A pair of n-tuples $(a_1,\ldots,a_n)$ and $(b_1,\ldots,b_n)$ where $a_k<b_k$ hold for each k, give the open rectangle

$$\square(a_\bullet,b_\bullet)=\{x=(x_1,\ldots,x_n)\,:\;a_k<x_k<b_k\}$$

Its n-dimensional volume is defined as the product of the differences $\{b_k-a_k\}$, i.e.

$$\text{vol}_n(\square(a_\bullet,b_\bullet))=\prod(b_k-a_k) \tag{0.1}$$

Starting from this, Lebegue introduced the family of *null sets*. By definition a subset A in $\mathbf{R}^n$ is a null set of there to every $\epsilon>0$ exists some denumerable family of open cubes $\{\square_\nu\}$ whose union contains A while

$$\sum \text{vol}_n(\square_\nu)<\epsilon$$

Keeping n fixed the family of null-sets in $\mathbf{R}^n$ is denoted by $\mathcal{N}$. In Lebesgue theory one identifies various object which only differ an a null set. For example, a pair of real-valued functions f and g defined in $\mathbf{R}^n$ are said to be *equal almost everywhere* if they only take distinct values on a null set.

Inspired by Lebegue's notion of null set, a more general notion of thin sets was introduced by Hausdorff in 1907. Given a postive number $0<\alpha<n-1$ one says that a set A in $\mathbf{R}^n$ has Hausdorff dimension zero of or order α if there to every $\epsilon>0$ exists some denumerable family of open cubes $\{\square_\nu\}$ whose union contains A while

$$\sum \text{vol}_n(\square_\nu)^\alpha<\epsilon$$

The family of sets A for which this holds is denoted by $\mathcal{N}_\alpha$. The reader can check that

$$0<\alpha<\beta \implies \mathcal{N}_\alpha\subset\mathcal{N}_\beta$$

In § xx we prove Stokes Theorem for bounded open subsets Ω in $\mathbf{R}^n$ where the boundary $\partial\Omega$ satiosfies a certain regularity condition expressed via Hausdorff dimensions of order $n-1$. Further familes of "thin sets" appear with important applications. Consider for example a closed subset E of the unit circle T. If $0<t\leq 1$ we get the open subset $E(t)$ in T whicoh consists of points $e^{i\theta}$ such that

$$\min_{e^{i\alpha}\in E}|e^{i\theta}-e^{i\alpha}|<\delta$$

The 1-dimensional Lebegue theory gives the function

$$\phi_E(t)=\text{vol}_1(E(t))$$

Now $t\mapsto\phi_E(t)$ is a continuous function on of t. Here E is a null set in the sense of Lebesgue if and only if $\phi_E(t)\to 0$ as $t\to 0$. We can impose the stronger condition that the integral

$$\int_0^1\frac{\phi_E(t)}{t}\,dt<\infty$$

Lebesgue theory.

The theory was created by Henri Lebesgue in his monograph *Lecons sur l'integration et la recherche des fonctions primitives* from 1904. Let us also remark that Vitali and de la Vallé Poussin later found some simplifications of Lebesgue's original proofs. We shall work in euclidian spaces $\mathbf{R}^n$ where n is some positive integer and points are $x = (x_1, \ldots, x_n)$. Keeping n fixed we construct partitions of $\mathbf{R}^n$ as follows:

0.1 Dyadic grids. For each non-negative integer N we consider the family $\mathcal{D}_N$ of cubes:

$$\square_N(k_\bullet) = \{x\colon\ k_\nu \cdot 2^{-N} \leq x_\nu < (k_\nu + 1) \cdot 2^{-N}\} \tag{0.1.1}$$

where $k_\bullet = (k_1, \ldots, k_n)$ are n-tuples of integers. Then

$$\{\square_N(k_\bullet)\} \quad : \ k_\bullet \in \mathbf{Z}^n$$

are pairwise disjoint subsets of $\mathbf{R}^n$ whose union taken over all $k_\bullet \in \mathbf{Z}^n$ is equal to $\mathbf{R}^n$. Notice that every cube in $\mathcal{D}_N$ is the disjoint union of 2^n many cubes from $\mathcal{D}_{N+1}$.

Let us now consider a bounded open subset Ω of $\mathbf{R}^n$, i.e Ω is open and contained in some ball $\{|x| < r\}$ of radius r centered at the origin. Denote by $\mathcal{D}_1(\Omega)$ the family of cubes from $\mathcal{D}_1$ which are contained in Ω. In the next step, $\mathcal{D}_2(\Omega)$ is the family of cubes $\square \in \mathcal{D}_2$ which are disjoint from the union of cubes in $\mathcal{D}_1(\Omega)$. By induction we get a sequence $\{\mathcal{D}_k(\Omega) : \ k = 1, 2, \ldots\}$, where $\mathcal{D}_{k+1}(\Omega)$ consists of cubes in $\mathcal{D}_{k+1}$ which are contained in Ω and disjoint from the union of cubes taken from $\mathcal{D}_1(\Omega), \ldots, \mathcal{D}_k(\Omega)$.

Now the union of all cubes from $\{\mathcal{D}_k(\Omega) : \ k = 1, 2, \ldots\}$ is equal to Ω. To see this we pick a point $x \in \Omega$. For every $k \geq 0$ the grid of degree N gives a unique cube $\square_k(x) \in \mathcal{D}_N$ which contains x. Next, since Ω is bounded the reader can check that there exists a *unique smallest integer* N such that

$$\square_N(x) \subset \Omega$$

and then the construction above entails that $\square_N(x)$ appears in the family $\mathcal{D}_N(\Omega)$. For every $k \geq 0$ we denote by $\rho_k(\Omega)$ the number of cubes in the family $\mathcal{D}_k(\Omega)$. The n-dimensional volume of Ω is defined by

$$\mathrm{vol}_n(\Omega) = \sum_{k=1}^{\infty} \rho_k(\Omega) \cdot 2^{-kn} \tag{0.1.2}$$

We can express the series in another way. For every $m \geq 1$ we denote by $\rho^*_m(\Omega)$ the number of cubes in $\mathcal{D}_m$ which are contained in Ω. The reader may check that

$$\rho^*_m(\Omega) \cdot 2^{-mN} = \sum_{k=1}^{k=m} \rho_k(\Omega) \cdot 2^{-kn} \implies$$

$$\mathrm{vol}_n(\Omega) = \lim_{m \to \infty} \rho^*_m(\Omega) \cdot 2^{-mn} \tag{0.1.3}$$

Example. A pair of n-tuples $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ where $a_k < b_k$ hold for each k, give the open rectangle

$$\square(a_\bullet, b_\bullet) = \{x = (x_1, \ldots, x_n) : \ a_k < x_k < b_k\}$$

Then its n-dimensional volume is given by the product

$$\mathrm{vol}_n(\square(a_\bullet, b_\bullet)) = \prod (b_k - a_k) \tag{0.1.4}$$

To see this one uses binary expansions of real numbers. Given a pair $a < b$ and a positive integer N chosen so large that $2^{-N} < b - a$ we find a pair of integers $k_*(N) < k^*(N)$ so that

$$(k_*(N) - 1) \cdot 2^{-N} \leq a < k_*(N) \cdot 2^{-N} \quad : \quad k^*(N) 2^{-N} < b \leq (k^*(N) + 1) 2^{-N}$$

We leave as an exercise to check that these binary expansions of real numbers entail that the volume of $\square(a_\bullet, b_\bullet)$ from (0.1.4) is equal to that defined by (0.1.3)..

0.2 The Lebesgue measure of compact sets. For each non-negative integer N we have the family $\overline{\mathcal{D}}_N$ of closed dyadic cubes

$$\overline{\square}_N(k_\bullet) = \{x\colon\ k_\nu \cdot 2^{-N} \leq x_\nu \leq (k_\nu + 1) \cdot 2^{-N}\}$$

Let E be a compact subset of $\mathbf{R}^n$. If $N \geq 0$ we denote by $\rho^*_N(E)$ the number of closed cubes in $\overline{\mathcal{D}}_N$ which have *a non-empty intersection* with E.

Exercise. Show that the sequence

$$N \mapsto \rho^*_N(E) \cdot 2^{-Nn}$$

is a non-increasing function of N. Since every non-decreasing seque ce of positive real numbers have a limit there exists the number

(0.2.1) $$\lim_{N\to\infty} \rho^*_N(E) \cdot 2^{-Nn}$$

Following de Valle Poussin this limit is denoted by $\mathrm{vol}_n(E)$ and called the n-dimensional Lebesgue measure of the compact set E.

0.3 Some limit formulas. Let Ω be a bounded open set. The positive series (0.1.2) converges so if $\epsilon > 0$ we find an integer N such that

(i) $$\mathrm{vol}_n(\Omega) < \epsilon + \sum_{k=1}^{k=N} \rho_k(\Omega) \cdot 2^{-kn}$$

Next, each $1 \leq k \leq N$ gives the finite family $\mathcal{D}_k(\Omega)$. If $0 < a < 1$ and $\square \in \mathcal{D}_k(\Omega)$ we construct the compact cube $\overline{a \cdot \square}$ as in (xxx) above. The union of cubes $\{\overline{a \cdot \square}\}$ taken from the families $\mathcal{D}_1(\Omega), \ldots, \mathcal{D}_N(\Omega)$ is a compact subset of Ω denoted by $K_N(a)$, Now one has

(ii) $$\mathrm{vol}_n(K_N(a)) = a^n \cdot \sum_{k=1}^{k=N} \rho_k(\Omega) \cdot 2^{-kn} > a^n \cdot (\mathrm{vol}_n(\Omega) - \epsilon)$$

Above $\epsilon > 0$ is arbitrary and a can be chosen arbitrary close to one. From this the reader can check that

(0.3.1) $$\mathrm{vol}_n(\Omega) = \sup_{K \subset \Omega} \mathrm{vol}_n(K)$$

with the supremum taken over compact subets of Ω.

0.3.2 Exercise. Deduce from the above that if E is a compact et and Ω some bounded open set which contains E, then

(i) $$\mathrm{vol}_n(\Omega) = \mathrm{vol}_n(\Omega \setminus E) + \mathrm{vol}_n(E)$$

Show also that

(ii) $$\mathrm{vol}_n(E) = \inf_{E \subset \Omega} \mathrm{vol}_n(\Omega)$$

where the infimum now is taken over open sets which contain E. In partocular, put

(iii) $$E(\delta) = \{x :\ \mathrm{dist}(x, E) < \delta\}$$

Then $\{E(\delta)\}$ are open sets and from (ii) the reader should check that

(iv) $$\mathrm{vol}_n(E) = \lim_{\delta \to 0} \mathrm{vol}_n(E(\delta))$$

0.5 Lebesgue measurable sets.

Let A be a bounded subset of $\mathbf{R}^n$. The outer measure is defined by:

$$|A|^* = \inf_{A\subset\Omega} \mathrm{vol}_n(\Omega)$$

where the infimum is taken over open sets containing A. The inner measure is defined by

$$|A|_* = \sup_{E\subset A} |E|^*$$

where the supremum is taken over compact subsets of A.

0.4.1 Definition. *A bounded set A is Lebesgue measurable if*

$$|A|_* = |A|^*$$

and then this common number is denoted by $|A|_n$ and called the Lebesgue measure of A.

0.4.2 σ-additivity. Let $\{A_\nu\}$ be a denumarable sequence of pairwise disjoint and measurable sets which all are contained in some ball $\{|x| < r\}$. Then their union is measurable and one has the equality

$$(*) \qquad |\cup A_\nu|_n = \sum |A_\nu|_n$$

To prove (*) one proceeds as follows: Let $\epsilon > 0$. For every ν we can find an open set U_ν which contains A_ν while

$$|U_\nu|_n < |A_\nu|_n + 2^{-\nu} \cdot \epsilon$$

Now $\Omega = \cup U_\nu$ is an open set which contains $\cup A_\nu$ and hence

$$\text{(i)} \qquad |\cup A_\nu|^* \leq |\Omega|_n \leq \sum |U_\nu|_n < \sum |A_\nu|_n + \epsilon$$

Next, for each ν we find a compact subset K_ν of A_ν such that

$$\text{(ii)} \qquad |A_\nu|_n < |K_\nu|_n + 2^{-\nu} \cdot \epsilon$$

For every $N \geq 1$ $K_1 \cup \ldots \cup K_N$ is a compact subset of $\cup A_\nu$ which gives

$$\text{(iii)} \qquad |\cup A_\nu|_* \geq \sum_{nu=1}^{\nu=N} |K_\nu|_n$$

where we used the additivity in (§ xx) applied to the pairwise disjoint compact sets $\{K_\nu\}$. From (ii) and (iii) we see that

$$\text{(iv)} \qquad \sum_{nu=1}^{\nu=N} |A_\nu|_n < |\cup A_\nu|_* + \epsilon$$

Here (iv) hold for every N and hence

$$\text{(v)} \qquad \sum_{nu=1}^{\infty} |A_\nu|_n < |\cup A_\nu|_* + \epsilon$$

Since ϵ can be arbitrary small we conclude that (i) and (v) give (*).

0.5 Lebesgue points.

Let A be a bounded and measurable set. For each $x \in A$ and every $\delta > 0$ we put

$$A_\delta(x) = \min_{\square} \frac{\mathrm{vol}_n(A \cap \square)}{\mathrm{vol}_n(\square)}$$

where the minimum is taken over all open cubes which contain x and have volume $\leq \delta^n$. Keeping x fixed the reader can check that the function

$$\delta \mapsto A_\delta(x)$$

is non-decreasing. Following Lebesgue we put

$$\mathfrak{Leb}(A) = \{x \in A : \lim_{\delta \to 0} A_\delta(x) = 1\} \tag{0.5.1}$$

A major result due to Lebesgue asserts that

$$A \setminus \mathfrak{Leb}A) \in \mathcal{N} \tag{0.5.2}$$

To prove (0.5.2) we need a *Covering Lemma* due to Vitali which goes as follows: Let Ω be a bounded open set in $\mathbf{R}^n$ and $\mathcal{V} = \{\square_\alpha\}$ a family of open cubes contained in Ω. Then one constructs a sequence of cubes form $\mathcal{V}$ as follows: Put

$$\rho_1 = \max_{\square \in \mathcal{V}} \text{vol}_n(\square)$$

and pick $\square_1^*$ so that is volume is $> 2\rho_1/3$. In the next step

$$\rho_2 = \max_{\square \in \mathcal{V}[1]} \text{vol}_n(\square)$$

where $\mathcal{V}[1]$ is the family of cubes from $\mathcal{V}$ which are disjoint from the closed cube $\overline{\square}_1$. Pick $\square_2^*$ in this family so that

$$\text{vol}_n(\square_2^*) > 2/3 \cdot \rho_2$$

One continues by induction, i.e. after k steps we have found cubes $\square_1^*, \ldots, \square_k^*$ where the closed cubes are disjoint and then

$$\rho_{k+1} = \max_{\square \in \mathcal{V}[k]} \text{vol}_n(\square)$$

where $\mathcal{V}[k]$ is the family of $\mathcal{V}$-cubes which are contained in open complement of $\overline{\square}_1^* \cup \ldots \cup \overline{\square}_k^*$.

Since the $\mathcal{V}$-cubes stay in a fixed bounded set it is clear that

$$\lim_{k \to \infty} \rho_k = 0$$

Next, for every positive integer N we put

$$S_N = \overline{\square}_1^* \cup \ldots \cup \overline{\square}_N^* \bigcup_{\nu > N} 3 \cdot \square_\nu^*$$

where $3 \cdot \square_\nu^*$ denote the cubes with the same center as $\square_\nu^*$ while the sides are three times larger. It is clear that the sets $\{S_N\}$ decrease with N and we put

$$S_* = \cap S_N$$

0.5.3 Definition. *Let $\mathcal{V}$ be a family of open cubes as above. A point $x \in \Omega$ is said to be covered by $\mathcal{V}$ in the sense of Vitali if there for every $\epsilon > 0$ exists a cube $\square \in \mathcal{V}$ whos volume is $< \epsilon$ and $x \in \square$.*

0.5.4 Vitali's covering theorem. *The set of points which are covered by $\mathcal{V}$ in Vitali's sense is contained in the set S_*.*

Proof. Let x be covered in Vitali's sense. We must show that $x \in S_N$ for every positive integer. If x already belongs to $\overline{\square}_1^* \cup \ldots \cup \overline{\square}_N^*$ we are done. If x is outside this set the assumption that it is covered in Vitali sense gives a small open cube $\square \in \mathcal{V}$ such that

$$x \in \square \quad \& \quad \square \in \mathcal{V}[N]$$

Next , since the ρ-numbers above tend to zero it follows that $\square$ cannot belong to $\mathcal{V}[M]$ for every $M > N$, and we find the *smallest* integer $M \geq N$ such that

$$\square \in \mathcal{V}[M] \quad \& \quad \square \cap \square_{M+1}^* \neq \emptyset \tag{i}$$

The first inclusion in (i) entails that

$$|\square|_n \leq \rho_{M+1} \implies \text{vol}_n(\square_{M+1}^*) > 2/3 \cdot |\square|_n \tag{ii}$$

At the same time we have the non-empty intersection in (i) and by drawing a figure the reader can check that (ii) entials that

(iii) $$\square \subset 3 \cdot \square^*_{M+1}$$

The consgruction of S_N means that (iii) gives

$$\square \subset S_N$$

and since x belongs to $\square$ we have proved the requested inclusion $x \in S_N$.

0.5.5 Proof of Lebegue's theorem.

Let A be a bounded and measurable set. Fix some $0 < \rho < 1$ and for every $\delta > 0$ we put

(i) $$S_\delta = \{x \in A : \; A_\delta(x) < \rho\}$$

We leave it to the reader to show that S_δ is a relatively open subset of A, i.e.

(ii) $$S_\delta = A \cap U_\delta$$

hold for some open set ${}_\delta$. In particular S_δ is measurable and we also notice that these sets decrease with δ. It follows from (§ xx) that

$$S_* = \bigcap_{\delta>0} S_\delta$$

is measurable. We shall prove that S_* is a mull set. To obtain this we consider some $\epsilon > 0$ and find an open set Ω which contains S_* while

(iii) $$|\Omega|_n < |S_*|_n + (1-\rho) \cdot \epsilon$$

Next, consider a point $x \in S_*$. We find $\delta > 0$ such that every open cube which contains x and has volume $< \delta^n$ is contained in Ω. The construction of S_δ gives therefore an open cube $\square$ which contains x where

(iv) $$\square \subset \Omega \quad \& \quad \mathrm{vol}_n(A \cap \square) \leq \rho \cdot |\square|_n$$

Let $\mathcal{V}$ be the family of such open cubes. If $x \in S_*$ we can find $\square$-cubes as above where $\delta 0$ is arbitrary small. Hence each $x \in S_*$ is covered by $\mathcal{V}$ in the sense of Vitali. Vitali's Theorem gives therefore sequence of pairwise disjoint cubes $\{\square_\nu\}$ in $\mathcal{V}$ such that

(v) $$S_* = \overline{\square}^*_1 \cup \ldots \cup \overline{\square}^*_N \bigcup_{\nu>N} 3 \cdot \square^*_\nu$$

hold for every $N \geq 1$. Since $S_* \subset A$ it follows from (v) and σ-additivity that

(vi) $$|S_*|_n \leq \sum |\overline{\square}^*_\nu \cap A|_n \leq \rho \cdot \sum |\overline{\square}^*_\nu|_n \leq \rho \cdot |\Omega|_n$$

where the last inequality holds because the cubes $\{\square^*_\nu\}$ are disjoint. Togehter (iii) and (vi) give

$$(1-\rho) \cdot |S_*|_n < \rho \cdot (1-\rho) \cdot \epsilon$$

Since $\rho < 1$ this entails that

$$|S_*|_n < \epsilon$$

Here $\epsilon > 0$ was arbitrary which proves that S_* is a null set. Finally, from (i) the set S_* depends on ρ and let us denote it by $S_*(\rho)$. Put

$$S^* = \bigcup_{m \geq 2} S_*(1 - 1/m)$$

B y § xx this denumerable union is again a null set, and at this stage the reader can check that the previous constructions give the inclusion

$$A \setminus S^* \subset \mathfrak{Leb}(A)$$

and Lebegue's Theorem follows.

We prove (0.xx) in § xx and establish also a similar result for Lebesgue measurable sets, and more generally bounded Lebesgue measurable functions. A consequence of this is that every bounded Lebesgue measurable function f is the pointwise limit outside a null set of a sequence of Lipschitz continuous functions defined by mean values of f over smaller and smaller cubes centered at the point where pointwise convergence takes place. In § 3.xx we also show that outside a null set an arbitrary bounded measurable function f differs from an everywhere defined function f_* given as the pointwise limit of a non-increasing sequence of upper semi-continuus functions. This means that f_* belongs to the second Baire class of functions. So when one agrees to identify Lebegue measurable functions which are equal outside a null set, then this class is reduced to a quite restricted family of everywhere defined functions. Roughly speaking, the existence of Lebesgue points implies that no "phantoms" occur in Lebesgue theory.

0.2.9 Non-measurable sets. There exist non-measurable sets. One must apply the Axiom of Choice to exhibit examples. Using this axiom there exists a subset E of the open interval $(0,1)$ with the property that the sets $\{E+q\}$ are disjoint when q runs over the set of rational numbers ehile

$$(0;1) \subset \bigcup_{q\in Q} E+q$$

Now E cannot be measurable. Namely, sincr the right hand side is a union of denumerable sets and the lebegue measure of $E+q$ is equal to that of E, it would first follow that

$$\mathrm{vol}_1(E) > 0$$

But this gives a contradiction. Namely, the disjointness would entail that if $q_1, \dots, q_N$ is a set of rational numbers between 0 and 1 then

$$N \cdot \mathrm{vol}_1(E) = \sum_{\nu=1}^{\nu=N} \mathrm{vol}_1(E+q_\nu) \leq 2$$

where the last equality holds since every set $E+q_\nu \subset [0,2]$. Since N can be arbitrary large we must have

$$\mathrm{vol}_1(E) = 0$$

and from this we conclude that E cannot be measurable,

bigskip

0.2.1 Unbounded open sets. If Ω is an open unbounded set we still get the unique partition formed by the disjoint subsets $\{D_k(\Omega)\}$. One says that Ω has a finite measure if every $D_k(\Omega)$ is a finite union and the series

$$\sum \mu_k(\Omega) \cdot 2^{-kn} < \infty$$

Outer and inner measures. Starting from the construction of n-dimensional measures of open sets one constructs outer and inner measures of an arbitrary bounded subset A of $\mathbf{R}^n$. The outer measure is defined by:

$$|A|^* = \min_{A \subset \Omega} \operatorname{vol}_n(\Omega)$$

In particular the outer measure is defined for every compact set E and the inner measure of A is defined by

$$|A|_* = \max_{E \subset A} |E|^*$$

where the maximum is taken over compact subsets of A.

