FH Aachen
University of Applied Sciences

Campus Jülich

Department „Medical Engineering and Technomathematics“

Degree Program: Computer Science

Application of the Johnson-Lindenstrauss Lemma for distortion-minimizing embedding of points from a high-dimensional space into a low-dimensional Euclidean space

Term paper by Morteza Montahaee

1. Examiner: Prof. Dr. rer. Nat. Martin Reißel

Aachen, May 2023

Table of contents

1 Introduction

Data can be modeled in several procedures for further processing. Sometimes they can be described as m data records with n features in a relational database. However, they can also be mined in the course of data mining or a numerical analysis. The dimensionality of the data strongly affects resource consumption (runtime and storage requirements) during processing. Using a data mapping can contribute to the reduction of dimensionality without loss of essential data features and thus to the reduction of resource loads during data processing.

1.1 Motivation

An important conceptual component of dimensionality reduction occurs in the Johnson-Lindenstrauss lemma. It states that any n data points in higher-dimensional Euclidean space can be mapped to a lower dimension k without exceeding a factor $1 \pm \varepsilon$ of distortion of the Euclidean distance between any two points (Figure 1* [L. JP009]).

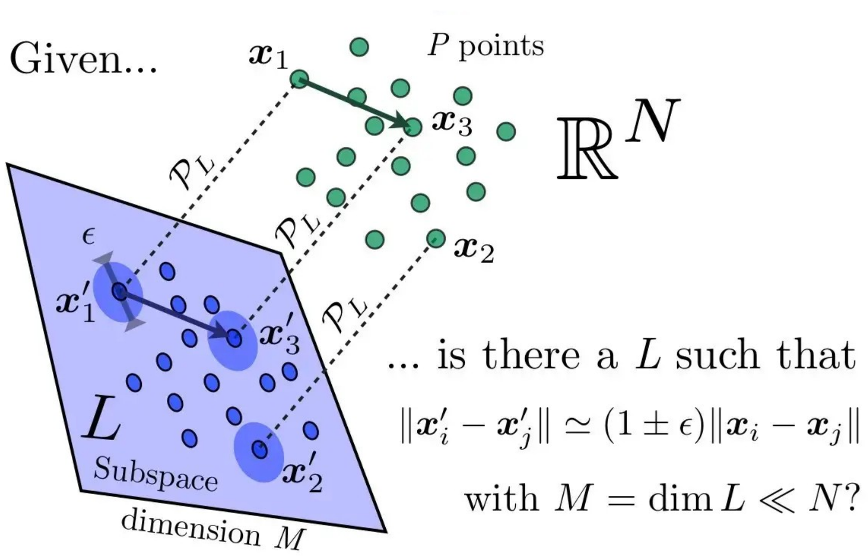

Figure 1*: A linear dimensionality reduction.

---

* Bildquelle von Laurent Jacques

2 Preliminary to Measure Theory

The sets whose measure we can define by virtue of the preceding ideas we will call measurable sets; we do this without intending to imply that it is not possible to assign a measure to other sets.

E. Borel, 1898

We start with the brief explanation of some elementary principles that form the basis for the theory that follows I the next section, namely measure theory. The key idea in calculating the volume or measure of a subset $\Re^d (d \geq 1)$ is to approximate this set by unifying another set whose topological geometry is easier to understand or whose volumes are easier to calculate. The word „Volumen" appropriately refers here to the measure in $\Re^d$. For example, this corresponds to the length for $d=1$ or area for $d=2$. Following this approach, we will use rectangles and cubes as the essential building blocks of the theory of measure. In other words, it is considered $\Re$ in Intervals and in $\Re^d (d \geq 2)$ products of intervals and thus the well-definedness1 of a volume in any dimension can be guaranteed as a result of the production of all interval lengths under Cartesian product.

As standard notation [RAR0051] we assume the following:

1. A point $x \in \Re^d$ consists of a tuple of real numbers.:
$$x = (x_1, x_2, \ldots, x_d), \qquad x_i \in \Re, \forall i \in \{1, \ldots, d\}$$
2. The addition between two points as well as multiplication of a point with a real number are defined component-wise:
$$x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_d + y_d); \lambda x = (\lambda x_1, \lambda x_2, \ldots, \lambda x_d), \qquad \forall\, x, y \in \Re^d, \lambda \in \Re$$
3. The norm of x is denoted with $|x|$ as standard the Euclidean norm as follows:
$$|x| = (x_1^2 + x_2^2 + \ldots, x_d^2)^{\left(\frac{1}{2}\right)}$$
4. The set $\Re^d$ of d-tuples with the above properties (1- 3) form a **vector space** over the reals.
5. From 2 and 3, the distance between two points simply follows under the $|x - y|$.
6. The absolute complement of a set $E \subset \Re^d$ is defined by
$$E^c = \{x \in \Re^d : x \notin E\}$$
7. For sets $E, F \subset \Re^d$ the relative complement (difference) is consequently defined by
$$E \setminus F = \{x \in \Re^d : x \in E, x \notin F\} = E \cap F^c$$
8. The distance between two sets $E, F \subset \Re^d$ is defined by
$$d(E, F) = \inf|x - y| = \max_{d \in \Re} \{d \leqslant |x - y|\}$$

2.1 Open, Closed, and Compact Sets

Definitions 2.1.1

1. The **open ball** of $\Re^d$ is defined by the center *x* and the radius *r*:
$$B_r(x)=\{x\in\Re^d : |y-x|<r\}$$
2. A set $E\subset\Re^d$ is open if the following holds:
$$\forall x\in E\,\exists\, r>0 : B_r(x)\subset E$$
3. A set $E\subset\Re^d$ is closed if its complement is open.

4. A set $E\subset\Re^d$ is countable if it is finite or there exists a bijective[1] function $f:\mathbb{N}\to E$, where $\mathbb{N}$ the set of natural numbers.

5. A set $E\subset\Re^d$ is bounded if it lies in a ball with a finite radius.

6. A bounded set $E\subset\Re^d$ is compact if it is also closed. In other words, $E\subset\Re^d$ is compact if the following statement is fulfilled:

   Whenever *E* lies in the union of open sets, then *E* lies in the union of finitely many of these sets. This is also called the covering theorem[2].
$$E\subset\bigcup_j G_j\to E\subset\bigcup_{j=1}^{N} G\alpha_j$$
7. A point $x\in\Re^d$ is **Limit Point** of $E\subset\Re^d$, if the following holds true:
$$\forall\, r>0\,\exists\, y\neq x : y\in E\cap B_r(x)$$
8. A point $x\in E\subset\Re^d$ is an isolated point, if the following is fulfilled::
$$\exists\, r>0 : \{x\}=B_r(x)\cap E$$
9. A point $x\in E\subset\Re^d$ is an inner point, if the following holds:
$$\exists\, r>0 : B_r(x)\subset E$$
10. A set $\mathring{E}\subseteq E\subset\Re^d$ is called the **inner** set of *E*, if the following holds true:
$$\mathring{E}=\{x\in E : B_r(x)\subset E\,;\, r>0\}$$
11. A **Closure** [FJIR001] $\bar{E}$ of the set $E\subset\Re^d$ is defined by
$$\bar{E}=\{x\in\Re^d : x\in E\vee x\ \mathit{ist\ limit\ Punkt\ der}\ E\}$$
12. The Closure of a set is thus closed, and a closed set contains all its limit points.
13. **Boundary** of a set $E\subset\Re^d$, ∂E is the set of points that lie in $\bar{E}$ but at the same time do not belong to the inner set/space of *E*, $\mathring{E}$ and

1 Bijectivity means that for all $e\in E$ there is exactly a $n\in\mathbb{N}$ mit $f(n)=e$. So the elements of E und $\mathbb{N}$ are paired.

2 It is also called Heine -Borel’s theorem.

$$\partial E = \bar{E} \setminus \breve{E} = [\breve{E} \cup (E^c)]^c$$

14. A set $E \subset \mathfrak{R}^d$ is **perfect**, if the set E has no isolated point.

15. A set $E \subset \mathfrak{R}^d$ is c**onvex**, if $E = \{x + t(y - x) |\, x,\ y \in E \wedge t \in [0,\ 1]\}$.

Lemma 2.1.2

Let F, $K \subset \mathfrak{R}^d$ be closed, compact and both disjoint, then *d(F, K) > 0.*

Proof: Since F is closed, there exists for each point $x \in K, \delta_x > 0$ with $d(x, F) > 3\delta_x{}^{(*)}$. Because K is compact, and K is covered by $\bigcup_{x\in K}^{\infty} B_{2\delta x}(x)$, we can determine a partial cover $\bigcup_{j=1}^{N} B_{2\delta_j}(x_j)^{(**)}$. If now $\delta = \min(\delta_1, ..., \delta_N)$, then $d(K, F) > \delta > 0$. In In this context, then for each $x \in K$, $y \in F$ and a j $|x_j - x| \leq 2\delta_j{}^{(**)}$ and $|y - x_j| \geq 3\delta_j{}^{(*)}$ *respectively*. Therefore

$$|y - x| \geq |y - x_j| - |x_j - x| \geq 3\delta_j - 2\delta_j \geq \delta. \blacksquare$$

2.2 Rectangles in $\mathfrak{R}^d$[1]

Given a closed rectangle $R \subset \mathfrak{R}^d$ defined by the product of *d* one-dimensional closed and bounded intervals

$$R = \prod_{j=1}^{n} [a_j, b_j] = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d],$$

where $a_j \leqslant b_j \in \mathfrak{R}$ *;* for all $j = 1, 2, \ldots, d$. This can also be reformulated as follows:

$$R = \{(x_1, x_2, \ldots, x_d) \in \mathfrak{R}^d : a_j \leqslant x_j \leqslant b_j; \quad j = 1, 2, \ldots, d\}.$$

2.3 Remarks

1. A rectangle here is closed and its sides are parallel to the coordinate axes. Precisely, rectangles in $\mathfrak{R}$ (Figure 2) are the closed bounded intervals, while they are in $\mathfrak{R}^2$ (Figure 3) the usual four-sided rectangles and in $\mathfrak{R}^3$ (Figure 4) the closed parallelepipeds.

2. We assume that the side lengths of the rectangle *R*, are $b_1 - a_1, \ldots, b_d - a_d$.

3. The volume of the rectangle *R*, |*R*| is defined [RAR0051] by
$$|R| = (a_1 - b_1) \times (a_2 - b_2) \times \cdots \times (a_d - b_d).$$
Here, the term „Volume“ corresponds to the length for *d = 1*, and the area for d= 2.

4. An open rectangle is the product of open intervals and therefore the inner set/space of the rectangle *R*, $\mathring{R}$ is defined by
$$\mathring{R} = (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d).$$

5. A cube Q is also considered as a rectangle with $b_1 - a_1 = b_2 - a_2 = \cdots = b_d - a_d = l$ and thus $|Q| = l^d$.

1 A generalization of the rectangle and the cuboid to any dimensions number is called a hyper rectangle in geometry.

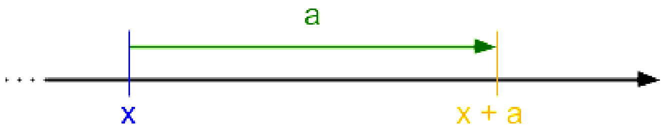

Figure 2: A rectangle in $\Re$

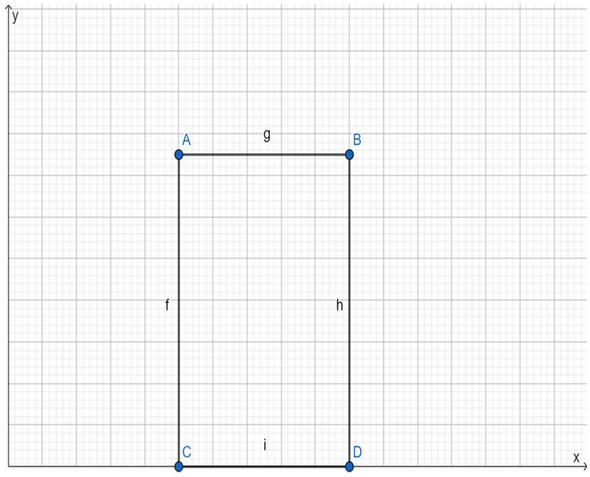

Figure 3: A rectangle in $\mathbb{R}^2$

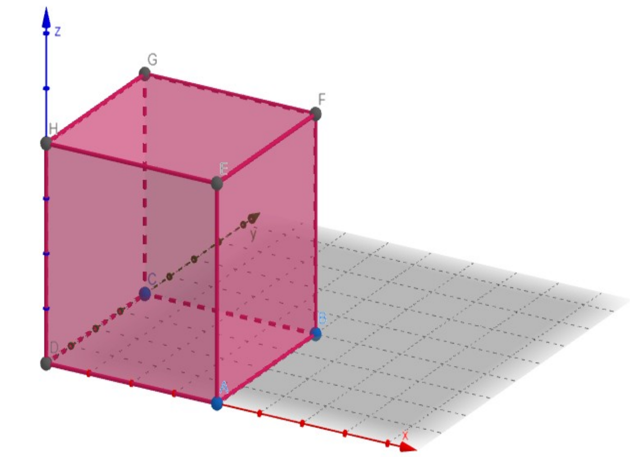

Figure 4: A rectangle in $\mathbb{R}^3$

Lemma 2.3.1

If a rectangle R is the union of finitely many rectangles, then the following equation is valid:

$$R = \bigcup_{k=1}^{N} R_k \rightarrow |R| = \sum_{k=1}^{N} |R_k|$$

Proof: We consider the grid, which is created by an extension of the edges or sides of all rectangles $R_1,\ldots, R_N$ (Figure 5). This construction results in finitely many rectangles $\widetilde{R}_1, \widetilde{R}_2, \ldots, \widetilde{R}_d$ *and a partition* $J_1,\ldots,J_N$ *of positive integers between one and* M, so that the two following unions remain under almost disjoint property $R = \bigcup_{j=1}^{N} \widetilde{R}_j \wedge R_k = \bigcup_{j \in J_k}^{N} \widetilde{R}_j$ (Figure 6).

For example, for the rectangle R is $|R| = \sum_{j=1}^{M} |\widetilde{R}_j|$, because the grid actually partitions the sides of R and each $\widetilde{R}_j$ is made up of the products of the intervals in the divisions. So when we sum the volumes of, we are summing the corresponding products of the length of the resulting intervals. Since this also holds to the other rectangles $R_1,\ldots, R_N$ gilt, we conclude the following:

$$|R| = \sum_{j=1}^{M} |\widetilde{R}_j| = \sum_{k=1}^{N} \sum_{j \in J_k} |\widetilde{R}_j| = \sum_{k=1}^{N} |R_k| .$$

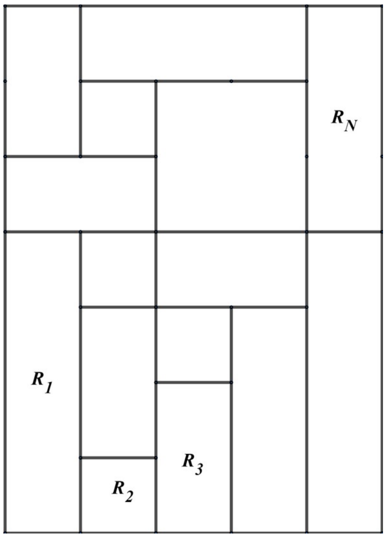

Figure 5: Creation of the grid through the rectangles

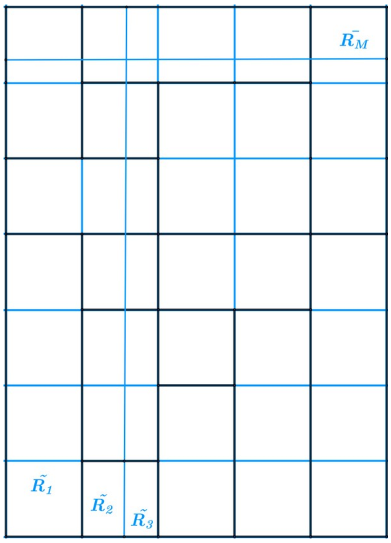

Figure 6: A constructed grid

Lemma 2.3.2

Let R, $R_1,\ldots, R_N$ *be rectangles defined as above* with $R \subset \bigcup_{k=1}^{N} R_k$, then: $|R| \leq \sum_{k=1}^{N} |R_k|$.

Proof: The main idea here is to reconstruct the grid this time by extending all sides of the rectangles $R_1,\ldots, R_N$ but also the rectangle R. Thus, the associated J_k sets (see the proof for the last lemma) must no longer be disjoint from each other. Otherwise, everything will simply go as follows:

$$0 \leq \sum_{\substack{j=1 \\ j\neq l}}^{M} |\widetilde{R}_j \cap \widetilde{R}_l| \leq |\hat{R}| - |R| = \sum_{k=1}^{N} |R_k| - |R| \rightarrow |R| \leq \sum_{k=1}^{N} |R_k|$$

In the following section, the structure of open sets with respect to cubes/cubes is described in concrete terms.

Theorem 2.3.3

Every open subset G of $\mathfrak{R}^1$ can be uniquely represented as a countable union of disjoint open intervals.

Proof: For each $x \in G$ we denote I_x as the largest interval that contains x and is in G. Since G is open, x lies in a smaller (not necessarily trivial) interval. For this reason $a_x < x < b_x$, where a_x and b_x are defined with possibly unbounded values by:

$$a_x = \inf\{a < x : (a, x) \subset G\} \quad \wedge \quad b_x = \sup\{b > x : (x, b) \subset G\}$$

If we now set $I_x = (a_x, b_x)$, we then have due to the construction $x \in I_x \wedge I_x \subset G$. This results:

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

We now assume that there would be at least two intervals I_x, I_y that intersect themselves, thus their narrowing, which is also an open interval, lies in G *and contains* x. *Because* I_x is maximal, it must $(I_x \cup I_y) \subset I_x$. The same must apply to I_y i.e. $(I_x \cup I_y) \subset I_y$. This can only happen if $I_x = I_y$. Therefore, any two different intervals in the set family $\Lambda = \{I_x\}_{x \in G}$ consists only of countably many distinct intervals. However, this is easily recognizable because every open interval I_x

1 The real set $\mathfrak{R}$ is not countable. An elegant proof of this has already been written by Mr. Georg Cantor.

contains a rational number[2]. Because different intervals are disjoint, they must contain different rational numbers and are therefore countable. This completes the proof.▪

Corollary 2.3.4

Regarding disjoint open intervals $I_j s$, if $G = \bigcup_{j=1}^{\infty} I_j$ is open, then measure of G *is equal with* $\sum_{j=1}^{\infty} |I_j|$.

Since this representation is unambiguous, we can take this as the definition of measure.

We would then state that whenever G_1 and G_2 are open and disjoint, the measure of their union is the sum of their measures.

Theorem 2.3.5

Every open subset G of $\Re^d$, $d \geq 1$, can be represented as a countable union of almost disjoint closed cubes.

Proof: We need to construct a countable set family $\aleph$ of closed cubes, whose interiors are disjoint and that is $G = \bigcup_{\chi \in \aleph} \chi$.

As a first step, we consider the grid in $\Re^d$, which arises from an extension of the sides or edges with length 1, where their nodes are integers. In other words, we consider the simple grid, which is parallel to the $\mathbb{Z}^d$-Axis, more precisely, the grid, which is generated by a planar graph or grid graph $\mathbb{Z}^d$. In addition, we will also use the grids that are generated from cubes with edge lengths 2^{-N}. These in turn have come to an end through successive bisection of the original grid. The cubes are either accepted or rejected as part of $\aleph$ in the original grid generation process according to the following rule (Figure 7). The rule is described as follows:

- If χ lies completely in G, it is accepted..
- If χ intersects both G and G^C, it is tentatively accepted.
- If χ is entirely contained in G^C, it will be rejected.

In a second step (Figure 8) we divide the cubes tentatively accepted into 2^d cubes with the edge length $\frac{1}{2}$; Then we repeat our procedure by accepting the smaller cubes if they lie completely in G, by tentative accepting them if they cut both G *and* G^C, or by rejecting them if they only lie in G^C lie. This procedure is repeated arbitrarily due to the construction and thus the resulting set family $\aleph$ from all permissible cubes is countable and at the same time consists of almost disjoint cubes.

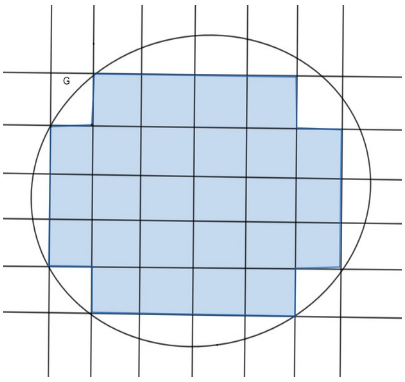

Figure 7: The first step in Grid generation process for an open space in $\mathbb{R}^2$.

