|_2 + \|\sigma_n(f) - \sigma_n(g)\|_2 + \|\sigma_n(g) - g\|_2 \\
&= \|f - \sigma_n(f)\|_2 + \|\sigma_n(g) - g\|_2 \\
&\to 0\,.
\end{aligned}$$

This shows that $\|f - g\|_2 = 0$.

Next, note that Cauchy-Schwartz's inequality and the fact that $\|f - g\|_2 = 0$ imply that

$$\begin{aligned}
\int_0^{2\pi} |f - g| &= \|f - g\|_1 \\
&= \|(f - g)\,1\|_1 \\
&\le \|f - g\|_2\,\|1\|_2 \\
&= \|f - g\|_2 \left(\int_0^{2\pi} 1^2\right)^{\frac{1}{2}} \\
&= \sqrt{2\pi}\,\|f - g\|_2 \\
&= 0\,.
\end{aligned}$$

(b) $\Rightarrow$ (a): For each integer $n \in \mathbb{Z}$, we have

$$\begin{aligned}
|c_n(f) - c_n(g)| &= |c_n(f - g)| \\
&= \left|\frac{1}{2\pi}\int_0^{2\pi} (f - g)\,\overline{e_n}\right| \\
&\le \frac{1}{2\pi}\int_0^{2\pi} |f - g|\,|\overline{e_n}| \\
&\le \frac{1}{2\pi}\int_0^{2\pi} |f - g| \\
&= 0\,.
\end{aligned}$$

This completes the proof. $\square$

Corollary. Uniqueness Theorem for Fourier Series of Continuous Functions. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be continuous and 2π periodic.*

Then the following statements are equivalent:

(a) *For all $n \in \mathbb{Z}$, we have*

$$c_n(f) = c_n(g)\,.$$

(b) *We have*

$$f = g\,.$$

Proof.
This follows immediately from the previous uniqueness theorem since for continuous functions $f, g : \mathbb{R} \to \mathbb{R}$, we know that $\|f - g\|_1 = \int_0^{2\pi} |f - g| = 0$ if and only if $f = g$. $\square$