0.2.2 Definition. *A set A in $\mathbf{R}^n$ has a finite Lebesgue measure if $|A|_* = |A|^* < \infty$ and this common number is denoted by $|A|_n$.*

0.2.3 Null sets. A set is called a null set if its outer measure is zero. The family of null sets is denoted by $\mathcal{N}(\mathbf{R})^n$. A property is said to hold almost everywhere if it is valid outside a null set. In Lebesgue theory one is often content to establish a result "almost everywhere", i.e. null sets are reduntant. Recall that a set S is of type G_δ if it is the intersection of an decreasing sequence of open sets $\{\Omega_n\}$. It is a set of type F_σ if it is given as the union of a decreasing familly of compact sets.

0.2.4 Exercise. Show that if A is Lebesgue measurable and $\{E_\nu\}$ an increasing sequence of compact subsets such that $|E_\nu|_n \to |A|_n$, then $A \setminus \cup E_\nu$ is a null-set. Similarly, if $\{\Omega_\nu\}$ an decreasing sequence of open sets containing A and $|\Omega_\nu|_n \to |A|_n$ then $\cap \Omega_\nu \setminus A$ is a null set. Hence every measurable set differs from a set of type F_σ by a null set, and similarly from a set of type G_δ.

0.2.5 Lebesgue points. Let A be a measurable set. A point $x \in A$ has unit density if

$$\lim_{|\square|_n \to 0} \frac{\operatorname{vol}_n(\square \setminus A)}{\operatorname{vol}_n(\square)} = 0 \tag{0.5.1}$$

where the limit is taken over cubes which contain x. It is not requested that x is the center of the cubes and their sides need not be parallell to the coordinate axis. The set where (0.5.1) holds is denoted by $\mathfrak{Leb}(A)$. A major result due to Lebesgue asserts the following:

0.2.6 Theorem. *For every measurable set A it follows that $A \setminus \mathfrak{Leb}(A)$ is a null set.*

This fundamental result is proved in § 3.

0.2.7 Measurable functions. Let $f(x)$ be a real-valued function defined in some open and bounded subset Ω of $\mathbf{R}^n$. It is measurable in the sense of Lebesuge if the sets

$$A_f(a;b) = \{x \in \Omega \colon a \leq f(x) < b\}$$

are measurable for all pairs of real numbers $a < b$. The function f has Lebesgue value a at a point $x_0 \in \Omega$ if

$$\lim_{|\square|_n \to 0} \frac{|A_f(a-\epsilon, a+\epsilon) \cap \square|_n}{|\square|_n|} = 1 \tag{0.7.1}$$

hold for every $\epsilon > 0$ where the limit of cubes is taken as in (0.5.1). It is obvious that if (0.7.1) hold for some real number $a <$, then it is uniquely determined and we set $\mathfrak{Leb}f(x_0) = a$. In § xx we show that the measurable function f has a Lebesgue value almost everywhere. More preciely, there exists a null set W of Ω such that $\mathfrak{Leb}f$ is defined in $\Omega \setminus W$ and is equal to f in this set. So up to null sets every measurable function f is recaptured from its Lebesgue function, i.e. ignoring a null set f can always be chosen so that its values coincide with its Lebesgue values.

0.2.8 Fubini's theorem. Let n snd m be a pair of positive integers and A a bounded and measurable subset of $\mathbf{R}^{n+m}$ whose points are written as (x, y) with $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$. To every x we get the slice

$$A(x) = \{y \colon (x, y) \in A\}$$

In §§ we prove that $A(x)$ are measurable sets in $\mathbf{R}^m$ for all x outside a null set and the almost everywhere defined function

$$x \mapsto |A(x)|_m$$

is measurable. After we have learnt how to construct Lebegue integrals we also prove the equality

$$\int_{\mathbf{R}^n} |A(x)|_m \, dx = |A|_{n+m} \tag{0.8.1}$$

0.2.9 Non-measurable sets. There exist non-measurable sets but in practice one does not encounter these ugly sets. In fact, one must appeal to the Axiom of Choice to exhibit examples. Using this axiom there exists a subset E of the open interval $(0, 1)$ with the property that the sets $\{E + q\}$ are disjoint when q runs over the set of rational numbers. This set cannot be measurable. The reason is that the disjointness would entail that if $q_1, \ldots, q_N$ is a set of rational numbers between 0 and 1 then

$$N \cdot |E|_* = \sum_{\nu=1}^{\nu=N} = |E + q_\nu|_* \leq 2$$

where the last equality holds since each set $E + q_\nu \subset [0, 2]$. Since N can be arbitrary large we have $|E|_* = 0$. So if E is measurable we have $|E| = |E|_* = 0$, i.e E is a null set. Since $E + q$ is translate of E they also give null sets for every rational number. But a denumerable union of null sets is a null set which gives a contradaction since $\cup\,(E + q)$ with the union taken over all rational numbers is the whole real line.

II. Abstract measure theory

Introduction.

To get some intuition we present the theory using concepts from probability theory. A sample space consists of a pair $(\Omega, \mathcal{B})$ where Ω is a set and $\mathcal{B}$ a Boolean σ-algebra of subsets. Thus, if $\{A_\nu\}$ is a denumarable family of sets in $\mathcal{B}$ indexed by non-negative integers, then $\cap A_\nu$ and $\cup A_\nu$ stay in $\mathcal{B}$. Moreover Ω belongs to $\mathcal{B}$ and one agrees that its complementary empty-set also belongs to the Boolean σ-algebra.

0.1 Measurable functions. A real valued function f defined on Ω is measurable if the sets

(0.1.1) $\qquad f^{-1}(-\infty, a)$ and $f^{-1}(-\infty, a]$ both belong to $\mathcal{B}$ $\quad : \quad$ a any real number

The class of such functions is denoted by $\mathcal{M}$.

Exercise. Use that $\mathcal{B}$ is a σ-algebra to show that a real-valued function f is measurable if the sets in (0.1.1) belong to $\mathcal{B}$ for every rational number a. Next, let f and g be a pair of measurable functions and show that $f+g$ is measurable. The hint is that for every real number a one has the set-theortetic equality:

$$\{f + g < a\} = \bigcup_{q \in Q} \{f < q\} \cap \{g < a - q\}$$

where the union is taken over all rational numbers. In a similar way one proves that $\{f + g \leq a\}$ are measurable sets for every real number a. Conclude that $\mathcal{M}$ is a vector space of the real number field.

0.2 Probability measures. A probability measure μ is a σ-additive map

$$\mu\colon\ \mathcal{B} \to [0, 1]$$

which assigns a numer $0 \leq \mu(A) \leq 1$ for every set A in $\mathcal{B}$, and $\mu(\Omega) = 1$. The σ-additivity means that if $\{A_\nu\}$ is a sequence of pairwise disjoint sets in $\mathcal{B}$, then

$$\mu(\cup A\nu) = \sum \mu(A_\nu)$$

The family of probablility measures is denoted by $\mathbf{P}(\mathcal{B})$.

0.3 Nullsets. Let μ be a probability measure. A subset F of Ω is a null set with respect to μ if there to each $\epsilon > 0$ exists $A \in \mathcal{B}$ such that

$$F \subset A \quad \& \quad \mu(A) < \epsilon$$

The class of null-sets is denoted by $\mathcal{N}_\mu$. By σ-additivity $\mathcal{N}_\mu$ is stable under a denumarable union, i.e.

$$\{F_\nu\} \subset \mathcal{N}_\mu \implies \cup F_\nu \in \mathcal{N}_\mu$$

Notice that one does not require that a null-set belongs to $\mathcal{B}$. If A_1 and A_2 is a pair of sets in $\mathcal{B}$ one says that they are μ-equivalent if the Boolean difference

$$A_1 \setminus A_2 \cup A_2 \setminus A_1 \in \mathcal{N}(\mu)$$

When it holds we write

$$A_2 \overset{\mu}{\simeq} A_1$$

0.4 μ-measurable sets. Let $B \subset \Omega$ be an arbitrary subset. Its outer - resp. inner μ-measure are defined by

$$\mu^*(B) = \min_{A \in \mathcal{B}} \mu(A) \quad : B \subset A \quad : \quad \mu_*(B) = \max_{A \in \mathcal{B}} \mu(A) \quad : \ A \subset B$$

Notice that both the maxium and the minium can be attained. For example, let $\{A_n\}$ be a sequence in $\mathcal{B}$ where $B \subset A_n$ hold for every n while

$$\lim \mu(A_n) = \mu^*(B)$$

Put

$$A^* = \cap A_n$$

By σ-additivity the reader can chek that

$$\mu(A^*) = \mu^*(B)$$

Here A^* belongs to $\mathcal{B}$. It is not unique but if A^{**} is anopther set which contains B and (x) holds, then the reader can verify that A^* and A^{**} are μ-equivalent. In the same way one finds A_* where $A_* \subset B$ and

$$\mu_*(B) = \mu(A_*)$$

The set B is called μ-measurable if

$$\mu_*(B) = \mu^*(B)$$

and then this common number is denoted by $\mu(B)$. Let $\mathfrak{M}_\mu$ be the family if all μ-measurable subsets of Ω. It is clear that we have the inclusion

$$\mathcal{B} \subset \mathfrak{M}_\mu$$

0.5 Exercise. From the above a set B is μ-measurable if and only if there exists some $B_* \in \mathcal{B}$ such that the Bolean difference of B and B_* belongs to $\mathcal{N}(\mu)$. Conclude that $\mathfrak{M}_\mu$ is the Boolean σ-algebra of subsets of Ω generated by $\mathcal{B}$ and $\mathcal{N}_\mu$. Show also that the map

$$B \mapsto \mu(B)$$

is σ-additive on $\mathfrak{M}_\mu$.

0.6 μ-measurable functions.

A real-valued function f on Ω is μ-measurable if

$$f^{-1}(-\infty, a) \text{ and } f^{-1}(-\infty, a] \text{ both belong to } \mathfrak{M}_\mu \quad : \quad a \in \mathbf{R}$$

Denote this class by $\mathcal{M}_\mu$. Since $\mathcal{B} \subset \mathfrak{M}_\mu$ one has the inclusion

$$\mathcal{M} \subset \mathcal{M}_\mu \tag{0.6.1}$$

0.6.2 Teorem. *Let $f \in \mathcal{M}_\mu$. Then there exists a null set $F \in \mathcal{N}_\mu$ and some $f_* \in \mathcal{M}$ such that $f = f_*$ in $\Omega \setminus F$. Thus, after modifying a μ-measurable function on a null set with respect to μ it becomes "truly measurable".*

Proof. Let $\{q_\nu\}_1^\infty$ enumerate the set Q of rational numbers. The exercise in (0.5) gives for each q_ν a pair of disjoint sets $F_\nu \in \mathcal{B}$ and $G_\nu \in \mathcal{N}_\mu$ such that

$$f^{-1}(-\infty, q_\nu) = F_\nu \cup G_\nu$$

We can also regard the inverse image of the singleton sets and find for every q_ν a disjoint pair $H_\nu \in \mathcal{B}$ and $S_\nu \in \mathcal{N}_\mu$

$$f^{-1}(\{q_\nu\}) = H_\nu \cup S_\nu$$

Put

$$G^* = \cup G_\nu \bigcup \cup S_\nu$$

Here $G^* \in \mathcal{N}_\mu$ which gives the existence of a decreasing sequence of sets $\{W_\nu\}$ in $\mathcal{B}$ such that

$$\mu(W_\nu) < 2^{-\nu} \quad \& \quad G^* \subset \cap W_\nu$$

Now $W_* = \cap W_\nu$ belons to $\mathcal{B}$ and it is also a null-set for μ. Finally, define the function f_* by

$$f_*(\omega) = f(\omega) \quad : \omega \in \Omega \setminus W_* \quad \text{and} \quad f_*|W_* = 0$$

By this construction $f_* \in \mathcal{M}_{\mathcal{B}}$ and f_* differs from f on a null set.

0.6.3 Equivalence classes of μ-measurable functions. A pair f, g in $\mathcal{M}_\mu$ are called equivalent if they are equal outside a null set. Theorem 0.6.2 shows that every equivalence class can be represented by a $\mathcal{B}$-measurable function.

0.7 Convergene in μ-measure.

On $\mathcal{M}_\mu$ there exists a metric defined by

$$d(f,g)=\sum_{k=1}^\infty 2^{-k}\cdot\frac{\mu\{|f-g|\geq 2^{-k}\}}{1+\mu\{|f-g|\geq 2^{-k}\}} \tag{0.7.1}$$

Let $\{f_n\}$ be a sequence in $\mathcal{M}_\mu$ and g a function in $\mathcal{M}_\mu$ such that

$$\lim_{n\to\infty}\ \mu(\{|f_n-g|\geq\epsilon\}]=0$$

hold for every $\epsilon>0$. The reader can check that this is equivalent to the condition that

$$\lim d(f_n,g)=0 \tag{0.7.2}$$

and when this holds we write

$$f_n\overset{\mu}{\to} g \tag{0.7.3}$$

0.8 Convergence almost everywhere.

A sequence of μ-measurable functions $\{f_n\}$ is said to converge μ-almost everywhere to a limit function g if there exists a null set $F\in\mathcal{N}_\mu$ such that every f_n is defined in $\Omega\setminus F$ and one has a pointwise limit:

$$\lim_{n\to\infty} f_n(\omega)=g(\omega):\ \omega\in\Omega\setminus F \tag{0.8.1}$$

When (0.8.1) holds we write

$$f_n\overset{a.e}{\to} g \tag{0.8.2}$$

Almost everywhere convergence means that pointwise convergence holds outside a null set F for μ. So the limit function g above is only defined in $\Omega\setminus F$. But when we regard μ-measurable functions one can ignore null sets for μ and the following hold:

0.8.3 Proposition. *If g is an almost everywhere limit of a sequence $\{f_\nu\}$ in $\mathcal{M}_\mu$ then $g\in\mathcal{M}_\mu$.*

Proof. For a real number a we put

$$\{g\leq a\}=\{w\in\Omega\setminus F:\ g(w)\leq a\}$$

and when N is a positive integer

$$A_N(a)=\{w\in\Omega\setminus F:\ \max_{\nu\geq N} f_\nu(w)\leq a\}$$

It is clear that

$$A_N(\epsilon)\subset\{g\leq a\}\quad:\quad N=1,2,\ldots \tag{i}$$

Conversely, if δ is a positive number the pointwise convergence entails that

$$\{g\leq a\}\subset\bigcup_{N\geq 1} A_N(a+\delta) \tag{ii}$$

Notice that the sets $A_N(a)$ increase with N and let $A^*(a)$ be the union. Then (i-ii) give the following inclusions hold for every positive integer m:

$$A^*(a)\subset\{g\leq a\}\subset A^*_N(a+1/m)$$

This implies that

$$\{g\leq a\}=\bigcap_{m\geq 1} A^*_N(a+1/m)$$

and the reader may verify that the right hand side is a μ-measurable set. Hence $\{g\leq a\}\in\mathcal{M}_\mu$. In a similar fashion one shows that $\{g\geq b\}\in\mathcal{M}_\mu$ hold for every real number and the reader can conclude that g is μ-measurable.

Examples show that convergence in measure need not imply convergence almost everywhere. But one has the following:

0.8.4 Proposition. *Let $\{f_n\}$ be a sequence of measurable functions which converges in μ-measure to a μ-measurable function g. Then there exists a subsequence*

$$\{g_k = f_{\nu_k} \,:\, \nu_1 < \nu_2 < \ldots\} \quad \& \quad g_k \overset{a.e}{\to} g$$

Proof. Regarding the sequence $\{f_n - g\}$ it suffices to prove the result when $g = 0$. We first find an integer ν_1 such that

$$\nu \geq \nu_1 \implies \mu(\{|f_\nu \geq 2^{-1}\} \leq 2^{-1}$$

Next, we find $\nu_2 > \nu_1$ such that

$$\nu \geq \nu_2 \implies \mu(\{|f_\nu \geq 2^{-2}\} \leq 2^{-2}$$

Inductively we get a sequence $\nu_1 < \nu_2 \ldots$ where

$$\nu \geq \nu_k \implies \mu(\{|f_\nu \geq 2^{-k}\} \leq 2^{-k}$$

Then, the sequence $\{g_k = f_{\nu_k}\}$ converges almost everywhere to the zero function. To see this we set:

$$g_k^* = \max_{\nu \geq k} |g_\nu|$$

The inductive construction above gives

$$\text{(i)} \qquad \mu(\{g_k^* \geq 2^{-k}\} \leq \sum_{\nu \geq k} \mu(\{|g_\nu| \geq 2^{-k}\} \leq \sum_{\nu \geq k} 2^{-\nu} = 2^{-k+1}$$

Here $g_1^* \geq g_2^* \geq \ldots$ is a monotone sequence and (i) shows that

$$g_k^* \overset{a.e}{\to} 0$$

From this it is clear that $\{g_k\}$ also converges almost everywhere to zero.

0.8.5 Exercise Let $\{f_n\}$ be a sequence in $\mathcal{M}_\mu$ such that

$$\lim_{n,m} \mu(\{|f_n - f_m| \geq \epsilon\}) = 0$$

hold for each $\epsilon > 0$, where (m, n) tend to ∞, i.e. $\{f_n\}$ is a Cauchy sequence with respect to the metric defined in (0.7.1). Prove by a similar method as in Proposition 0.8.4 that there exists a subsequence $\{g_k = f_{n_k}\}$ which converges almost everywhere to a μ-measurable limit function f_*. Show also that f_* is unique in $\mathcal{M}_\mu$ since the almost everywhere convergence of the subsequence entails that $f_n \overset{\mu}{\to} f_*$. The conclusion is that $\mathcal{M}_\mu$ is a *complete metric space* under (0.7.1).

1. Integrals

1.1 Elementary functions. As in § 0 we are given $(\Omega, \mathcal{B})$. Every set $B \in \mathcal{B}$ gives the characteristic function χ_B which is 1 on B and 0 in $\Omega \setminus B$. An elementary function is a finite linear sum

$$(*) \qquad F = \sum a_k \cdot \chi_{B_k}$$

where $B_1, \ldots, B_N$ is a finite set of disjoint sets in $\mathcal{B}$ whose union is Ω and and $a_1, \ldots, a_N$ are real numbers. Let $E(\mathcal{B})$ denote the family of all elementary functions. This is a a linear space of functions on Ω. For let F be given by (*) and

$$(*) \qquad G = \sum b_m \cdot \chi_{C_m}$$

is another elementary function.. Now the double indexed sets $\{B_k \cap C_m\}$ is a disjount partion of Ω and gives the elementary function

$$\sum (a_k + b_m) \cdot \chi_{B_k \cap C_m}$$

which obviously is equal to $F+G$. In a similar fashion the product $F \cdot G$ is the elementary function defined by

$$\sum a_k \cdot b_m \cdot \chi_{B_k \cap C_m}$$

Next, let F be a bounded $\mathcal{B}$-measurable function i.e. there exists a constant K such that

$$-K \leq F(\omega) \leq K$$

hold for all $\omega \in \Omega$. Then F can be *uniformly approximated* by elementary functions. Namely, if $N \geq 1$ we set

$$B_N(\nu) = \{-K + K \cdot \nu \cdot 2^{-N} \leq F < -K + K \cdot (\nu+1) \cdot 2^{-N}\} \quad : 0 \leq \nu \leq 2 \cdot (2^N - 2)$$

$$\text{and} \quad B_N(2 \cdot 2^N - 1) = \{K - 2^{-N} \leq F \leq K\}$$

We get the elementary function

$$F_* = \sum_{\nu=0}^{2 \cdot 2^N - 1} (-K + \nu \cdot 2^{-2N}) \cdot \chi_{B_N(\nu)}$$

It is clear that

$$F_* \leq F \leq F_* + 2^{-N}$$

Since N can be arbitrary large this shows that F is uniformly approximated by elementary functions.

1.2 Integrals of elementary functions. Let μ be a probability measure. To each elementary function

$$F = \sum a_k \cdot \chi_{B_k}$$

we assign the μ-integral

$$(1.2.1) \qquad \int F \cdot d\mu = \sum a_k \cdot \mu(B_k)$$

It is clear that (1.2.1) yields an additive map from $E(\mathcal{B})$ into the complex number where we regard complex-valued functions. The maximum-norm $||F||_\infty$ is the maximum of $\{|a_k|\}$. Since μ is a probability measure the triangle inequality gives

$$(1.2.2) \qquad |\int F \cdot d\mu\,| \leq ||F||_\infty$$

1.3 Integrals of bounded functions. If $K > 0$ is a positive number we denote by $\mathcal{M}_\mu(K)$ the class of μ-measurable functions f such that the absolute value $|f|$ is $\leq K$ almost everywhere. So in the equivalence class we can take f to be a function in $\mathcal{M}_\mathcal{B}$ where this everywhere defined function has maximum norm K at most. By (1.1) f can be uniformly approximated by a sequence $\{F_n\}$ from $E(\mathcal{B})$ with maximum norms $\leq K$. For each pair n, m the triangle inequality gives

$$(1.3.1) \qquad |\int F_n \cdot d\mu - \int F_m \cdot d\mu| \leq \int |F_n - F_m| \cdot d\mu \leq ||F_n - F_m||_\infty$$

Since $||F_n - f||_\infty \to 0$ it follows that (1.3.1) tends to zero as n and m increase. Hence the evaluated integrals $\{\int F_n \cdot d\mu\}$ is a Cauchy sequence of complex numbers and there exists a limit

$$(1.3.2) \qquad \lim_{n\to\infty} \int F_n \cdot d\mu$$

Moreover, (1.2.1) implies that this limit is intrinsic, i.e. independent of the chosen sequence $\{F_n\}$ which approximates f uniformly. The limit (2) is called the μ-integral of f and is denoted by

$$(1.3.3) \qquad \int f \cdot d\mu$$

1.4 Absolutely integrable functions. For each positive real number K we define the the truncation operator T_K from $\mathcal{M}_\mu$ to $\mathcal{M}_\mu(K)$ by:

$$T_K(f)(x) = f(x) : \ |f(x)| \leq K \quad \text{and} \quad T_K(f)(x) = 0 : \ |f(x)| > K$$

When $f \in \mathcal{M}_\mu$ we get bounded functions $\{T_K(f) : \ K > 0\}$. We can also take absolute values and notice that

$$(1) \qquad K_1 < K_2 \implies |T_{K_1}(f)| \leq |T_{K_2}(f)|$$

From (1) the μ-integrals of the non-negative functions $\{|T_K(f)|\}$ increase with K, i.e.

$$(2) \qquad K_1 < K_2 \implies \int |T_{K_1}(f)| \cdot d\mu \leq \int |T_{K_2}(f)| \cdot d\mu$$

We can impose the condition that the non-decreasing sequence (2) is bounded. This leads to

1.5 Definition. *A function $f \in \mathcal{M}_\mu$ is said to be absolutely integrable if there exists a constant C such that*

$$(*) \qquad \int |T_{K_1}(f)| \cdot d\mu \leq C \quad \text{for all} \quad K > 0$$

The class if these functions is denoted $L^1(\mu)$.