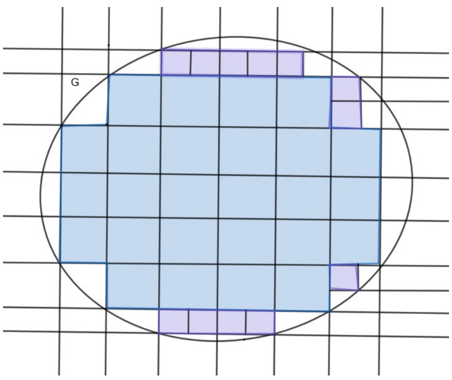

Figure 8: The second step in Grid generation process for an open space in $\mathbb{R}^2$.

2 The set of rationale numbers is dense in $\Re$

To confirm that the union of the cubes still lies in G, we note that for $x \in G$ there is a cube with an edge length 2^{-N} (obtained from successive halving or bisections of the original grid), which contains x and which lies completely in G. This cube is either allowed or lies in a cube that had been widely allowed. With this it follows that the union of all cubes in $\aleph$ covers G. So if $G = \bigcup_{j=1}^{\infty} R_j$, where the rectangles R_j are almost disjoint. Due to the knowledge available, the following measure size for G is to be shown: $G = \sum_{j=1}^{\infty} |R_j|$.This is understandable, since the boundary volume of a rectangle should be zero, and overlap or intersection of the rectangles is not added to the volume G. We note that the decomposition into cubes described above is not unique. In addition, independence between the sum of the decomposition is not immediately apparent. In a sense, volume and area are for $\Re^d$ with $d \geq 2$ even if it comes to open spaces .▪

Preface: The general theory (measure theory) developed in the next section actually contains a concept of volume that corresponds to the decompositions of open spaces in the last two sentences, and holds to all dimensions (i.e. for all open spaces $\Re^d$ with $d \geq 1$).

2.4 Outer Measure and its Properties

The concept of outer measure is the first of two important concepts that one needs to develop a measure theory. We start with the definition and basic properties for an outer measure.

The outer/exterior measure [RAR0051], as its name suggests, should describe the volume of a set or a space $E \subseteq \Re^d$ through outer approximation. The space is covered by cuboids or cubes and when the cover becomes finer, there should not be more overlap of the cubes. Thus, the volume of E should approach the sum of the volumes of the cubes more. The exact definition of the outer measure is as follows:

Definition 2.4.1 Outer Measure

For a subspace E in $\Re^d$, an **outer measure „m_*“** is defined by

(1) $$m_*(E) = \inf \sum_{j=1}^{\infty} |Q_j|$$

Here, the infimum is calculated over all countable coverings $E \subset \bigcup_{j=1}^{\infty} Q_j$ by means of closed cuboids.

Remark 2.4.2

The outer measure is always non-negative but can be infinite. $(0 \leq m_*(E) \leq \infty)$.

Properties 2.4.3

1. The outer measure of a point $p \in \Re^d$ is zero. This becomes very clear when we consider a point as a cube with the volume zero, which of course covers.
2. The outer measure of the empty space is also zero.
3. The outer measure of a closed cube is equal to its volume.

Proof: We assume that Q is a closed cube in $\Re^d$. Since Q itself covers, we must have $0 \leq m_*(Q) \leq |Q|$. Therefore, it is enough to show the inverse inequality namely $m_*(Q) \geq |Q|$. We consider an arbitrary covering by cubes Q, $|Q| \subset \bigcup_{j=1}^{\infty} Q_j$, and note that it is only necessary to prove the following inequality:

(2) $$|Q| \leq \sum_{j=1}^{\infty} |Q_j|$$

For a fixed $\varepsilon > 0$ we choose for each j an open cube S_j, which contains Q_j, and $|S_j| \leq (1+\varepsilon)|Q_j|$. From the open covering $\bigcup_{j=1}^{\infty} S_j$ of the compact space Q we can determine a finite sub-covering, which after possible renumbering of the rectangles can be described as $|Q| \subset \bigcup_{j=1}^{N} S_j$. If we take the closure or Closure of the cubes S_j, we can conclude from Lemma 2.2 sc, that $|Q| \leq \sum_{j=1}^{N} |S_j|$. From this it follows,

$$|Q| \leq (1+\varepsilon) \sum_{j=1}^{N} |Q_j| \leq (1+\varepsilon) \sum_{j=1}^{\infty} |Q_j|.$$

Since ε is arbitrary, it follows that the inequality (2) is fulfilled and thus $m_*(Q) \geq |Q|$.▪

4. Let $E_1 \subset E_2 \subset \mathfrak{R}^d$, is then $m_*(E_1) \leq m_*(E_2)$(this is called **Monotony** for outer measure).

Proof: This inequality follows from the observation or fact that every covering of E_2 by a countable collection of cubes means the same covering of E_1. n particular, it follows from the monotony that every bounded subspace (subspace) of $\mathfrak{R}^d$ has a finite outer measure.

5. If Q is an open cube, it still holds $m_*(Q) = |Q|$.

Proof: Since Q is covered by its own closure $\overline{Q}$, and $|\overline{Q}| = |Q|$, it follows directly that $0 \leq m_*(Q) \leq |Q|$. To prove the reverse inequality, we note the following:

If Q_0 is a closed cube, which lies in Q, then $m_*(Q_0) \leq m_*(Q)$(Monotony). This inequality is due to the fact that every covering of the space Q by countable closed cubes also means a covering of Q_0. Therefore, it is $|Q_0| \leq m_*(Q)$ and because the choice Q_0 with a volume $|Q|$ can be closed to s desired, we must then have $m_*(Q) \geq |Q|$.▪

6. The outer measure of a rectangle R is equal to its volume.

As argued in property 3, it is easy to see $m_*(R) \geq |R|$. To prove the reverse inequality, we now consider a grid $\mathfrak{R}^d$, which arises from cubes with the edge length $\frac{1}{k}$. If then $\mathbb{Q}$ consists of the (finite) collection of all cubes that lie in R, and $\mathbb{Q}$' consists of the (finite) collection of all cubes that intersect the complement of R, R^C, we first note, $R \subset \bigcup_{Q \in \mathbb{Q} \cup \mathbb{Q}'} Q$. A simple argument results from $\sum_{Q \in \mathbb{Q}} |Q| \leq |R|$.In addition, there are cubes[1] in $\mathbb{Q}$' with a volume k^{-d}, so that

$$\sum_{Q \in \mathbb{Q}'} |Q| = O\left(\frac{1}{k}\right).$$

This follows $\sum_{Q \in (\mathbb{Q} \cup \mathbb{Q}')} |Q| \leq |R| + O\left(\frac{1}{k}\right)$. Now as $k \to \infty$, we get $m_*(R) \leq |R|$.▪

7. The outer measure of $\mathfrak{R}^d$ is infinite.

Proof: This follows from the fact that every closure of $\mathfrak{R}^d$ is the same as a closure of any cube $Q \subset \mathfrak{R}^d$ and therefore $m_*(\mathfrak{R}^d) \geq |Q|$. Since Q can have an arbitrarily large volume, must be $m_*(\mathfrak{R}^d) = \infty$.▪

8. Let be $E = \bigcup_{j=1}^{\infty} E_j$, is then $m_*(E) \leq \sum_{j=1}^{\infty} m_*(E_j)$ (Countable Subadditivity).

1 The notation $f(x) = O(g(x))$ means that for a constant C and all x in a specified range of values $|f(x)| \geq C|g(x)|$. In this case there is a Ck^{d-1} cube, als $k \to \infty$.

Proof: First, we assume that $m_*(E_j) < \infty$, otherwise the inequality is clearly fulfilled. For each $\varepsilon > 0$, it follows from the definition of the outer measure for each j a covering $E_j \subset \bigcup_{j=1}^{\infty} Q_{k,j}$, which is created by closed cubes with $\sum_{k=1}^{\infty} |Q_{k,j}| \leq m_*(E_j) + \frac{\varepsilon}{2^j}$. It follows that $E \subset \bigcup_{j,k=1}^{\infty} Q_{k,j}$ forms a covering for the space E by closed cubes. Therefore

$$m_*(E) \leq \sum_{j,k=1}^{\infty} |Q_{k,j}| = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} |Q_{k,j}| \leq \sum_{j=1}^{\infty} \left(m_*(E_j) + \frac{\varepsilon}{2^j} \right) = \sum_{j=1}^{\infty} m_*(E_j) + \varepsilon.$$

Since this is true for every ε > 0 , the proof is finished.▪

9. Let $E \subset \Re^d$, is then $m_*(E) = \inf m_*(G)$, where the infimum covers all open set G, which contained E.

Proof: The monotony property directly provides the inequality $m_*(E) \leq \inf m_*(G)$.To show the reverse inequality, we take ε > 0 and choose cubes Q_j, which $E \subset \bigcup_{j=1}^{\infty} Q_j$ with $\sum_{k=1}^{\infty} |Q_j| \leq m_*(E_j) + \frac{\varepsilon}{2^j}$ fulfill. Let $\hat{Q}_j$ be an open cube with $|\hat{Q}_j| \leq |Q_j| + \frac{\varepsilon}{2^{j+1}}$, which contains Qj. Therefor $G = \bigcup_{j=1}^{\infty} \hat{Q}_j$ is open and from the Countable Sub additivity (8th)Property follows

$$m_*(G) \leq \sum_{j=1}^{\infty} |m_*(\hat{Q}_j)| = \sum_{j=1}^{\infty} |\hat{Q}_j| \leq \sum_{j=1}^{\infty} \left(|Q_j| + \frac{\varepsilon}{2^{j+1}} \right) \leq \sum_{j=1}^{\infty} |Q_j| + \frac{\varepsilon}{2} \leq m_*(E) + \varepsilon.$$

As aresult $\inf m_*(O) \leq m_*(E)$ thus the proof is completed.▪

10. Let $E = E_1 \cup E_2$ and $d(E_1, E_2) > 0$, then we have $m_*(E) = m_*(E_1) + m_*(E_2)$.

Proof: From the Countable Subadditivity(8th)Property we already know $m_*(E) \leq m_*(E_1) + m_*(E_2)$ and therefore the reverse inequality remains to be shown. To do this, we first choose $\delta > 0$ with $\delta < d(E_1, E_2)$. Next, we determine a covering $E \subset \bigcup_{j=1}^{\infty} Q_j$, which is created by closed cubes with $\sum_{k=1}^{\infty} |Q_j| \leq m_*(E_j) + \varepsilon$. After the subdivision of the cubes Q_j we can assume that each Q_J has a diameter less than δ. In this case, each Q_j can intersect at most one of the two sets or spaces E_1 or E_2. If we denote with J_1 and J_2 the set of indices j, for which Q_j intersect E_1 and E_2, then the intersection $J_1 \cap J_2$ is empty, and so we have

$$E_1 \subset \bigcup_{j \in J_1}^{\infty} Q_j \quad \wedge \quad E_2 \subset \bigcup_{j \in J_2}^{\infty} Q_j.$$

This follows

$$m_*(E_1) + m_*(E_2) \leq \sum_{j \in J_1}^{\infty} |Q_j| + \sum_{j \in J_2}^{\infty} |Q_j| \leq \sum_{j=1}^{\infty} |Q_j| \leq m_*(E_j) + \varepsilon.$$

Since this is true for every $\varepsilon > 0$, the proof is completed.▪

11. Let the set E from the countable union of almost disjoint cubes $E = \bigcup_{j=1}^{\infty} Q_j$, then

$$m_*(E) = \sum_{j=1}^{\infty} |Q_j| .$$

Proof: Let $\tilde{Q}_j$ be a strictly contained by Qj cube, for which $|Q_j| \leq |\tilde{Q}_j| + \frac{\varepsilon}{2^j}$ with a fixed ε > 0 holds. This results in the fact that for each N cubes $\tilde{Q}_1, \tilde{Q}_2, \ldots, \tilde{Q}_N$ are disjoint. With regard to the distance of one from another cube and multiple application of property 10, it results that

$$m_*\left(\bigcup_{j=1}^{N} Q_j\right) = \sum_{j=1}^{N} |\widetilde{Q}_j| \geq \sum_{j=1}^{N} \left(|Q_j| - \frac{\varepsilon}{2^j}\right).$$

Since $\bigcup_{j=1}^{N} \widetilde{Q}_j \subset E$, we conclude for each positive integer N, that $m_*(E) \geq \sum_{j=1}^{N} |\widetilde{Q}_j| - \varepsilon$.

As $N \to \infty$, forms for each $\varepsilon > 0$ the inequality $\sum_{j=1}^{\infty} |Q_j| \leq m_*(E) + \varepsilon$ and thus $\sum_{j=1}^{\infty} |Q_j| \leq m_*(E)$. With the combination of the countable sub-additivity property, the desired equation is thus achieved.▪

Remark 2.4.4

The last property (11th property) shows that if a set can be decomposed into almost disjoint cubes, its outer measure can be calculated as the sum of the volumes of the cubes. In particular, it follows from Theorem 2.3.5 that the outer measure of an open set is equal to the sum of the volumes of the cubes in a decomposition. This also simultaneously provides a proof for the independence of the sum from the decomposition.

Remark 2.4.5

Despite the last two properties (10th and 11th properties), one cannot generally conclude that the following statement holds to the disjoint union of two subsets or subspaces of :

$$m_*(E_1 \cup E_2) = m_*(E_1) + m_*(E_2).$$

2.5 Measurable Sets and Lebesgue Measure

The concept of measurability is limited to a collection of subsets in , for which the outer measure fulfills all desired properties including additivity (countable additivity) for disjoint unions of sets.

Definition 2.5.1

A subset E of $\mathfrak{R}^d$ is called Lebesgue-measurable or simply measurable if for every $\varepsilon > 0$ there exists an open set O with the following properties:

$$E \subset G \quad \wedge \quad m_*(G \setminus E) \leq \varepsilon.$$

This is comparable to the ninth property in 2.4.3, which holds to all sets E.

If E is measurable, we define its Lebesgue measure (or measure) *m(E)* by $m(E) = m_*(E)$. Thus, a Lebesgue measure inherits all properties of the outer measure in 2.4.3.

Properties 2.5.2

1. Every open set $G \subset \mathfrak{R}^d$ is measurable.

Proof: It follows directly from Definition 2.5.1.

2. If $m_*(E) = 0$, then E is measurable. In particular, if F is a subset of a set with outer measure 0, then F is consequently measurable.

Proof: According to Property 9 in 2.4.3, we know that for every $\varepsilon > 0$ there exists an open set G with $E \subset G \ \wedge \ m_*(G) \leq m_*(E) + \varepsilon \leq \varepsilon$. Since $(G \setminus E) \subset G$, the Monotony Property (4th Property in 2.4.3) yields $m_*(G \setminus E) < \varepsilon$.▪

3. A countable union of measurable sets is again measurable.

Proof: Let $E = \bigcup_{j=1}^{\infty} E_j$, where each E_j is measurable. Given $\varepsilon > 0$. We can then for each index j an open set O_j with $E_j \subset G_j \ \wedge \ m_*(G_j \setminus E_j) < \frac{\varepsilon}{2^j}$.Therefore, the union $G = \bigcup_{j=1}^{\infty} G_j$ is open, $E \subset G$, and

$(G \setminus E) \subset \bigcup_{j=1}^{\infty} (G_j \setminus E_j)$. Now from Monotony and Subadditivity properties of the outer measure (Properties 4 and 8 in 2.4.3) follows

$$m_*(G \setminus E) \leq \sum_{j=1}^{\infty} m_*(G_j \setminus E_j) \leq \varepsilon. ▪$$

4. Closed sets are measurable.

Proof: We show instead that compact sets are measurable, because every closed set F can be described as the union of compact sets. So $F = \bigcup_{k=1}^{\infty} (F \cap B_k)$. Here we denote B_k as a closed ball with radius k and origin as the center. With the application of the last property (Property 3), the proof is then complete. So let F be compact (so that in particular $m_*(F) < \varepsilon$ is fulfilled). According to the 9th property of the outer measure, we can for every $\varepsilon > 0$ choose an open set G with $F \subset G \;\wedge\; m_*(G) \leq m_*(F) + \varepsilon$. Since F is closed, the difference $G \setminus F$ is open. According to Theorem 2.3.5, we can capture this difference as a countable union of almost disjoint cubes

$$(G \setminus F = \bigcup_{j=1}^{\infty} Q_j).$$

For a fixed N, the finite union $K = \bigcup_{j=1}^{N} Q_j$ is compact. From Lemma 2.1.2 it follows $d(K,F) > 0$. Since $(K \cup F) \subset G$, it follows from the Properties 4, 10, and 11 in 2.4.3 that

$$m_*(G) \geq m_*(F) + m_*(K) = m_*(F) + \sum_{j=1}^{N} m_*(Q_j).$$

Consequently $\sum_{j=1}^{N} m_*(Q_j) \leq m_*(G) - m_*(F) \leq \varepsilon$. The last inequality also holds to $N \to \infty$.
Using the Subadditivity property of the outer measure, one finally finds

$$m_*(G \setminus F) \sum_{j=1}^{N} m_*(Q_j) \leq \varepsilon. ▪$$

5. The complement of a measurable set is again measurable.

Proof: If E is measurable, we can then for every positive integer n choose an open set G_n with $E \subset G_n \quad \wedge \quad m_*(G_n \setminus E) \leq \frac{1}{n}(*)$. The complement G_n^c is closed and therefore, from the last (4th) property, it is measurable. From the 3rd property, it follows that the countable union $S = \bigcup_{n=1}^{\infty} G_n^c$ is also measurable. We note with (*) that $S \subset E^c$ and $(E^c \setminus S) \subset (G_n \setminus E)$ with $m_*(E^c \setminus S) \leq \frac{1}{n}$ for all n exist. As a result with $n \to \infty$

$m_*(E^c \setminus S) = 0$ and according to the 2nd property $E^c \setminus S$ is therefore measurable. Since S and $(E^c \setminus S)$ are both measurable, their union, namely E^c is again measurable according to the 3rd property.▪

6. A countable intersection of measurable sets is again measurable.

Proof: From the Properties 3, and 5 and De Morgan's laws for measurable E_j follows

$$\bigcap_{j=1}^{\infty} E_j = \left(\bigcup_{j=1}^{\infty} E_j^c \right)^c. ▪$$

Theorem 2.5.3

Let $E_1, E_2, \ldots$, be disjoint measurable sets and $E = \bigcup_{j=1}^{\infty} E_j$, then $m(E) = \sum_{j=1}^{\infty} m(E_j)$.

Proof: First, we further assume that each E_j s bounded. Consequently, for each j by means of the definition of measurability to E_j^c a closed subset F_j of E_j $(F_j \subseteq E_j)$ with $m_*(E_j \setminus F_j) \leq \frac{\varepsilon}{2^j}$. For each fixed N the set F_1, …, F_N are compact (due to the boundedness of the E_j) and disjoint, so that $m(\bigcup_{j=1}^{N} F_j) = \sum_{j=1}^{N} m(F_j)$. From $\bigcup_{j=1}^{N} F_j \subset E$ it follows

$$m(E) \geq \sum_{j=1}^{N} m(F_j) \geq \sum_{j=1}^{N} m(E_j) - \varepsilon.$$

So $m(E) \geq \sum_{j=1}^{\infty} m(E_j)$, as $N \to \infty^{(*)}$. According to the Subadditivity Property (8th Property in 2.4. 3) the reverse inequality holds $m(E) \leq \sum_{j=1}^{\infty} m(E_j)$, which leads to the equation $m(E) = \sum_{j=1}^{\infty} m(E_j)$.

We choose in general an arbitrary sequence of cubes $\{Q_k\}_{k=1}^{\infty}$, which grows to $\mathfrak{R}^d$ in the sense that $\forall\, k \geq 1,\ Q_k \subset Q_{k+1}$ and $\bigcup_{k=1}^{\infty} Q_k = \mathfrak{R}^d$. We then set $S_k = Q_k \setminus Q_{k-1}$ with $S_1 = Q_1$ for $k \geq 2$. We now set $E_{j,k} = E_j \cap S_k$ as measurable sets, so tha $E = \bigcup_{j,k}^{\infty} E_{j,k}$. We notice that each $E_{j,k}$ is bounded and the last union is disjoint. In addition, each union $E_j = \bigcup_{k=1}^{\infty} E_{j,k}$ is also disjoint. With these properties and by remembering (*) it follows that

$$m(E) = \sum_{j,k}^{\infty} m(E_{j,k}) = \sum_{j}^{\infty} \sum_{k}^{\infty} m(E_{j,k}) = \sum_{j=1}^{\infty} m(E_j). \blacksquare$$

Remark 2.5.4

With Theorem 2.5.3, the countable additivity of the Lebesgue measure on measurable sets is established. This result presents us with the necessary connection between the following:

- Our primary notation of volume is given by outer measure,
- The exact concept of measurable sets,
- The countably infinite operations that are allowed on these sets.

Definition 2.5.5

- Let E_1, E_2,… be a growing and countable collection (i.e. $E = \bigcup_{k=1}^{\infty} E_k$ with $E_k \subset E_{k+1}$ for all $k \geq 1$, we will mark this from now on with $E_k \nearrow E$
- Let E_1, E_2,… be a collection decreasing to E (i.e. $E = \bigcap_{k=1}^{\infty} E_k$ with $E_k \supset E_{k+1}$ for all $k \geq 1$, we will mark this from now on with $E_k \searrow E$.

Corollary 2.5.6

(a) If $E_k \nearrow E$, is then $m(E) = \lim_{N \to \infty} m(E_N)$.

(b) If $E_k \searrow E$ and $\exists\, k \in \mathbb{N}$: $m(E_k) < \infty$, then we have $m(E) = \lim_{N \to \infty} m(E_N)$.

(a) *Proof:* We set $G_k = E_k \setminus E_{k-1}$ with $G_1 = E_1$ for $k \geq 2$. We note in their construction that the sets G_k are measurable , disjoint, and that $E = \bigcup_{k=1}^{\infty} G_k$. This results in

$$m(E) = \sum_{k=1}^{\infty} m(G_k) = \lim_{N \to \infty} \sum_{k}^{N} m(G_k) = \lim_{N \to \infty} m(\bigcup_{k=1}^{N} G_k) = \lim_{N \to \infty} m(E_N) \ .\blacksquare$$

(b) *Proof:* We can of course assume that $m(E_1) < \infty$. We set $G_k = E_k \setminus E_{k+1}$ with $E_1 = E \cup \{\bigcup_{k=1}^{\infty} G_k\}$ as a disjoint union of measurable sets. Consequently, we find that

$$m(E_1) = m(E) + \lim_{N\to\infty} \sum_{k}^{N} (m(E_k) - m(E_{k+1})) = m(E) + m(E_1) - \lim_{N\to\infty} m(E_N).$$

Because $m(E_1) < \infty$, we clearly see $m(E) = \lim_{N\to\infty} m(E_N)$.▪

Remark 2.5.7

The second part of the corollary (b) will be omitted without the condition $\exists\, k \in \mathbb{N}\colon m(E_k) < \infty$. This can be confirmed by setting $E_n = (n, \infty) \subset \mathfrak{R}$ (for all $n \in \mathbb{N}$).

Theorem 2.5.8

Let $E \subseteq \mathfrak{R}^d$ measurable, the following holds for all $\varepsilon > 0$:

i. There exists an open set G with $E \subset G$ and $m(G \setminus E) \le \varepsilon$.
ii. There exists a closed set F with $F \subset E$ and $m(E \setminus F) \le \varepsilon$.
iii. There exists a compact set K with $K \subset E$ and $m(E \setminus K) \le \varepsilon$, if $m(E) < \varepsilon$.
iv. There exists a finite union of closed cubes $F = \bigcup_{j=1}^{N} Q_j$ with $m(E \,\Delta\, F) \le \varepsilon$, if $m(E) < \varepsilon$.

The notation $E \,\Delta\, F$ stands for the symmetric difference between the sets E and F as follows

$$E \,\Delta\, F = (E \setminus F) \cup (F \setminus E)$$

(i) Proof: This follows directly from the definition of measurability.▪

(ii) Proof: We know that E^c is measurable. So there exists an open set G with $E^c \subset G \;\wedge\; m_*(G \setminus E^c) \le \varepsilon$. If $F = G^c$, is then F a closed set with $F = G^c \subset E \;\wedge\; E \setminus F = E \cap G = G \setminus E^c$. Thus $m(E \setminus F) \le \varepsilon$.▪

(iii) Proof: First, we choose a closed set $F \subset E$ with $m(E \setminus F) \le \frac{\varepsilon}{2}$. If for each positive integer n, B_n is the ball with the origin as the center. We define compact sets $K_n = (F \cap B_k)$. Thus $E \setminus K_n$ is a sequence of measurable sets that decreasing $E \setminus F$. Since $m(E) < \infty$, we find for all large enough n, $m(E \setminus K_n) \le \varepsilon$.▪

(iv) Proof: Let $\{Q_k\}_{k=1}^{\infty}$ be a family of closed cubes with $E \subset \bigcup_{j=1}^{\infty} Q_j \;\wedge\; \sum_{j=1}^{\infty} |Q_j| \le m(E) + \frac{\varepsilon}{2}$. Since $m(E) < \infty$, the sum sequence converges and therefore there exists an $N > 0$ with $\sum_{j=N+1}^{\infty} |Q_j| \le \frac{\varepsilon}{2}$. If $F = \bigcup_{j=1}^{N} Q_j$, then

$$m(E \,\Delta\, F) = m(E \setminus F) + m(F \setminus E) \le m\Big(\bigcup_{j=N+1}^{\infty} Q_j\Big) + m\Big(\bigcup_{j=1}^{\infty} Q_j \setminus E\Big) \le \sum_{j=N+1}^{\infty} |Q_j| + \sum_{j=1}^{\infty} |Q_j| \le \varepsilon.▪$$

Invariance properties 2.5.9

(a) Translation invariance:

Let $E \subset \mathfrak{R}^d$ be measurable and $h \in \mathfrak{R}^d$, then the set $E_h = E + h = \{x+h \mid x \in E\}$ is also measurable and $m(E_h) = m(E + h) = m(E)$.

This is easier to see from the observation of a cube. According to the properties of the outer measure (Properties 1, 3, 5, 10 in 2.4.3) we have $m_*(E_h) = m_*(E)$.

To show measurability E_h, we consider under the assumption that the set E is measurable, an open set $G \supset E$ with $m_*(G \setminus E) \le \varepsilon$. Consequently, the set $G_h \supset E_h$ with $m_*(G_h \setminus E_h) \le \varepsilon$ is open.

(b) Relative dilation invariance:

Let $\delta > 0$ and $\delta E = \{\delta x \mid x \in E\}$, we can determine in the same way as above that δE is measurable whenever E is measurable. And it holds $m(\delta E) = \delta^d m(E)$

(c) Reflection invariance:

If $-E = \{-x \mid x \in E\}$, is then $m(-E) = m(E)$, whenever E measurable is.

2.6 Measurable Functions

The idea of a characteristic function is a good starting point for a measurable function. This is defined [RAR0051] by

$$\chi_E(x) := \begin{cases} 1 & \text{falls } x \in E \\ 0 & \text{falls } x \notin E \end{cases}$$

The next step is a transition to the functions that form the basis for integration theory. For the Riemann integral1, it is actually the class of step functions, each given as a finite sum

$$f = \sum_{k=1}^{N} a_k \chi_{R_k},$$

Where the R_k are rectangles and a_k are constant.

The Lebesgue integrals correspond to an even more general concept, which can be represented by the so-called *simple functions*. *Simple function* is a finite sum

$$f = \sum_{k=1}^{N} a_k \chi_{E_k},$$

where each E_k is a measurable set with finite measure and a_ks are constant.

We only consider real-valued functions[1] on $\Re^d$ i.e., $f\colon \Re^d \to \Re$. We denote the function f as finite-valued f, if $-\infty < f(x) < \infty$ for all $x \in \Re^d$.

Definition 2.6.1

A function f, which is defined on a subset $E \subset \Re^d$, is called ***measurable*** if for each $a \in \Re$ the set $f^{-1}([-\infty, a)) = \{x \in E \mid f(x) < a\} := \{f(x) < a\}$ is measurable. Alternatively, the measurability is defined by $f^{-1}([-\infty, a]) = \{x \in E \mid f(x) < a\} := \{f(x) \leq a\}$. Equivalently, one can instead link the measurability of a function with measurability of the upper sets:

f is measurable if and only if the set $\{x \in E \mid f(x) < a\} := \{f(x) \leq a\}$ is measurable for all a.

In fact $\{f(x) \leq a\} = \bigcap_{k=1}^{\infty} \left\{f(x) < a + \frac{1}{k}\right\}$, one remembers that a countable intersection of measurable sets is again measurable (6. Eigenschaft 2.5.2). Apparently, a finite-valued function is measurable if and only if the set $\{b < f(x) \leq a\}$ is measurable.

We consider that it also holds analogously for other direction $\{f(x) < a\} = \bigcup_{k=1}^{\infty} \left\{f(x) \leq a - \frac{1}{k}\right\}$.

Properties 2.6.2

1. $-f(x)$ is measurable whenever $f(x)$ is measurable.

Proof: This follows directly from the above definition and the fifth property (2.5.2) when $f(x)$ is replaced by $-f(x)$ i.e., $\{-f(x) \leq a\} = \{f(x) \geq -a\} = \{f(x) < -a\}^C$.▪

1 In more general terms, a measurable function can be mapped by two given measurable spaces.

2. The finite-valued function *f(x)* is measurable if and only if $f^{-1}(G)$ is for each open set *G*, masurable, and when $f^{-1}(F)$ is measurable for each closed set *F*.

Proof: These also follow directly from the above definition and can even be extended with the assumption of measurability $f^{-1}(\infty)$ und $f^{-1}(-\infty)$.▪

3. i) Let *f(x)* on $\Re^d$ be continues[1], then *f(x) is* measurable.
   ii) Furthermore, if *f(x)* is finite-valued and *g(x)* is a continuous function on $\Re$, then *gof(x)* is also measurable.

Proof: i) Since the set $G = (-\infty, a) \subset \Re$ is open, we must first show that $f^{-1}(G)$ is also open. Let $p \in \Re^d$ und $f(p) \in G$. Since *G* is open, there exist $\varepsilon > 0$, so that $y = f(x) \in G$, if $d_{\Re}(f(x), f(p)) = |f(x) - f(p)| < \varepsilon$. On the other hand, since *f* is continuous at the point *p*, there exists a $\delta > 0$, so that $d_{\Re}(f(x), f(p)) < \varepsilon$ for $d_{\Re^d}(x, p) = |x - x_0| < \delta$. From this it follows that $x \in f^{-1}(G)$, as soon as $d_{\Re^d}(x, p) < \delta$. Thus $f^{-1}(G) \coloneqq \Phi$ is open.

With the first property 2.5.2, the proof for measurability of f(x) is achieved. For the second part *ii)* is analogous to the first part. It should be noted that finite-valuedness is a necessary condition for the second statement.▪

4. Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions, then
$$\sup_n f_n(x), \quad \inf_n f_n(x), \quad \limsup_{n\to\infty} f_n(x), \quad \text{und} \quad \liminf_{n\to\infty} f_n(x)$$
are measurable.

Proof: For $\sup_n f_n(x)$ the proof follows directly from $\{\sup_n f_n > a\} = \bigcup_n \{f_n > a\}$. This can also apply to $\inf_n f_n(x)$, because $\inf_n f_n(x) = -\sup_n(-f_n(x))$. For *limsup* and *liminf*, it follows that
$$\limsup_{n\to\infty} f_n(x) = \inf_k\{\sup_{n\geq k} f_n\} \quad \text{und} \quad \liminf_{n\to\infty} f_n(x) = \sup_k\{\inf_{n\geq k} f_n\}.▪$$

5. Let $\{f_n\}_{n=1}^{\infty}$ be a collection of Measurable Functions and $\lim_{n\to\infty} f_n(x) = f(x)$, then f is measurable.

Proof: The proof follows immediately from the last property, because
$$\limsup_{n\to\infty} f_n(x) = f(x) = \liminf_{n\to\infty} f_n(x).▪$$

6. If *f* and *g* are measurable functions, then
   (i) The f_k are also measurable for $k \geq 1$.
   (ii) *f* +*g* and *fg* are measurable if the two functions *f* and *g* are finite-valued.

Proof: For (i) we note that

- if the exponent is not even, then $\{f^k > a\} = \{f > a^{1/k}\}$.
- if the exponent is even, then $\{f^k > a\} = \{f > a^{1/k}\} \cup \{f < -a^{1/k}\}$.▪
-

For (ii) we note that $\{f + g > a\} = \bigcup_{r\in\mathbb{Q}} \{f > a - r\} \cap \{g > r\}$, where $\mathbb{Q}$ is denoted as the set of rational numbers. Consequently, it follows that
$$fg = \frac{1}{4}[(f+g)^2 - (f-g)^2].▪$$

1 A function *f(x)* ,defined on $\Omega \subset \Re^d$, called as continues at the $x_0 \in \Omega$, if for each $\varepsilon > 0$ some $\delta > 0$ exists, so that $|f(x) - f(x_0)| < \varepsilon$, whenever $|x - x_0| < \delta$ for $x \in \Omega$.

Definition 2.6.3

We call two functions *f* and *g*, defined on $E \subset \Re^d$ almost everywhere equal with $f(x) = g(x)$ a.e. $x \in E$, when the set $\{x \in E \mid f(x) \neq g(x)\}$ has a measure of zero. So in general, a property, or statement is "almost everywhere (a.e.)", if it is true despite the exception of a set with a measure of zero.

Remark 2.6.4

- With the above definition, it is easy to see that if a function *f* is measurable and *f* = *g* a.e., then *g* is also measurable. This follows directly from the fact that $\{f < a\}$ und $\{g < a\}$ differs from a set with a measure of zero. Thus, all upper properties (1-6) can also be expanded by applying a.e.
- If *f* and *g* are almost everywhere defined on a measurable subset $E \subset \Re^d$ then the functions *f* + *g* and *fg* can only be defined on the intersection of the domains of definition of *f* and *g*. Since the union of two sets with a measure of zero again has a measure of zero, *f* + *g* is almost everywhere defined on *E*.

Theorem 2.6.5

If *f* is a non-negative measurable function on $\Re^d$, then there exists an increasing sequence of non-negative *simple functions* $\{\varphi_k\}_{k=1}^{\infty}$, which converges pointwise to *f*. That is:

$$\varphi_k(x) \leq \varphi_{k+1}(x) \quad \text{und} \quad \lim_{k \to \infty} \varphi_k = f(x); \quad \forall x \in \Re^d .$$

Proof: Let Q_N be be a cube under a truncation around a $N \geq 1$ at the origin. We define

$$F_N(x) := \begin{cases} f(x) & \text{falls } x \in Q_N \text{ und } f(x) \leq N \\ N & \text{falls } x \in Q_N \text{ und } f(x) > N \\ 0 & \text{sonst} \end{cases}$$

So the $F_N(x) \to f(x)$ for all $x \in E \subset \Re^d$ als $N \to \infty$. Now we divide the range of values of F_N, namely $[0, N]$ as defined below:

For fixed N, $M \geq 1$, $E_{l,M} = \left\{ x \in Q_N \mid \frac{l}{M} < F_N(x) \leq \frac{l+1}{M};\ 0 \leq l < NM \right\}$. So we can have a simple function $F_{N,M}(x) = \sum_l \frac{l}{M} \chi_{E_{l,M}}(x)$, which $0 \leq F_N(x) - F_{N,M}(x) \leq \frac{1}{M}$ for all x in $\Re^d$.

By choosing $N = M = 2^k$ (mit $k > 1$; $k \in \mathbb{N}$) and setting $\varphi_k = F_{2^k, 2^k}$ it results for all $x \in \Re^d$, $0 \leq F_N(x) - \varphi_k(x) \leq \frac{1}{2^k}$. In fact $\varphi_{k-1} = F_{2^{k-1}, 2^{k-1}} \leq F_{2^k, 2^{k-1}} \leq F_{2^k, 2^k} = \varphi_k$. Thus, due to the increasing of the sequence φ_k all desired properties in the theorem are fulfilled.▪

Remark 2.6.6

We consider that the result of the last theorem holds to non-negative functions, which as an extended-valued function* in the case of a limit $+\infty$ are well-defined. This can apply analogously to a non-negative function with extended limit $-\infty$.

Theorem 2.6.7

Let *f* be a measurable function on $\Re^d$, then there exists a sequence of *simple functions* $\{\varphi_k\}_{k=1}^{\infty}$, which satisfy $|\varphi_k(x)| \leq \varphi_{k+1}(x)$ and $\lim_{k \to \infty} \varphi_k = f(x); \quad \forall x \in \Re^d$. In particular $|\varphi_k| \leq |f(x)|$

* A extended-valued function is some function, whose value range is extended to infinity.

(boundaries of the f) apply for all x and k.

Proof: Let $f := f(x) = f^+(x) - f^-(x)$ be a decomposition with $f^+(x) = max\{f(x), 0\}$ and $f^-(x) = max\{-f(x), 0\}$. Thanks to the last theorem, we know that there exist increasing sequences of non-negative *simple functions* $\{\varphi_k^{(1)}\}_{k=1}^{\infty}$ und $\{\varphi_k^{(2)}\}_{k=1}^{\infty}$, that converge pointwise to the non-negative functions f^+ bzw. f^- . Setting $\varphi_k(x) = \varphi_k^{(1)}(x) - \varphi_k^{(2)}(x)$, it follows that

$$\lim_{k\to\infty} \varphi_k = \lim_{k\to\infty} \varphi_k^{(1)}(x) - \lim_{k\to\infty} \varphi_k^{(2)}(x) = f^+ - f^- = f; \quad \forall x \in \Re^d .$$

Finally, we note that $|f| = |f^+ - f^-| = f; \quad \forall x \in \Re^d$ and for each x $\varphi_k^{(1)}(x) . \varphi_k^{(2)}(x) = 0$. The last equation follows directly from the fact that at most one of the two $f^+(x)$, $f^-(x)$ is not equal to zero, $0 \le \varphi_k^{(1)} \le f^{(+)}(x)$, and $0 \le \varphi_k^{(2)} \le f^{(-)}(x)$. Thus $|\varphi_k(x)| = \varphi_k^{(1)}(x) + \varphi_k^{(2)}(x)$ obviously results as an increasing sequence.▪

2.7 The Lebesgue Integral

We proceed in four stages step by step integration.

2.7.1 Simple Functions

We remember from chapter 2.5 the so-called s*imple function* $\varphi(x) = \sum_{k=1}^{N} a_k \chi_{E_k}(x)$, where

$\varphi(x) = \sum_{k=1}^{N} a_k \chi_{E_k}(x)$, where the measurable set E_k with finite measure and a_k are constant.

Definition 2.7.1.1

Let φ be a *simple functions* with canonical form $\varphi(x) = \sum_{k=1}^{M} c_k \chi_{F_k}(x)$. The Lebesgue Integral of φ is defined by $\int_{\Re^d} \varphi(x)\, dx = \sum_{k=1}^{M} c_k m(F_k)$, where $F_k = \{x \in \Re^d \,|\, \varphi(x) = c_k\}$ represents the desired canonical form. Furthermore, if the set $E \subset \Re^d$ is measurable with a finite measure size, then $\varphi(x)\chi_E(x)$ is also a *simple function* with the following definition:

$$\int_E \varphi(x)\, dx = \int \varphi(x)\chi_E(x)\, dx.$$

To clarify the choice of the Lebesgue measure, the *dx* is sometimes replaced by a *dm(x)* in the definition. That is, $\int_{\Re^d} \varphi(x)\, dm(x)$.

In the following, we will make propositions about the properties of *simple functions* without stating their proofs.

Proposition 2.7.1.2

The integral of *simple functions* defined above fulfills the following properties:

i. (Representation independence) If $\varphi(x) = \sum_{k=1}^{N} a_k \chi_{E_k}(x)$ is any representation of φ, then $\int \varphi = \sum_{k=1}^{N} a_k m(E_k)$.

ii. (Linearity) If φ und ψ are s*imple functions* and $a, b \in \Re$, then
$$\int (a\varphi + b\psi) = a\int \varphi + b\int \psi$$

iii. (Additivity) If E and F are disjoint subsets in the $\Re^d$ with finite measure, then

$$\int_{E\cup F} \varphi = \int_E \varphi + \int_F \varphi .$$

iv. (Monotony) If $\varphi < \psi$ are *simple functions*, then:

$$\int \varphi \le \int \psi$$

v. (Triangle inequality) If φ is a *simple function*, then it also holds for $|\varphi|$ and

$$\left|\int \varphi\right| \le \int |\varphi|.$$

2.7.2 Bounded Functions

Definition 2.7.2.1

A support of measurable function *f* is defined by

$$\mathrm{supp}(f) = \{x | f(x) \neq 0\}.$$

Furthermore, the function f is supported on a set $E \subset \Re^d$, if *f(x) = 0* whenever $x \notin E$.