1.6 Integrals of $L^1(\mu)$-functions. Let f belong to $L^1(\mu)$. From the construction of the T-operators it is clear that if $K_2 > K_1 > 0$ then

$$(i) \qquad |\int T_{K_2}(f) \cdot d\mu - \int T_{K_1}(f) \cdot d\mu| \leq \int |T_{K_2}(f)| \cdot d\mu - \int |T_{K_1}(f)| \cdot d\mu$$

Since $f \in L^1(\mu)$ the non-decreasing sequence $\{\int |T_{K_2}(f)| \cdot d\mu\}$ is bounded above and therefore convergent. Since a convergent sequence also is a Cauchy sequence, it follows from (i) that

$$\lim_{K_2, K_1} \int T_{K_2}(f) \cdot d\mu - \int T_{K_1}(f) \cdot d\mu = 0$$

as K_1 and K_2 tend to $+\infty$. Since Cauchy sequences of real numbers converge there exists a limit

$$(*) \qquad \lim_{K\to\infty} \int T_K(f) \cdot d\mu$$

This limit is denoted by $\int f \cdot d\mu$ and is called the μ-integral of f.

1.7 Exercise. Show that the integral on $L^1(d\mu)$ is additive, i.e.

$$\int f d\mu + \int g d\mu = \int (f+g) d\mu \quad : \quad f, g \in L^1(d\mu)$$

1.8 The L^1-norm. The linear space $L^1(d\mu)$ becomes a normed space when we set

$$||f||_1 = \int |f[d\mu \quad : \quad f \in L^1(d\mu$$

1.9 Exercise. Show that $E(\mathcal{B})$ is a dense subspace of $L^1(\mu)$.

1.10 Convergence in the L^1-norm.. By definition a sequence $\{f_n\}$ converges in the L^1-norm to a limit function g in $L^1(\mu)$ when

(i) $$\lim_{n\to\infty} ||f_n - g||_1 = 0$$

If $\epsilon > 0$ we notice that

(ii) $$\mu(\{|f_n - g| \geq \epsilon\} \leq \epsilon^{-1} \cdot ||f_n - g||_1$$

Hence (i) implies that

(3) $$\lim_{n\to\infty} \mu(\{|f_n - g| \geq \epsilon\} = 0$$

Here ϵ is arbitrary small which means that $f_n \xrightarrow{\mu} g$. Thus L^1-convergence implies convergence in μ-measure.

1.11 Warning. The converse is not true. For example, let μ be the Lebesgue measure on $[0, 1]$. To each $n \geq 1$ we define $f_n(x)$ to be n if $0 \leq x \leq \frac{1}{n}$ and otherwise zero. Then

$$\int_{=}^{1} f_n(x) \cdot dx = 1$$

hold for all n. At the same time $f_n(x) \to 0$ for every $x > 0$, i.e. the sequence converges almost every where to zero and hence also in measure. But if we restrict the attention to bounded functions a converse result holds.

1.12 Proposition *Let $K > 0$ be fixed. Then a sequence $\{f_\nu\}$ in $\mathcal{M}_\mu(K)$ converges to a limit function g in the L^1-norm if and only if the sequence converges in measure to g.*

Proof. Consider a pair f, g in $\mathcal{M}_\mu(K)$. If $\epsilon > 0$ we get the measurable set

$$E_\epsilon(f, g) = \{|f - g| \geq \epsilon\}$$

Since the maximum norm $||f - g||_\infty \leq 2K$ it follows that

(i) $$||f - g||_1 = \int_{E_\epsilon(f,g)} |f - g| d\mu + \int_{\Omega\setminus E_\epsilon(f,g)} |f - g| d\mu \leq 2K \cdot \mu(E_\epsilon(f, g)) + \epsilon$$

If $f_\nu \xrightarrow{\mu} g$ where $\{f_\nu\}$ in $\mathcal{M}_\mu(K)$ one has

(ii) $$\lim_{\nu\to\infty} \mu(E_\epsilon(f_\nu, g)) = 0 \quad \text{hold for each} \quad \epsilon > 0$$

Hence (i) implies that $||f_n - g||_1||_1 \to 0$. Since we already proved that L^1-convergence entails convergence in measure we get the equivalence assertion in Proposition 1.12 follows.

1.13 Dominated convergence theorem

Let $\{f_n\}$ be a sequence in $L^1(\mu)$ which converges in μ-measure to a limit function g. The example in 1.11 shows that convergence need not hold in the L^1-norm. To compensate for this we impose a certain bound so that the situation is essentially the same as in Proposition 1.12

1.14 Theorem *Let $\{f_n\}$ be a sequence in $L^1(\mu)$ where $f_n \xrightarrow{\mu} g$ holds. Assume in addition that there exists a non-negative $\phi \in L^1(\mu)$ such that*

(1) $$|f_n| \leq \phi$$

hold almost everywhere for each n. Then the limit function g belongs to $L^1(\mu)$ and

(2) $$\lim_{\nu\to\infty} ||f_n - g||_1 = 0$$

1.15 Exercise Prove Theorem 1.14. The hint is to apply truncation operators to ϕ so that the L^1-norm of $\phi - T_K(\phi)$ are small and then use Proposition 1.12.

2. A completeness theorem.

Let $\{f_n\}$ be Cauchy sequence in the L^1-norm, i.e.

$$(*) \qquad \lim_{n,m\to\infty} \int |f_n - f_m| \cdot d\mu = 0$$

2.1 Theorem *When (*) holds there exists a unique L^1-function f_* such that*

$$\lim_{n,m\to\infty} \int |f_n - f_*| \cdot d\mu = 0$$

Hence the normed space $L^1(\mu)$ is complete, i.e. it is a Banach space.

Proof. From (*) it follows that there exists a subsequence $\{n_k\}$ such that if $g_k = f_{n_k}$ then

$$\text{(i)} \qquad ||g_{k+1} - g_k||_1 \leq 2^{-k}$$

Since $\sum 2^{-k} < \infty$ the non-negative function

$$G = |g_1| + \sum_{k=1}^{\infty} |g_{k+1} - g_k|$$

is integrable. It is clear that

$$\text{(ii)} \qquad |g_k| \leq G \quad : k = 1, 2, \ldots$$

Next, (i) entails that $\{g_k\}$ is a Cauchy sequence in $L^1(\mu)$ and hence it is also a Cauchy sequence in the complete metric space $\mathcal{M}_\mu$. It follows that there exists a μ-measurable function g where

$$g_k \xrightarrow{\mu} g$$

Next, (ii) enable us to apply Theorem 1.14 which gives

$$||g_k - g||_1 \to 0$$

Above $\{g_k\}$ is a subsequence of $\{f_n\}$ and the triangle inequality shows that $||f_n - g||_1 \to 0$ which shows that g is the requested limit in $L^1(\mu)$ of the given Cauchy-sequence $\{f_n\}$.

3. Signed measures and the Radon-Nikodym theorem.

Let $(\Omega, \mathcal{B})$ be a sample space. Consider an additive real-valued map $\mu\colon \mathcal{B} \to \mathbf{R}$ which may take negative values and for which there exists a constant C such that

(*) $$-C \leq \mu(A) \leq C \quad : \quad A \in \mathcal{B}$$

This uniform bound and additivity imply that if $\{A_\nu\}$ is a sequence of disjoint sets in $\mathcal{B}$ then

$$\sum |\mu(A_\nu)| < 2C$$

Hence the series $\sum \mu(A_\nu)$ is absolutely convergent. We say that μ is σ-additive if

$$\sum \mu(A_\nu) = \mu(\cup A_\nu)$$

hold for every sequence of disjoint sets in $\mathcal{B}$ and refer to μ as a *signed* measure. If $\mu(A) \geq 0$ for all $A \in \mathcal{B}$ we say that μ is a positive measure.

3.1 Definition *Two positive measures μ and ν are mutually singular if there exist null sets $A \in \mathcal{N}_\mu$ and $B \in \mathcal{N}_\nu$ such that*

$$\mu(B) = \mu(\Omega) \quad : \quad \nu(A) = \nu(\Omega)$$

When this holds we write $\mu \perp \nu$.

3.2 Hahn's Theorem *Every signed measure μ has a unique decomposition*

$$\mu = \mu_+ - \mu_- \quad \text{where} \quad \mu_, \mu_- \quad \text{are both positive and} \quad \mu_+ \perp \mu_-$$

Proof. A set $A \in \mathcal{B}$ is called a μ-positive set if

(i) $$E \subset A \quad \& \quad E \in \mathcal{B} \implies \mu(E) \geq 0$$

Denote this class by $P_+(\mu)$. Obviously the union of two μ-positive sets is again μ-positive and by σ-additivity there exists a $A^* \in P_+(\mu)$ such that

(ii) $$\mu(A^*) = \max_{A \in P_+(\mu)} \mu(A)$$

Sublemma. One has $\mu(B) \leq 0$ for each $B \subset \Omega \setminus A^*$.

Proof of Sublemma. We argue by contradiction. Suppose there is some $B_0 \subset \Omega \setminus A^*$ with $\mu(B_0) = \delta > 0$. The maximality of $\mu(A^*)$ implies that B_0 does not belong to $P_+(\mu)$ which gives some $\delta_1 > 0$ such that

(i) $$-\delta_1 = \min_{E \subset B_0} \mu(E)$$

Choose $E_1 \subset B_0$ with $\mu(E_1) \leq -\delta_1/2$ and set $B_1 = B \setminus E_1$. Now $\mu(B_1) \geq \delta + \delta_1/2$ and exactly as above we get a negative number

$$-\delta_2 = \min_{E \subset B_1} \mu(E)$$

Choose $E_2 \subset B_1$ with $\mu(E_2) \leq -\delta_2/2$. Inductively we get a decreasing sequence of sets

$$B_\nu = B_0 \setminus (E_1 \cup \ldots \cup E_\nu)$$

where $\{E_\nu\}$ are disjoint. Moreover, we have a sequence $\{\delta_\nu\}$ of positive numbers where

(ii) $$-\delta_\nu = \min_{E \subset B_\nu} \mu(E) \quad \text{and} \quad \mu(E_{\nu+1}) \leq -\delta_\nu/2$$

Since μ is a signed measure there is a constant A such that

(iii) $$\mu(E_1) + \ldots + \mu(E_N) \geq -A \quad \text{for all} \quad N \geq 1$$

It follows that

(iv) $$\delta_1 + \ldots + \delta_N \leq 2A \quad \text{for all} \quad N \geq 1$$

Hence the positive series $\sum \delta_\nu$ is convergent which gives $\delta_\nu \to 0$ as $\nu \to +\infty$. Put

(v) $$B_* = \cap B_\nu$$

For each $\nu \geq 1$ the inclusion $B_* \subset B_N$ and the definition of δ_ν from (ii) give:

(vi) $$\min_{E \subset B_*} \mu(E) \geq -\delta_\nu$$

This hold for every ν and since $\delta_\nu \to 0$ the minimum in (vi) is ≥ 0 which means that B_* belongs to $P_+(\mu)$. At the same time the construction above gives

(vii) $$\mu(B_\nu \geq \delta_0 + \frac{1}{2}(\delta_1 + \ldots + \delta_\nu) \geq \delta_0$$

for every ν. By σ-additivity we have

$$\mu(B_*) = \lim_{\nu \to \infty} \mu(B_\nu)$$

and then (vii) entails that $\mu(B_*) > 0$. This contradicts the (1) since A^* and B_* are disjoint and the Sublemma is proved.

Final part of the proof . The Sublemma gives positive measures μ_+ and μ_- defined by

$$\mu_+(E) = \mu(E \cap A^*) \quad : \quad \mu_-(E) = -\mu(E \cap (\Omega \setminus A^*)$$

We see that $\mu_+ \perp \mu_-$ and $\mu = \mu_+ - \mu_-$. This proves the existence of at least one Hahn-decompostion. The proof of *uniqueness* of such a decomposition is left to the reader.

3.3 Radon-Nikodym derivatives.

Let μ be a positive measure and consider a non-negative function $f \in L^1(d\mu)$. This gives a positive measure defined by the σ-additive map

$$E \mapsto \int_E f \cdot d\mu \quad : E \in \mathcal{B}$$

Denote this positive measure by $f \cdot \mu$. If $E \in \mathcal{N}_\mu$ the construction of μ-integrals implies that $\int_E f \cdot d\mu = 0$. Hence one has the inclusion

(3.1) $$\mathcal{N}_\mu \subset \mathcal{N}_{f \cdot \mu}$$

In general, a positive measure ν is called *absolutely continuous* with respect to μ if

$$\mathcal{N}_\mu \subset \mathcal{N}_\nu$$

It turns out that such positive measures are of the form $f \cdot \mu$ with $f \in L^1(\mu)$.

3.4 Theorem. *Let μ be a positive measure. Then every positive measure ν which is absolutely continuous with respect to μ is of the form $f \cdot \mu$ for a unique non-negative function $f \in L^1(d\mu)$.*

Proof. For each pair (k, N) where N is a positive integer and k a non-negative integer we consider the following two signed measures

(i) $$\nu - k2^{-N} \cdot \mu \quad \text{and} \quad (k+1)2^{-N} \cdot \mu - \nu$$

The Hahn decomposition applied to $\nu - k2^{-N} \cdot \mu$ gives a maximal set

(ii) $$S_N(k) \in P_+(\nu - k2^{-N} \cdot \mu)$$

and similarly we find a maximal set

(iii) $$T_N(k) \in P_+((k+1)2^{-N} \cdot \mu - \nu)$$

If N is fixed the measure $(\nu - k2^{-N}) - (\nu - (k+1)2^{-N}) = 2^{-N} \cdot \mu \geq 0$. This implies that

(iv) $$S_N(k+1) \subset S_N(k) \quad \text{for all} \quad k = 0, 1, \ldots$$

Moreover, since ν is absolutely continuous with respect to μ it is clear that:

(v) $$\cap_{k\geq 1} S_N(k) = \emptyset$$

Finally the reader may observe that

(vi) $$S_N(k) \setminus S_N(k+1) = S_N(k) \cap T_N(k)$$

Let us put

(vii) $$W_N(k) = S_N(k) \cap T_N(k)$$

From (vi) it follows that $\{W_N(k)\}$ is a family of disjoint subsets of Ω and we notice that

(viii) $$k2^{-N}\mu(E) \leq \nu(E) \leq (k+1)2^{-N}\mu(E) \quad : E \subset W_N(k)$$

The reader may also verify the set-theoretic equality

(ix) $$W_N(k) = W_{N+1}(2k) \cup W_{N+1}(2k+1)$$

for all pairs k and N, Next we construct a sequence of functions by:

(x) $$f_N = \sum_{k=1}^{\infty} k2^{-N} \cdot \chi_{W_N(k)} \quad : N = 1, 2 \ldots$$

Using (ix) the reader may verify that $\{f_N\}$ increase with N and (viii) entails that

(xi) $$\int_E f_N \cdot d\mu \leq \nu(E) \leq \int_E f_N \cdot d\mu + 2^{-N} \cdot \mu(E) \quad : E \in \mathcal{B}$$

Now $\{f_N\}$ is a non-decreasing sequence and there exists a limit function

$$f_* = \lim_{N\to\infty} f_N$$

where the convergence holds almost everywhere and it is celar that (xi) implies that

$$\nu(E) = \lim_{N\to\infty} \int_E f_n \cdot d\mu = \int_E f_* \cdot d\mu \quad : E \in \mathcal{B}$$

This gives $\nu = f_* \cdot d\mu$ and Theorem 3.3, is proved.

3.4 A general decomposition.

Let μ be a positive measure. For any other positive measure ν there exists a unique decomposition of ν into a sum of one measure which is singular with respect to μ, while the other term is given by an $L^1(d\mu)$-function. More precisely one has:

3.5 Theorem *Given a positive measure μ every other positive measure ν is of the form*

$$\nu = \nu_s + f d\mu \quad : \quad \nu_s \perp \mu \quad \text{and} \quad f \in L^1(d\mu)$$

Proof To find the singular part ν_s we put

(1) $$M = \max \nu(A) \quad : \quad A \in \mathcal{N}_\mu$$

By σ-additivity there exists some $A_* \in \mathcal{N}_\mu$ such that $\nu(A_*) = M$. Define the measure ν_* by:

(2) $$\nu_*(E) = \nu(A_* \cap E)$$

Here $\nu_* \perp \mu$. Put $\gamma = \nu - \nu_*$. The construction of ν_* gives

(3) $$A \in \mathcal{N}_\mu \implies \gamma(A) = 0$$

Then Theorem 3.5 gives $\gamma = f \cdot \mu$ for some $f \in L^1(\mu)$ and Theorem 3.5 follows.

3.6 The Vitali-Hahn-Saks theorem.

Let $\Omega, \mathcal{B})$ be a probability space and μ a probability measure which gives the σ-algebra $\mathcal{N}(\mu)$ of null-sets. A pair of μ-measurable sets are identified when their Boolean difference is a null-set. If $A \in \mathcal{B}$ then A_* denotes its equivalence class. Let X denote the space whole points are such equivalence classes of subdsets from $\mathcal{B}$. A metric is defined on X by

$$d(A_*, B_*) = \mu(A \setminus B) + \mu(B \setminus A)$$

3.6.1 Exercise. Show that X is a complete metric space.

Next, an additive real-valued function λ defined on $\mathcal{B}$ is called μ-continuous if

$$\lim_{\mu(A) \to 0} \lambda(A) = 0$$

3.6.2 Theorem. *Let $\{\lambda_n\}$ be a sequece of additive μ-continuous functions such that*

(*)
$$\lim_{n \to \infty} \lambda_n(A) = \lambda_*(A)$$

exist for every $A \in \mathcal{B}$. Then, for each $\epsilon > 0$ there exists δ such that

(**)
$$\mu(A) \leq \delta \implies \max_n |\lambda_n(A)| \leq \epsilon$$

Proof. For each $\epsilon > 0$ and every pair of integers n, m we get the closed sets in X:

$$\Sigma_{n,m} = \{A \in \mathcal{B} : \ |\lambda_n(A) - \lambda_m(A)| \leq \epsilon/2\}$$

Set

$$\Sigma_p^* = \bigcap_{n,m \geq p} \Sigma_{n,m}$$

Then (*) entails that $\cup \Sigma_p^* = X$. Exercise 3.6.1 and Baire's category theorem gives some p and $\delta > 0$ such that

$$\mu(A) < \delta \implies A \in \Sigma_p^*$$

Shrinking δ if necessary

$$\mu(A) < \delta \implies \max_{n \leq p_*} \lambda_n(A) \leq \epsilon/2$$

Now the reader can check that (**) holds.

3.6.3 Exercise. Show that if $\lambda_*(A)$ is the limit value in (*) then λ_* is additive and μ-continuous.

Measure theory

Contents

Strul ...

The sections 0.1-0.7 give background for topics in the subsequent sections § 1-9. In § 0.4 we expose the construction by Weierstrass of a continuous function which fails to have a derivative at every point on the real line. More precisely if $0 < b < 1/5$ then Weierstrass proved that the convergent cosine series

$$f(x) = \sum_{k=1}^{\infty} b^k \cdot \cos($$

is a continuous function which fails to have a derivative at every point as soon as

$$k > .....$$

Introduction

Modern integration theory started in 1894 when Stieltjes defined integrals of the form

$$(*) \qquad \int_a^b g(x) \cdot df(x)$$

where g is a real-valued continuous function and f is continuous and *non-decreasing* on the compact interval $[a, b]$. Stieltjes defined (*) as the limit of sums:

$$(**) \qquad \sum g(x_\nu) \cdot \big[f(x_{\nu+1} - f(x_\nu)\big]$$

where $a = x_0 < \ldots < x_N = b$ and $\max\,(x_{\nu+1} - x_\nu)$ tends to zero. The *uniform continuity* of g on $[a, b]$ gives a limit exactly as for the ordinary Riemann integral. Namely, consider the modulus of continuity:

$$\omega_g(\delta) = \max_{x_1, x_2} |g(x_1) - g(x_2)| \quad : \text{ maximum taken over pairs } |x_2 - x_1| \leq \delta$$

and put $M = f(b) - f(a)$. Let $\{x_\nu\}$ and $\{y_j\}$ be two δ-partitions of $[a, b]$, i.e.

$$\max x - \nu + 1 - x_\nu \leq \delta \quad \& \quad \max y_{j+1} - y_j \leq \delta$$

hold. The the reader cab check that the Stieltjes' sums in (**) using the two partitions differ in absolute value by a quantity which is $\leq M \cdot \omega_g(\delta)$. Hence the the uniform continuity of g entails that the .limit in (**) exists as $\delta \to 0$.

In the special case when f is a C^1-function, i.e. the derivative $f(x)$ exists as a continuous function, the Stieltjes' integral (*) is equal to the ordinary Riemann integral

$$\int_a^b g(x) f'(x) \cdot dx$$

To see this one uses Rolle's mean-value formnula which shows that every sum in (**) can be expressed as

$$\sum g(x_\nu) \cdot f'(\theta_\nu) \cdot (x_{\nu+1} - (x_\nu)$$

where $x_\nu < \theta_\nu < x_{\nu+1}$ hold for every ν.

Soon after Stieltjes' construction another method was introduced by Emile Borel. His idea was to decompose the range of g. To begin with Borel used the elementary fact from topology that every open subset on the real line is the disjoint union of at most denumerable open intervals. If $\{(a_\nu, b_\nu)\}$ is a family of disjoint open subintervals of (a, b) the variation of f over their open union Ω is defined by

$$\mathrm{Var}_f(\Omega) = \sum f(b_\nu) - f(a_\nu)$$

Since f is non-decreasing the termes of the series are non-negative and the sum is majorised by $f(b) - f(a)$. More generally, if Ω is an open subset of an open interval (a, b) and F a closed subset of (a, b), Borel defined the variation of f over $\Omega \cap F$ to be

$$\mathrm{Var}_f(\Omega) - \mathrm{Var}_f((a, b) \setminus F)$$

Let us then consider a continuous function g on $[a, b]$. For every pair $\xi < \eta$ the continunity of g entails that the set

$$(a, b) \cap \{\xi \leq g < \eta\}$$

is the intersection of a closed and an open set on which the variation of f can be defined. Choose $M > 0$ so that the range of g is contained in the open interval $(-M, M)$. For every sequence

$$-M = \xi_0 < \xi_1 < \ldots < \xi_N = M$$

one defines

$$\text{(i)} \qquad \sum_{\nu=0}^{\nu=N_1} \xi_\nu \cdot \mathrm{Var}_f((a, b) \cap \{\xi_\nu \leq g < \xi_{\nu+1}\}$$

0.1 Exercise. Show by partial summations that (i) differs from (*) by a quantiity which does not exceed

$$(b - a) \cdot \max_\nu (\xi_{\nu+1} - \xi_\nu)$$

'The merit in Borel's construction of integrals that it enable us to define integrals where the g-function no longer need to be continuous. A first example is when g is a bounded function given as a *pointwise limit* of continuous functions. Thus, suppose that $\{g_n\}$ is a sequence of uniformly bounded continuous functions on some open interval (a, b) and that

$$\text{(i)} \qquad \lim_{n \to \infty} g_n(s) = g(s) \quad : \quad a < s < b$$

With f as before there exists the Borel integral

$$\text{(0.2)} \qquad \int_a^b g(x) \cdot df(x)$$

To find (0.2), Borel argued as follows. For every n we put

$$g_n^*(x) = \max_{\nu \geq n} g_\nu(x)$$

Since the g-functions are bounded it is clear that the sets

$$(a,b) \cap \{g_n^*(x) < \eta\}$$

are open for every real number η. Exactly as above the reader can check that that there exist f-variations:

$$\mathrm{Var}_f((a,b) \cap \{\xi \leq g_n^* < \eta\}$$

for all pairs $\xi < \eta$ and from this one contructs the Borel integrals

$$\int g_n^*(x) \cdot df(x) \quad : \; n = 1, 2, \ldots$$

Now $\{g_n^*\}$ is a non-increasing sequence of functions which converges pointwise to g. It follows first that the integrals (xx) form a non-increasing sequence of real numbers and since $\{g_n\}$ by assumption is uniformly bounded the sequence in (xx) converges to a limit which Borel defined as the iintegral of the pair g and f.