Remark 2.7.2.2

An important result from the last chapter (Theorem 2.6.7) states that if the function *f* is bounded by *M* and supported on the set *E*, then there exists a sequence $\{\varphi_n\}_{n=1}^{\infty}$ of *simple functions*, which are bounded by *M* and supported on *E*, so that for all $x \in \Re^d$, $\varphi_N(x) \to f(x)$.

Lemma 2.7.2.3

Let *f* be bounded functions and supported on a set with finite measure $E \subset \Re^d$. then there exists a sequence $\{\varphi_n\}_{n=1}^{\infty}$ of *simple functions* which are bounded by *M* and supported on *E*, such that $\varphi_n(x) \to f(x)$ for a.e. x.

i. The $\lim\limits_{n\to\infty} \int \varphi_n$ exist.

ii. If $f = 0$ a.e., then $\lim\limits_{n\to\infty} \int \varphi_n = 0$.

Proof: The Proof[¬] is omitted due to prior knowledge such as about Egorov's Theorem and about about Littelwood's principles.

Definition 2.7.2.4

For a bounded functions, which is supported on a set $E \subset \Re^d$ with finite measure , its **Lebesgue Integral** is defined by:

$$\int f(x)\,dx = \lim_{n\to\infty} \int \varphi_n(x)\,dx.$$

Here $\{\varphi_n\}_{n=1}^{\infty}$ is a sequence of *simple functions* with the following properties

- $|\varphi_n(x)| \le M$.
- φ_n is supported on the supp(*f*).
- $\varphi_N(x) \to f(x)$ for a.e. x.

In the following, we will make again some propositions about the properties of *simple functions* without stating their proofs.

¬ Readers can browse prior knowledge and the proof in the book „ Princeton Lectures in Analysis“ by M.Stein and Shakarchi for their own interest.

Proposition 2.7.2.5

If f and g are bounded functions and supported on the set with finite measure, then the following properties apply:

i. (Linearity) If $a, b \in \Re$, then
$$\int (a\, f + b\, g) = a \int f \;+ b \int g\,.$$
ii. (Additivity) If E and F are disjoint subsets in the $\Re^d$ with finite measure, then
$$\int_{E \cup F} f = \int_E f \;+ \int_F f\,.$$
iii. (Monotony) If $f \;<\; g$, then
$$\int f \;\leq \int g\,.$$
iv. (Triangle inequality) If $|f\,|$ is bounded and supported on a set with finite measure, then
$$\left|\int f\right| \leq \int |f\,|.$$

Theorem 2.7.2.6 (Bounded Convergence)

Let $\{f_n\}$ be a sequence of measurable functions, which are bounded by M and supported on a set $E \subset \Re^d$ with finite measure. Furthermore, let $f_n(x) \rightarrow f(x)$ for a.e. x als $n \rightarrow \infty$. Then f is a measurable function, which is bounded, supported on the set E, and
$$\int |f_n \text{ - } f| \rightarrow 0 \;\text{ bzw. }\; \int |f_n| \rightarrow f \;\text{ als } n \rightarrow \infty.$$

Remark 2.7.2.7

The last theorem can be used to show that Riemann-integrabl[1] functions are also Lebesgue integrable.

2.7.3 Non-Negative Functions

We now continue with the integral of functions that are measurable, non-negative but not necessarily bounded (extended-valued function*) [RAR0051].

Definition 2.7.3.1

Let f be a function as described above. Its (extended) **Lebesgue Integral** is defined by,
$$\int f(x)\, dx \;=\; \sup_g \int g(x)\, dx,$$
where the function g ($0 \leq g \leq f$) is bounded and supported on a set with finite measure size. Thus, the integral of the function f is finite or infinite. In the case $\int f(x)\, dx < \infty$ the function f is called **Lebesgue integrable** or simply **integrable**.

Analogous to the previous development steps for the definition of the Lebesgue integral, the properties for non-negative functions will be partly repeated.

Proposition 2.7.3.2

The integral of non-negative measurable functions has the following properties.

i. (Linearity) Let $a, b \in \Re$, then
$$\int (a\, f + b\, g) = a \int f \;+ b \int g\,.$$
ii. (Additivity) Let E and F be disjoint subsets in $\Re^d$ with finite measure size, then
$$\int_{E \cup F} f = \int_E f \;+ \int_F f\,.$$
iii. (Monotony) If $f \;<\; g$, then the following is true :
$$\int f \;\leq \int g\,.$$

iv. Let g be integrable and $0 \leq g \leq f$, then it holds for f as well.
v. Let f be integrable, then $f(x) < \infty$ a.e. x.
vi. Let $\int f = 0$, then $f(x) = 0$ almost for every(a.e.) x.

Lemma 2.7.3.3 (Fatou)

Let $\{f_n\}$ be a sequence of measurable functions with $f_n \geq 0$. If $\lim_{n \to \infty} f_n(x) = f(x)$ for a.e. x, then

$$\int f \leq \liminf_{n \to \infty} \int f_n.$$

Corollary 2.7.3.4

Let f be a non-negative measurable function, and $\{f_n\}$ a sequence of non-negative measurable functions with $f_n(x) \leq f(x)$ und $\lim_{n \to \infty} f_n(x) = f(x)$ for a.e. x, then the following will hold.

$$\lim_{n \to \infty} \int f_n = f$$

Corollary 2.7.3.5 (Monotone Convergence Theorem)

Let $\{f_n\}$ be a sequence of measurable functions with $f_n \nearrow f$(2.5). Then the following is to hold.

$$\lim_{n \to \infty} \int f_n = f$$

Corollary 2.7.3.6

Consider $\sum_{k=1}^{\infty} a_k(x)$, where $a_k(x) \geq 0$ is measurable for each $k \geq 1$. Then it holds that

$$\int \sum_{k=1}^{\infty} a_k(x)dx = \sum_{k=1}^{\infty} \int a_k(x)dx.$$

Furthermore, let $\sum_{k=1}^{\infty} \int a_k(x)dx$ be finite. Then the sum $\sum_{k=1}^{\infty} a_k(x)$ converges for a.e. x.

2.7.4 Integrable Functions (general case)

Definition 2.7.4.1

In the sense of the last section (2.7.3) we denote a real-valued function f as **Lebesgue integrable** (or simply **integrable**), if the non-negative function $|f|$ is integrable.

Remark 2.7.4.2

In fact, to realize the above definition, the function f is decomposed as follows:

$$f := f(x) = f^+(x) - f^-(x) \text{ mit } f^+(x) = max\{f(x), 0\} \text{ und } f^-(x) = max\{-f(x), 0\}.$$

Now that the two non-negative functions $f^+, f^- \leq |f|$, both are integrable whenever the function f is integrable and thus the Lebesgue integral of function f is well-defined by the following

$$\int f = \int f^+ - \int f^-.$$

Proposition 2.7.4.3

The integral of Lebesgue integrable functions is linear, additive, monotone and satisfies the triangle inequality.

Proposition 2.7.4.4

Let f be an integrable function on $\Re^d$. Then for each $\varepsilon > 0$ the following holds:

i. There exists a set of finite measure size B (e.g a Ball), so that

$$\int_{B^C} |f| < \varepsilon.$$

ii. There is a $\delta > 0$, so that $\int_{B^C} |f| < \varepsilon$ whenever $m(E) < \delta$.

Theorem 2.7.4.5

Let $\{f_n\}$ be a sequence of measurable functions with $\lim_{n \to \infty} f_n(x) = f(x)$ for a.e. *x*. Furthermore , if $\left| f_n(x) \right| < g(x)$, consequently it holds that

$$\int |f_n - f| \to 0 \text{ bzw. } \int f_n \to f \text{ als } n \to \infty.$$

Remark 2.7.4.6

In general, we can consider $\Re^d$ as a product $\Re^d = \Re^{d_1} \times \Re^{d_2}$, where $d = d_1 + d_2$, und $d_1, d_2 \geq 1$. Thus, a point in the $\Re^d$ can be considered in the form *(x, y)* with $x \in \Re^{d_1}$, $y \in \Re^{d_2}$. Through such decomposition of the $\Re^d$, the general concept of slice (Figure 9), which arises by fixing a variable, becomes understandably imaginable. For a function $f \in \Re^{d_1} \times \Re^{d_2}$, the slice of *f* with respect to the variable $y \in \Re^{d_2}$ the function f^y on $x \in \Re^{d_1}$ is defined by $f^y = f(x, y)$. Similarly, the slice of f for a fixed $x \in \Re^{d_1}$ is defined by $f_x = f(x, y)$.

In the case of a set $E \subset \Re^{d_1} \times \Re^{d_2}$ its slices are defined by $E^y = \{ x \in \Re^{d_1} \mid (x, y) \in E \}$ and $E_x = \{ y \in \Re^{d_2} \mid (x, y) \in E \}$

Theorem 2.7.4.7†

Let *f(x, y)* be integrable on $\Re^{d_2} \times \Re^{d_2}$, then for a.e. $y \in \Re^{d_2}$

i. The slice f^y is integrable on $\Re^{d_1}$.

ii. The function $F^y = \int_{\Re^{d_1}} f^y(x)\, dx$ is integrable on $\Re^{d_2}$.

iii. $\int_{\Re^{d_2}} \left(\int_{\Re^{d_1}} f(x, y)\, dx \right) dy = \int_{\Re^d} f.f^y$

It is clear that the theorem is symmetric w.r.t *x* and *y*, so we can analogously say the same for the slice f_x on $\Re^{d_2}$ for a.e. $x \in \Re^{d_1}$ (i.e. integrability f_x on $\Re^{d_2}$ for a.e. $x \in \Re^{d_1}$, so integrability of $\int_{\Re^{d_2}} f_x(y)\, dy$ on $\Re^{d_1}$ and $\int_{\Re^{d_1}} \left(\int_{\Re^{d_2}} f(x, y)\, dy \right) dx = \int_{\Re^d} f$

In particular, the theorem states that the integral *f* on $\Re^d$ can be calculated by an iteration of small dimensional integrals, i.e.

$$\int_{\Re^{d_2}} \left(\int_{\Re^{d_1}} f(x, y)\, dx \right) dy = \int_{\Re^{d_1}} \left(\int_{\Re^{d_1}} f(x, y)\, dy \right) dx = \int_{\Re^d} f.$$

Proof: We will refrain from the proof due to necessary basic knowledge of calculus of variations. However, a good proof can be found by the reader with basic knowledge in the book "Princeton Lectures in Analysis" by M.Stein and Shakarchi [RAR0051].

† The Theorem was developed by Fubini and is also known by this name.

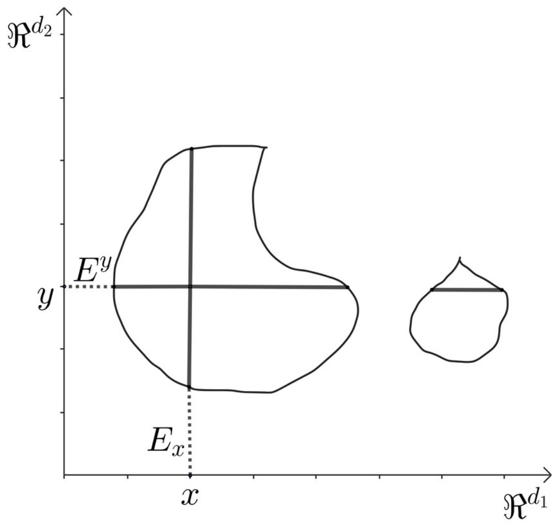

Figure 9: Slices E^y and E_x (for fix x and y) in a set E.

Theorem 2.7.4.8

Let *f(x, y)* be a non-negative measurable function on $\Re^{d_2} \times \Re^{d_2}$, then for a.e. $y \in \Re^{d_2}$ the followings will be holde:

i. The slice f^y is measurable on $\Re^{d_1}$.

ii. The function $F^y = \int\limits_{\Re^{d_1}} f^y(x)\, dx$ is measurable on $\Re^{d_2}$.

iii. $\int\limits_{\Re^{d_2}} \big(\int\limits_{\Re^{d_1}} f(x, y)\, dx \big)\, dy = \int\limits_{\Re^{d}} f(x, y)\, dx\, dy$ in the sense of an extension.

In practice, this theorem[†] is often used in association with Fubini's theorem. In fact, for a given measurable function f on $\Re^d$ calculation of $\int\limits_{\Re^d} f$ is asked. To justify the iterated integration, we first apply the theorem for $|f|$. This allows us to calculate the iterated integrals of the non-negative functions unhindered. If these are finite, the theorem guarantees that the function f is integrable, namely $\int |f| < \infty$.

Proof: We consider the following truncation:

$$f_k(x, y) := \begin{cases} f(x, y) & \text{falls } \|(x, y)\|, f(x) < k \\ 0 & \text{sonst} \end{cases}$$

Each f_k is integrable, and by the first part (i) of Fubini's theorem there exists a set $E \subset \Re^{d_2}$ with measure 0, so that the slice $f_k(x)$ for all $y \in E_k^c$ is measurable. When using $E = \bigcup\limits_{k=1} E_k$, we now find that $f^y(x)$ for all $y \in E^c$ and all k is measurable. Furthermore, $m(E) = 0$. Since $f_k^y \nearrow f^y$, it follows from Monotone Convergence Theorem for each $y \notin E$:

$$F_k^y = \int\limits_{\Re^{d_1}} f_k^y(x)\, dx \nearrow \int\limits_{\Re^{d_1}} f^y(x)\, dx = F^y \text{ als } k \to \infty.$$

† The Theorem was first formulated by Tonelli.

By Fubini's theorem, F_y^k is measurable for all $y \in E^c$ and therefore it is equal F_y. With another application of Monotone Convergence, it follows that

(1) $$\int_{\mathfrak{R}^{d_2}} \left(\int_{\mathfrak{R}^{d_1}} f_k(x, y)\, dx \right) dy \to \int_{\mathfrak{R}^{d_2}} \left(\int_{\mathfrak{R}^{d_1}} f(x, y)\, dx \right) dy .$$

From the last part of Fubini's theorem (iii) we already know that

(2) $$\int_{\mathfrak{R}^{d_2}} \left(\int_{\mathfrak{R}^{d_1}} f_k(x, y)\, dx \right) dy = \int_{\mathfrak{R}^{d}} f_k .$$

With the Monotone Convergence Theorem, it follows that

(3) $$\int_{\mathfrak{R}^{d}} f_k \to \int_{\mathfrak{R}^{d}} f .$$

With the combination (1), (2), and (3) the proof is complete.▪

Corollary 2.7.4.9

Let E be a measurable set in $\mathfrak{R}^{d_2} \times \mathfrak{R}^{d_2}$, then for a.e. $y \in \mathfrak{R}^{d_2}$ the slice

$$E^y = \left\{ x \in \mathfrak{R}^{d_1} \,\middle|\, (x, y) \in E \right\}$$

is a measurable subset in $\mathfrak{R}^{d_1}$. Furthermore, $m(E^y)$ is a measurable function of y and

$$m(E) = \int_{\mathfrak{R}^{d_2}} m(E^y)\, dy .$$

Proof: This follows directly by setting the function $f = \chi_E$ in the first part of the last sentence.

Proposition 2.7.4.10

Let $E = E_1 \times E_2$ (as **production set)** with $E_j \subset \mathfrak{R}^{d_j}$ und $m_*(E_2) > 0$ be measurable. Then. E_1 is measurable.

Proof: With the Corollary 2.7.4.9 it follows for a.e. $y \in \mathfrak{R}^2$, that the slice function

$$f^y := \left(\chi_{E_1 \times E_2} \right)^y (x) = \chi_{E_1}(x) \chi_{E_2}(y)$$

is measurable as a function of x. The existence of such y is justified with $m_*(E_2) > 0$. In fact, one shows for the subset $F \subset \mathfrak{R}^{d_2}$, where the slice E^y is measurable. Therefor $m(F^c) = 0$. From the decomposition E_2 into two separate subsets $(E_2 \cap F)$ und $(E_2 \cap F^c)$ it follows that

$$0 < m_*(E_2) \le m_*(E_2 \cap F) + m_*(E_2 \cap F^c) = m_*(E_2 \cap F).▪$$

Lemma 2.7.4.11

Let $E_1 \subset \mathfrak{R}^{d_1}$ and $E_2 \subset \mathfrak{R}^{d_2}$. Understanding $m_*(E_1 \times E_2) = 0$, if the outer measure of one of the E_j is zero. We conclude that

$$m_*(E_1 \times E_2) \le m_*(E_1) m_*(E_2).$$

Proposition 2.7.4.12

Let the sets $E_1 \subset \mathfrak{R}^{d_1}$ and $E_2 \subset \mathfrak{R}^{d_2}$ be sets measurable. This then also holds to $E = E_1 \times E_2$. With the understanding that $m_*(E_1 \times E_2) = 0$, if outer measure of one of the E_j is zero.
Furthermore, we have consequently

$$m(E) = m(E_1) m(E_2).$$

Proof: In the case of measurability of *E*, the assertion about $m(E)$ directly follows from the Corollary 2.7.4.9. Since each set E_j is measurable, there exist sets $G_j \subset \Re^{d_j}$ as a countable intersection of open sets* with $G_j \subset E_j$ and $m_*(G_j \setminus E_j) = 0$ for each $j = 1,2$. The last Equation follows from the measurability of the E_j and therefore the $G = (G_1 \times G_2) \in (\Re^{d_1} \times \Re^{d_2})$ and consequently

$$G = (G_1 \times G_2) \setminus (E_1 \times E_2) \subset ((G_1 \setminus E_1) \times G_2) \cup (G_1 \times (G_2 \setminus E_2)).$$

By the last lemma we find that $m_*(G \setminus E) = 0$ and thus *E* is measurable.▪

Corollary 2.7.4.13

Let *f* be a measurable function on $\Re^{d_1}$. The function $\tilde{f}$, defined by $\tilde{f}(x, y) = f(x)$ is measurable in $\Re^{d_1} \times \Re^{d_2}$.

Proof: We assume that it is a real-valued function f and simultaneously find that the $E_1 = \left\{ x \in \Re^{d_1} \,\middle|\, f(x) < a,\, a \in \Re \right\}$ is measurable. Since

$$\left\{ (x, y) \in \Re^{d_1} \times \Re^{d_2} \,\middle|\, \tilde{f}(x, y) < a \right\} = E_1 \times \Re^{d_2},$$

the last proposition shows that $\forall\, a \in \Re$ die $E_1 \times \Re^{d_2}$ is measurable. Therefor the $\tilde{f}(x, y)$ is a measurable function on $\Re^{d_1} \times \Re^{d_2}$.▪

Corollary 2.7.4.14

Let *f(x)* be a non-negative function on $\Re^d$ and $A = \left\{ (x, y) \in \Re^d \times \Re \,\middle|\, 0 \le y \le f(x) \right\}$. Then the following statements can hold:

i. *f* is measurable on $\Re^d$, if and only if *A* measurable on $\Re^{d+1}$.
ii. If the condition in (i) is fulfilled man, then

$$\int_{\Re^d} f(x)\, dx = m(A).$$

Proof: If *f* is measurable on $\Re^d$, the last Proposition, guarantees that the function

$$F(x, y) = y - f(x)$$

is measurable on $\Re^{d+1}$. This makes it immediately clear that the set $A = \{ y \ge 0 \} \cap \{ F(x) \le 0 \}$ is also measurable.

Conversely, let it be *A* measurable set. We find that for each $x \in \Re^d$ the slice $A_x = \{ y \in \Re \mid (x, y) \in A \}$ is a closed segmentation, namely $A_x = [0, f(x)]$. In Following the Corollary 2.7.4.9, it follows with the exchange of the role x and y the measurability of $m(A_x) = f(x)$.As a result we have that

$$m(E) = \int \chi_A(x, y)\, dx\, dy = \int_{\Re^{d_1}} m(A)\, dx = \int_{\Re^{d_1}} f(x)\, dx.▪$$

Remark 2.7.4.15

Generally it can easily be shown that for a measurable function *f* on $\Re^d$ the function $\tilde{f}(x, y) = f(x - y)$ on $\Re^d \times \Re^d$ is measurable.

* Such set (G_δ) correspond to the most basic so-called Borel set.

2.8 The Brunn-Minkowski Inequality

Definition 2.8.1

Let $A, B \subset \mathfrak{R}^d$ be measurable, their sum[2] (Figure 10[3]) is defined by

$$A + B = \{ x \in \mathfrak{R}^d \mid x = x' + x''; \, x' \in A \wedge x'' \in B \}.$$

Remark 2.8.2

Let the $A, B \subset \mathfrak{R}^d$, and $A + B$ be measurable. We can see from a simple example of Cantor sets like $A = B = C = \left\{ \sum_{j=1}^{\infty} a_k 3^{-k} \,\middle|\, a_k \in \{0, 2\}, \forall k \right\}$, that $m(A) = m(B) = 0$ but $m(A+B) \neq 0$, because $A + B = C + C = \left\{ \sum_{j=1}^{\infty} (a_k + b_k) 3^{-k} \,\middle|\, a_k, b_k \in \{0, 2\}, \forall k \right\} \supseteq [0, 2]$. This makes it easy to see that it is not possible to determine an upper bound for $m(A + B)$ by combination of $m(a)$ and $m(B)$. In the following, we will show that, on the contrary, a lower bound for $m(A + B)$ can be calculated in the form of $m(a)$ and $m(B)$.

Remark 2.8.3

Let $A, B \subset \mathfrak{R}^d$ be measurable, it does not necessarily follow that $A + B$ is also measurable. As a concrete example, we choose in $\mathfrak{R}^2$, $A = \{0\} \times [0, 1] \wedge B = \aleph \times \{0\}$, where $\aleph$ is a non-measurable set in $\mathfrak{R}$, thus $A + B = \aleph \times \aleph$ is not measurable.

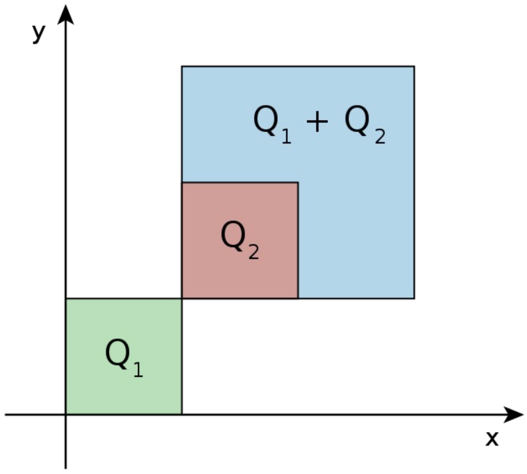

Figure 10[3]:Minkowski Addition $Q_1 + Q_2 = [0, 1]^2 + [1, 2]^2 = [1, 3]^2$.

Lemma 2.8.4

Let x_j be non-negative numbers,$(x_j \geq 0)$, then $\frac{1}{d} \sum_{j=1}^{d} x_j \geq (\prod_{j=1}^{d} x_j)^{\frac{1}{d}}$ holds for every j.

Proof: This can be easily shown by successively replacing elements. We rewrite the inequality $\alpha = \frac{x_1 + x_2 + \cdots + x_d}{d} \geq \sqrt[d]{x_1 x_2 \cdots x_d} = \beta$, and consider that equality should only hold for $x_1 = x_2 = \cdots = x_d$. If all numbers are not equal, then x_i, x_j exist with $x_i < \alpha < x_j$. When replacing x_i by α and x_j by $(x_i + x_j - \alpha)$, the arithmetic mean remains unchanged in contrast to the geometric mean. In fact, this increases the geometric mean, because

$$\alpha(x_i + x_j - \alpha) - x_i x_j = (\alpha - x_i)(x_j - \alpha) > 0.$$

2 I also called Minkowski Addition in some references.

3 Image source from Wikipedia.

If all numbers are still not equal, we repeat the above replacement. After at most *(d - 1)* such steps, all numbers are replaced by α, while the geometric mean continues to increase. Thus, after the last step, we get

$$\sqrt[d]{x_1 x_2 \cdots x_d} = \sqrt[d]{\alpha\alpha \cdots \alpha} = \alpha. \blacksquare$$

Theorem 2.8.5

Let $A, B \subset \Re^d$ and their sum $A + B$ be measurable, then it holds

$$(m(A+B))^{\frac{1}{d}} \geq (m(A))^{\frac{1}{d}} + (m(B))^{\frac{1}{d}}. \qquad (3)$$

Proof: First, we assume that A and B are rectangles with edge lengths $\{a_j\}_{j=1}^d$ and $\{b_j\}_{j=1}^d$, the

$$\Big(\prod_{j=1}^{d}(a_j+b_j)\Big)^{\frac{1}{d}} \geq \Big(\prod_{j=1}^{d} a_j\Big)^{\frac{1}{d}} + \Big(\prod_{j=1}^{d} b_j\Big)^{\frac{1}{d}} \qquad (2)$$

needs to be shown. By homogeneity of the rectangles, the last inequality can be reduced to the special case, namely $(a_j+b_j)=1$ for each j, by replacing a_j, b_j by $\lambda_j a_j$, $\lambda_j b_j$ mit $\lambda_j > 0$. Thus, we only need the $\lambda_j = (a_j + b_j)^{-1} = \frac{1}{(a_j + b_j)}$. However, this is more illustrative by dividing both sides of inequality (2) with $\Big(\prod_{j=1}^{d}(a_j+b_j)\Big)^{\frac{1}{d}} \neq 0$. That means

$$1 \geq \Big(\prod_{j=1}^{d} \frac{a_j}{a_j+b_j}\Big)^{\frac{1}{d}} + \Big(\prod_{j=1}^{d} \frac{b_j}{a_j+b_j}\Big)^{\frac{1}{d}}$$

With this reduction, inequality (2) is achieved as a direct consequence of the arithmetic-geometric inequality $\frac{1}{d}\sum_{j=1}^{d} x_j \geq \Big(\prod_{j=1}^{d} a_j\Big)^{\frac{1}{d}}$, $x_j \geq 0$ by setting $x_j = a_j$ and $x_j = b_j$. So

$$\Big(\prod_{j=1}^{d} \frac{a_j}{a_j+b_j}\Big)^{\frac{1}{d}} \leq \frac{1}{d}\sum_{j=1}^{d} \frac{a_j}{a_j+b_j}, \text{ und } \Big(\prod_{j=1}^{d} \frac{b_j}{a_j+b_j}\Big)^{\frac{1}{d}} \leq \frac{1}{d}\sum_{j=1}^{d} \frac{b_j}{a_j+b_j},$$

and using Lemma 2.8.5

$$\Big(\prod_{j=1}^{d} \frac{a_j}{a_j+b_j}\Big)^{\frac{1}{d}} + \Big(\prod_{j=1}^{d} \frac{b_j}{a_j+b_j}\Big)^{\frac{1}{d}} \leq \frac{1}{d}\sum_{j=1}^{d} \frac{a_j}{a_j+b_j} + \frac{1}{d}\sum_{j=1}^{d} \frac{b_j}{a_j+b_j} = \frac{1}{d}\sum_{j=1}^{d} \frac{a_j+b_j}{a_j+b_j} = 1.$$

Next, we consider the case where the sets A and B are the union of finitely many rectangles with disjoint interiors. We will prove this case by induction on the number of rectangles in A and B, namely n. It is important to note that the inequality (1) will remain unchanged with the independent translation (shift) of the sets A and B. In fact, replacing A and B by $A_h = A + h$ and $B_h = B + h'$ leads to replacing $A + B$ by $A + B + h + h'$ and thus the corresponding measures do not change (translation invariance). We now choose for the composition of A as a collection of disjoint rectangle pairs R_1 and R_2 and find that R_1 and R_2 can be separated from each other by a coordinate representation of a hyperplane. For this reason, we can assume for some j that R_1 and R_2 lie in $A_- = A \cap \{x_j \leq 0\}$ or in $A_+ = A \cap \{x_j \geq 0\}$ by suitable h. It follows that A_- and A_+ at most *(n – 1)* rectangles. Then we adjust the translation of the set B so that $B_- = B \cap \{x_j \leq 0\}$ and $B_+ = B \cap \{x_j \geq 0\}$ satisfy the equation $\frac{m(B_\pm)}{m(B)} = \frac{m(A_\pm)}{m(A)}$.

However, it holds that $A + B \supset (A_+ + B_+) \cup (A_- + B_-)$, and the union on the right side is necessarily disjoint (the two parts of the union lie in different half-spaces). From the induction hypothesis, it follows

$$m(A_+ + B_+) \geq (m(A_+)^{\frac{1}{d}} + m(B_+)^{\frac{1}{d}})^d, \text{ und } m(A_- + B_-) \geq (m(A_-)^{\frac{1}{d}} + m(B_-)^{\frac{1}{d}})^d, \text{ that}$$

$$m(A + B) \geq m(A_+ + B_+) + m(A_- + B_-) \geq \left[m(A_+)^{\frac{1}{d}} + m(B_+)^{\frac{1}{d}}\right]^d + \left[m(A_-)^{\frac{1}{d}} + m(B_-)^{\frac{1}{d}}\right]^d$$

$$= m(A_+)\left[1 + \left(\frac{m(B_+)}{m(A_+)}\right)^{\frac{1}{d}}\right]^d + m(A_-)\left[1 + \left(\frac{m(B_-)}{m(A_-)}\right)^{\frac{1}{d}}\right]^d = (m(A_+) + m(A_-))\left[1 + \left(\frac{m(B)}{m(A)}\right)^{\frac{1}{d}}\right]^d$$

$$= \left[m(A)^{\frac{1}{d}} + m(B)^{\frac{1}{d}}\right]^d.$$

Thus, inequality (1) results if A and B are both finite unions of rectangles with disjoint interiors. This immediately leads to the result when A and B are open sets with finite measure. In fact, according to theorem 2.3.5, for each $\varepsilon > 0$, we can find unions of almost disjoint rectangles A_ε and B_ε, such that $A_\varepsilon \subset A$, $B_\varepsilon \subset B$, with $m(A) \leq m(A_\varepsilon) + \varepsilon$, and $m(B) \leq m(B_\varepsilon) + \varepsilon$. Since $A + B \supset A_\varepsilon + B_\varepsilon$, inequality (1) results from a transition to a limit for A_ε and B_ε as $\varepsilon \to 0$:

$$m(A + B) \geq m(A_\varepsilon + B_\varepsilon) \geq \left[m(A_\varepsilon)^{\frac{1}{d}} + m(B_\varepsilon)^{\frac{1}{d}}\right]^d \geq \left[(m(A) - \varepsilon)^{\frac{1}{d}} + (m(B) - \varepsilon)^{\frac{1}{d}}\right]^d$$

$$= \left[m(A)^{\frac{1}{d}} + m(B)^{\frac{1}{d}}\right]^d, \text{ als } \varepsilon \to 0.$$

Starting from this, we can lead to the case where A and B are arbitrary compact sets. We note that $A + B$ is then also compact. Furthermore, by definition $A^\varepsilon = \{x \in \mathfrak{R}^d \,|\, d(x, A) < \varepsilon\}$ is an open set with $A^\varepsilon \searrow A$, als $\varepsilon \to 0$.

Thus, from Lemma 2.5.6(b) $m(A^\varepsilon) \to m(A)$, als $\varepsilon \to 0$. Similarly B^ε und $(A + B)^\varepsilon$ are defined. We note that $A + B \subset A^\varepsilon + B^\varepsilon \subset (A + B)^{2\varepsilon}$. Assuming that $m(A + B) < \infty$ for each $\rho > 0$ there exist $\varepsilon > 0$, so that

$$\rho + m(A + B) \geq m\left((A + B)^{2\varepsilon}\right) \geq m(A^\varepsilon + B^\varepsilon) \geq \left[m(A^\varepsilon)^{\frac{1}{d}} + m(B^\varepsilon)^{\frac{1}{d}}\right]^d \geq \left[m(A)^{\frac{1}{d}} + m(B)^{\frac{1}{d}}\right]^d,$$

Thus the inequality (1) is obtained again, as $\rho \to 0$.

ow let A, B, and $A + B$ with $m(A + B) < \infty$ (so consequently $m(A)$, $m(B) < \infty$) be measurable, there exists by suitable approximation to the inside A and B, namely $K_1 \subset A$, $K_2 \subset B$according to part *(iii)* of theorem 2.5.8 with $m(A \setminus K_1)$, $m(B \setminus K_2) < \varepsilon$. Then is $K_1 + K_2 \subset A + B$, and

$$m(A + B) \geq m(K_1 + K_2) \geq \left[m(K_1)^{\frac{1}{d}} + m(K_2)^{\frac{1}{d}}\right]^d \geq \left[(m(A) - \varepsilon)^{\frac{1}{d}} + (m(B) - \varepsilon)^{\frac{1}{d}}\right]^d$$

$$= \left[m(A)^{\frac{1}{d}} + m(B)^{\frac{1}{d}}\right]^d, \text{ als } \varepsilon \to 0. \blacksquare$$

Remark 2.8.6

For a convex set A (2.1.15) is described with a $\lambda > 0$ under reference to the definition (above) of the set $\lambda A = \{\lambda x \,|\, x \in A\}$, From the measure $A + \lambda A = (1 + \lambda)A$ namely $m(A + \lambda A) = (1 + \lambda)^d \, m(A)$, Minkowski's inequality is again confirmed under

$$(1 + \lambda)^d \geq 1 + \lambda^d \text{ und } \lambda > 0.$$

Theorem 2.8.7

Let $E \in \Re^d$. It holds for the slice (2.7.4.6) $E_a(E_{x=a}), E_b,$ und E_c mit $a < b < c$

$$m(E_b) = \min\{m(E_a), m(E_c)\}.$$

Proof: For the case $m(E_a) = 0$ oder $m(E_c) = 0$ the statement is trivial. Otherwise, let $t = \dfrac{b-a}{c-a} \to t \in (0, 1)$. With the monotony and subadditivity properties of the measure E_b (2.4), the convexity of E, with the above defined $b = a + t(c-a)$ and finally with the Minkowski inequality it result that

$$\begin{aligned}
(m(E_b))^{\overbrace{n}^{\frac{1}{d-1}}} &\geq m((1-t)E_a + t\ E_c)^{\frac{1}{n}} \geq m((1-t)E_a)^{\frac{1}{n}} + m(t\ E_c)^{\frac{1}{n}} \\
&= \left[(1-t)^n m(E_a)\right]^{\frac{1}{n}} + \left[t^n\ m(E_c)\right]^{\frac{1}{n}} \\
&= (1-t)(m(E_a))^{\frac{1}{n}} + t\ (m(E_c))^{\frac{1}{n}} \\
&\geq (1-t)\left(\min\{m(E_a), m(E_c)\}\right)^{\frac{1}{n}} + t\ \left(\min\{m(E_a), m(E_c)\}\right)^{\frac{1}{n}}\ \} \\
&= \left(\min\{m(E_a), m(E_c)\}\right)^{\frac{1}{n}}. \blacksquare
\end{aligned}$$

Corollary 2.8.8

Let $A, B \subset \Re^d$ and their sum $A + B$ measurable, then it holds

$$m\left(\frac{(A+B)}{2}\right) \geq \sqrt{m(A)m(B)}.$$

Proof: It follows directly from Theorem 2.8.6 and the arithmetic-geometric inequality.

$$\begin{aligned}
m\left(\frac{(A+B)}{2}\right)^{\frac{1}{n}} &\geq m\left(\frac{A}{2} + \frac{B}{2}\right)^{\frac{1}{n}} = m\left(\frac{(A)}{2}\right)^{\frac{1}{n}} + m\left(\frac{(B)}{2}\right)^{\frac{1}{n}} = \frac{(m(A))^{\frac{1}{n}} + (m(B))^{\frac{1}{n}}}{2} \\
&\geq \sqrt{(m(A))^{\frac{1}{n}}\ (m(B))^{\frac{1}{n}}}. \blacksquare
\end{aligned}$$

2.9 Rectifiable Curve

Definition 2.9.1

We consider a continuous curve $\Gamma = \{z(t) = (x(t), y(t)) \mid t \in [a, b]\}$ with continuous real-valued functions x and y of t (in the XOY plane). The length of curve Γ is often approximated as the supremum of the lengths of all polygonal lines (Figure 11), which are successively connected by finitely many points of the curve Γ.

$$L(\Gamma) = \sup_{\{t_j\}} \sum_{j=1}^{N} \left| z(t_j) - z(t_{j-1}) \right|.$$

If the sum of all lengths is bounded, i.e. there is a $M < \infty$ for each partition, so that

$$\sum_{j=1}^{N} \left| z(t_j) - z(t_{j-1}) \right| \leq M,$$

the curve Γ is **rectifiable** [RAR0051].

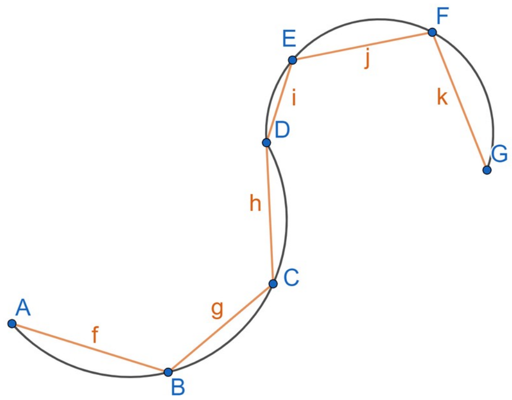

Figure 11: Approximation of a rectifiable curve by (polygonal) lines *f, g, h, i, j* and *k.*

Remark 2.9.2

When the function *z(t)* or *x(t)* and *y(t)* are also continuously differentiable, the length of the curve is defined as follows

$$L = \int_a^b \left[(x'(t))^2 + (y'(t))^2 \right]^{\frac{1}{2}} dt^1.$$

Definition 2.9.3

Let F(t) be a complex-valued function defined on $[a,b]$ mit $a = t_0 < t_1 < \cdots < t_N = b$ as a Partition. The **variation** on this partition is defined by

$$\sum_{j=1}^{N} \left| F(t_j) - F(t_{j-1}) \right|.$$

Furthermore, the function F *is called* **bounded variation**, if the variations of the function over all partitions are bounded. In other words, there exists a $M < \infty$ for each partition, so that

$$\sum_{j=1}^{N} \left| F(t_j) - F(t_{j-1}) \right| \leq M.$$

Remark 2.9.4

The requirement of continuity of the function *F(t)* only comes into play when it is referred to as a curve namely $F(t) = z(t) = x(t) + iy(t)$ Also given is a partition $\tilde{\wp}$[2] by $a = \tilde{t}_0 < \tilde{t}_1 < \cdots < \tilde{t}_N = b$ and as a refinement of another partition $\wp$ by $a = t_0 < t_1 < \cdots < t_N = b$. Thus, the variation of F on F on $\tilde{\wp}$ is greater or equal to the variation of F on $\wp$.

Theorem 2.9.5

A curve $(x(t), y(t))$ parametrized by is rectifiable for $t \in [a, b]$ only if both functions *x(t)* and *y(t)* are of bounded variation.

Proof: It follows directly from the definition of the $F(t) = x(t) + iy(t)$ and the telescopic sum rule for $l_j = F(t_j) - F(t_{j-1}) = (x(t_j) - x(t_{j-1})) + i(y(t_j) - y(t_{j-1}))$, when a and b are real

$$|a + ib| \leq |a| + |b| \leq 2|a + ib|. \blacksquare$$

Definition 2.9.6

- The **total variation** of a function F auf $[a, x]$, wobei $a \leq x \leq b$, is defined by

$$T_F(a, x) = \sup_{\{t_j\}} \sum_{j=1}^{N} \left| F(t_j) - F(t_{j-1}) \right|,$$

1 The condition of continuous differentiability for *x(t)* and *y(t)* can be satisfied by absolute continuity, because *z(t)* is now differentiable for a.e. $t \in [a, b]$.

2 A partition $\tilde{\wp}$ of $[a, b]$ is a refinement of a Partition $\wp$ on $[a, b]$, if every point is in the $\wp$, is also in the $\tilde{\wp}$.

where it is the supremum over all partitions of $[a, x]$.

- The **positive variation** of a function F auf $[a, x]$ is defined by

$$P_F(a, x) = \sup_{\{t_j\}^+} \sum_{j=1}^{N} F(t_j) - F(t_{j-1}),$$

where it is the supremum over all partitions $[a, x]$ with $F(t_j) \geq F(t_{j-1})$.

- The **negative variation** of a function F auf $[a, x]$ is defined by

$$N_F(a, x) = \sup_{\{t_j\}^-} \sum_{j=1}^{N} \left[F(t_j) - F(t_{j-1})\right],$$

where it is the supremum over all partitions $[a, x]$ with $F(t_j) \leq F(t_{j-1})$.

Lemma 2.9.7

Let F be a real-valued function and one of bounded variation on $[a, b]$, it holds for all $x \in [a, b]$

i. $F(x) - F(a) = P_F(a, x) - N_F(a, x)$

ii. $T_F(a, x) = P_F(a, x) - N_F(a, x)$

Definition 2.9.8

A curve $z(t)$ parametrized by $(x(t), y(t))$ is called **simple** for $t \in [a, b]$, if the mapping $t \mapsto z(t)$ injective[1] for $t \in [a, b]$. If it holds $t \in [a, b)$ with $z(a) = z(b)$, is called $z(t)$ a **closed simple** curve. Generally, a curve is called **quasi-simple** (Figure 12), if the mapping $z(t)$ is injective for t in the complement[2] of finitely many points of the interval $[a, b]$.