The Baire algebra $\mathcal{B}^\infty[a,b]$. By transfinite induction, Baire constructed a space of bounded real-valued functions on a given compact interval $[a,b]$ which is stable under poinitwise limits and contain the space of bounded continuous functions. This gives an algebra of bounded real-valued functions denoted by $\mathcal{B}^\infty[a,b]$ and called the Baire algebra. Borel proved that for every such function g the variations

$$\mathrm{Var}_f(\{\xi \leq g < \eta)$$

exist when f is a non-decreasing continuous function on $[a,b]$ and in this way one defines the Borel integrals by a limit as in Exercise 0.1 for every bounded Baire functions.

Remark. The set-theoretic constructions which lead to Baire functions are presented in great detail in Baire's original work. Borel gave later gave an abstract theory where measures and their integrals are defined on sample spaces $(\Omega, \mathcal{B})$ where $\mathcal{B}$ is a Boolean σ-algebra of subsets of a set Ω. This abstract theory is exposed in § 1. The crucial point in Borel's approach is that transfinite induction leads to the *Borel algebra* which by defitiion is the smallest Boolean σ-algebra if subsets on the real line which contains all open intervals. Moroever, for every continuous and non-decreasing function f and every bounded Borel set U one can define the variation of f over U.

Lebesgue theory.

By a null set on the real line one means a set E such that for every $\epsilon > 0$ there exists a family of open intervals (a_ν, b_ν) whose union contains E while

$$\sum xxx < \epsilon$$

In his thesis from 1902, Henri Lebesgue introduced a class $\mathcal{L}$ of sets as follows: If E is a bounded subset of $\mathbf{R}$ its outer Lebesgue meaure is defined by

$$|E|^* = \inf_{E \subset U} |U|$$

where U are bounded open sets. The inner measure is defined by

$$|E|_* = \max_{K \subset E} |K|$$

where K are compact subsets of E. When $|E|^* = |E|_*$ one says that E is measurable in the sense of Lebesgue and the common number is called the Lebesgue measure. Next, a bounded real-valued function f defined on some compact interval $[a,b]$ is called Lebesgue measurable if the sets

$$[a,b] \cap \{a \leq f < b\}$$

belong to $\mathcal{L}$ for every pair $a < b$. By decomposing the range of a bounded Lebesgue measurable function f defined in a compact interval $[a, b]$ one defines the Lebesgue integral

$$\int_a^b f(x)\,dx$$

0.2.9 Non-measurable sets. There exist non-measurable sets but in practice one does not encounter these ugly sets. In fact, one must appeal to the Axiom of Choice to exhibit examples. Using this axiom there exists a subset E of the open interval $(0, 1)$ with the property that the sets $\{E+q\}$ are disjoint when q runs over the set of rational numbers. This set cannot be measurable. The reason is that the disjointness would entail that if $q_1, \ldots, q_N$ is a set of rational numbers between 0 and 1 then

$$N \cdot |E|_* = \sum_{\nu=1}^{\nu=N} = |E + q_\nu|_* \leq 2$$

where the last equality holds since each set $E + q_\nu \subset [0, 2]$. Since N can be arbitrary large we have $|E|_* = 0$. So if E is measurable we have $|E| = |E|_* = 0$, i.e E is a null set. Since $E + q$ is translate of E they also give null sets for every rational number. But a denumerable union of null sets is a null set which gives a contradaction since $\cup\,(E + q)$ with the union taken over all rational numbers is the whole real line.

A major result in the Lebesgue theory is the *almost everywhere existence of Lebesgue points.* More precisely, in his thesis Lebegue proved that when f as above is bounded and Lebesgue measurable on some open interval (a, b), then

$$\lim_{h,k} \frac{1}{h+k} \cdot \int_{x-k}^{x+h} |f(s) - f(x)|\,ds = 0$$

hold for almost evety $a < x < b$, where the limit is taken while h and k tend to zero. Let $\mathfrak{Leb}(f)$ denote the set of x for which (xx) holds. So Lebesgue proved that $(a, b) \setminus \mathfrak{Leb}(f)$ is a null set. The proof of this result is given in § xx and there we also present Lebesgue's extension to higher dimenbsion where one regards the n-dimensional lebegue measure in $\mathbf{R}^n$.

Existence of derivatives. Another major result due to Lebesgue is the almost everywhere existence of derivatives when f is a non-decreasing function defined on some interval (a, b) on the real line. To begin with such a function f may contain jumps. Namely, for evey $a < x < b$ we put

$$f_*(x) = \lim_{\xi < x} f(\xi) \quad \& \quad f^*(x) = \lim_{\eta > x} f(\eta)$$

where the limits are taken as $\xi \to x$ and $\eta \to x$. Since f is non-decreasing we have

$$f_*(x) \leq f(x) \leq f^*(x)$$

for every x. Moreover, the reader can check that f is continuous at x if and only if $f_*(x) = f^*(x)$. If the strict inequality $f^*(x) > f_*(x)$ occurs, the function f has a jump discontinuity at x. By assumption f is bounded, i.e. $f(b) - f(a)$ is finite. From this the reader can check that jumps can only occur at a denumerable set of points $\{x_\nu\}$. For every such jump we define the function $s_\nu(x)$ which is zero if $x < x_\nu$ and equal to $f^*(x_\nu) - f_*(x_\nu)$ if $x \geq x_\nu$. Then the function

$$s_f(x) = \sum s_\nu(x)$$

is non-decreasing and the reader can check that

$$f - s_f(x) = f_0(x)$$

is a non-negative and non-decreasing continuous function. one refers to $s_f(x)$ as the jump-function assocaiged to f. In § xx we prove that there exists a null set E such that both s_f and f_0 have ordinary derivatives at every $x \in (a, b) \setminus E$, and here the derivative of s_f is zero.

Remark. Alreadt the case for jump-functions is quite remarkable. More precisely, let $\{x_\nu\}$ be an arbitrary denumerable subset of (a,b) and $\{\rho_\nu\}$ a sequence of positive numbers where $\sum \rho_\nu < \infty$. Now we get the non-decreasing function

$$S(x) = \sum \rho_\nu \cdot s_\nu(x)$$

The fact that S has an ordinary derivative which is zero outside a null set is not evident. In fact, this fact constitutes one of the high points in Lebesgue theory. The proof is given in § xx.

Cantor sets and their Lebesgue functions. Starting for set-theoretic consgructions by Cantor, one gets a quite extensive family of non-decreasing contionuous functions which have ordinary derivatives equal to zero outside a null-set, and yet they are not reduced to contants.

Lebegue's theorem is proved in § xx. An extension of this result applies to arbitrary continous and real-valued functions f defined on open intervals (a,b). More precisely one has the *Denjoy-Young Theorem.* Since this result illustrates the philosopuhy in measure theory we announce it already in this introduction. Let $f(x)$ be a real-valued continuous function on an open interval (a,b). For each $a < x < b$ we set

$$D^*(x) = \limsup \frac{f(x+h) - f(x-k)}{h+k} \quad \& \quad D_*(x) = \liminf \frac{f(x+h) - f(x-k)}{h+k}$$

where h and k are positive when we pass to the limits above. It is clear that f has a derivative in the usual sense at a point x if and only if $D_*(x) = D^*(x)$. The Denjoy-Young result is:

Theorem. *Outside a (possibly empty) null-set E of (a,b) the following two possibilities occur for each $x \in (a,b) \setminus E$: Either there exists a common finite limit $D^*(x) = D_*(x)$ or else one has*

$$D^*(x) = +\infty \quad \text{and} \quad D_*(x) = -\infty$$

The material devoted to the 1-dimensional case in Chapter 1 is quite demanding and often more involved compared to the general theory which is elegant but the proofs are often quite straightforward and rely only upon "clever set-theoretic constructions". But the role of general measure theory cannot be underestimated. For example, in functional analysis measures are often used to construct dual spaces and also perform integrals where the integrand in general can be an operator-valued function. This is for example used for such crucial resuilts as Hilbert's spectral theorem for self-adjoint operators on infinite-dimensional Hilbert spaces. A more recent use of measure theory is due to Laurent Schwartz who introduced distributions which has become a fundamental tool in PDE-theory. Every distribution in an euclidian space is for example locally a finite sum of distribution derivstives of Riesz measures.

Abstract measure theory

Here one studies σ-additive measures defined on an abstract Bolean σ-algebra of sets. This elegant and powerful theory is due to Emile Borel and was initiated in his thesis from 1895 and completed in later work. The material in § 1 is based upon the article [Borel] from 1908. One starts from a sample space Ω and a Boolean σ-algebra $\mathcal{B}$ of subsets. The pair $(\Omega, \mathcal{B})$ gives a class of σ-additive measures and for each specific measure μ there exists the family $\mathcal{N}_\mu$ of its null sets. One proceeds to study μ-measurable functions which are identified when they agree outside a null set. This leads to the construction of integrals for every μ-measurable function. Borel defined integrals of a μ-measurable function f, using a a partition of its values. Namely, one first regards μ-measures of inverse images of sets $\{a \leq f < b\}$ for all pairs of real numbers $a < b$, and after a limit where $\max(b-a) \to 0$ one gets the Borel integral denoted by

$$\int f \cdot d\mu$$

Consider as an example the case when the σ-additive measure μ is non-negative and has total mass one, i.e. $\mu(\Omega) = 1$. Let f be a μ-measurable function whose range is confined to an interval

$[0, A]$ with > 0, i.e. f is non-negative and is $\leq A$ on Ω. Now there exists a non-decreasing function on the real s-line defined by

$$\mu_f(s) = \mu\{f \leq s\}$$

It is zero if $s < 0$ and $\mu_f(A) = 1$. The σ-additivity of μ entails that the function is right continuous, i.e.

$$\lim_{\epsilon \to 0} \mu_f(s + \epsilon) = \mu_f(s)$$

hold for every s, where $\epsilon > 0$ during the passage to the limit. In § x we show the equality

$$(*) \qquad \int_\Omega f \cdot d\mu = \int_0^A (A - \mu_f(s))\, ds$$

The integral in the right hand side can be computed in a similar fashion as the usual Riemann integral which we explain later on. The merit in Borel's construction of integrals it that they enjoy "robust properties" which can be seen already in the case of integrals of continuous functions on on the real x-line, taken with respect to the ordinary measure dx. Via Borel's ingenious idea to perform partitions of the range of a continuous function, the approximations of $\int f(x)\, dx$ appear with a rate which does not dened upon the modulus of continuity of f. It is this improved rate of approximation which makes it possible to construct integrals of functions in a far more extensive class compared to those which appear in Riemann integrals.

Riesz' representation theorem. An important contributions was given by F. Riesz around 1910. He found a representation of the dual space of $C^0(S)$ when S is a compact Hausdorff space. To begin with Riesz introduced the vector space $\mathcal{E}(S)$ of real-valued bounded functions on S which are **R**-linear combinations of characteristic functions taken over sets of the form $U \cap F$ where $U \subset S$ are open and $F \subset S$ are closed. Under the maximum norm this gives a normed vector space and let $\mathcal{R}(S)$ be its completion, i.e. vectors in $\mathcal{R}(S)$ are bounded real-valued functions on S which appear as a uniform limit of a sequence in $\mathcal{E}(S)$. It is easily verified that $C^0(S)$ appears as a closed subspace of $\mathcal{R}(S)$. The Hahn-Banach theorem yields a surjective map

$$\text{(i)} \qquad \mathcal{R}(S)^* \to C^0(S)^*$$

i.e. every continuous linear functional on $C^0(S)$ has a norm-preserving extension to $\mathcal{R}(S)$. Next, let $\mathfrak{B}(S)$ be the Boolean algebra of subsets of S generated by open subsets. If L belongs to the dual space $\mathcal{E}(S)^*$ we get an additive measure μ defined on $\mathfrak{B}(S)$ by

$$\mu(\Omega) = L(\chi_\Omega)$$

where χ_Ω is the characteristic function of a set $\Omega \in \mathfrak{B}(S)$. Exactly as in Borel's constuctions there exist integrals

$$\int_S f \cdot d\mu \quad : f \in C^0(S)$$

which arise after a decomposition of the range of f. The surjective map in (i) entails that every $\lambda \in C^0(S)^*$ is obtained via the integrals in the right hand side above, where the additive measure μ on $\mathfrak{B}(S)$ is found after a norm-preserving extension of λ. But this norm-presering extension is not unique because $C^0(S)$ is a proper subspace of $\mathcal{E}(S)$. Riesz proved that a *unique norm-preserving extension exists* via an additive measure μ satisfying a certain regularity condition which goes as follows: First, for every set $\Omega \in \mathfrak{B}(S)$ the total variation of μ over Ω is defined by

$$|\mu|(\Omega) = \max_{A \subset \Omega} \mu(A) + |\mu(\Omega \setminus A)|$$

with the maximum taken over A-sets in $\mathfrak{B}(S)$. With this notation μ is regular in the sense of Riesz if

$$(*) \qquad \max_{K \subset U} |\mu|(U \setminus K) = 0$$

hold for every open set U, where the maximum is taken over compact subsets of U. In 1§ xx we prove that for every additive measure μ as above there exists a unique regular measure μ_* such that

$$\int_S f \cdot d\mu = \int_S f \cdot d\mu_* \quad : f \in C^0(S)$$

From this it follows that $C^=(S)^*$ can be identified with the space of additive measures on $\mathfrak{B}(S)$ satisfying (*). In the special case when S is a compact subset of $\mathbf{R}^n$ for some $n \geq 1$ this leads to a fairly concrete description of $C^0(S)^*$ which is exposed in § xx.

Lebesgue theory. In § 3 we expose results due to Henri Lebesgue whose pioneering work from 1902 has become a veritable cornerstone in analysis. Here the notion of *Lebesgue points* plays a central role. A major discovery by Lebesgue is the *almost everywhere existence of Lebesgue points.* We remark that the gemoetry in euclidian spaces is needed to arrive at this result, i.e. Lebesgue's theorem cannot be incorporated in the abstract measure theory. The notion of *null-sets* in Lebesgue's sense is explained in § 3. Here we recall that a subset A of $\mathbf{R}^n$ is a null-set if there to every $\epsilon > 0$ is possible to cover A by a family of open balls $\{B_\nu\}$ such that

$$\sum \mathrm{vol}_(B_\nu) < \epsilon$$

where we refer to the ordinary n-dimensional volume of every open ball. Lebesgue proved that if E is a compact set, then $E \setminus \mathcal{L}(E)$ is a null-set. Here a point $p \in E$ is a Lebesgue point if

$$\lim_{\epsilon \to 0} \frac{B_p(\epsilon) \setminus E}{\epsilon^n} = 0$$

where $B_p(\epsilon)$ denote open balls of radius ϵ centered at p. The reader may notice that (*) means that the compact et E is "fat" form a measure theoretic point of view" around p. The fact that $\mathcal{L}(E) \setminus E$ is a null set even in the case when E has no interior points is quite remarkable.

Hausdorff dimensions. Inspired by Lebegue's work, Hausdorff introduced a more extensive procedure to measure the size of sets which in general are null sets. Let $\alpha > 0$ be a positive real number. For each bounded set E in $\mathbf{R}^n$ we perform the following condtructions: First, if $\delta > 0$ we can cover E by a finite family of open balls $\{B_k\}$ where every ball has radius $\leq \delta$ and set

$$H_{\alpha,\delta}(E) = \min \sum \mathrm{rad}(B_k)^\alpha$$

with the minimum taken over families of balls with radius $\leq \delta$. These H-numbers increase with δ and we put

$$H_\alpha(E) = \lim_{\delta \to 0} H_{\alpha,\delta}(E)$$

This number can be zero, finite or infinite. With $\alpha 0 n$ one finds that $H_n(E) = 0$ ifn and only if E is a null set in Lebegue's sense. When $0 < \alpha < n$ these Hausdorff dimensions are used to investigate "thin sets" which are not covered by ordinary Lebesgue theory. In potentail theory other notions of thin sets appear. For example, we regard compact subsets E of the closed unit interval and study integrals

$$\int_E \log \frac{1}{|x-t|}\, dt$$

which can be defined for every $0 \leq x \leq 1$.

Other topics. In § 5 we prove Stokes Theorem for domains in $\mathbf{R}^n$ where irregular points on the boundary can appear. In § 6 and 7 we study the Hardy-Littlewood maximal function and the Rademacher functions, and § 8 is devoted to the construction of the Brownian motion where the main result is the almost everywhere continuity of individual Brownian paths.

Historic comments.

A problem which played an important role to develop general measure theory appeared in the article *Sur la distribution de l'ectrecité à la surface des conducteurs fermés et des conducteurs ouverts* from 1886 by G. Robin. He asked if every compact set E in $\mathbf{C}$ without interior points carries a distribution of mass whose logarithmic potential function is constant on E. Robin's problem led to the notion of sets of absolute harmonic measure zero and other thin sets. Modern integration theory started in work by Stieltjes from 1894 in connection with moment problems. He considered the class of non-decreasing continuous functions $F(x)$ defined on $x \geq 0$ where $F(0) = 0$ and $F(x)|to1$ as $x \to +\infty$, and the integrals

$$m_k(F) = \int_0^\infty x^k \cdot dF(x) < \infty$$

for every positive integer k. If G is another function of this form and the moments $m_k(F) = m_k(G)$ for every $k \geq 1$, the question arises if the two functions are identical. Stieltjes' found examples where $F \neq G$ even if they have the same moments. This "negative answer" illustrates that the family of all probablity measures on $x \geq 0$ is quite ample. In § xx we shall construct examples where Stieltjes' phenomenon appears.

In addition to pioneering work by Stieltjes, Borel and Lebesgue, important contributions are also due to Vitali, F. Riesz and Norbert Wiener. Some results about convergence were established by Vitali in the article *Sulle funzione integrali* [Atti. R. acad. Sci. di Torino: 1905] which put earlier results by Lebesgue in a broader context. Staying in an euclidian space $\mathbf{R}^n$, F. Riesz constructed a class of measures in $\mathbf{R}^n$ where Borel's abstract theory is applied to the σ-algebra of Borel sets in $\mathbf{R}^n$, i.e. the smallest σ-algebra which contains all open and closed sets. One consequence of Riesz' theory goes as follows: The solution of the Dirichlet problem in the unit disc yields a 1-1 correspondence between the class of positive harmonic functions $H(x, y)$ in the open unit disc $D = \{x^2+y^2 < 1\}$ normalised so that $H(0, 0) = 1$, and the class of probability measures supported by the unit circle $T = \{x^2 + y^2 = 1\}$. A notable point is that the class contains measures which are singular with respect to the angular Lebesgue measure on T. A more general construction due to Wiener works for every bounded open set Ω even if the Dirichlet problem cannot be solved for every continuous boundary function f in $C^0(\partial\Omega)$. More precisely, for each $z \in \Omega$, Wiener constructed a distinguished probability measure $\mathfrak{m}_z^\Omega$ supported by $\partial\Omega$ such that each $f \in C^0(\partial\Omega)$ produces a unique harmonic function W_f defined in Ω by

$$W_f(z) = \int_{\partial\Omega} f(\zeta) \cdot \mathfrak{m}_z^\Omega(\zeta)$$

Wiener's result is proved in the chapter about harmonic and subharmonic functions. In Ergodic Theory measures are used frequently. The literature is extensive and current research active but we shall not enter a discussion of this important and very extensive subject.

Infinite products of measure spaces. A method to construct integrals over infinite product spaces was carried out by Daniell in [Dan]. It is logically transparent but has a disadvantage because one cannot so easily call upon existing theorems in Lebesgue theory from the finite dimensional case. To overcome this, Norbert Wiener used *binary expansions* in order to extend the Lebesgue measure to an infinite product of $\mathbf{R}$. The crucial point is the construction of a measure preserving map from the the two-dimensional unit square $\square = \{(x, y) \colon\ 0 < x, y < 1\}$ onto the interval $(0, 1)$ which goes as follows: Every real number $0 < x < 1$ has a binary expansion:

$$x = \alpha_1/2 + \alpha_2/4 + \dots \qquad \alpha_k = 0 \text{ or } 1$$

Similarly $0 < y < 1$ has a binary expansion

$$y = \beta_1/2 + \beta_2/4 + \dots \qquad \beta_k = 0 \text{ or } 1$$

These expansions are unique when one prescribes that the expansion never terminates with a sequence where the number 1 appears. For example, one represents $1/4$ by $0, 1, 0, , 0 \dots$ while the sequence $0, 0, 1, 1, 1, \ldots$ is not allowed. A pair of sequences as above is arranged into a single sequence which is used to express a real number $0 < t < 1$ whose binary series is:

$$(*) \qquad t(x,y) = \alpha_1/2 + \beta_2/4 + \alpha_2/8 + \beta_2/16 + \ldots$$

This yields a bijective map from $\square$ onto $(0,1)$ which is *measure preserving*, i.e. if $F(x,y)$ is a continuous function of (x,y) then

$$\iint_\square F(x,y) \cdot dxdy = \int_0^1 f(t) \cdot dt \quad \text{where} \quad f(t(x,y)) = F(x,y)$$

In this way Lebesgue theory in dimension two can be recaptured from the one-dimensional case. Similar measure preserving transformations from $\mathbf{R}^n$ into $\mathbf{R}$ exist for every $n \geq 2$ and once this is achieved it is not difficult to construct a measure preserving transformation from a denumerable product to the real t-line. Details of this construction appear in the text-book [Paley-Wiener].

Another issue are product spaces with infinitely many factors where one seeks to extend the Fubini theorem. Here a major contribution is due to Jessen in the work *Bidrag til intergalteorin for funktioner av uendlig mange variabler* [Copenhagen 1930]. His article *The theory of integration in a space of infinitely many variables* [Acta math. 1933] is a veritable classic with a wide range of applications, foremost in probability theory.

The Vitali-Hahn-Saks theorem. During the development of functional analysis in a broad sense which started around 1930, original results by Vitali were put into the context of topological vector spaces and quite general results were established by Saks in the article *Integration in abstract metric spaces.* [Duke Math. J. Vol 4: 1938]. Here one regards a Boolean σ-algebra $\mathcal{B}$ of subsets of a sample space Ω and μ is a non-negative σ-additive meaasure which assigns real numbers to sets in $\mathcal{B}$. In § xx we introduce the family of μ-continuous additive functions and prove a result concerned with uniform convergence which ressembles the Banach-Steinhaus theorem in Frechet spaces. As expected the Baire category theorem plays an essential role in the proofs.

Probability theory. A feeling for measure theory arises via probabilistic interpretations. For example, if μ and ν is a pair of Riesz measures on the real oine which both are non-negstive and have mass one, then there exists the convolution $\mu * \nu$. If μ and ν describe probability distributions of two random variables, then $\mu * \nu$ is the probability distribution of their sum under the condition that they are independent. In § 8 we study the Brownian motion. Here we first expose Wiener's construction of a stochastic process in continuous time with independent increments while the outcome at each time-value is a normally distributed random variable. To analyze individual Brownian paths one needs a sample space which by Wiener's construction is found by a denumerable infinite product of "ordinary" measure spaces. The whole study culminates in Theorem 9 in from § 8.1 which is one of the major results in this chapter. In § 8.2 we have also included the construction of a parabolic PDE-system which governs a stochastic vector-valued process in continuous time. This has been inserted since such equations occur in many applications, foremost in quantum mechanics but also in the nowadays popular subject referred to as "matheamatics of finance".

Ohter applications of measure theory. The text-book [Weil] by Andre Weil employs measure theoretic results to establish reciprocity theorems and advanced results in class-field theory. In the appendix devoted to functional analysis we establish the existence of Haar measures on locally compact groups which in general are non-commutative.

Integral inequalities and distribution theory. The study of various L^p-estimates is an extensive subject where Fourier analysis appears together with various geometric constructions such as Whitney coverings. The reader may consult a number of excellent text-books by Elias Stein and co-authors which treat the theory about singular integrals. Here we only give a brief account about the Hardy-Littlewood maximal function in § xx. Passing to differentiability one is led to distribution theory. In fact, every distributions in $\mathbf{R}^n$ is locally expressed by finite sums of distribution derivatives of Riesz measures. So measure theory is essential in order to grasp the theory of distributions.

Let us finish the introduction by describing some results expressed via measure theoretic concepts but the proofs require "hard analysis".

Hitting probability for the Brownian motion. The following result is due to the late Björn Dahlberg:

Let U be a bounded open set in $\mathbf{R}^n$ defined by $\{F(x) < 0\}$ where $F(x)$ is a Lipschitz continuous function and $\partial U = \{F(x) = 0\}$. Then the class of null sets in ∂U for the Brownian motion which starts at some point $x_0 \in U$ is equal to the class of subsets of ∂U whose $(n-1)$-dimensional Hausdorff measure are zero.

Above $E \subset \partial U$ is a null set for the Brownian motion if the probability to hit E before the Brownian path reaches ∂U at points outside E is equal to zero. For this discovery Dahlberg was awarded the Salem Prize in 197?.

Carleson' convergence theorem in L^2. In 1965 Lennart Carleson proved the following:

Let $\{a_n\}$ be a sequence of complex numbers whose ℓ^2-norm is finite, i.e. $\sum |a_n|^2$ is finite. To every $N \geq 1$ one constructs Fourier's partial sum:

$$S_N(x) = \sum_{\nu=-N}^{\nu=N} a_n e^{inx}$$

Then $\lim_{N\to\infty} S_N(x)$ exists for all $0 \leq x \leq 2\pi$ except for a set of Lebesgue measure zero.

This result is one of the greatest achievements ever in mathematical analysis. In 2006 Carleson was awarded the Abel Prize for this theorem together with his contributions in other areas of mathematics. Let us only remark that Carleson's theroem came as surpsie even to specialists since Kolmogorov had contructrd an absolutely integrable function f on the unit circle whose partial sums of its Fourier series are *everywhere divergent.*

0.1 The Borel-Riesz theory.

Let $\mathcal{B}$ be a Boolean algebra of sets and denote by $\mathbf{ba}(\mathcal{B})$ the space of additive and bounded real-valued functions on $\mathcal{B}$. Thus, one considers maps

$$E \mapsto \mu(E)$$

Additivity means that $\mu(\cup E_\nu) = \sum \mu(E_\nu)$ when $\{E_\nu\}$ is a finite family of disjoint sets in $\mathcal{B}$, and boundedness means that there exists a constant C such that

$$|\mu(E)| \leq C \quad : E \in \mathcal{B} \tag{i}$$

Define the non-negative function μ_+ by

$$\mu_+(E) = \max_{A \subset E} \mu(A)$$

where we declare that $\mu(\emptyset) = 0$.

Exercise. Show that μ_+ is additive, i.e.

$$\mu_+(E \cup F) = \mu_+(E) + \mu_+(F) \tag{0.0.1}$$

hold for every pair of disjoint sets in $\mathcal{B}$.

Now $\mu_+ - \mu$ is a non-negative additive function denoted by μ_- and we have

$$\mu = \mu_+ - \mu_- \tag{0.0.2}$$

Hence μ is represented as the difference of two non-negative and additive functions. If $E \in \mathcal{B}$ the total variation of μ over E is defined by

$$|\mu|(E) = \mu_+(E) + \mu_-(E)$$

The reader may verify that (i) above gives

$$|\mu|(E) \leq 2C$$

for every $E \in \mathcal{B}$.

Let us now consider a compact topological space S which in addition is Hausdorff. Denote by $\mathcal{B}_S$ the Boolean algebra generated by open sets in S. The reader can check that every set in $\mathcal{B}_S$ is finite union of disjoint sets of the form $U \cap F$ where U are open and F are closed. Denote by $\chi(S)$ the linear space of real-valued functions on S given by $\mathbf{R}$-linear combinations of characteristic functions of sets in $\mathcal{B}_S$, i.e. functions of the form

$$\phi = \sum c_k \cdot \chi_{E_k}$$

where $\{E_k\}$ is a finite family of disjoint subsets from $\mathcal{B}_S$ and $\{c_k\}$ are real constants. To each pair $\phi \in \chi(S)$ and $\mu \in \mathbf{ba}(\mathcal{B}_S)$ one defines the integral

$$\int \phi \cdot d\mu = \sum c_k \cdot \mu(E_k)$$

Exercise. Prove additivity, i.e. that

$$\int (\phi + \psi) \cdot d\mu = \int \phi \cdot d\mu + \int \psi \cdot d\mu \quad : \phi, \psi \in \chi(S)$$

A hint is that one can represent the two functions in (i) with respect to a common partition of S, i.e. there exists a familly of disjoint sets $\{E_k\}$ so that

$$\phi = \sum c_k \cdot \chi_{E_k} \quad : \psi = \sum d_k \cdot \chi_{E_k}$$

where $\{c_k\}$ and $\{d_k\}$ are constants.

0.1.2 Borel functions and their integrals. Via the maximum norm $\chi(S)$ is a normed vector space. If $\mu \in \mathbf{ba}(\mathcal{B}_S)$ its total variation is denoted by $||\mu||$. The construction of integrals gives

$$|\int \phi \cdot d\mu| \leq |\phi|_S| \cdot ||\mu||$$

where $|\phi|_S$ is the maximum norm of ϕ. The closure of $\chi(S)$ taken in the normed space of bounded real-valued functions on S yields a Banach space denoted by $\mathfrak{B}(S)$. For each $f \in \mathfrak{B}(S)$ a μ-integral is defined as follows: By definition there exists a sequence $\{\phi_n\}$ from $\chi(S)$ such that $|f - \phi_n|_S \to 0$. When n, m are positive integers we have

$$|\int \phi_n \cdot d\mu - \int \phi_m \cdot d\mu| \leq |\phi_n - \phi_m|_S \cdot ||\mu||$$

The uniform convergence $\phi_n \to f$ entails that $\{|\phi_n - \phi_m|_S\}$ is a Cauchy-sequence and hence there exists

$$\lim_{n\to\infty} \int \phi_n \cdot d\mu$$

It is clear that the last limit only depends on f and not upon the particular sequence in $\chi(S)$ which approximates f uniformly. The common limit is defined as the μ-integral of f, and

$$f \mapsto \int f \cdot d\mu$$

gives a linear functional on the normed vector space $\mathfrak{B}(S)$. This identifies $\mathbf{ba}(\mathcal{B}_S)$ with a subspace of the dual space $\mathfrak{B}(S)^*$. Conversely, let L be a continuous linear functional on $\mathfrak{B}(S)$ which means that there exists a constant C such that

$$|L(f)| \leq C \cdot |f|_S \quad : f \in \mathfrak{B}(S)$$

Now L yields a unique additive measure μ where

$$\mu(E) = L(\chi_E) \quad : E \in \mathcal{B}_S$$

We leave it to the reader that this entails that

$$L(f) = \int f \cdot d\mu \quad : f \in \mathfrak{B}(S)$$

Hence one has an isomorphism

(*) $$\mathfrak{B}(S)^* \simeq \mathbf{ba}(\mathcal{B}_S)$$

0.1.3 The dual of $C^0(S)$. Tbe space of continuous and real-valued functions on the compact space S is a subspace of $\mathfrak{B}(S)$. To see this we take some $f \in C^0(S)$ and notice that for every pair of real numbers $a < b$, the set

$$\{s \colon a \leq f(s) < b\}$$

belongs to $\mathcal{B}_S$. If $\epsilon > 0$ we choose an integer N so that $N^{-1} > \epsilon$ and introduce the sets

$$E_k^f(N) = \{s \colon \frac{k}{N} \leq f(s) < \frac{k+1}{N}\}$$

where k are integers. Since the range of f is compact it follows that $E_k^f(N) \neq \emptyset$ only hold for a finite family of integers k, and we get a function in $\chi(S)$ defined by

$$\phi = \sum \frac{k}{N} \cdot \chi_{E_k^f(N)}$$

Here the maximum norm $|f - \phi|_S < \epsilon$ and we conclude that f belongs to $\mathfrak{B}(S)$. The inclusion $C^0(S) \subset \mathfrak{B}(S)$ means that each $\mu \in \mathbf{ba}(\mathcal{B}_S)$ restricts to a continuous linear functional on the normed vector space $C^0(S)$. Hence there exists a map

$$\mathbf{ba}(\mathcal{B}_S) \to C^0(S)^*$$

The Hahn-Banach Theorem shows that this map is surjective which gives an isomorphism

(*) $$C^0(S)^* \simeq \frac{\mathbf{ba}(\mathcal{B}_S)}{C^0(S)^\perp}$$

where $C^0(S)^\perp$ is the space of $\mu \in \mathbf{ba}(\mathcal{B}_S)$ such that

$$\int f \cdot d\mu = 0 \quad : f \in C^0(S)$$

Riesz' representation theorem.

A measure μ in $\mathbf{ba}(\mathcal{B}_S)$ is said to be regular if the following hold for every set $A \in \mathcal{B}_S$:

(0.1..4) $$\min |\mu|(U \setminus A) = 0 \quad \& \quad \min |\mu|(A \setminus K) = 0$$

where the minimum is taken over open sets U which contain A, respectively compact subsets K. The family of these measures is denoted by $\mathfrak{M}(S)$.

0.1.5 Theorem. *For each $\mu \in \mathbf{ba}(\mathcal{B}_S)$ there exists a unique $\mu^* \in \mathfrak{M}(S)$ such that*

(*) $$\int f \cdot d\mu = \int f \cdot d\mu^* \quad : f \in C^0(X)$$

The uniqueness part. Let us first show that if a regular measure μ^* exists above then it is unique. This amounts to show that if a measure $\gamma \in \mathfrak{M}(S)$ is such that

$$\int f \cdot d\gamma = 0$$

for every $f \in C^0(S)$ then $\gamma = 0$. To show this we argue by contradiction, i.e. suppose that $\gamma(A) \neq 0$ for some $A \in \mathcal{B}_S$. Multiplying γ with a constant we can assume that $\gamma(A) = 1$. Let $0 < \epsilon 1/4$ and by (0.1.4) we find a compact subset F of A and an open set U which contains A such that the total variation

(i) $$|\gamma|(U \setminus F) < \epsilon$$

Notice that (i) and the equality $\gamma(A) = 1$ give

(ii) $$\gamma(F) \geq 1 - \epsilon$$

Next, there exists $f \in C^0(S)$ whose range is contained in $[0,1]$ while $f = 1$ on F and zero on $S \setminus U$. It follows that

$$\int f \cdot d\gamma \geq \mu(F) - |\int_{S \setminus F} f d\gamma| \geq \mu(F) - |\gamma|(U \setminus F) \geq 1 - 2\epsilon > 1/2$$

This contradiction proves the uniqueness part in Theorem 0.1.5.

The existence part. We are given μ and by (0.0.2) we can write $\mu = \mu_+ - \mu_-$. By additivity it suffices to prove the existence of regular measures μ_+^* and μ_-^* in Theorem 0.1.5. So we can assume that μ from the start is non-negative and let L be the associated linear functional from (0.1.2). For every open set U we set

(i) $$\mu_*(U) = \max_f L(f)$$

with the maximum taken over all continuous functions with range in $[0,1]$ and compact support contained in U. Next, if K is a pair of compact sets we define

(i) $$\mu_*(K) = \min_U L(f)$$

with the minimum taken over continous functions with range in $[0,1]$ and $f = 1$ on K. We claim that

$$\mu_*(K) + \mu_*(X \setminus K = 1$$

for every compact set. To see thid we take $\epsilon > 0$ and (i) gives some f with compact support in $X \setminus K$ while

$$\mu_*(X \setminus K) < L(f) + \epsilon$$

Now $1-f$ is identically one on K and hence $L(1-f) \geq \mu_*(K)$ which shows that (*) is majorised by

$$L(f) + L(1-f) + \epsilon = 1 + \epsilon$$

Since ϵ was arbitrary themleft hand side in (xx) is ≤ 1. Conversely, we find some f which is identically one on K and

$$\mu_*(K) > L(f) - \epsilon$$

Now $\{f \leq 1-\epsilon\}$ is disjoint from K® and we can find a cotinuous function g with range on $[0,1]$ which is equal to one on this set while $g = 0$ on $\{f \geq 1-\epsilon/2\}$. It follows that $\mu(X \setminus K) \geq L(g)$ and together with (xx) we have

$$\mu_*(K) + \mu_*(S \setminus K) > L(f) + L(g) - \epsilon$$

At the same time we notice that

$$f + g \geq 1 - \epsilon$$

holds in x and since L was non-negative we get

$$L(f) + L(g)?L(f+g) \geq 1\epsilon$$

Since ϵ was arbitrary small we see that (x) and (xx) gives the opposite inequality in (xx).

Exercise. Conclude via (*) that μ^* is an additive and regular measure on S.

There remains to prove the equality

$$(*) \qquad L(f) = \int f \cdot d\mu_*$$

for every continuous function. Since every $f \in C^0(S)$ is the difference of two non-negative functions $f_+ = \max(f,0)$ and $f_- = -\min(f,0)$, additivity implies that it suffices to show (*) when f is non-negative. For every partition $0 = a_0 < a_1 < \ldots < a_M$ where a_M is strictly larger than the maximum of f, the reader should check that there exists a sequence $\{\xi_k\}$ with $a_k \leq \xi_k \leq a_{k+1}$ and

$$L(f) = \sum \xi_k \cdot \mu(\{a_k \leq f < a_{k+1}\})$$

At the same time we have

$$\int f \cdot d\mu_* \leq \sum a_{k+1} \cdot \mu_*(\{a_k \leq f < a_{k+1}\})$$

The construction of μ_* entials that

$$\mu_*(\{a_k \leq f < a_{k+1}\}) \leq \mu(\{a_k \leq f < a_{k+1}\})$$

for every k. So if $\delta = \max a - k + 1 - a_k$ we get

$$\int f \cdot d\mu_* \leq \delta \cdot \sum \xi_k \cdot \mu(\{a_k \leq f < a_{k+1}\})$$

and the last term is majorised by $\mu(S)$ which is a given finitei constant. Passing to refined partions with $\delta \to 0$ we conclude that

$$\int f \cdot d\mu_* \leq L(f)$$

To prove the opposite inequality we shall employ special partitions. We are given f and say that a real number β is L-critical if there exists a positive number c such that

$$\mu(\{\beta - \epsilon \leq f \leq \beta + \epsilon\}) \geq c$$

By additivity and the finiteness of $\mu(S)$ it is clear that for every $c > 0$ the set of β for which (x) hold is finite. From this observation the reader can check that for every $\delta > 0$ there is a partition $\{a_\nu\}$ where $a_{k+1} - a_k \leq \delta$ hold for every k and none of the a-numbers are L-critical. Next, we notice that

$$L(f) \leq \delta + \sum a_k \cdot \mu(\{a_k \leq f < a_{k+1}\})$$

The careful choice of the a-numbers gives a sequence $\{b_k\}$ with $a_k < b_k < b_{k+1} < a_{k+1}$ and

$$\sum a_k \cdot \mu(\{a_k \leq f < a_{k+1}\}) \leq \delta + \sum a_k \cdot \mu(\{b_k \leq f \leq b_{k+1}\}) \implies$$

$$\text{(i)} \qquad L(f) \leq 2 \cdot \delta + a_k \cdot \mu(\{b_k \leq f \leq b_{k+1}\})$$

Next, for every k we can find a continuous function g_k with range in $[0,1]$ which is identically one on the compact set $\{b_k \leq f \leq b_{k+1}\}$ and at the same time has compact support in $\{a_k \leq f < a_{k+1}\}$. Now the reader can check that the construction of μ_* entails that

$$\text{(ii)} \qquad \int f \cdot d\mu_* \geq \sum a_k \cdot \mu(\{b_k \leq f \leq b_{k+1}\})$$

Together with (i) it follows that

$$L(f) \leq 2 \cdot \delta \int f \cdot d\mu_*$$

Above δ can be chosen arbitary small which proves the equality (*) and finishes the proof of Theorem xx.

0.1.7 σ-additivity and the algebra $\widehat{\mathcal{B}}_S$

Above we started from the Bolean algebra $\mathcal{B}_S$ generated by open subsets of S. By general set-theory there exists a unique smallest σ-algebra $\widehat{\mathcal{B}_S}$ which contains $\mathcal{B}_S$. Sets in this family are called Borel sets. Let us remark that one can construct $\widehat{\mathcal{B}_S}$ by transfinite induction using ordinal numbers. We shall not pursue this in detail but refer to original work by Baire and Borel where these constructions are carried out in detail. The reader may also consult the text-book [Kol] by Kolmogorov for a detailed account about the contruction of σ-algebras and extensions of σ-additive measures. A major fact is the following:

0.1.8 Theorem. *Each $\mu \in \mathfrak{M}(S)$ extends in a unique way to a σ-additive measure $\widehat{\mu}$ on $\widehat{\mathcal{B}_S}$. Conversely, every σ-additive measure on $\widehat{\mathcal{B}_S}$ is equal to $\widehat{\mu}$ for a unique Riesz measure.*

In § 4 we prove this theorem in detail when $S = \mathbf{R}^n$ for some $n \geq 1$.

0.1.9 Null sets. Let μ be a non-negative Riesz measure on the compact space S. A subset G is a null set with respect to μ if

$$\min_{G \subset U} \mu(U) = 0$$

where U are open sets containing G. Let $\mathcal{N}(\mu)$ be the family of these null sets. Additivity entails that a finite union of null sets again is null set. Less obvious is that a denumerable union of null sets is a null set. To show this we consider s sequence $\{G_1, G_2, \ldots\}$ in $\mathcal{N}(\mu)$. Let $\epsilon > 0$ and for each n we find an open set U_n containing G_n such that

$$\mu(U_n) \leq 2^{-n}\epsilon$$

Now $U^* = \cup U_n$ contains $\cup G_n$. Since μ is regular there exists a compact subset F of U^* such that $\mu(U^* \setminus F) < \epsilon$. Now $\{U_n\}$ is an open covering of F and Heine-Borel's Lemma gives some integer N such that the union of $U_1, \ldots, U_N$ contains F. It follows that

$$\mu(U^*) = \mu(U^* \setminus F) + \mu(F) \leq \epsilon + \mu(U_1 \cup \ldots \cup U_N) \leq \epsilon + \sum_{k=1}^{k=N} |\, \mu(U_k) \leq 2\epsilon$$

Since ϵ is arbitrsty small we conclude that $\cup G_n$ is a null set.

In § II where abstract measure theory is exposed, the result above is the starting point for the study of "almost everywhere defined functions" with respect to μ which leads to the space $L^1(\mu)$ of functions which are integrable with respect to μ.

0.2 Lebesgue theory.

The theory was created by Henri Lebesgue and presented in his monograph *Lecons sur l'integration et la recherche des fonctions primitives* from 1904. We expose some basic facts from his work and remark that Vitali and de la Vallé Poussin later gave some simplifications of Lebesgue's original proofs. A major result in § III is the existence of *Lebesgue points* which implies that measurable functions in Lebesgue's sense are realized in a concrete way. For example, a bounded Lebesgue measurable function f is the pointwise limit outside a null set of a sequence of Lipschitz continuous functions defined by mean values of f over smaller and smaller cubes centered at the point where pointwise convergence takes place. In § 3.xx we also show that outside a null set an arbitrary bounded measurable function f differs from an everywhere defined function f_* given as the pointwise limit of a non-increasing sequence of upper semi-continuus functions which means that f_* belongs to the second Baire class of functions. So when one agrees to identify Lebegue measurable functions which are equal outside a null set, then this class is reduced to a restricted family of everywhere defined functions.