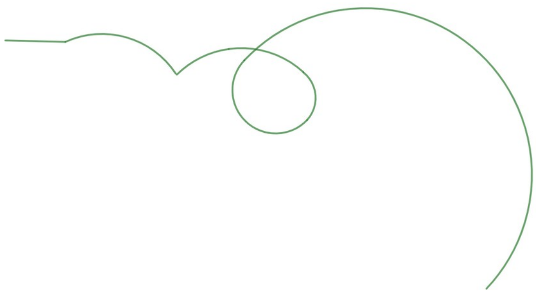

Figure 12: A quasi-simple curve.

Definition 2.9.9

Let $K \in \Re^2$ be a compact set (like Γ above). We represent an open set K^δ as a set that contains all points of the set K at a distance less than δ (Figure 13)

$$K^\delta = \left\{x \in \Re^2 \,\middle|\, d(x, K) < \delta\right\}.$$

We define **Minkowski content**† of the set K with

$$\mu(K) = \lim_{\delta \to 0} \frac{m(K^\delta)}{2\delta},$$

when the above limit exist.

Remark 2.9.10

For a compact set K, an equivalence to the Minkowski contents can be represented with the help of the infimum and supremum of the limit, namely for the two values

1 A mapping $f\colon X \to Y$ is called injective, if for each element y of the target set Y ther is at most one element x of the initial set X. This yields $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$

† This is one dimensional (1D) Minkowski content.

$$\mu^{*}(K) = \limsup_{\delta \to 0} \frac{m(K^{\delta})}{2\delta} \text{ and } \mu_{*}(K) = \liminf_{\delta \to 0} \frac{m(K^{\delta})}{2\delta}$$

$\mu_{*}(K) = \mu^{*}(K) < \infty$ must hold.

Theorem 2.9.11

Let $\Gamma = \{z(t) = (x(t), y(t)) \mid t \in [a, b]\}$ be a quasi-simple curve. The Minkowski content of the Γ exist if and only if the Γ is rectifiable. If this is the case, then $\mu(\Gamma) = L$ (2.9.1).

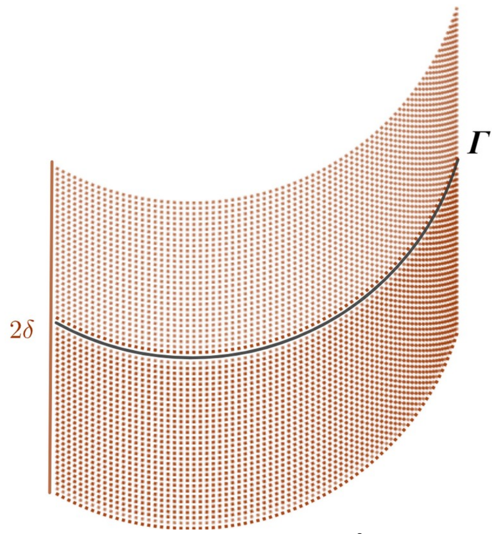

Figure 13: The curve Γ and the set Γ^{δ}.

Proposition 2.9.12

Let $\Gamma = \{z(t) = (x(t), y(t)) \mid t \in [a, b]\}$ be a quasi-simple curve. If $\mu_{*}(\Gamma) < \infty$, then the curve is rectifiable and its length $L \le \mu_{*}(\Gamma)$.

Lemma 2.9.13‡

Let $\Gamma = \{z(t) = (x(t), y(t)) \mid t \in [a, b]\}$ be any curve, and $\Delta = |z(b) - z(a)|$ the distance between its endpoints. Then it holds that

$$m(\Gamma^{\delta}) \ge 2\delta\Delta.$$

Proposition 2.9.14

Let $\Gamma = \{z(t) = (x(t), y(t)) \mid t \in [a, b]\}$ be a rectifiable curve with length L. It holds that

$$\mu^{*}(\Gamma) \le L.$$

Remarks 2.9.15

i. There is also the so-called gamma function, which is defined for $s > 0$ as follows.

$$\Gamma(s) = \int_0^{\infty} e^{-x} x^{s-1}\, dx.$$

The integral can be approximated with the so-called Stirling's formula as follows

$$\Gamma(s) = e^{s \log(s)} e^{-s} \frac{\sqrt{2\pi}}{\sqrt{s}} \left(1 + \sum_{j=1}^{N} a_j s^{-j} + O(s^{-N})\right).$$

‡ To better understand the background of this statement, a prior knowledge of the segmentation under total variation is helpful, which is dispensed with due to the core topic of our letter.

Here a_i, (e.g $a_1 = \frac{1}{2}$) are real constants for each $N > 0$, when $s \in S_\delta = \{s : |arg\, s| \leq \pi - \delta\}$ [•]

ii. With regard to the length of a curve (in Remark 2.9.2) , a parametrization can be established around variable $t \in [a, b]$ and thus around the set $s = s(t) = L(a, t)$ by the pair $\tilde{z}(s) = \tilde{x}(s) + i\,\tilde{y}(s)$, when $\tilde{z}(s) = z(t)$ for $s = s(t)$. Here, It is clear that $|\tilde{z}(s_1) - \tilde{z}(s_2)| \leq |s_1 - s_2|$(distance between two points $\leq$ length of the portion) of the curve. This parameterization is referred to as arc length parameterization.

Theorem 2.9.16

Let $z(t) = (x(t), y(t)), \forall t \in [a, b]$, be a rectifiable curve with length L. Consider the arc length parameterization $\tilde{z}(s) = (\tilde{x}(t), \tilde{y}(t))$described above, then $\tilde{x}(t)$ and $\tilde{y}(t)$ are absolutely continuous, $|\tilde{z}'(s)| = 1$ a.e $s \in [0, L]$, and

$$L = \int_0^L \left[(\tilde{x}'(s))^2 + (\tilde{y}'(s))^2\right]^{\frac{1}{2}} ds.$$

Proof: We note that $|\tilde{z}(s_1) - \tilde{z}(s_2)| \leq |s_1 - s_2|$. This immediately results in the absolute continuity of $\tilde{z}(s)$ and therefore the function is differentiable for a.e $s \in [0, L]$. Moreover, dividing this inequality by $|s_1 - s_2|$ leads to $|\tilde{z}'(s)| \leq 1$ a.e $s \in [0, L]$. Now by definition the totale variation of $\tilde{z}(s)$ equals L, indeed

$$L = T_{\tilde{Z}}(a, x) = \int_0^L |\tilde{Z}'(s)|\, ds = \int_0^L \left[(\tilde{x}'(s))^2 + (\tilde{y}'(s))^2\right]^{\frac{1}{2}} ds.$$

The second equation is only possible if $|\tilde{z}'(s)| = 1$ a.e $[0, L]$.▪

2.10 The Isoperimetric Inequality

With the knowledge from the last chapter, we begin to understand and prove the meaning of the inequality first in a 2-dimensional area.

Remark 2.10.1

The Isoperimetric Inequality states that among all curves of a certain length, the circle contains the maximum area [RAR0051].

In preparation for the proof of the inequality, we assume that $\Omega \in \mathfrak{R}^2$a bounded open set, whose boundary $\overline{\Omega} \setminus \Omega$ is a rectifiable curve with a length $\ell(\Gamma)$. Whether Γ is a closed simple curve does not matter.

Theorem 2.10.2 (Isoperimetric Inequality in $\mathfrak{R}^2$)

he Isoperimetric $4\pi m(\Omega) < (\ell(\Gamma))^2$.

Proof: For each $\delta > 0$ we consider the outer set

$$\Omega_+(\delta) = \left\{x \in \mathfrak{R}^2 \,\middle|\, d(x, \overline{\Omega}) < \delta\right\}$$

and the inner set

$$\Omega_-(\delta) = \left\{x \in \mathfrak{R}^2 \,\middle|\, d(x, \Omega^C) \geq \delta\right\}.$$

This results in $\Omega_-(\delta) \subset \Omega \subset \Omega_+(\delta)$ (Figure 14). We now note that for

$\Gamma^\delta = \left\{x \in \mathfrak{R}^2 \,\middle|\, d(x, \Gamma) < \delta\right\}$

[•] For any non-negative complex number $z = r\,e^{i\theta} = r(\cos\theta + i\sin\theta)$, $\theta \in \mathfrak{R}$ is called argument of z and it is often denoted by *arg z*.

(1) $$\Omega_+(\delta) = \Omega_-(\delta) \cup \Gamma^\delta$$

as a disjoint union. Furthermore, let $D(\delta) = \{x \in \Re^2 \,|\, |x| < \delta\}$ be the open ball (disk with the origin as the center and radius δ). This results in

$$\begin{cases} \Omega_+(\delta) \supset \Omega + D(\delta), \\ \Omega \supset \Omega_-(\delta) + D(\delta) \end{cases}$$

With the Brunn-Minkowski inequality, the first conclusion results in

$$m\left(\Omega_+(\delta)\right) \geq \left[\left(m(\Omega)\right)^{\frac{1}{2}} + m\left(D(\delta)\right)^{\frac{1}{2}}\right]^2.$$

Since $m\left(D(\delta)\right) = \pi\delta^2$ and $(A + B)^2 \geq A^2 + 2AB$(whenever $AB \geq 0$), we find that

(2) $$m\left(\Omega_+(\delta)\right) \geq m(\Omega) + 2\pi^{\frac{1}{2}}\delta m(\Omega)^{\frac{1}{2}}.$$

Analogous to the first conclusion, the second results in

$$m(\Omega) \geq m(\Omega_-(\delta)) + 2\pi^{\frac{1}{2}}\delta m(\Omega_-(\delta))^{\frac{1}{2}}$$

or equivalent:

(3) $$-m(\Omega_-(\delta)) \geq 2\pi^{\frac{1}{2}}\delta m(\Omega_-(\delta))^{\frac{1}{2}} - m(\Omega).$$

Now follows from the (1)

$$m\left(\Omega_+(\delta)\right) - m(\Omega_-(\delta)) = m\left(\Gamma^\delta\right),$$

and with the upper inequalities (2) and (3) you get

$$m\left(\Gamma^\delta\right) \geq 2\pi^{\frac{1}{2}}\delta\left(m(\Omega)^{\frac{1}{2}} + m(\Omega_-(\delta))^{\frac{1}{2}}\right).$$

Now you get by dividing 2δ on both sides of the last inequality and at the same time noticing that $\Omega_-(\delta) \nearrow \Omega$, as $\delta \to 0$

$$\mu^*(\Gamma^\delta) \geq \pi^{\frac{1}{2}}\left(2m(\Omega)^{\frac{1}{2}}\right).$$

With the Proposition 2.9.14 ($\ell(\Gamma) \geq \mu^*(\Gamma)$) it results that

$$\ell(\Gamma) \geq 2\left(\pi m(\Omega)\right)^{\frac{1}{2}}.\blacksquare$$

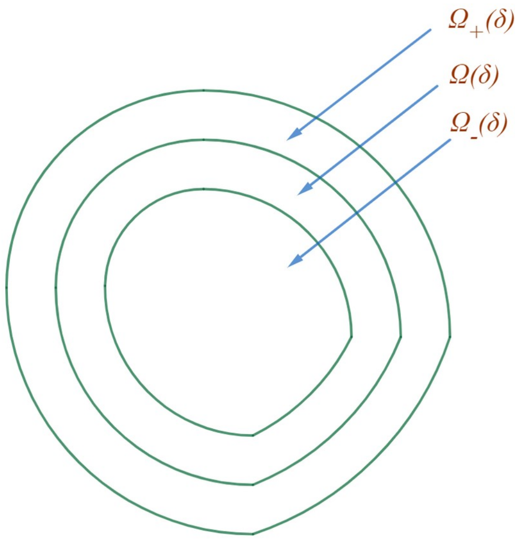

Figure 14: The sets Ω_ , Ω, and Ω₊ ∈ ℝ²

Remarks 2.10.3

i. he same result will even apply without assumption for the edge, i.e. as a rectifiable curve. Indeed , the proof shows that for every bounded open set with a boundary Γ it holds that

$$4\pi m(\Omega) < \left(\mu^*(\Gamma^\delta)\right)^2.$$

ii. In general, the measure of any ball $B \in \Re^d$ can be represented by invariance properties (Translation and Dilatation) of a measurable set. Also $m(B) = v_d r^d$, where the constants $v_d = m(B_1)$ and $B_1 = \{x \in \Re^d \,|\, |x| < 1\}$. In addition, with the Corollary 2.7.4.14 it can be shown that

$$v_2 = \int_{-1}^{1} (1 - x^2)^{\frac{1}{2}}\, dx = \pi.$$

Using the same procedure, the constant for the ball B is found as follows:

$$v_d = 2\int_0^1 (1 - x^2)^{\frac{(d-1)}{2}}\, dx = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2} + 1)},$$

where $\Gamma(z)$ is the Gamma function☼.

Corollary 2.10.5

Let be a curve in in $\Re^d$ as a continuous mapping $t \mapsto z(t)$ from an interval $[a, b] \to \Re^d$. Analogous to the conditions for rectifiability of curves, given in Theorem 2.9.5 and Theorem 2.9.16, it also applies to this curve. In addition, according to the Definition 2.9.9 the Minkowski content $\mu(K)$of a compact set $K \in \Re^d$ holds in the case of existence of the limit $\mu(K) = \lim_{\delta \to 0} \frac{m(K^\delta)}{m_{d-1}(B(\delta))}$, where $m_{d-1}(B(\delta))$ is the measure (in the $\Re^{d-1}$)of Ball with $B_\delta = \{x \in \Re^{d-1} \,|\, |x| < \delta\}$.

Finally, corresponding statements according to Propositions Propositionen 2.9.12 and 2.9.14 apply to the curve described above.

☼ An Equivalent for even dimension can be $\Gamma(\frac{d}{2} + 1)$ by the definition of function as $\prod_{j=0}^{(d/2)-1} (\frac{d}{2} - j)$ and for odd dimension wit taking real number $\frac{d}{2}$, $\Gamma(\frac{d}{2} + 1) = \frac{\sqrt{\pi}(d!!)}{2^{\frac{(d+1)}{2}}}$ with $d!! = \prod_{j=0}^{[d/2]} (d - 2j)$.

3 Johnson-Lindenstrauss Lemma

We start first with necessary Principles of probability theory, before we come to the description of the main lemma or the application of the lemma in stochastic.

3.1 Principles of probability theory

Definition 3.1.1

a) The **sample space** Ω(signed by the **omega set**) is the set of all possible outcomes, which is referred to as a random outcome of an experiment or trial (random experiment).

b) A set $\mathcal{F}$ (often also denoted with $\sum$) as a collection of subsets of the set Ω is (meets the conditions for) **σ-Algebra**[1], if

(i) $\mathcal{F}$ is a non-empty set.

(ii) For each $X \in \mathcal{F}$ the set $\overline{X} = (\Omega \setminus X) \in \mathcal{F}$.

(iii) For each $X, Y \in \mathcal{F}$ the union is $(X \cup Y) \in \mathcal{F}$.

c) A **probability measure** is a mapping $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$ (mapping of the probability) to events and has the following properties [LVSM006][2]:

(i) $\mathbb{P}[\Omega] = 1$.

(ii) (Additivity) For two disjoint sets $X, Y \in \mathcal{F}$ it holds that $\mathbb{P}[X \cup Y] = \mathbb{P}[X] + \mathbb{P}[Y]$.

d) A **probability space** is a triple *(Ω, $\mathcal{F}$, P)*, where *Ω* is a sample space, $\mathcal{F}$ is a σ-algebra, defined over *Ω*, and *P* is a probability measure.

e) A **random variable** *f* is a mapping from *Ω* into a set $\mathcal{G}$. We note that the probability that the random variable takes a value in a given subset of values is well-defined. Formally expressed, this means that for each subset $U \subseteq \mathcal{G}$ applies: $f^{-1}(U) \in \mathcal{F}$. That is, $P[f \in U] = P[f^{-1}(U)]$ is defined. For example, let a random variable be defined by $X: \Omega \rightarrow \Re$ on a probability space *(Ω, $\mathcal{F}$, P)*. We can ask questions, .how likely is it that the value of X is equal to two? This corresponds to the probability of the event $\{\omega \mid X(\omega) = 2\}$ which is often referred to as *P(X = 2)*.

f) We say that *X* is a **continuous** random variable if there is a non-negative function *f(x)* that is defined for all real $x \in (-\infty, \infty)$ and has the property that for each set B of real numbers with respect to the Lebesgue measure λ the following holds [SMR019]:

$$P\{X \in B\} = \mathbb{P}[B] = \int_B f(x)\, d\lambda(x).$$

Here, the function *f(x)* is referred to as **probability density function**(pdf) of the random variable *X* and $\mathbb{P}[B]$ as its **distribution function** (Figure 15).

g) If a real random variable *X* has a probability density function *f(x)* and the **expected value** exists, it can be calculated as follows:

$$\mathbb{E}[X] = x \int_B f(x)\, d\lambda(x).$$

In the case of **discrete** random variable, it is calculated by $\mathbb{E}[X] = \sum_{x:p(x)>0} xp(x)$.

h) Let $(\Omega, \sum, P)$[3] be a probability space and *X* a random variable on this space. The **variance** is defined as follows: $\mathbb{E}[X] = \mu$, $Var(X) := \mathbb{E}[(X-\mu)^2]$.

1 Also Borel set, described in the last chapter (page 27), is a σ-Algebra.

2 The probability measure is a special case of Lebesgue measure and thus it has the properties.

3 The $\sum$ is a σ-Algebra around of the base set Ω.

i) For a vector $X = (X_1, X_2, \cdots, X_n)^T$,whose entries are random variables, a **covariance matrix** is defined as follows [SMR019]:

$$\Re^{n\times n} \ni K_{XX} = \left[K_{X_i X_j}\right]_{1\le i,j\le n} \text{ mit } K_{X_i X_j} = cov[X_i, X_j] = \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right].$$

j) A continuous random variable X has a **normal distribution** with expected value $\mathbb{E}[X] = \mu$ and variance $Var(X) = \sigma^2$, written as $X \sim \mathcal{N}(\mu, \sigma^2)$, if X has the following probability density function (pdf) [ARGB007]:

$$\forall x \in (-\infty, \infty), f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right).$$

Analogous to the one-dimensional normal distribution, the **multidimensional** normal distribution is defined as follows (for real random variable $X \sim \mathcal{N}_d(\mu, \Sigma)$):

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right).$$

Here, Σ is a regular covariance matrix, which is symmetric positive defined[$].

The symmetry here means for every two points $x, y \in \Re^d$ not only $|x| = |y|$ but also the probability density function at their position are equal (i.e. $f(x) = f(y)$).

The **projection** of the normal distribution in each direction is a one-dimensional normal distribution. If you hold ***d*** variables $X_1, X_2, \cdots, X_d$ through one-dimensional normal distribution, this leads to a point $(X_1, X_2, \cdots, X_d)^T$ as a multidimensional normal distribution.