The crucial point in Lebesgue theory is to define the n-dimensional volume of a bounded open set. To achieve this we use *binary expansions* of real numbers. If $n \geq 1$ and $\square$ is a cube placed in $\mathbf{R}^n$ it has 2^n many corner points and its size is measured by the positive number a where the distance between two adjacent corner points is a, and the n-dimensional volume is a^n. To construct the n-dimensional measure of an open set Ω in $\mathbf{R}^n$ we employ binary expansions of real numbers x:

(i)
$$x = [x] + \sum_{k=1}^{\infty} \epsilon_k(x) 2^{-k}$$

where $[x]$ is the largest integer $\leq x$ and each ϵ is 0 or 1. The expansion is unique under the rule that the the sequence $\{\epsilon_k(x)\}$ never terminates with 1 all the time. For example, the number $x = 1/8$ has $\epsilon_3(x) = 1$ while all other ϵ-numbers are zero. So it is not permitted to represent 1/8 by the geometric series $1/16 + 1/32 + 1/64 + \ldots$. If $N = 0$ we put $\xi_0(x) = [x]$ and if $N \geq 1$ we put

(ii)
$$\xi_N(x) = [x] + \sum_{k=1}^{k=N} \epsilon_k(x) 2^{-k} \implies \xi_N(x) \leq x < \xi_N(x) + 2^{-N}$$

Next, if $n \geq 2$ and $p = (x_1, \ldots, x_n)$ is a point in $\mathbf{R}^n$ one takes the binary expansion of each coordinate and set $\xi_N(p) = (\xi_N(x_1), \ldots, \xi_N(x_n))$. For each non-negative integer N one associates the half-open cube

$$\square_N(p) = \xi_N(p) + \square_N^* \quad : \quad \square_N^* = \{(x_1, \ldots, x_n) : \ 0 \leq x_k < 2^{-N}\}$$

The reader may verify that

(iii)
$$q \in \square_N(p) \implies \xi_N(q) = \xi_N(p)$$

and the last equality entails that $\square_N(q) = \square_N(p)$. Let us now consider a bounded open set Ω in $\mathbf{R}^n$. To each $p \in \Omega$ we find the unique smallest non-negative integer $N_\Omega(p)$ such that $\square_{N_\Omega(p)} \subset \Omega$. If $k \geq 0$ we put

(iv)
$$D_k(\Omega) = \{p \in \Omega : \ N_\Omega(p) = k\}$$

Here (iii) entails that $\{D_k(\Omega)\}$ are disjoint subsets of Ω. Since Ω is open it is clear that the union is equal to Ω. Moreover (iii) entails that when $D_k(\Omega \neq \emptyset$ then it is a union of pairwise disjoint half-open cubes:

$$D_k(\Omega) = \cup \square_N(p_\nu)$$

Since Ω is bounded every such union is finite. Denote by $\mu_k(\Omega)$ the number of cubes in this disjoint union. The n-dimensional measure is defined by

(*)
$$\text{vol}_n(\Omega) = \sum_{k=0}^{\infty} \mu_k(\Omega) \cdot 2^{-kn}$$

0.2.1 Unbounded open sets. If Ω is an open unbounded set we still get the unique partition formed by the disjoint subsets $\{D_k(\Omega)\}$. One says that Ω has a finite measure if every $D_k(\Omega)$ is a finite union and the series

$$\sum \mu_k(\Omega) \cdot 2^{-kn} < \infty$$

Outer and inner measures. Starting from the construction of n-dimensional measures of open sets one constructs outer and inner measures of an arbitrary bounded subset A of $\mathbf{R}^n$. The outer measure is defined by:

$$|A|^* = \min_{A \subset \Omega} \mathrm{vol}_n(\Omega)$$

In particular the outer measure is defined for every compact set E and the inner measure of A is defined by

$$|A|_* = \max_{E \subset A} |E|^*$$

where the maximum is taken over compact subsets of A.

0.2.2 Definition. *A set A in $\mathbf{R}^n$ has a finite Lebesgue measure if $|A|_* = |A|^* < \infty$ and this common number is denoted by $|A|_n$.*

0.2.3 Null sets. A set is called a null set if its outer measure is zero. The family of null sets is denoted by $\mathcal{N}(\mathbf{R})^n$. A property is said to hold almost everywhere if it is valid outside a null set. In Lebesgue theory one is often content to establish a result "almost everywhere", i.e. null sets are reduntant. Recall that a set S is of type G_δ if it is the intersection of an decreasing sequence of open sets $\{\Omega_n\}$. It is a set of type F_σ if it is given as the union of a decreasing familly of compact sets.

0.2.4 Exercise. Show that if A is Lebesgue measurable and $\{E_\nu\}$ an increasing sequence of compact subsets such that $|E_\nu|_n \to |A|_n$, then $A \setminus \cup E_\nu$ is a null-set. Similarly, if $\{\Omega_\nu\}$ an decreasing sequence of open sets containing A and $|\Omega_\nu|_n \to |A|_n$ then $\cap \Omega_\nu \setminus A$ is a null set. Hence every measurable set differs from a set of type F_σ by a null set, and similarly from a set of type G_δ.

0.2.5 Lebesgue points. Let A be a measurable set. A point $x \in A$ has unit density if

$$\lim_{|\square|_n \to 0} \frac{\mathrm{vol}_n(\square \setminus A)}{\mathrm{vol}_n(\square)} = 0 \tag{0.5.1}$$

where the limit is taken over cubes which contain x. It is not requested that x is the center of the cubes and their sides need not be parallell to the coordinate axis. The set where (0.5.1) holds is denoted by $\mathfrak{Leb}(A)$. A major result due to Lebesgue asserts the following:

0.2.6 Theorem. *For every measurable set A it follows that $A \setminus \mathfrak{Leb}(A)$ is a null set.*

This fundamental result is proved in § 3.

0.2.7 Measurable functions. Let $f(x)$ be a real-valued function defined in some open and bounded subset Ω of $\mathbf{R}^n$. It is measurable in the sense of Lebesuge if the sets

$$A_f(a; b) = \{x \in \Omega \colon a \leq f(x) < b\}$$

are measurable for all pairs of real numbers $a < b$. The function f has Lebesgue value a at a point $x_0 \in \Omega$ if

$$\lim_{|\square|_n \to 0} \frac{|A_f(a - \epsilon, a + \epsilon) \cap \square|_n}{|\square|_n|} = 1 \tag{0.7.1}$$

hold for every $\epsilon > 0$ where the limit of cubes is taken as in (0.5.1). It is obvious that if (0.7.1) hold for some real number $a <$, then it is uniquely determined and we set $\mathfrak{Leb} f(x_0) = a$. In § xx we show that the measurable function f has a Lebesgue value almost everywhere. More preciely, there exists a null set W of Ω such that $\mathfrak{Leb} f$ is defined in $\Omega \setminus W$ and is equal to f in this set. So up to null sets every measurable function f is recaptured from its Lebesgue function, i.e. ignoring a null set f can always be chosen so that its values coincide with its Lebesgue values.

0.2.8 Fubini's theorem. Let n snd m be a pair of positive integers and A a bounded and measurable subset of $\mathbf{R}^{n+m}$ whose points are written as (x,y) with $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$. To every x we get the slice

$$A(x) = \{y\colon (x,y) \in A\}$$

In §§ we prove that $A(x)$ are measurable sets in $\mathbf{R}^m$ for all x outside a null set and the almost everywhere defined function

$$x \mapsto |A(x)|_m$$

is measurable. After we have learnt how to construct Lebegue integrals we also prove the equality

$$\int_{\mathbf{R}^n} |A(x)|_m \, dx = |A|_{n+m} \tag{0.8.1}$$

0.2.9 Non-measurable sets. There exist non-measurable sets but in practice one does not encounter these ugly sets. In fact, one must appeal to the Axiom of Choice to exhibit examples. Using this axiom there exists a subset E of the open interval $(0,1)$ with the property that the sets $\{E+q\}$ are disjoint when q runs over the set of rational numbers. This set cannot be measurable. The reason is that the disjointness would entail that if $q_1, \ldots, q_N$ is a set of rational numbers between 0 and 1 then

$$N \cdot |E|_* = \sum_{\nu=1}^{\nu=N} = |E + q_\nu|_* \leq 2$$

where the last equality holds since each set $E + q_\nu \subset [0,2]$. Since N can be arbitrary large we have $|E|_* = 0$. So if E is measurable we have $|E| = |E|_* = 0$, i.e E is a null set. Since $E + q$ is translate of E they also give null sets for every rational number. But a denumerable union of null sets is a null set which gives a contradaction since $\cup (E + q)$ with the union taken over all rational numbers is the whole real line.

0.3 Riesz measures in $\mathbf{R}^n$.

To each integer $N \geq 0$ we have the lattice $2^{-N} \cdot \mathbf{Z}^n$ of points $p = (p_1, \ldots, p_n) \in \mathbf{R}^n$ such that $2^N p_\nu$ are integers for every $1 \leq \nu \leq n$. With the notstions from 0.2.4 we obtain for each $p \in 2^{-N} \cdot \mathbf{Z}^n$ the half-open cube

$$\square_N(p) = p + \square_N^*$$

Set

(0.3.1) $$\mathcal{D}_N = \{\square_N(p) \colon \, p \in 2^{-N}\mathbf{Z}^n\}$$

Notice that each cube in $\mathcal{D}_N$ is the union of 2^n many disjoint cubes from $\mathcal{D}_{N+1}$. Suppose that non-negative numbers $\{\mu(\square_N(p))\}$ are assigned for all $N \geq 0$ and every $p \in \mathbf{Z}^n$. The assignement is additive if

$$\mu(\square_N(p)) = \sum \mu(\square_{N+1}(q))$$

where the sum is taken over the 2^n many cubes in $\mathcal{D}_{N+1}$ contained in $\square_N(p)$. Given an additive assignement we obtain a μ-measure for every bounded open set Ω as follows: To each $k \geq 0$ we have the family $D_k(\Omega)$ from (0.2.XX)- Put

$$\mu_k(\Omega) = \sum \mu(\square_N(p))$$

with the sum taken over the4 finite disjoint family of cubes in $D_k(\Omega)$. Now we define

$$\mu(\Omega) = \sum_{k=0}^{\infty} \mu_k(\Omega)$$

Additivity implies that if $\{\Omega_\nu\}$ is a finite family of disjoint bounded open sets then

$$\mu(\cup\, \Omega_\nu) = \sum \mu_k(\Omega_\nu)$$

Next, if E is a compact set we set

$$\mu(E) = \min_{E \subset \Omega} \mu(\Omega)$$

0.3.1 The regularity condition. If $N \geq 0$ and $0 < \delta < 1$ one has the compact cube

$$\square_N[\delta] = \{x \colon 0 \leq x_k \leq \delta \cdot 2^{-N}\}$$

To each $p \in \mathbf{Z}^n$ we get the compact cube

$$\square_N(p)[\delta] = p + \square_N[\delta]$$

0.3.2 Definition. *An additive μ-measures is regular if the following hold for every $N \geq 0$ and every lattice point $p \in \mathbf{Z}^n$:*

(0.9.2) $$\mu(\square_N(p)) = \lim_{\delta \to 1} \mu(\square_N(p)[\delta])$$

The class of such regular measures is denoteb by $\mathfrak{M}(\mathbf{R}^n)$ and called Riesz measures in $\mathbf{R}^n$.

It turns out that (0.3.2) and additivity enable us to define μ-measures in a similar fashion as in Lebegue's case. For example, (0.3.2) entails that when Ω is a bounded open set then

$$\mu(\Omega) = \sup_{E \subset \Omega} \mu(E)$$

where the supremum is taken over compact subsets. Next, if A is a bounded set one defines its outer μ-measure by

$$\mu^*(A) = \min_{A \subset \Omega} \mu(\Omega)$$

Similarly the inner measure is defined by

$$\mu_*(A) = \max_{A \subset E} \mu(E)$$

One says that A is μ-measurable if $\mu^*(A) = \mu_*(A)$ and the common number is denoted by $\mu(A)$.

0.3.4 Remark. In § IV we expose the theory about Riesz measures where μ-measurable functions and their μ-integrals are constructed. We shall learn that there exists an extensive class of non-negative Riesz messures whose mass is concentrated to null sets in Lebesgue's sense. The construction of Riesz measure is not easy to grasp. One reason is that apart from the additivity the regularity introduces a condition which cannot be predicted in a straightforward manner since one needs the axiom of choice to obtain the whole class of non-negative Riesz measures which correspond to non-negative linear functionals on the space of real-valued and continuous functions in $\mathbf{R}^n$ having compact support. Important examples of a geometric origin appear in connection with Stokes Theorem in § V.

0.4 1-dimensional study.

A construction due to Weierstrass gave considerable motivation for the development measure theory.

0.4.1 Weierstrass example. Let $0 < b < 1$ be a real number and k a positive integer. Define $f(x)$ by the series

$$f(x) = \sum_{n=1}^{\infty} b^n \cdot \cos(k^n \pi x)$$

Since $b < 1$ the series of cosine-functions converges uniformly so f is a continuous and 2π-periodic. Let $b = 1/5$ and the integer $k > 5$ is so large that

$$(*) \qquad \frac{3\pi}{k/5 - 1} \leq \frac{1}{4}$$

When (*) holds we will show that limes superior of inifinitesmal difference quotients of f are everywhere $+\infty$, and similarly limes inferior are everywhere $-\infty$. To prove this we proceed as follows: To each $x > 0$ and every positive integer n we find the integer $\{k^n x\}$ such that

$$k^n x = \{k^n x\} + q_n \quad \text{where} \quad -1/2 \leq q_n < 1/2$$

Fix some $m \geq 2$ and put

$$(1) \qquad h_m = \frac{1 - q_m}{k^m}$$

If $n \geq m$ we obtain

$$(2) \qquad k^n(x + h_m) = k^{n-m}(\{k^m x\} + q_m + 1 - q_m) = k^{n-m}(\{k^m x\} + 1)$$

When $\{k^m x\}$ is an odd integer we take $n = m$ above and get

$$\cos(k^m \pi(x + h_m)) = \cos \pi(\{k^m x\} + 1) = 1$$

At the same time

$$\cos(k^m \pi x) = \cos \pi(\{k^m x\} + q_m) = \cos(\pi(1 + q_m)) \leq 0$$

where the last equality follows since the absolute value $|q_m| \leq 1/2$. Hence

$$(3) \qquad \cos(k^m \pi(x + h_m)) - \cos(k^m \pi x) \geq 1$$

Next, consider the partial sum difference

$$D_{m-1}(x) = \sum_{n=1}^{n=m-1} b^n \cdot [\cos(k^n \pi(x + h_m)) - \cos(k^n \pi x)]$$

The derivative of the cosine-function is the sine-function whose maximum norm is ≤ 1. Hence Rolle's theorem gives the inequality:

$$|\cos(k^n \pi(x + h_m)) - \cos(k^n \pi x)| \leq k^n \pi h_m \quad \text{for all} \quad 1 \leq n \leq m - 1 \implies$$

$$|D_{m-1}(x)| \leq \pi h_m \cdot \sum_{n=1}^{n=m-1} b^n \cdot k^n = \pi h_m \cdot \frac{(bk)^m - 1}{bk - 1} \leq$$

$$(4) \qquad \pi h_m \cdot \frac{(bk)^m}{bk - 1} = \pi(1 - q_m) \frac{b^m}{bk - 1} \leq \frac{3\pi}{2(bk - 1)} \cdot b^m$$

where the last inequality follows since $|q_m| \leq 1/2$. Next, since the maximum norm of the cosine function is ≤ 1 the triangle inequality gives

$$(5) \qquad \sum_{n=m+1}^{\infty} b^n \cdot [\cos(k^n \pi(x + h_m)) - \cos(k^n \pi x)] \leq 2 \sum_{n=m+1}^{\infty} b^n \cdot = \frac{2}{1 - b} \cdot b^{m+1}$$

A summation over n and the triangle inequality therefore give:

$$f(x+h_m) - fx) \geq b^m - \frac{3\pi}{2(bk-1)} \cdot b^m - \frac{2}{1-b} \cdot b^{m+1} \tag{6}$$

Here $b = 1/5$ and together eith (*) we see (6) gives

$$f(x+h_m) - f(x) \geq \frac{b^m}{4} \tag{7}$$

Inserting h_m from (1) above we obtain

$$\frac{f(x+h_m) - f(x)}{h_m} \geq \frac{(bk)^m}{1-q_m} \geq \frac{1}{6} \cdot (bk)^m \tag{8}$$

where $1 - q_m \geq 1/2$ was used.

Conclusion. Since $bk > 1$ the right hand side tends to $+\infty$ when k increases and at the same time $h_m \to 0$. Following Weierstrass we take some $0 < x < 1$ and get the sequence of integers $\{k^2 x\}$ as $k = 1, 2, \ldots$. It is clear that we find arbitrary large integers k such that $\{k^2 x\}$ is an odd integer and (8) entails that

$$\limsup_{h \to 0} \frac{f(x+h) - f(x)}{h} = +\infty$$

At this stage we leave it to the reader to verify that the limes inferior is $-\infty$ when we instead use integers k such that $\{k^2 x\}$ are even integers. So Weierstrass cosine-series gives a continuous function which fails to have derivatives at every x.

0.4.2 Existence of derivatives.

Weierstrass' example led to find of sufficient conditions in order that a function has derivatives. Lebesgue proved that if $f(x)$ is a continuous and monotone function, i.e. non-increasing or non-decreasing, then f has an ordinary derivative almost everywhere. This is shown in § 1 where we also prove a general result due to Denjoy and Young which applies to an arbitrary real-valued continuous function $f(x)$ defined on some open interval (a, b) and was announced in the introdution. The specific example by Weierstrass is no accident. In fact, later work by Dini, Mazurkiewics and Banach demonstrated that in the family of all continuos functions defined on the unit interval $[0, 1]$, the generic function fails to have derivatives at all points $0 \leq x \leq 1$. From this point of view the discoveries by Lebesgue were invaluable.

0.4.3 Absolute continuity. Let $f(x)$ be a real-valued continuous function on $[0, 1]$. To each sequence $0 = a_0 < a_1 < \ldots < a_N = 1$ we set

$$\mathrm{Var}_f(a_\bullet) = \sum |f(a_{\nu+1}) - f(a_\nu)|$$

Set

$$\rho(a_\bullet) = \max_\nu a_{\nu+1} - a_\nu)$$

One says that f is absolutely continuous if there to every $\epsilon > 0$ exists $\delta > 0$ such that the following impliction hold for all

$$\rho(a_\bullet) < \delta \implies \mathrm{Var}_f(a_\bullet) < \epsilon$$

0.4.4 Theorem. *If f is absolutely continuous then its derivative $f'(x)$ exists almost everywhere and the following hold:*

$$\int_0^1 |f'(x)| < \infty \quad \text{and} \quad f(x) = f(0) + \int_0^x f'(t)\, dt \quad : \quad 0 < x \leq 1$$

0.4.5 Remark. The theorem asserts that an absolutly continuous function is the primitive of its almost everywhere existing derivative. Conversely, if ϕ is an L^1-function on $[0,1]$ in Lebegue's sense, then

$$\Phi(x) = \int_0^x \phi(t)\,dt$$

is absolutely continuous and for every Lebegue point of ϕ, the derivative $\Phi'(x)$ exists and is equal to $\phi(x)$. The proof is given in § xx after we have introduced the space $L^1(0,1)$ of Lebesgue meausurable functions on $[0,1]$.

0.5 Çantor sets.

On the unit interval $[0,1]$ closed Cantor sets arise by a succesive removal of open intervals. A general construction goes as follows: In the first step an open interval $J_1 = (a_1, b_1)$ is removed where $0 < a_1 < b_1 < 1$. In the second step one removes an open interval $J_{12} = (a_{12} < b_{12})$ from $(0, a_1)$ and an open interval $J_{22} = (a_{22}, b_{22})$ from $(b_1, 1)$. One continues and after n steps $1+2+\ldots+2^n$ many disjoint open intervals have been removed and there remains a disjoint union of closed intervals $\{[a_\nu^{(n)}, b_\nu^{(n)}]\colon 1 \leq \nu \leq 2^{n+1}\}$. Here $a_1^{(n)} = 0$ and $b_{2^n}^{(n)} = 1$ while

$$a_\nu^{(n)} < b_\nu^{(n)} < a_{\nu+1}^{(n)}$$

In step $n+1$ one removes 2^{n+1} open subintervals from each of the closed intervals above. The union of all removed open intervals gives an open set U whose closed complement is called a Cantor set. The Cantor set is of null-type arises if the sum of the lengths of the removed open intervals is equal to one which means that the closed complement $\mathcal{C} = [0,1] \setminus U$ is a null-set in Lebesgue's sense.

Lebesgue functions. To every Cantor set of null-type Lebegue constructed a continuous non-decreasing function $L(x)$ where $L(0) = 0$ and $L(1) = 1$ while L stays constant on the removed open intervals. The L-function is found via a limit of piecewise linear functions $\{f_n(x)\}$. Namely, after n steps we have the closed intervals $\{[a_\nu^{(n)}, b_\nu^{(n)}\}\colon 1 \leq \nu \leq 2^{n+1}$ above. Now

$$f_n(x) = (\nu-1)2^{-n-1} + 2^{-n-1} \cdot \frac{x - a_\nu^{(n)}}{b_\nu^{(n)} - a_\nu^{(n)}} \quad : a_\nu^{(n)} \leq x \leq b_\nu^{(n)}$$

while $f_n(x)$ is constant on open intervals which have been removed up to the n:th step. So each f_n is a non-decreasing piecewise linear function. From the construction it is clear that the maximum norms

$$\max_x |f_n(x) - f_{n-1}(x)| \leq 2^{-n}$$

for every n. Hence $\{f_n\}$ converge uniformly to a continuous function which yields the Lebesgue function associated to the Cantor set. Since $\mathcal{C}$ is a null-set the derivative of the Lebesgue function $L(x)$ is zero almost everywhere and yet L is a non-constant function. This leads to an example of a singular measure. More precisely there exists a non-negative and σ-additive measure $\mu_{\mathcal{C}}$ on $[0,1]$ where the mass over an interval $[a,b]$ is the ordinary Riemann integral of L taken over this interval. So Cantor's set-theoretic constructions give a family of Riesz measures on $[0,1]$ which carry their mass on a null-set.

0.5.1 Arithmetic properties. For certain applications one seeks special Cantor sets $\mathcal{C}$ whose associated singular measure $\mu_{\mathcal{C}}$ has special properties. One such property is the behaviour of its Fourier series. In §xx we expose a construction by Salem which shows that for every number $0 < p < 1/2$ there exists a Cantor set $\mathcal{C}$ if null-type such that $\mu_{\mathcal{C}}$ satisfies

$$\Big|\int_0^1 e^{2\pi i n x}\, d\mu_F(x)\Big| \leq C \cdot |n|^{-p}$$

for all n and a constant C.

0.6 Examples from function theory.

Measure theoretic assertions tend to be vague when results only asserts that some "nice property" holds almost everywhere, while the actual null-set on which the property is not described. To give an example we consider a continuous and real-valued function $u(\theta)$ on the unit circle. Define $v(r,\phi)$ when $0 \leq r < 1$ and $0 \leq \phi \leq 2\pi$ by

$$v(r,\phi) = \frac{r}{\pi} \int_0^{2\pi} \frac{\sin(\theta - \phi) \cdot u(\theta)}{1 + r^2 - 2r\cos(\theta - \phi)} \cdot d\theta$$

It means that v is the conjugate harmonic function in the unit disc of the harmonic extension of u. The Brothers Riesz' Theorem asserts that

$$(*) \qquad \lim_{r\to 1} v(r,\phi) = v^*(\phi)$$

exists almost everywhere, i.e. for all ϕ outside a null-set on $[0,2\pi]$. However, the precise description of this null-set in terms of the given continuous function u cannot be attained in general. So one must be content with assertion that radial limits exist almost everywhere.

0.6.1 A boundary value problem Let $f(x,y)$ be a real-valued and continuously differentiable function defined in an open rectangle $\square = \{0 < x < a\} \times \{-1 < y < 1\}$. The first order partial derivatives f'_x and f'_y may satisfy integrability conditions. For a given $p \geq 1$ we impose the L^p-condition:

$$\text{(i)} \qquad \iint_\square \left(|f'_x(x,y)|^p + |f'_y(x,y)|^p\right) dxdy < \infty$$

A result due to to Hardy, Littlewood and Paley asserts that if $p > 2$ then $f(x,y)$ extends to a continuous function on the half-open rectangle $\{0 \leq x < a\} \times \{-1 < y < 1\}$, i.e. the L^p-condition ensures that f extends to define a continuous function on boundary line where $x = 0$. The case when $p = 2$ leads to a more involved situation. The condition that f'_x and f'_y are square integrable does not ensure that f is a bounded function in $\square$. Examples arise via Riemann's conformal mapping theorem which is explained in § X from Chapter VI. However, the L^2-condition in (i) implies that

$$\text{(ii)} \qquad \iint_\square |f(x,y)|^q \, dxdy < \infty \quad \text{hold for all} \quad q > 2$$

There remains to study the eventual existence of limits

$$\text{(iii)} \qquad \lim_{x\to 0} f(x,y) \quad : \quad -1 < y < 1$$

when (i) holds with $p = 2$. When f is harmonic in $\square$ and the L^2-condition holds in (i) precise limit results due to Beurling appear in Special Topics § xx. Here advanced analytic function theory is needed to attain precise measure theoretic results.

0.6.2 Measures in a logarithmic scale. Let E be a closed null-set on the unit circle T. The complement consists of open intervals $\{\omega_\nu = \{e^{i\alpha_\nu} < e^{i\theta} < e^{i\beta_\nu}\}$ and the condition that E us a null-set means that

$$\sum \beta_\nu - \alpha_\nu = 2\pi$$

Next, to each $t > 0$ we have the closet set $E(t)$ of points $e^{i\theta}$ such that

$$\min_{e^{is}\in E} |e^{i\theta} - e^{is}| \leq t$$

Let $\phi_E(t)$ denote the linear Lebesgue measure of $E(t)$. Now one can impose the condition that

$$(*) \qquad \int_0^1 \frac{\phi_E(t)}{t} \, dt < \infty$$

Example show that this is a more restrictive than the sole condition that E is a null-set. It turns out that (*) is a crucial condition in several situations related to analytic function theory. In § xx we prove a result due to Carleson which shows that (*) entails that there exist non-vanishing analytic functions $f(z)$ in the disc algebra $A(D)$ which are zero on E and at the same time can be made differentiable of any order $N \geq 1$. Conversely a result due to Beurling shows that if (*) diverges and $f \in A(D)$ is Hölder continuous of some order > 0 on T, then f cannot be zero on E unless it is identically zero. The proofs rely upon constructions using Herglotz integral and a measure theoretic result which goes as follows: Given the null set E we have the complementary open intervals $\{\omega_\nu\}$. Denote by ℓ_ν the length of ω_ν.

0.6.3 Proposition. *The integral (*) converges if and only if*

$$(**) \qquad \sum \ell_\nu \cdot \log \ell_\nu < \infty$$

Proof. Arrange the intervals so that $\{\ell_1 \geq \ell_2 \geq \ldots\}$ and if N is a non-negative integer we denote by S_N the number of intervals such that $2^{-N-1} < \ell_\nu \leq 2^{-N}$. It is easily seen that (**) converges if and only if

$$\text{(i)} \qquad \sum_{n=0}^{\infty} 2^{-N} \cdot N \cdot S_N < \infty$$

For each $t > 0$ we set

$$\text{(ii)} \qquad \psi(t) = \sum_{\ell_\nu \leq t} \ell_\nu$$

It is clear that

$$\text{(iii)} \qquad 2^{-N-1} S_N \leq \psi(2^{-N}) - \psi(2^{-N-1}) \leq 2^{-N} S_N$$

Hence (i) converges if and only if

$$\text{(iii)} \qquad \sum_{n=0}^{\infty} N \cdot (\psi(2^{-N}) - \psi(2^{-N-1}))$$

Abel's partial sum formula entails that (iii) converges if and only if

$$\text{(iv)} \qquad \sum_{n=0}^{\infty} \psi(2^{-N}) < \infty$$

Moreover, the reader can check that (iv) holds if an only if the integral

$$\text{(v)} \qquad J = \int_0^1 \frac{\psi(t)}{t}\, dt < \infty$$

To finish the proof there remains to show that the convergence of (*) and (v) are equivalent. One direction is easy for if $t > 0$ and $\ell_\nu \leq t$ then ω_ν is contained in $E(2t)$. It follows that

$$\psi(t) \leq \phi_E(2t)$$

Using the variable subsitution $t \mapsto 2t$ we conclude that the convergence of (*) entails that of (v). To get the converse we shall majorize $\phi_E(t)$. With $t > 0$ kept fixed we take an integer $N \geq 0$ and denote by F_N the union of ω-intervals such that $2^N t < \ell_\nu \leq 2^{N+1} t$ which gives

$$\text{(vi)} \qquad \phi_E(t) \leq \psi(t) + \sum_{N=0}^{\infty} |F_N \cap E(t)|_1$$

Notice that $|\omega \cap E(t)| \leq t$ for each $\omega \in F_N$. Hence the last sum in (vi) is majorized by

$$t \cdot \sum_{N=0}^{\infty} \psi(2^{N+1}t) - \psi(2^N t)$$

The summation only extends as long as $2^N \leq 2\pi$ and is therefore taken over a finite set of integers whose sum obviously is majorized by $(2\pi - \psi(t))t$. We conclude that

$$\phi_E(t) \leq (1-t)\psi(t) + 2\pi t \leq \psi(t) + 2\pi t$$

and then the convergence in (v) entails that of (*).

0.7 Rational series

We shall consider series of the form:

$$(*) \qquad \sum \frac{A_\nu}{z - a_\nu}$$

Before general measure theory was developed such series were studied by Poincaré, Goursat and Pringheim who used them to construct functions via lacunary expansions. The general theory about the series (*) is foremost due Emile Borel who introduced a class of generalised analytic functions by extending Cauchy's concept *de fonction monogène*. A full account of Borel's work would lead us too far so the interested reader may consult his text-book [Borel] for details. Non-trivial situations arise when the sequence $\{a_\nu\}$ is everywhere dense in an open subset Ω of $\mathbf{C}$, or even in the whole complex plane.

0.7.1 A special case. Suppose there exists some $\gamma > 1/2$ such that

$$\sum |A_\nu| \cdot \nu^\gamma < \infty \tag{1}$$

For a given $\ell > 0$ we consider the open discs $\{D_\nu\}$ of radius $\ell \cdot \nu^{-\gamma}$ centered at the points $\{a_\nu\}$. The 2-dimensional area of the domain given by the union of there discs is majorized by:

$$\pi \cdot \ell^2 \cdot \sum \nu^{-2\gamma} \tag{2}$$

Let $C_\ell = \mathbf{C} \setminus \cup D_\nu$ be the closed complement. These sets increase as $\ell \to 0$ and (2) shows that the complement of $\cup C_\ell$ has planar measure zero, i.e. it is a null-set in the complex plane. Next, if $z \in C_\ell$ the series (*) is absolutely convergent since

$$\frac{1}{|z - a_\nu|} \leq \ell^{-1} \cdot \nu^\gamma$$

hold for each ν and (1) gives therefore the absolute convergence

$$\sum \frac{|A_\nu|}{|z - a_\nu|} \leq \ell^{-1}$$

When the sequence $\{a_\nu\}$ is everywhere dense the interior of C_ℓ is empty so the resulting function

$$f(z) = \sum \frac{A_\nu}{z - a_\nu} \tag{*}$$

is defined on every closed set C_ℓ so it cannot be regarded as an analytic function. But f is the limit of meromorphic functions

$$f_m(z) = \sum_{\nu=1}^{\nu=m} \frac{A_\nu}{z - a_\nu} \tag{*}$$

where the convergence holds uniformly on C_ℓ for every fixed ν and one may therefore expect some analyticity properties of f. Let us say that f is of quasi-analytic type if it cannot vanish identically on any Jordan arc γ situated in C_ℓ for some $\ell > 0$. The condition for f to be quasi-analytic depends on $\{A_\nu\}$. Even if $\{A_\nu\}$ decreases quite rapidly there may exist non-quasi-analytic series. Examples where

$$|A_\nu \leq e^{-n^\alpha} \quad : \quad 0 < \alpha < 1/2$$

are given in [Denjoy 1922]. However, if $\{A_\nu\}$ decreases sufficiently rapidly then quasi-analyticity holds. The following sufficiency result was proved by Carleman in [Carleman 1922):

0.7.2 Theorem. *Suppose there exists $\epsilon > 0$ such that*

$$|A_n| \leq e^{-(\gamma+\epsilon)\cdot n \cdot \log n} \quad : \quad n \geq 1$$

Then $f(z)$ is quasi-analytic in C_ℓ for every $\ell > 0$.

The proof employs analytic function theory and is given in § XX from Chapter III. Carleman's book [xxx] contains several examples including necessity conditions for certain classes of series in (*). The interested reader should also consult [Wolff] where J. Wolff constructs series as above when $\{a_\nu\}$ is everywhere dense and (*) converges at every point outside this denumerable set.

Another case. Consider the series (*) under the condition that

$$\sum_{\nu=1}^{\infty} \sqrt{|A_\nu|} < \infty \tag{***}$$

Let us take a dense set $\{a_\nu\}$ in some open set Ω and an analytic Jordan curve Γ given by the image under a bijective and real analytic function $t \mapsto z(t)$ from the unit interval $[0,1]$. When all the a-points are outside Γ we can define the partial sums

$$S_N^\Gamma(t) = \sum_{\nu=1}^{\nu=N} \frac{A_\nu}{z(t) - a_\nu} \tag{1}$$

We say that the infinite series (*) converges uniformly on Γ if $\{S_N^\Gamma(t)\}_1^\infty$ converges uniformly on $[0,1]$. Borel proved that (***) implies that this uniform convergence holds on *almost every closed line segment* in the complex plane. To be precise, if we represent a line segment by its two end-points (z_0, z_1) then there exists a nullset $\mathcal{N}$ in $\mathbf{C}^2$ such that the series

$$t \mapsto \sum \frac{A_\nu}{tz_0 + (1-t)z_1 - a_\nu}$$

converges uniformly for $0 \leq t \leq 1$ whenever (z_0, z_1) is outside $\mathcal{N}$. Next, relax the condition of uniform convergence on $[0,1]$ to the weaker condition that $\{S_N^\Gamma(t)\}$ converge pointwise for all t outside a null set on $[0,1]$. With this relaxed notion of convergence Borel proved the following:

0.7.3 Theorem. *Assume that* $\{A_\nu\}$ *satisfies*

$$\sum_{\nu=1}^{\infty} |A_\nu|^{\frac{2}{3}} < \infty$$

Then the partial sum sequence $\{S_N^\Gamma(t)\}$ *converges almost every along every real-analytic Jordan arc* Γ *in* $\mathbf{C}$.

The results above illustrate that the study of series (*) is a rich subject with wide range of applications and many important problems remain unsolved. See [Borcea. et.al.] for an account.

I. Measure theory on the real line.

Introduction. The modern theory of integration started in 1894 when Stieltjes published the article *Recherches sur les fractions continues* which contains a wealth of new ideas; among others, a new conceptual integral which was used to study *moment problems.* Given a sequence of non-negative real numbers $\{c_n\}_0^\infty$ one asks if there exists a non decreasing function $f(t)$ defined on the non-negative real axis $t \geq 0$ such that

$$(*) \qquad c_n = \int_0^\infty t^n \cdot df(t) \quad \text{hold for all } n = 0, 1, \ldots$$

To analyze this problem Stieltjes constructed integrals of the form:

$$(**) \qquad \int_a^b g(x) \cdot df(x)$$

where $f(x)$ is a continuous and non-decreasing function and $g(x)$ is a continuous function on the closed interval $[a, b]$. Stieltjes defined the integral (**) as the limit of sums:

$$(***) \qquad \sum g(x_\nu) \cdot \left[f(x_{\nu+1} - f(x_\nu)\right]$$

where $a = x_0 < \ldots < x_N = b$ and $\max(x_{\nu+1} - x_\nu)$ tends to zero. The *uniform continuity* of g on $[a, b]$ gives a limit exactly as for the ordinary Riemann integral. Namely, consider the modulus of continuity:

$$\omega_g(\delta) = \max_{x_1, x_2} |g(x_1) - g(x_2)| \quad : \text{maximum taken over pairs } |x_2 - x_1| \leq \delta$$

Put $M = f(b) - f(a)$. If $\delta > 0$ and $x_{\nu+1} - x_\nu \leq \delta$ hold for every ν in the partition which defines (***), then Stieltjes proved that this sum differs from a limit by a quantity which is $\leq M \cdot \omega_g(\delta)$. Hence the integral (**) exists since uniform continuity of g entails that

$$\lim_{\delta \to 0} \omega_g(\delta) = 0$$

Remark. Stieltjes's beautiful discoveries about the moment problem will not be treated in these notes. We remark only that Stieltjes found a necessary and sufficient condition on the sequence $\{c_n\}$ in order that there exists a function f such that (*) above holds expressed by a condition on the series of continued fractions associated to $\{c_n\}$.

Exercise. Let f be a C^1-function, i.e. the derivative $f(x)$ exists as a continuous function. Show that (**) is equal to the ordinary Riemann integral

$$\int_a^b g(x) f'(x) \cdot dx$$

If g also is a C^1-function then partial integration gives

$$(1) \qquad \int_a^b g(x) f'(x) = g(b)f(b) - g(a)f(a) \int_a^b f(x) g'(x) dx$$

Verify that the partial integration formula (1) remains valid for Stieltjes' integrals, i.e. if f and g are continuous and non-decreasing on $[a, b]$ then partial summations give the equality:

$$\int_a^b f(x) \cdot dg(x) + \int_a^b f(x) \cdot dg(x) = f(b)g(b) - f(a)g(a)$$

0.1 Functions of bounded variation. A continuous function $f(x)$ on $[a, b]$ has a bounded variation if there exists a constant M such that

$$\sum_{\nu=0}^{\nu=N-1} \big|f(x_{\nu+1}) - f(x_\nu)\big| \leq M$$

for all partitions $a = x_0 < x_1 < \ldots < x_N = b$. The Stieltjes integrals

$$\int_a^b g(x) \cdot df(x)$$

are again defined for every continuous function $g(x)$. To prove this one uses the following result:

0.2 Decomposition Lemma. *Let f have a bounded variation. Then there exists a unique pair of continuous functions (f_*, f^*) where f^* is non-decreasing and f_* is non-increasing such that*

$$f(x) = f^*(x) - f_*(x) \quad \text{where} \quad f^*(a) = f(a) \quad \text{and } f_*(a) = 0$$

Exercise. Prove this result. The hint is to define $f^*(x)$ for every $a \leq x \leq b$ by

$$f^*(x) = f(a) + \max \sum f(\xi_\nu) - f(\eta_\nu)$$

where the maximum is taken over sequences $0 \leq \eta_1 < \xi_1 < \eta_2 < \ldots < \eta_N < \xi_N \leq x$.

0.3 Borel's construction of integrals

.

Stieltjes' integral can be computed in another way using the variation of f over a certain family of subsets of $[a, b]$. This fundamental method was introduced by Emile Borel in 1895. Recall that a bounded and open subset Ω of $\mathbf{R}$ is a unique union of pairwise disjoint open interval $\{(a_\nu, b_\nu)\}$. In general this family is infinite but it is at most denumerable. Let $f(x)$ be continuous and non-decreasing. The variation of f over Ω is defined by

$$\mathrm{Var}_f(\Omega) = \sum_{\nu=1}^{\infty} \big[f(b_\nu) - f(a_\nu)\big]$$

If E is a closed subset of $[a, b]$ we put:

(i) $$\mathrm{Var}_f(\Omega \cap E) = \mathrm{Var}_f(\Omega \cap E) - \mathrm{Var}_f(\Omega \setminus E)$$

Next, consider a continuous function g on $[a, b]$. For each ν we get the following subset of $[a, b]$:

$$S_g[a_\nu, b_\nu) = \{x \quad : \quad a_\nu \leq g(x) < b_\nu\}$$

Since $\{g < b_\nu\}$ is open and $\{g \geq a_\nu\}$ is closed, it follows from (i) that the variation of f is defined over these S_g-sets. Suppose that the range $g([a, b])$ is contained in the interval $[-A, A]$ for some $A > 0$. For every sequence

$$-A \leq \xi_0 < \xi_1 < \ldots < \xi_N = A$$

we set

$$S_*(\xi_\bullet) = \sum \xi_\nu \cdot \mathrm{Var}_f(S_g[\xi_\nu, \xi_{\nu+1}) \quad \text{and} \quad S^*(\xi_\bullet) = \sum \xi_{\nu+1} \cdot \mathrm{Var}_f(S_g[\xi_\nu, \xi_{\nu+1})$$

When $M = f(b) - f(a)$ we see that

(0.3.1) $$S^*(\xi_\bullet) - S_*(\xi_\bullet) \leq M \cdot \max_\nu (\xi_{\nu+1} - \xi_\nu)$$

Hence there exists

(*) $$S = \lim S_*(\xi_\bullet)$$

where the limit is taken over arbitrary partitions of $[-A, A]$ such that

$$\max_\nu \ \xi_{\nu+1} - \xi_\nu \to 0$$

Abel's summation formula entails that the limit (*) is equal to the Stieltjes' integral, i.e.:

$$S = \int_a^b g(x) \cdot df(x) \tag{0.3.2}$$

This equality paves the way to constructions of integrals where the regularity of g can be relaxed. The reason is that Borel's limit is more *robust* as compared with the construction by Stieljes. For let $A > 0$ and consider the class $\mathcal{C}_A$ of continuous functions $g(x)$ on $[a, b]$ whose range is contained in $[-A, A]$. To every ξ-sequence $-A = \xi_0 < \xi_1 < \ldots < \xi_N = A$ we set

$$\delta(\xi_\bullet) = \max_\nu \xi_{\nu+1} - \xi_\nu$$

By (0.3.1) the lower- resp. the over-sums S_* and S^* differ from the limit S in (*) by number majorized by $A \cdot \delta(\xi_\bullet)$. This hold for all g-functions in $\mathcal{C}_A$. So the rate of convergence depends upon $\delta(\xi_\bullet)$ and not upon the chosen g-function in $\mathcal{C}_A$, i.e. the rate of convergence in Borel's construction is *independent* of the modulus of continuity of an individual g-function.

0.4 Borel functions.

The robust limit above enable us to construct integrals where the g-functions need not be continuous. In fact, there exists the integral

$$\int_a^b g(x) df(x) \tag{0.4.1}$$

under the sole assumption that g is a *bounded and Borel measurable function.* This means that $g(x)$ is a function on $[a, b]$ with some bounded range $[-A, A]$ and for each pair of real numbers $\eta < \xi$ the set $\{\eta \leq g < \xi\}$ belongs to the Borel algebra $\mathfrak{B}$ which by definition is the smallest Boolean σ-algebra of subsets of $\mathbf{R}$ containing all half-open intervals. To define (0.4.1) for a bounded Borel function g one first constructs the variation of f over an arbitrary Borel set. Once this is done the existence of limits which give (0.4.1) follow exactly as in the case when g is a continuous function.

0.5 Baire classes.

The class $\mathcal{B}$ of Borel measurable functions arises from increasing families of functions. First $\mathcal{B}_0$ is the family of continuous functions. Next, a function g is of the first Baire class if it is equal to the *pointwise limit* of some sequence of continuous functions. This gives the class $\mathcal{B}_1$. Next we get the class $\mathcal{B}_2$ which consists of pointwise limits of $\mathcal{B}_1$-functions. One proceeds by an induction over positive integers and arrive at the class

$$\mathcal{B}_\infty = \cup B_n$$

However, this does not stop the process via pointwise convergence. The reason is that there exist sequences $\{g_k\}$ of functions in B_w which have a pointwise limit function which does not belong to B_∞, i.e. the limit function does not belong to B_w for a given integer w. To get a family of functions which is stable under pointwise limits one must continue and define $\mathcal{B}$-classes by an induction using *ordinal numbers.* We shall not pursue this any further but refer to the extensive literature about Baire's construction of all Borel functions on the real line. See for example [XXX].

0.6 Monotone functions and their Riesz measures

.

On an interval $[a, b]$ we consider a non-decreasing and continuous functions $f(x)$. We have the σ-algebra $\mathfrak{B}[a, b]$ of Borel sets in $[a, b]$. Consider a denumerable family of disjoint half-open intervals $\{[a_\nu, b_\nu)\}$ on $[a, b]$. Then the positive series

$$\sum f(b_\nu) - f(a_\nu) \tag{1}$$

converges. Next, since $\mathfrak{B}[a,b]$ is the smallest σ-algebra which contains half-open subintervals of $[a,b]$, the convergence in (1) and general set-theoretic arguments imply that the variation of f is defined over every Borel set S, i.e. we obtain a σ-additive measure μ_f on $\mathfrak{B}[a,b]$ where

$$\mu_f(S) = \text{Var}_f(S)$$

holds for every Borel set S in $[a,b]$. More precisely,

$$S \mapsto \mu_f(S) \tag{2}$$

yields a σ-additive map.

A converse result. Suppose that we have a non-negative σ-additive map

$$S \mapsto \mu(S)$$

from $\mathfrak{B}[a,b]$ into the set of non-negative real numbers. Thus, if μ is normalised so that its mass on $[a,b]$ is one, then it is a probability measure on the sample space formed by $[a,b]$ and $\mathfrak{B}[a,b]$ in the sense of abstract measure theory. To μ we associate the non-decreasing function $f(x)$ defined by

$$f(x) = \mu([a,x]) \quad : a \leq x \leq b$$

Exercise. Prove the equality

$$\mu = \mu_f \tag{1}$$

under the assumption that μ has no *discrete point masses*, i.e. that the μ-mass is zero on every singleton set. An equivalent condition is that

$$\mu(\eta, \xi) = \mu([\eta, \xi]) \tag{2}$$

hold for all pairs $\leq \eta < \xi \leq b$. When μ satisfies this condition one says that μ has no atoms or refer to an atomless measure. Hence we can conclude:

0.6.1 Proposition. *There exists a 1-1 correspondence between the class of non-decreasing continuous functions on $[a,b]$ and the class of σ-additive and non-negative measurs on $[a,b]$ without atoms.*

0.6.2 Discrete measures and jump functions. Let $\{x_k\}$ be a sequence of points on $\mathbf{R}$ indexed by positive integers. No special assumption is imposed. For example, the sequence may be some enumeration of all rational numbers. If $\{p_k\}$ is a sequence of positive numbers such that the series $\sum p_k < +\infty$ then we get the jump-fuction

$$s(x) = \sum p_k \cdot H_{x_k}(x) \tag{1}$$

Above we introduced Heaviside functions, i.e. for every real number x_* we define $H_{x_*}(x) = 1$ if $x \geq x_*$ and 0 when $x < x_*$. To the s-function corresponds the discrete measure which assigns the mass p_k at every x_k.

This yields a 1-1 correspondence between non-decreasing jump functions and non-negative discrete measures. A non-trivial result due to Lebesgue and Vitali goes as follows:

0.6.3 Theorem. *Every jump-function $s(x)$ has an ordinary derivative whole value is zero outside a nullset.*

We prove this in § 1.15 below. Let us now consider an arbitrary σ-additive and non-negative measure μ defined on $\mathfrak{B}[a,b]$. It may have point masses. Since the total mass of μ over $[a,b]$ is finite the sum of all mass assigned to the atoms is finite, i.e. we get a discrete part given by

$$\mu_d = \sum p_k \cdot \delta_{x_k}$$

Here $\sum p_k < \infty$ and $\{x_k\}$ is at most a denumerable subset of $[a,b]$. The difference $\mu - \mu_d$ has no atoms and corresponds to a non-decreasing continuos function f as above. So here $\mu - \mu_d = \mu_f$ and one refers to μ_f as the continuous part of μ. It is clear that this decomposition of μ is unique.