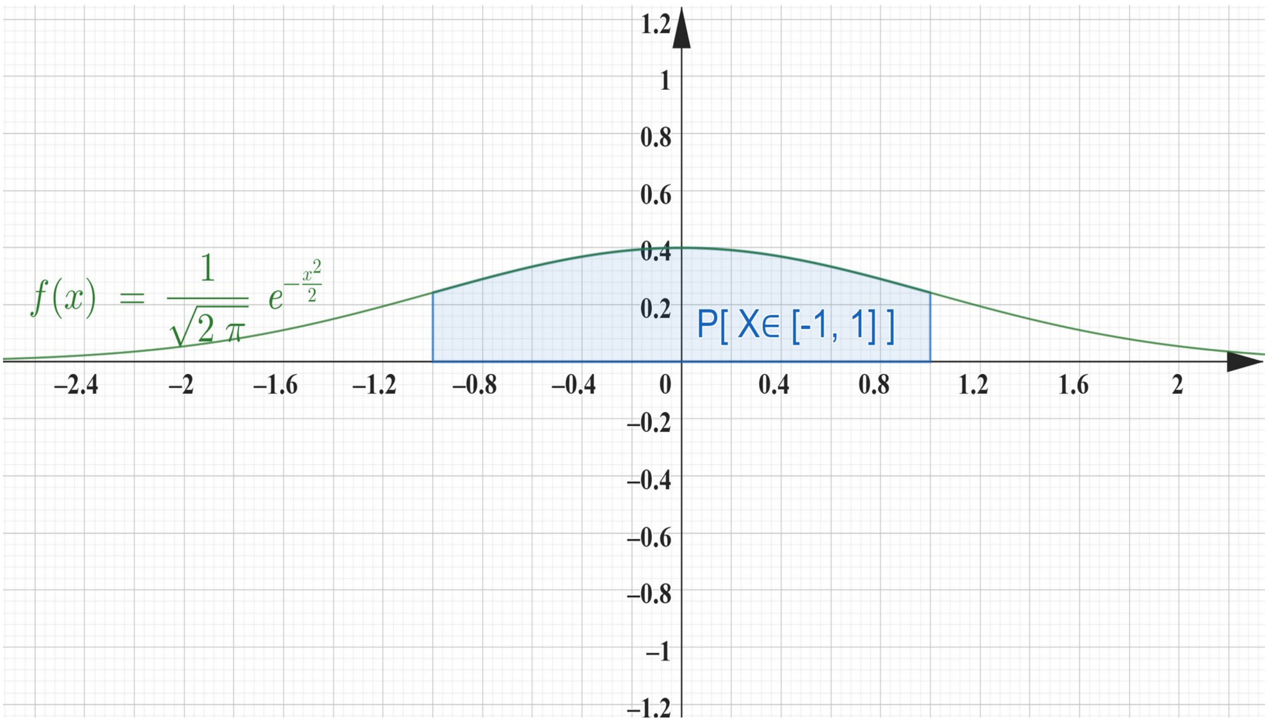

Figure 15:The standard normal distribution function for a 2-dimensional random variable.

3.2 Affine Geometry

While convex sets (2.1.15) contain the line segment between any two points, affine sets contain the entire line between any two points [OGUE010]. Therefore, the natural setting for convex sets is an

---

$ A Matrix $M \in \Re^{n\times n}$ is symmetric positive definite $\Leftrightarrow \forall x \in \Re^n \setminus \{0\},\ x^T M x > 0$.

affine subset of a vector space (Preliminary to Measure Theory 4).

Definition 3.2.1

1. A non-empty subset A of a vector space E is called an affine set if for points x and y in A the line running through x and y is contained in A.
$$x,\ y \in A \Rightarrow \ell := \{x + t(x - y) = (1-t)x + y : t \in \Re\} \subseteq A.$$
2. Let **A** and **B** be affine sets in vector spaces **E** and **D**. A mapping **$F\colon A \to B$** is called an affine mapping if for all t ∈ R $F((1-t)x + ty) = (1-t)F(x) + tF(y)$ holds.
3. Let $\{x_k\}_{k=1}^{d}$ be a finite set of points in the vector space E. An affine combination of $\{x_k\}_{k=1}^{d}$ is any point $y = \sum_{k=1}^{d} \lambda_k x_k$ with $\sum_{k=1}^{d} \lambda_k = 1$.
4. Let B be a non-empty set in the vector space E. The affine hull (or span) of B is the set of all affine combinations of points from B, that is:
$$\left\{ \sum_{k=1}^{d} \lambda_k b_k \,\middle|\, b_k \in B,\ \sum_{k=1}^{d} \lambda_k = 1,\ d = 1, 2\cdots \right\}.$$
5. Let $A \subseteq E$ be an affine subset of the vector space E and $a \in A$ an arbitrary point, then $L := A - a = \{y - a \mid y \in A\}$ is a linear subspace of the vector space E, which is independent of the point $a \in A$. Consequently it holds that $A = a + L$ and $L := A - a = \{y - z \mid y, z \in A\}$†.

Remark 3.2.2

An affine space is also well-defined by a set **A** together with a vector space **V** under a transitive and free group operation of the additive group of **V** on the set **A**. Explicitly, the above definition means that the group operation is a mapping, generally referred to as addition, with the following properties [BM984]:

$$A \times V \to A$$
$$(a, v) \mapsto a + v,$$

1. Neutral element:
$\forall a \in A,\ a + 0 = a$, where 0 the zero vector in **V** ist.
2. Associativity:
$\forall v, w \in V, \forall a \in A, (a + v) + w = a + (v + w)$ (here the last addition is an addition in **V**)
3. Free and Transitive Operations:
$\forall a \in A$, the mapping $V \to A\colon v \mapsto a + v$ is a bijection (page4).
4. Existence of injective translations:
$\forall v \in V$, the mapping $A \to A\colon a \mapsto a + v$ is a bijection.
5. Subtraction:
$\forall a, b \in A$, there exists a unique $v \in V$, denoted as $b - a$, so that $b = a + v$.

Thus $\Re^n$, an affine space (as a vector space over itself) and its affine subspaces have the form $E = p + V$, where $V \subset \Re^n$ is a vector subspace.

† Readers can find for their own interest the proof of these statments in the book „Foundation of Optimization“ by Osman Güler in the chapter 4.1 ‚Affine Geometry.

Example:

Let $p = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \wedge V = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \middle| \begin{pmatrix} x \\ y \end{pmatrix} = t\begin{pmatrix} 2 \\ 3 \end{pmatrix}, t \in \Re \right\} \subset \Re^2$, then the affine subspace is :

$$E = p + V = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \middle| \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1+2t \\ -1+3t \end{pmatrix}, t \in \Re \right\} \subset \Re^2.$$

Lemma 3.2.3

For affine spaces D and E, an affine mapping in the form $F: D \to E$ with $F(x) = Ax + b$ can be represented, where the matrix-vector product „*Ax*“ plays the role of a linear transformation[↷] and the vector "*b*" plays the role of the translation.

Proof: Let t ∈ R and $x, y \in D$, then $F((1-t)x + ty) = M((1-t)x + ty) + b$

$$= (1-t)(Mx + b) + t(My + b) + (t - 1)b - tb + b = (1-t)F(x) + tF(y).$$ ▪

3.3 Concentration of Measure

The Concentration of Measure is another aspect of measure theory, which predominantly describes a mathematical phenomenon that often occurs in stochastics.

Definition 3.3.1

(a) A **metric** space is an ordered pair *(M, d)*, where *M* is a set and *d* is a **metric** on *M*, i.e. a function $d: M \times M \to \Re$, which fulfills the following prerequisites for all points $x, y, z \in M$ [RAR0052]:

i. The distance from a point to itself is zero: $d(x,x) = 0$.

ii. (Positivity) The distance between two different points is always positive:

If $x \neq y$, dann $d(x, y) > 0$.

iii. (Symmetry) The distance from x to y is always equal to the distance from y to *x*:

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

iv. The triangle inequality holds:

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

(b) Given two metric spaces (X, d_X) and (Y, d_Y), where d_X denotes the metric on the set X and d_Y denotes the metric on the set *Y*, a function $f: X \to Y$ is called **Lipschitz continuous** if there is a real constant K ≥ 0 such that for all x_1 and x_2 in *X*:

$d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2)$. Each such *K* is referred to **Lipschitz constant** for the function f and f is often also referred to as *K-Lipschitz*.

(c) Let *(X, d)* be a metric space with a measure μ on the Borel sets(see above) with $\mu(X) = 1$.

Let $\alpha(\varepsilon) = \sup\{ m(X \setminus B_\varepsilon) | B$ is a Borel set and $\mu(B) \geq 1/2 \}$, where the function $\alpha(.)$ is referred to as the **concentration rate** of the space *X*.

(d) In a metric space $M=(X,d_X)$, a set *V* is a neighborhood of a point *p*, if there is an open ball (2.1.1) with center *x* and radius $r > 0$, so that it is contained in *V* (Figure 16[‡]).

(e) A **spherical** probability measure is a probability measure on the sphere, i.e. a spherical

↷ Let V and W be vector spaces over a common fundamental field(set with well-definition for the addition and multiplication operations) K. A mapping $f: V \to W$ is called a linear mapping if for all $u, v \in V$ and $\lambda \in K$ the following condition applies: f(λu+v)=λf(u) + f(v).

‡ Image source from Wikipedia.

surface. It describes the distribution of random points [e)] on the sphere. The following are to be considered:

i. For the probability measure, the probability space is $(\Omega, |\,.\,|, P)$, where $|\,.\,|$ is the Euclidean norm (see page 3).

ii. This is a rotation invariant on the spherical surface $S^{(n-1)} := \{x \in \Re^n \mid |x| = 1\}$normalized. I.e. for a $A \subseteq S^{n-1}$ and a rotation matrix T it holds that $m(\widehat{A}) = m(TA) = m(A)$, wehre $TX = \{\alpha x \mid x \in X, \alpha \in [0, 1]\}$is a simple representation of X after the rotation (We remember that elements of a rotation matrix always lie between 0 and 1).

iii. We consider a function $f\colon S^{(n-1)} \to \Re$, and imagine that a probability density function is defined over the sphere. Let further

$$\mathbb{P}[f \le t] = \mathbb{P}\left[\left\{x \in S^{(n-1)} \mid f(x) \le t\right\}\right].$$

The **median** of the function f is denoted as ***med(f)*** and defined by $\sup_t \mathbb{P}[f \le t]$.

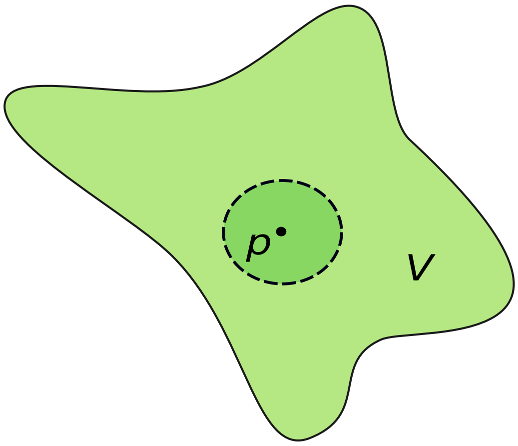

Figure 16‡: Set **V** is a neighbourhood pf the point **p**.

Lemma 3.3.2

Let $\mathcal{F}$ be a k-dimensional affine subspace, and $P_{\mathcal{F}}\colon \Re^d \to \mathcal{F}$ be a projection that maps each point $x \in \Re^d$ to its nearest neighborhood to $\mathcal{F}$, then P as the projection is a contraction (i.e., 1 *Lipschitz continuous*). In other words, the following statement holds:

$$|P_{\mathcal{F}}(p) - P_{\mathcal{F}}(q)| \le |p - q|.$$

Proof: Obviously, the projection $P_{\mathcal{F}}$ is an affine mapping between $\Re^d$ (Subtraction:) and $\mathcal{F}$. According to Lemma Lemma 3.2.3 in the last section, there exist a matrix $\boldsymbol{M}$ and a vector $\boldsymbol{b}$ for $P_{\mathcal{F}}(p) = o$ as follows: $P_{\mathcal{F}}(p) = Mp + b$.

With the definition of the projection, we get:

$$|(M-I)p + b| = |Mp + b - p| = |P_{\mathcal{F}}(p) - p| = |o - p| < \varepsilon^{(1)}.$$

From the triangle inequality, we know: $|(M-I)p + b| \le |(M-I)p| + |b|^{(2)}$.

From (1) and (2) together it follows: $|(M-I)p| \le |(M-I)p| + |b| \le \varepsilon \equiv |(M-I)p| \le 0$, as $\varepsilon \to 0$ (od. $|Mp| \le |p|$, as $\varepsilon \to 0$). Now from the last inequality it follows:

$$|P_{\mathcal{F}}(p) - P_{\mathcal{F}}(q)| \le |Mp + b - Mq - b| \le |M(p - q)| \le |p - q|. \blacksquare$$

Lemma 3.3.3

Für jede $\hat{a} \in T(A) = \widehat{A}$ (od. *TA above*) and $\hat{b} \in T(B) = \widehat{B}$ hlods

$$\left|\frac{\hat{a} + \hat{b}}{2}\right| \le 1 - \frac{t^2}{8}.$$

Proof: Let $\hat{a} = \alpha a$ and $\hat{b} = \beta b$ for $a \in A$ und $b \in B$(Figure 17) and for simplification we mark $|Om_1| = |m_1|$, $|Oa| = |a|$, and $|Ob| = |b|$, further for positive parameter t mit $t \le |a - b|$ it holds that

$$|m_1| = \left|\frac{a + b}{2}\right| = \sqrt{1^2 - \left|\frac{a - b}{2}\right|^2} = \sqrt{1^2 - \left(\frac{t}{2}\right)^2} <= \sqrt{1^2 - \frac{t^2}{8}} \le 1 - \frac{t^2}{8}(1).$$

In relation $\hat{a}$ and $\hat{b}$ we assume that $\alpha \le \beta$ holds. It is to be observed that the length $|\hat{a} + \hat{b}|$ is maximized when $\beta = 1$. Consequently, the triangle inequality yields that

$$\left|\frac{\hat{a} + \hat{b}}{2}\right| = \left|\frac{\alpha a + \beta b}{2}\right| = \left|\frac{\alpha a + b}{2}\right| \le \left|\frac{\alpha(a + b)}{2}\right| + \left|\frac{(1 - \alpha) b}{2}\right| \le \alpha\left(1 - \frac{t^2}{8}\right) + (1 - \alpha)\frac{1}{2} = \tau.$$

Obviously, τ is a convex combination of $\frac{1}{2}$ and $\left(1 - \frac{t^2}{8}\right)$. Particularly we conclude that

$$\tau \le \max\left\{\left(1 - \frac{t^2}{8}\right), \frac{1}{2}\right\} \le \left(1 - \frac{t^2}{8}\right), \text{ since } t \le 2. \blacksquare$$

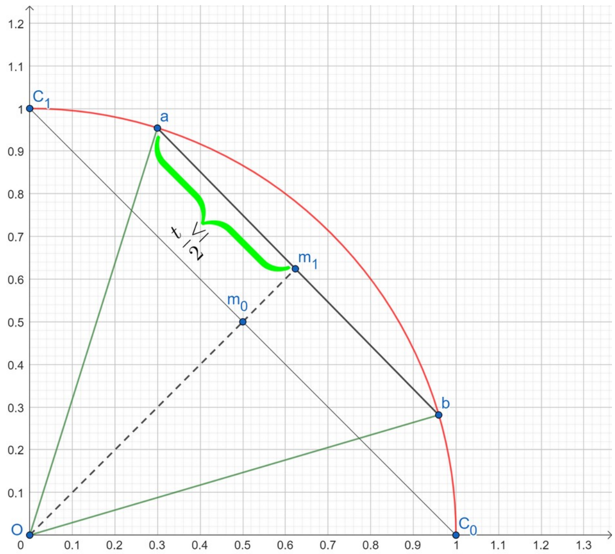

Figure 17: Part of the surface a spherical segment

Theorem 3.3.4 (Measure concentration on the sphere).

Let $A \subseteq S^{(n-1)}$ be a measurable set with $\mathbb{P}[A] \ge \frac{1}{2}$ and let A_t be the set of points of $S^{(n-1)}$ at a distance of at most t from A, where $t \le 2$. Then $1 - \mathbb{P}[A_t] \le 2\exp\left(\frac{-nt^2}{2}\right)$ [SHP018].

Proof: We make a slight deviation from the above assertion or the estimate to the proof, by converting the root of the Euler function to $\sqrt{\exp\left(\frac{-nt^2}{2}\right)} = \exp\left(\frac{-nt^2}{4}\right)$.

We first choose a random point p uniformly within the unit sphere B_1. Let ψ be the probability that p lies in $\hat{A}$. Obviously, $\mathrm{vol}(\hat{A}) = \psi\,\mathrm{vol}(B_1)$. We therefore consider the normalized point $q = \frac{p}{|p|}$. According to the definition of $\hat{A}$ Clearly, p lies in $\hat{A}$ if and only if q lies in A. Thus

$m(\widehat{A}) = \dfrac{\mathrm{vol}(\widehat{A})}{\mathrm{vol}(B_1)} = \psi = \mathbb{P}[p \in \widehat{A}] = \mathbb{P}[q \in A] = \mathbb{P}[A]$, because q s assumed to have a uniform distribution on the hydrosphere. Let $B = S^{(n-1)} \setminus A_t$ and $\widehat{B} = T\ B$, where $A_t = \{x \in \Re^n \mid d(x, A) < t\}$ as t- Inflation (an inner set) is (see Figure 18†). For all $a \in A$ und $b \in B$ we then have $\|a - b\| \geq t$. By applying the last lemma, we know that $\left(\dfrac{\widehat{A} + \widehat{B}}{2}\right)$ is included in the ball rB_1 with $r = \left(1 - \dfrac{t^2}{8}\right)$. We consider that

$$m(\widehat{A}) = m(rB_1) = \frac{\mathrm{vol}(rB_1)}{\mathrm{vol}(B_1)} = r^n = \left(1 - \frac{t^2}{8}\right)^n.$$

Using special case der Brunn-Minkowski inequality, it results

$$\left(1 - \frac{t^2}{8}\right)^n = m(rB_1) = m\left(\frac{\widehat{A} + \widehat{B}}{2}\right) \geq \sqrt{m(\widehat{A})m(\widehat{B})} = \sqrt{\mathbb{P}[A]\mathbb{P}[B]} \geq \sqrt{\frac{\mathbb{P}[B]}{2}}.$$

Thus, because of $1 - x \leq \dfrac{1}{e^x} \leq e^{-x}$(for $x \geq 0$) $\mathbb{P}[B] \leq 2\left(1 - \dfrac{t^2}{8}\right)^n \leq 2\exp\left(\dfrac{-2nt^2}{8}\right) = 2\exp\left(\dfrac{-t^2}{4}\right)$ ▪

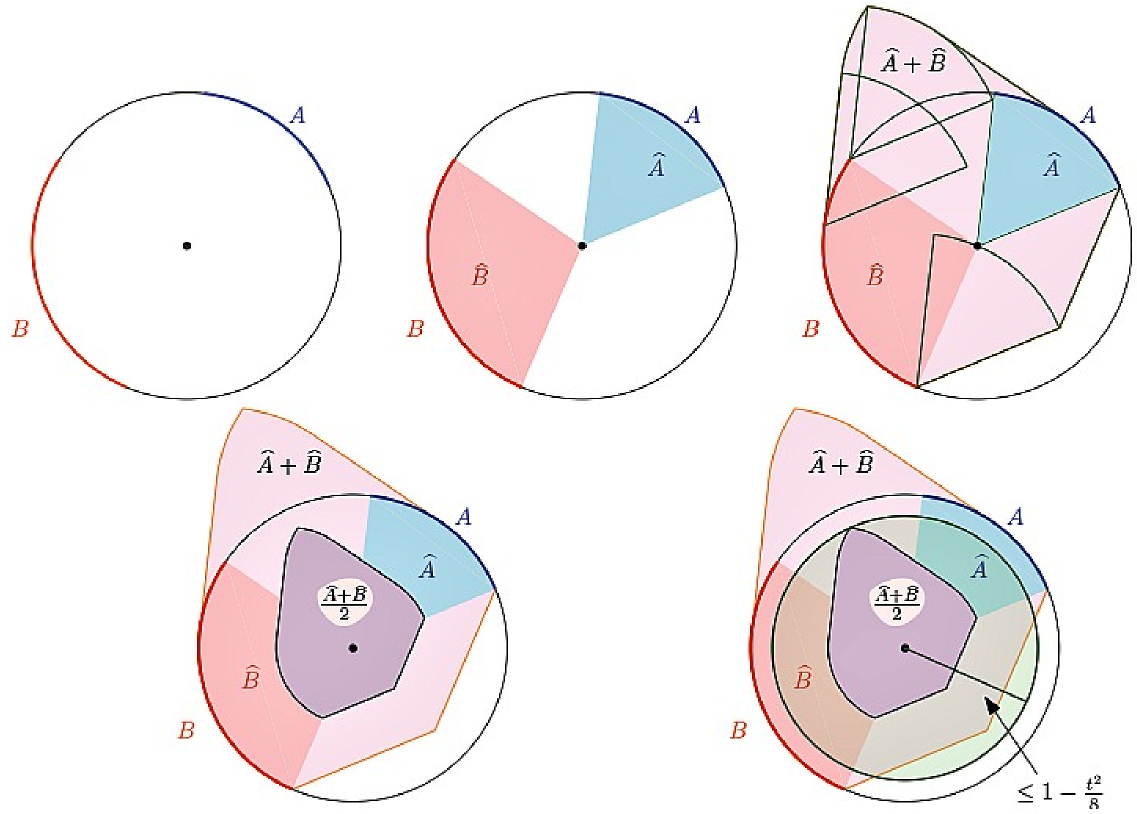

Figure 18†: Demonstration A and B,their rotations $\widehat{A}, \widehat{B}$ on the surface of the sphere $S^{(n-1)}$ in the Theorem 3.3.4.

Lemma 3.3.5

For the probability measure described above, the following is provided:

$$\mathbb{P}[f \leq \mathrm{med}(f)] \leq \frac{1}{2} \text{ and } \mathbb{P}[f > \mathrm{med}(f)] \leq \frac{1}{2}.$$

† The image source from the origin handbook of Sariel Har-Peled.

Proof: Because (2.6.1) $\{f(x) < \mathrm{med}(f)\} = \bigcup_{k=1}^{\infty}\left\{f(x) \le \mathrm{med}(f) - \frac{1}{k}\right\}$

$$= \bigcup_{k=1}^{\infty}\left(-\infty,\ \mathrm{med}(f) - \frac{1}{k}\right] = (-\infty,\ \mathrm{med}(f))$$

$= \bigcup_{k=1}^{\infty}\left(-\infty,\ \mathrm{med}(f) - \frac{1}{k}\right] = (-\infty,\ \mathrm{med}(f))$, we have

$$\mathbb{P}[f \le \mathrm{med}(f)] = \sup_{k \ge 1} \mathbb{P}[f \le \mathrm{med}(f) - \frac{1}{k}] \le \sup_{k \ge 1}\frac{1}{2} = \frac{1}{2}.$$

The second assertion follows analogously to the first through the following symmetric argumentation:

$$\{f(x) > a\} = \{f(x) \le a\}^C;\ \forall a \in \mathfrak{R}. \blacksquare$$

The following theorem is often also called „Lévy's isometric inequality" and deals with the functions that fulfill the first Lipschitz constant, and provides an upper estimate including medians.

Theorem 3.3.6 (Lévy's isometric inequality)

Let $f: S^{(n-1)} \to \mathfrak{R}$ be a 1-Lipschitz function. Then for all $t \in [0, 1]$:

$$\mathbb{P}[f > \mathrm{med}(f) + t] \le 2\exp\left(\frac{-nt^2}{2}\right) \text{ and } \mathbb{P}[f < \mathrm{med}(f) - t] \le 2\exp\left(\frac{-nt^2}{2}\right)$$

Proof: Let $A = \left\{x \in S^{(n-1)} \,\middle|\, f(x) \le \mathrm{med}(f)\right\}$. Remembering the last lemma (3.2.4) we have $P[A] \ge \frac{1}{2}$. We now consider a point $x \in A_t$, where A_t is a *t*-Inflation over $S^{(n-1)}$. Let $\eta(x)$ be the nearest point in A to x. We then have by definition of A_t $\|x - \eta(x)\| \le t$. On the other hand, from the 1-Lipschitz continuity of the function f and $\eta(x) \in A$:

$$f(x) - f(\eta(x)) \le \|x - \eta(x)\| \le t \to f(x) \le t + f(\eta(x)) \le t + \mathrm{med}(f).$$

With the measure concentration on the sphere theorem it results:

$$\mathbb{P}[f > \mathrm{med}(f) + t] = 1 - \mathbb{P}[f - \mathrm{med}(f) \le t] = 1 - \mathbb{P}[A_t] \le 2\exp\left(\frac{-nt^2}{2}\right).$$

The second assertion follows analogously to the first through the symmetric argumentation. ▪

3.4 The Johnson-Lindenstrauss Lemma

Lemma 3.4.1

For a unit vector $x \in S^{(n-1)}$ let $\xi(x) = (x_1^2, x_2^2, \ldots, x_k^2)^{\left(\frac{1}{2}\right)}$ be the length of the projection of x in the subspace formed by the first k coordinates . Furthermore, let x be a random vector generated from uniform distribution from $S^{(n-1)}$. Then the mentioned norm is referred to as sharply concentrated with the following property:

$$\mathbb{P}[\xi(x) \ge m + t] \le 2\exp\left(\frac{-nt^2}{2}\right) \text{ and } \mathbb{P}[\xi(x) \le m - t] \le 2\exp\left(\frac{-nt^2}{2}\right),$$

where the variable m depends on the number of dimensions of the original and subspace($m = m(n, k)$) and $t \in [0, 1]$. Furthermore, for $k \ge 10\ln(n)$, the variable $m \ge \frac{1}{2}\sqrt{\frac{k}{n}}$.

Proof: The orthogonal projection $P: \mathfrak{R}^n \to \mathfrak{R}^k$ given by $p(x_1, \ldots, x_n) = (x_1, \ldots, x_k)$ is according to the Lemma 3.3.2 1-Lipschitz (i.e., distances can only become smaller). As such $\xi(x) = |P(x)|$ is 1-Lipschitz, siche for eache *x, y* hold

$$|\xi(x) - \xi(y)| = ||P(x)| - |P(y)|| \leq |P(x) - P(y)| \leq |x - y|.$$

Lévy's isometric inequality provides the required overestimate with $m = med(f)$. Thus, we only need to prove the lower bound for *m*.

For a random $x = (x_1, \ldots, x_n) \in S^{(n-1)}$ we have $\mathbb{E}[|X|^2] = \mathbb{E}[1] = 1$. By linearity of the expectation function and symmetry, for each specific $1 \leq j \leq n$.

$$1 = \mathbb{E}[|X|^2] = \mathbb{E}\left[\sum_{i=1}^{n} x_i^2\right] = \sum_{i=1}^{n} \mathbb{E}[x_i^2] = n\mathbb{E}[x_j^2].$$

For $j = 1, \ldots, n$, as a result $\mathbb{E}[x_j^2] = \frac{1}{n}$ and thus again by linearity:

$$\mathbb{E}[(\xi(x))^2] = \mathbb{E}\left[\sum_{i=1}^{k} x_i^2\right] = \sum_{i=1}^{k} \mathbb{E}[x_i^2] = \sum_{i=1}^{k} \frac{1}{n} = \frac{k}{n}.$$

Next, we use that $\xi(x)$is concentrated to show that $\xi^2(x)$ is also relatively concentrated. Since $\forall x \in S^{(n-1)}$, $\xi(x) \leq 1$, fr each $t \geq 0$ then following holds

$$\frac{k}{n} = \mathbb{E}[\xi^2] = \mathbb{P}[\xi \leq m + t](m + t)^2 + \mathbb{P}[\xi \geq m + t](1) \leq (1)(m + t)^2 + 2\exp\left(\frac{-nt^2}{2}\right)^{(1)}.$$

If we now set $t = \sqrt{\frac{k}{5n}}$, it results with $k \geq 10\ln(n)$ that

$$2\exp\left(\frac{-nt^2}{2}\right) = 2\exp\left(\frac{-k}{10}\right) \leq \frac{2}{\exp(\ln(n))} = \frac{2}{n}^{(2)}.$$

Combining (1) and (2) we get that $\frac{k}{n} \leq \left(m + \sqrt{\frac{k}{5n}}\right)^2 + \frac{2}{n}$.

This means that $\sqrt{\frac{k-2}{n}} \leq \left(m + \sqrt{\frac{k}{5n}}\right)$ and thus it leads to

$$m \geq \sqrt{\frac{k-2}{n}} - \sqrt{\frac{k}{5n}} = \sqrt{\frac{1}{n}}\left(\frac{0.8k - 2}{\sqrt{k-2} + \sqrt{0.2k}}\right)^{(3)}.$$

Now we know that $\forall n \in \mathbb{N}$, $k \geq 2 \wedge k \geq 10\ln(n)$. In other words, ***k*** should be in the last expression $10\ln(2)$. However, already for a ***k*** with $k \geq 4.5 < 10\ln(2)$ the following inequality holds:

$$k' = k - 2; \sqrt{k-2} + \sqrt{0.2k} \leq 2\sqrt{0.8(k-2.5)} \equiv \sqrt{k' + 0.5} + \sqrt{0.2k' + 0.5} \leq 2\sqrt{0.8k''}$$

If we replace this (last) inequality in (3),the desired inequality yields directly

$$m \geq \frac{1}{2}\sqrt{\frac{k}{n}}. \blacksquare$$

Remark 3.4.2

- **projection of a Vector**

  Next, we want to argue that the projection of a fixed vector into a random k-dimensional subspace leads to a random vector whose length is highly concentrated. This would mean that we can perform a dimension reduction and still maintain distances between points that are important to us. For this purpose, we want to reverse the last lemma. Instead of randomly selecting a point x and projecting it onto the first k-dimensional space, we want to fix x and randomly select the k-dimensional subspace into which we project. However, we must carefully select this random k-dimensional space. In fact, if we rotate this random subspace through a transformation T so that it occupies the first k dimensions, the point T(x) must be uniformly distributed on the sphere in order to be able to use Lemma (3.4.1). Therefore, we want to select a random rotation of $\mathfrak{R}^n$. This maps the standard orthonormal basis into a randomly rotated orthonormal system. The choice of the subspace spanned (3.2.4) by the first ***k*** vectors of the rotated basis results in a random **k**-dimensional subspace. Such a rotation is an orthonormal matrix with determinant 1. We can generate such a matrix by randomly selecting a vector $e_1 \in S^{(n-1)}$ʋ. Next, we set as the first column of our rotation matrix and generate the other *n – 1* associated basis vectors by recursively generating *n - 1* orthonormal vectors in the space orthogonal to e_1 [SHP018].

- **Generation of a Random Point on the Sphere**

At this point, of course, the reader may wonder how to select a point uniformly from the unit sphere $S^{(n-1)}$. The idea is to select a point from the multidimensional normal distribution $X \sim \mathcal{N}_n(\mu=0, \sigma^2=1)$ and normalize it so that it has length 1. Since the multidimensional normal distribution has the density function $2\pi^{\frac{-n}{2}} \exp(\frac{-|X|^2}{2})$, which is symmetric (i.e., all points have the same distribution at distance ***r*** from the origin), it follows that this actually generates a point randomly and uniformly on $S^{(n-1)}$. Generating a vector with multidimensional normal distribution is nothing more than selecting each coordinate according to the normal distribution. Given a source of random numbers corresponding to the uniform distribution, this can be done with *O(1)* calculations per coordinate by using the Box-Muller transformation [BM58]. In total, any random vector can be generated in *O(n)* time. Since projecting an **n**-dimensional normal distribution into the lower-dimensional space results in a normal distribution, it follows that generating a random projection is nothing more than randomly selecting **n** vectors according to the multidimensional normal distribution $v_1, \ldots, v_n$. Then we orthonormalize them with Graham-Schmidt, which is nothing more than $\widehat{v_1} = \frac{v_1}{|v_1|}$ with normalized vector of $v_i - w_i$, where w_i is the projection of v_i onto the space spanned by $v_1, \ldots, v_{i-1}$ (3.2.4). By taking these vectors as columns of a matrix *A* with determinant either *1* or *-1*. We multiply one of the vectors by -1 if the determinant is -1. The resulting matrix is a random rotation matrix. We can now reformulate the last Lemma regarding the scenario as mentioned, in which the vector is fixed and the projection takes place in a random subspace [SHP018].

Lemma 3.4.3

Let $x \in S^{(n-1)}$ be any unit vector. Now consider a random k-dimensional subspace Ƒ and let $\xi(x) = (x_1^2, x_2^2, \ldots, x_k^2)^{(\frac{1}{2})}$ be the length of the projection of *x* into Ƒ. Then there is *m = m(n, k),* so that

$$\forall t \in [0,1],\ \mathbb{P}[\xi(x) \geq m + t] \leq 2\exp\left(\frac{-nt^2}{2}\right) \text{ and } \mathbb{P}[\xi(x) \leq m - t] \leq 2\exp\left(\frac{-nt^2}{2}\right),$$

Furthermore for $k \geq 10\ln(n)$ the variable $m \geq \frac{1}{2}\sqrt{\frac{k}{n}}$ holds [SHP018].

Proof: Let v_i be the *i-th* orthonormal vector with 1 at the *i-th* coordinate. Let *M* be a random

ʋ It is provided as orthogonal basis vector or first eigenvector of the orthogonal matrix versehen.

translation of the space, generated as described above. Obviously, for any fixed unit vector *x*, the vector *Mx* is uniformly distributed on the sphere. Now the *i-th* column of the matrix *M* is the random vector e_i and $M^T v_i = e_i$. Thus, we have

$$\langle Mx, v_i \rangle = (Mx)^T v_i = x^T (M)^T v_i = x^T e_i = \langle x, e_i \rangle.$$

In particular, if *Mx* is treated as a random vector and projected onto the first *k* coordinates, we have

$$\xi(x) = \sqrt{\sum_{i=1}^{k} \langle Mx, v_i \rangle^2} = \sqrt{\sum_{i=1}^{k} \langle Mx, e_i \rangle^2}.$$

But $e_1, \ldots, e_k$ form just an orthonormal basis of a random *k*-dimensional subspace. Thus, the right-hand expression in the last equation is the length of the projection of x into a random k-dimensional subspace. Therefore, the length of the projection of x into a random k-dimensional subspace has exactly the same distribution as the length of the projection of a random vector into the first k coordinates. The claim now follows from Lemma 3.4.1.▪

Definition 3.4.4

The mapping $f : \Re^n \to \Re^k$ is for $X \subseteq \Re^n$ ***Bilipschitz equivalence***, if the following condition is fulfilled,

$$\forall\, p, q \in X, \exists c > 0 \text{ s.t } c K^{-1} |p - q| \leq |f(p) - f(q)| \leq c |p - q|.$$

The smallest constant K^*,which fulfills the upper condition (i.e. $K^* = \min K = \frac{|f(p) - f(q)|}{c|p - q|}$) is called ***Distortion*** of the function *f* and is often denoted with *dist(f)*. The function *f* is marked here as a ***K-embedding*** of *X*. We emphasize that it is fundamentally analogous to the definition of the Lipschitz constant (3.3.(b)) that a *Bilipschitz* equivalence for *f: X → Y* exist, if there is a real constant K ≥ 1, so that for all x_1 and x_2 in *X* in the metric spaces d_X, d_Y the following holds

$$\frac{1}{K} d_X(x_1, x_2) \leq d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2).$$

In other words the first condition (being defined the Euclidean norm) is fulfilled if and only if

$$K = c^2 \text{ [NW018].}$$

Remark 3.4.5

Let $X \subseteq \Re^m$ be a set with *n* points, where *m* may be much larger than *n* can be($m \gg n$). In this case, due to the observation of the distances between the points in *X*, X can be considered as a set of points that lie in the affine subspace $\mathcal{F}$ spanned by the points of the set *X*. This subspace has the dimension n-1. Therefore, each point of *X* can be interpreted as an *(n - 1)*-dimensional point in $\mathcal{F}$[¥]. In other words, it can be assumed that the set of n points being observed lies in $\Re^n$ (more precisely in $\Re^{n-1}$) If $m < n$ the coordinates of the points in *X* can be replaced by zeros so that they lie in $\Re^n$.

Theorem 3.4.6 (Johnson-Lindenstrauss Lemma)

Given a set *X* with *n* points in a Euclidean space and an $\varepsilon \in (0, 1]$. Then there exists a *(1 + ε)-embedding* of *X* in $\Re^k$ with $k = O(\varepsilon^{-2} \log n)$.

Proof: According to the last Remark, it can be assumed that $X \subseteq \Re^n$. For $k = 200\varepsilon^{-2} \ln(n)$ we assume that $k < n$, $\mathcal{F}$ is a random k-dimensional linearer subspace of $\Re^n$ and $P_{\mathcal{F}} : \Re^n \to \mathcal{F}$, and the orthogonal projection operator from $\Re^n$ in the $\mathcal{F}$. Furthermore, let *m* be the number by which for $x \in S^{(n-1)}$, $|P_{\mathcal{F}}(x)|$ is concentrated, as in Lemma 3.4.3. For two points $x, y \in \Re^n$ the following

¥ An affine space is in fact a subspace of a projective space, which in turn arises from a vector space through an equivalence relation (not through a linear subspace).

experssion can be proven with a probability greater than or equal to $1 - \left(\frac{1}{n}\right)^2$, i.e.

$$\left(1 - \frac{\varepsilon}{3}\right)m|x - y| \leq \left|P_F(x) - P_F(y)\right| \leq \left(1 + \frac{\varepsilon}{3}\right)m|x - y|(\text{☼}).$$

Since there are totally $\binom{n}{2}$ possibilities for choosing a pair of points in X, and the $\left(\frac{1}{n}\right)^2$ is probability of a fixed pair for which the upper statement does not apply (i.e. $\left(\frac{1}{n}\right)\left(\frac{1}{n}\right)$), the probability for at least one such event (pairs that do not fulfill upper inequalities (statement)) is

$\binom{n}{2}\left(\frac{1}{n}\right)^2 = \frac{n(n-1)}{2}\left(\frac{1}{n^2}\right) = \frac{n-1}{2n}$. Thus, the probabilities for the counter-event, namely the upper statement should apply to all pairs, $1 - \frac{n-1}{2n} = \frac{n+1}{2n} > \frac{1}{2}$. In this case, the mapping P is a D-embedding of X in $\Re^k$ adoptable to the definition above $\frac{c}{D} \geq \left(1 - \frac{\varepsilon}{3}\right)m$ with $c = 1 + \frac{\varepsilon}{3}$.

Thus, the embedding is guaranteed by a constant as below

$$D \leq \frac{1 + \frac{\varepsilon}{3}}{(1 - \frac{\varepsilon}{3})} \leq 1 + \frac{\frac{2\varepsilon}{3}}{(1 - \frac{\varepsilon}{3})\} \geq \frac{2}{3}} \leq 1 + \varepsilon.$$

Regarding the linearity of the operator $P_F(.)$, we have for $u = x - y$, $P_F(u) = P_F(x) - P_F(y)$. Thus, the restriction becoms $\left(1 - \frac{\varepsilon}{3}\right)m|u| \leq \left|P_F(u)\right| \leq \left(1 + \frac{\varepsilon}{3}\right)m|u|$.

Since the projection is a linear operator, the restriction for each $\alpha > 0$ is equivalent to

$$\left(1 - \frac{\varepsilon}{3}\right)m|\alpha u| \leq \left|P_F(\alpha u)\right| \leq \left(1 + \frac{\varepsilon}{3}\right)m|\alpha u|.$$

Therefore, it can be assumed that $|u| = 1$ with the choice $\alpha = \frac{1}{|u|}$. Consequently it must shown that

$$\left||P_F(u)| - m\right| \leq \frac{m\varepsilon}{3}.$$

Let $f(u) = |P_F(u)|$. Since $m \geq \frac{1}{2}\sqrt{\frac{k}{n}}$ and $k = 200\varepsilon^{-2}\ln(n)$, for $t = \frac{\varepsilon m}{3}$ w.r.t. Lemma 3.4.1 (where the random space is swapped with the random vector), the probability that this is not fulfilled is bounded by

$$\mathbb{P}\left[\left|f(u) - m\right| \leq t\right] = \mathbb{P}\left[f(u) \leq m + t\right] + \mathbb{P}\left[f(u) \geq m - t\right] \leq 4\exp\left(\frac{-nt^2}{2}\right) = 4\exp\left(\frac{-nm^2\varepsilon^2}{18}\right)$$

$$\leq 4\exp\left(\frac{-k\varepsilon^2}{72}\right) = 4\left(e^{-\ln(n)}\right)^{\left(\frac{200}{72}\right)} = \frac{4}{n^{\frac{25}{9}}} < \frac{1}{n^2}. \blacksquare$$

Remark 3.4.7

We consider that the last inequality in the proof above is only valid for $n \geq 6$. Modifying, e.g. simplify $\left(1 - \frac{\varepsilon}{3}\right)$, $\left(1 + \frac{\varepsilon}{3}\right)$ into (☼) $\left(1 - \frac{\varepsilon}{2}\right)$ resp. $\left(1 + \frac{\varepsilon}{2}\right)$ the inequality will apply for a $n \geq 2$.

4 Appendix

4.1 List of Illustrations

4.2 Bibliography

[L. JP009]: L. Jacques, Presentation slides of Laurent Jacques, 2009
[NW018]: Nik Weaver, Lipschitz Algebras, 2018
[SHP018]: Sariel Har-Peled, The Johnson-Lindenstrauss Lemma, 2018
[RAR0052]: Elias M. Stein, Rami Shakarchi, Real Analysis, 2005
[BM984]: Marcel Berger, Problems in Geometry, 1984
[OGUE010]: Osman Güler, Foundations of Optimization, 2010
[ARGB007]: Alvin C. Rencher and G. Bruce Schaalje, Linear Model In Statistics, 2007
[SMR019]: Sheldon M. Ross, Introduction to Probability Models, 2019
[LVSM006]: LaValle, Steven M, Planning Algorithms, 2006
[FJIR001]: Frank Jones, Lebesgue Integration on Euclidean Space, 2001
[RAR0051]: Elias M. Stein, Rami Shakarchi, Real Analysis, 2005