0.7 Signed measures.

A continuous function f on $[a,b]$ has a bounded variation if there exists a constant C such that

$$\sum \big|f(\xi_{\nu+1}) - f(\xi_\nu)\big| \leq C$$

for all partitions $a = \xi_0 < \ldots < \xi_N = b$. The smallest number C for which this holds is called the total variation of f and is denoted by $V(f)$. A result due to Ascoli asserts the following:

0.8 Theorem. *Let f be a continuous function with bounded variation. Then there exists a unique pair of non-decreasing functions g, h such that*

$$f(x) = g(x) - h(x)$$

where $g(a) = f(a)$ and $h(a) = 0$.

Exercise. Prove this. Here is the hint is to construct the g-function. Put

$$\phi(x) = f(a) + \max \sum_{\nu 01}^{\nu=N} f(\xi_\nu - f(\eta_\nu)$$

where the maximum is taken over all sequences

$$a < \eta_1 < \xi_1 < \eta_2 < \ldots < \xi_{N-1} < \eta_N < \xi_N \leq x$$

Intuitively, we seek intervals in $[0,x]$ where f increases as much as possible, while eventual intervals where f decrease are omitted. The reader should verify that the ϕ-function is non-decreasing and continuous and that the function

$$x \mapsto f(x) - \phi(x)$$

is non-increasing. Now $g = f(a) + \phi(x)$ and $h = f(x) - \phi(x) - f(a)$. This proves the existence in Theorem 0.8 and the proof of uniqueness is left to the reader.

The signed measure μ_f. The decomposition above gives a signed measure μ_f defined by

$$\mu_f = \mu_g - \mu_h$$

In this decomposition there exists a Borel set S in $[a,b]$ such that

$$\mu_g([a,b]) = \mu_g(S) \quad \text{and} \quad \mu_h([a,b]) = \mu_h([a,b] \setminus S)$$

This means that μ_g and μ_h are orthogonal and one writes

$$\mu_g \perp \mu_h$$

Summing up we have the following

0.9 Theorem. *There exists a 1-1 correspondence between the family of signed Riesz measures μ on $[a,b]$ without atoms and the class of continuous functions $f(x)$ of bounded variation normalised with $f(a) = 0$.*

Let us remark that the results above are special cases of the general decomposition by Hahn which is proved in the section devoted to abstract measure theory.

0.10 Borel-Stieltjes integrals. The previous material shows that if f is a continuous function of bounded variation on $[a,b]$ then there exists an integral

$$(*) \qquad \int_a^b g(x) \cdot d\mu_f(x) = \int_a^b g(x) \cdot df(x)$$

for every bounded Borel function $g(x)$.

0.11 Weak limits of measures

Above we have seen that every signed Riesz measure μ on $[a,b]$ is a unique sum of an atomless measure ν_c and a discrete measure μ_d. It turns out that the class of discrete measures recapture all Riesz measures after suitable passage to the limit. Let us clarify this and begin with the case of non-negative measures. Consider a positive integer N and some N-tuple $a \leq \xi_1 < \ldots < \xi_N \leq b$

and a sequence $\{c_\nu \geq 0\}$ where $\sum c_\nu = 1$ To this we associate the discrete measure μ_N and obtain:

$$\int g \cdot d\mu_N = \sum c_\nu \cdot g(\xi_\nu) \quad \text{for every } g \in C^0[a,b]$$

Suppose now that there exists the limits:

$$(*) \qquad \lim_{N\to\infty} \int g \cdot d\mu_N \quad \text{for all} \quad g \in C^0[a,b]$$

When (*) holds we say that $\{\mu_N\}$ is weakly convergent. Now the following holds:

0.12 Theorem. *Let $\{\mu_N\}$ be weakly convergent. Then there exists a unique non-negative Riesz measure μ such that*

$$\lim_{N\to\infty} \int g \cdot d\mu_N = \int_a^b g \cdot d\mu$$

hold for all $g \in C^0[a,b]$.

Proof. The existence of limits mean that we have a linear map

$$g \mapsto \lim_{N\to\infty} \int g \cdot d\mu_N = \int_a^b g \cdot d\mu$$

on the vector space $C^0[a,b]$. Let us denote it by L. Notice that $g \geq 0$ gives $L(g) \geq 0$ and taking the identity function we have $L(1) = 1$. Now we construct a non-decreasing function $f(x)$ as follows: To each $a < x \leq b$ we consider the family $\mathcal{F}_x$ of continuos functions g such that $0 \leq g \leq 1$ and there exists some $\epsilon > 0$ with $g(y) = 0$ when $y \geq x - \epsilon$. In other words, g has a support given by a compact subset of the half-open interval $[a,x)$. Set

$$f(x) = \max_{g\in\mathcal{F}_x} L(g)$$

Here $f(a) = 0$ and it is obvious that f is non-decreasing. The reader should also verify that f is left continuous, i.e.

$$\lim_{\epsilon\to 0} f(x-\epsilon) = f(x)$$

hold for every $a < x \leq b$. There may exist a set $\{\eta_\nu\}$ where f is discontinuous. Here a jump at η_ν is given by a positive number

$$\rho_\nu = \lim_{\epsilon\to 0} f(\eta_\nu + \epsilon) - f(\eta_\nu)$$

To every such point we have the Heaviside function $H_\nu(x)$ and define the jump function

$$s(x) = \sum \rho_\nu \cdot H_\nu(x)$$

Then $f - s = f_*$ is continuous and we get the Riesz measure

$$\mu = \mu_d + \mu_{f_*}$$

where the discrete part

$$\mu_d = \sum \rho_\nu \cdot \delta_{\eta_\nu}$$

At this stage we leave it to the reader to verify the equality below for every continuous functions $g(x)$.

$$L(g) = \int_a^b g \cdot d\mu$$

Now Theorem 0.12 follows from the Riesz representation theorem for non-negative and continuous linear functionals on $C^0[a,b]$.

Remark. If X^* denotes the dual space of the Banach space $C^0[a,b]$ the sequence $\{\mu_N\}$ converge in the so called weak-star topology to μ. Conversely, let μ be a non-negative Riesz measure on $[a,b]$ of total mass one. We also assume that μ is continuous. So $\mu = \mu_f$ for some non-decreasing and continuous function f on $[a,b]$ with $f(a) = 0$. In this situation we can approximate μ weakly by a sequence of discrete measures as follows: For every positive integer $N \geq 2$ we define the discrete measure

$$\mu_N = \sum_{k=1}^{k=N} p_k \delta_{x_k}$$

where

$$p_k = f(a + \frac{k(b-a)}{N}) - f(a + \frac{(k-1)(b-a)}{N}) \quad \text{and} \quad x_k = a + \frac{k(b-a)}{N})$$

Exercise. Show that the sequence $\{\mu_N\}$ converges weakly to μ_f. The hint is to use uniform continuity of the g-functions in Theorem 0.12.

0.13 Cantor's function. Consider the unit interval $[0,1]$. Let $\mu_1 = \frac{1}{2}[\delta_0 + \delta_1]$. Next, we get

$$\mu_2 = \frac{1}{4}[\delta_0 + \delta_{1/3} + \delta_{2/3} + \delta_1]$$

$$\mu_3 = \frac{1}{8} \cdot [\delta_0 + \delta_{1/9} + \delta_{2/9} + \delta_{3/9} + [\delta_{6/9} + \delta_{7/9} + \delta_{8/9} + \delta_1]$$

and so on where the reader can recognize the inductive construction. Intuitively one removes one third of middle intervals at each step. For every integer $k \geq 4$ we get a discrete measure μ_k which assigns the point mass 2^{-k} at 2^k many points.

0.14 Exercise. Show that the sequence $\{\mu_k\}$ converges weakly to a limit measure μ_f where f is continuous and non-decreasing. The notable fact is that f is constant on many intervals. To begin with it is constant on the middle interval $(1(3, 2/3)$. It is also constant on the intervals $(1/9, 2/9)$ and $(7/9, 8/9)$. The total length on intervals where f is constant becomes

$$(*) \qquad \frac{1}{3} + 2 \cdot \frac{1}{3^2} + \ldots = \frac{1}{3} \cdot \sum_{\nu=0}^{\infty} (2/3)^{\nu} = 1$$

The support of the measure μ_f is the closed complement of the removed intervals above. The construction is remarkable since f has a vanishing derivative in each removed open interval and at the same time (*) holds while f increases from zero to $f(1) = 1$ so the function is not constant. Hence Cantor's function f violates the fundamental theorem of calculus, i.e. it is not recovered by an integral of its derivative. Moreover, there is defined intervals on the closed complement of the removed intervals above.

0.15 Absolutely continuous functions. Let f be a non-decreasing and continuous function on $[0,1]$ with $f(0) = 0$ and $f(1) = 1$. To each $0 < \delta < 1$ we consider open sets which consist of pairwise disjoint open intervals $\{(a_\nu, b_\nu)\}$ where

$$(*) \qquad \sum (b_\nu - a_\nu) < \delta$$

No condition is imposed on the number of these intervals. With δ fixed we set:

$$V_f(\delta) = \max \sum f(b_\nu) - f(a_\nu)$$

where the maximum is taken over all finite families $\{(a_\nu, b_\nu)\}$ for which (*) hold.

0.16 Definition. *The function f is called absolutely continuous if*

$$\lim_{\delta \to 0} V_f(\delta) = 0$$

The space of such functions is denoted by $AC^0[0,1]$.

It turns out that if $f \in AC^0[0,1]$ then it is recaptured from its derivative $f'(x)$. This relies upon the theorem by F. Riesz in the next section which shows that f has an ordinary derivative outside a null set in the sense of Lebesgue. After this one constructs the Lesbegue integral of f' to get a primitive function defined by

$$F(x) = \int_0^x f'(t) \cdot dt$$

The almost everywhere existence of Lebesgue points implies that the derivative of F is equal to that of f almost everywhere and there remains to consider the difference

$$g = f - F$$

Here g belongs to $AC^0[0,1]$ and its derivative is almost everywhere zero. From this one can show that g is identically zero and which gives the equality

(*) $$f(x) = \int_0^x f'(t) \cdot dt$$

In other words, every absolutely continuous function is recovered by its derivative.

0.17 Fubini's theorem. Let $\{f_n\}$ be a sequence of continuous and non-decreasing functions defined on some interval $[a, b]$. We also assume that every f_n is non-negative and that the positive series

$$\sum_{n=1}^{\infty} f_n(x)$$

converges for every x. The sum yields a non-decreasing function $s(x)$. As we shall explain in more detail later on, it follows that $s(x)$ is a sum of a non-decreasing function $s_*(x)$ and a discrete jump-function $s_d(x)$. Moreover, $s(x)$ has a derivative almost everywhere and the same is true for each f_n by Lebesgue's theorem to be proved in the next section. Hence there exists a null-set E_0 such that every f_n and s have derivatives at points outside E_0. In this situation Fubini proved that the sum of derivatives converges, i.e. when x is outside E_0 then

$$s'(x) = \lim_{N \to \infty} \sum_{n=1}^{n=N} f_n'(x)$$

The proof of Fubini's theorem occurs at the end of the next section under the headline Exercises and Examples.

0.18 Points of density.

Recall from Calculus that a bounded and open set U is the disjoint union of interval $\{(a_n, b_n)\}$. Its Lebesgue measure is defined by $\sum (b_n - a_n)$ and denoted by $|U|$. A subset S of the real line is a null-set if for every $\epsilon > 0$ there exists an open set U such that $S \subset U$ and $|U| < \epsilon$. One shows easily that every denumerable union of null-sets again is a null-set. Next, let E be a compact subset of $\mathbf{R}$. To each $\delta > 0$ we get the open set

$$E_\delta = \{x : \operatorname{dist}(x, E) < \delta\}$$

These sets decrease with δ and the Lebesgue measure of E is defined to be

(i) $$|E| = \lim_{\delta \to 0} |E_\delta|$$

Exercise. Show that a compact set E is a null-set if and only if (i) above is zero. The hint is that whenever U is an open set which contains E, then there exists $\delta > 0$ such that $E_\delta \subset U$.

Next, let E be compact and consider a point $x \in E$. It is called a point of density in E if:

(*) $$\lim_{h+k \to 0} \frac{\big|E \cap [x-k, x+h]\big|}{h+k} = 0$$

where h, k both tend to zero under the sole assumption that $h + k \to 0$.

0.19 Theorem. *Almost every $x \in E$ is a point of density.*

This result was established by Lebesgue. A simplification of the original proof was later found by Vitali using a general covering lemma which we announce below while the proof of Theorem 0.19 is postponed until § 0.23.

0.20 Vitali's covering theorem. Let F be an arbitrary bounded subset of $\mathbf{R}$. Let $\mathcal{V}$ be a family of open intervals with the properties: To every $x \in F$ and $\delta > 0$ there exists an interval $\omega \in \mathcal{V}$ such that $x \in \omega$ and $|\omega| < \delta$. In addition we assume that $\omega \subset (-A, A)$ hold for some $A > 0$ which is chosen so large that F is contained in a compact subset of $(-A, A)$. If $\omega = (a, b)$ is an interval we denote by $3 \cdot \omega$ the enlarged interval $(a - \ell, b + \ell)$ where $\ell = b - a$ is the length of ω so that $3 \cdot \omega$ has length 3ℓ. With this notation one has:

0.21 Theorem. *There exists a sequence of pairwise disjoint intervals $\{\omega_n\}$ in $\mathcal{V}$ which covers F in Vitali's sense which means that for every positive integer N one has the inclusion*

$$(*) \qquad F \subset \bar{\omega}_1 \cup \ldots \cup \bar{\omega}_N \bigcup_{n>N} 3 \cdot \omega_n$$

Proof. Put

$$\delta_1 = \sup_{\omega \in \mathcal{V}} |\omega|$$

Choose ω_1 with $|\omega_1| > 2\delta_1/3$. Let $\mathcal{V}_1$ be the subfamily of intervals $\omega \in \mathcal{V}$ which have empty intersection with the closed interval $\bar{\omega}_1$. Put

$$\delta_2 = \sup_{\omega \in \mathcal{V}_1} |\omega|$$

Choose $\omega_2 \in \mathcal{V}_1$ with $|\omega_2| > 2\delta_2/3$. In the next step $\mathcal{V}_2$ is the subfamily of intervals in $\mathcal{V}$ which have empty intersection with $\bar{\omega}_1 \cup \bar{\omega}_2$ and

$$\delta_3 = \sup_{\omega \in \mathcal{V}_2} |\omega|$$

Choose $\omega_3 \in \mathcal{V}_2$ with $|\omega_3| > 2\delta_3/3$. We continue in this way and get a sequence $\{\omega_n\}$ where the intervals by construction are pairwise disjoint. At this stage we leave it to the reader to verify that the sequence $\{\omega_n\}$ satisfies (*) in Theorem 0.21 for every positive integer N.

0.22 Exercise Since $\{\omega_\nu\}$ is a sequence of disjoint intervals which all are contained in a bounded set, it follows that

$$\sum \delta_n < \infty$$

Use this convergence and the inclusions in (*) for every positive integer N to show that the outer measure of F is majorized by:

$$|F|^* \leq \sum \delta_n$$

0.23 Proof of Theorem 0.19. For each integer $n \geq 2$ we denote by $E(n)$ the subset of E which consists of points such that

$$\text{(i)} \qquad \liminf_{h+k \to 0} \frac{\big|E \cap (x-k, x+h)\big|}{h+k} \leq 1 - \frac{1}{n}$$

Let $\mathcal{L}(E)$ be the set of points of density in E. It is clear that

$$E \setminus \mathcal{L}(E) = \bigcup_{n \geq 2} E(n)$$

We conclude that $E \setminus \mathcal{L}(E)$ is a null-set if $E(n)$ are null-sets for every $n \geq 2$. To prove this for a given n we proceed as follows: For each $\epsilon > 0$ we can choose $\delta > 0$ such that

$$\text{(ii)} \qquad |E_\delta \setminus E| < \frac{\epsilon}{n}$$

If $x \in E(N)$ the definition of limes inferior means that we can choose arbitrary small open intervals ω which contain x while $\omega \subset E_\delta$ and at the same time

$$\text{(iii)} \qquad \frac{\big|E \cap \omega\big|}{|\omega|} \leq 1 - \frac{1}{n}$$

It is clear that the family of such intervals is a Vitali covering of $E(n)$. Hence we find a sequence $\{\omega_\nu\}$ from Theorem XX and by the remark in XX the outer measure

$$|E(n)|^* \leq \sum |\omega_\nu|$$

Next, for each fixed ν we have

$$|\omega_\nu| = |\omega_\nu \cap E| + |\omega_\nu \cap (E_\delta \setminus E)|$$

Since $\{\omega_\nu\}$ are disjoint and (iii) above hold for every ν it follows that

$$\sum |\omega_\nu| \leq (1 - \frac{1}{n}) \cdot \sum |\omega_\nu| + |E_\delta \setminus E|$$

Hence (ii) gives

$$\sum |\omega_\nu| \leq \epsilon$$

Since ϵ was arbitrary we conclude that $E(n)$ is a null-set and Lebegue's theorem is proved.

0.24 Exercise. Use Lebegue's theorem to show that if F is a set such that there exists a constant $c < 1$ for which

$$\big|F \cap (a,b)\big|^* \leq c(b-a)$$

hold for every interval (a,b), then F is a null-set.

1.B. Derivatives of functions

In the text-book *Théorie de l'intégration* from 1904, Lebesgue proved that a monotone function defined in a real interval has an ordinary derivative outside a null-set. For an arbitrary continuous function a more general result was discovered by M. Young and Denjoy which goes as follows: Let $f(x)$ be a real-valued continuous function defined on some interval (a,b). For each $a<x<b$ we set

$$D^*(x)=\limsup_{h+k\to 0}\frac{f(x+h)-f(x-k)}{h+k}$$

where h and k are positive when we pass to the limes superior. Similarly

$$D_*(x)=\liminf_{h+k\to 0}\frac{f(x+h)-f(x-k)}{h+k}$$

1.1 Theorem. *Outside a (possibly empty) null-set E of (a,b) the following two possibilities occur for each $x\in(a,b)\setminus E$: Either there exists a common finite limit*

$$(*)\qquad D^*(x)=D_*(x)$$

Or else one has

$$(**)\qquad D^*(x)=+\infty\quad\text{and}\quad D_*(x)=-\infty$$

Remark. Above the pair $h,k)$ tends to zero under the sole condition that $h+k\to 0$. We can take $k=0$ or $h=0$ and consider one-sided limits:

$$\text{(i)}\qquad D^+(x)=\limsup_{h\to 0}\frac{f(x+h)-f(x)}{h}\quad\text{and}\quad d^+(x)=\liminf_{h\to 0}\frac{f(x+h)-f(x)}{h}$$

$$\text{(ii)}\qquad D_+(x)=\limsup_{k\to 0}\frac{f(x)-f(x-k)}{k}\quad\text{and}\quad d_+(x)=\liminf_{k\to 0}\frac{f(x)-f(x-k)}{k}$$

With these notations it is clear that:

$$D_*(x)\le d^+(x)\le D^+(x)\le D^*(x)$$

So the equality $D_*(x)=D^*(x)$ entails that f has an ordinary right derivative. Since

$$D_*(x)\le d_+(x)\le D_+(x)\le D^*(x)$$

we conclude that if (*) holds in the theorem, then f has a ordinary derivative at x. If (**) occurs at a point x, then the graph of f close to x is steep but may also change sign in a small interval around such a point. Take for example $x=0$ and let

$$f(x)=\sqrt{x}\quad\text{when}\quad x>0\quad:\quad f(x)=\sqrt{-x}\quad\text{when}\quad x<0$$

With $k=0$ and $h>0$ we see that $D^*(0)=+\infty$ and with $h=0$ and $k>0$ we see that $D_*(=)=-\infty$. Next, recall Weierstrass' construction of a continuous function $f(x)$ which fails to have an ordinary derivative at every point in the interval (a,b). The Denjoy-Young theorem shows that such a continuous function has a "turbulent" graph where $D^*(x)=+\infty$ and $D_*(x)=-\infty$ both hold for all x outside a null-set.

1.2 The case of monotone functions. If the continuous function f is non-increasing or non-decreasing, then case (**) cannot occur. So Theorem 1.1 implies that a monotone continuous function has an ordinary derivative almost everywhere. Concerning Theorem 1.1 its proof relies upon Lebesgue's result about points of density. The interested reader may consult Riesz' plenary talk at the IMU-congress in Zürich (1932) for a historic account about derivatives of functions on the real line and the subsequent proof follows Riesz' presentation in [ibid] closely.

1.3 Forward Riesz intervals. Let $g(x)$ be a real-valued and continuous function defined on some open interval (a,b), The forward Riesz set $\mathcal{F}_g$ consists of all points $a < x < b$ for which there exists some $x < y \leq b$ such that

$$(*) \qquad g(x) < g(y)$$

If g is non-decreasing then $\mathcal{F}_g = \emptyset$. Excluding this case, continuity entails that $\mathcal{F}_g$ is an open subset of (a,b) and hence a disjoint union of intervals

$$(1) \qquad \mathcal{F}_g = \cup\,(\alpha_\nu, \beta_\nu)$$

Each interval in (1) is called a *forward Riesz interval* of g. It may occur that some interval is of the form (α, b) i.e. b is a right end-point. Similarly a can be a left end-point. For example, if g from the start is strictly increasing then $\mathcal{F}_g = (a,b)$.

1.4 Proposition *For each forward Riesz interval* (α, β) *one has*

$$(*) \qquad g(\beta) = \max_{\alpha \leq x \leq \beta} g(x)$$

Proof. Assume the contrary. This gives some maximum point $\alpha \leq x^* < \beta$ for the g-function on the closed interval $[\alpha, \beta]$. Now $x^* \in \mathcal{F}_g$ which means that

$$\exists\, y > x^* \quad \text{and} \quad g(x^*) > g(y)$$

Since x^* is a maximum point over $[\alpha, \beta]$ we must have $y > \beta$. But then $\beta \in \mathcal{F}_g$ which is impossible since β was a boundary point of the open set $\mathcal{F}_g$.

1.5 Backward Riesz intervals Put

$$\mathcal{B}_+ g = \{a < x < b \colon \; \exists\, a < y < x \quad : \quad g(y) > g(x)\}$$

Again $\mathcal{B}_g$ is open and hence a disjoint union of open intervals (c_ν, d_ν). They are called backward Riesz intervals. By similar reasoning as above one shows that if (c,d) is a backward Riesz interval then

$$(**) \qquad g(c) = \max_{c \leq x \leq d} g(x)$$

1.6 A study of monotone functions.

Let $f(x)$ be a continuous and *non-decreasing* function on $[a,b]$. To each $a < x < b$ we set

$$(1) \qquad D^+(x) = \limsup_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

where limes superior is taken as $h > 0$ decrease to zero. The function $x \mapsto D^+(x)$ takes values in $[0, +\infty]$.

1.7 Proposition. *For each positive number C the following set-theoretic inclusion holds:*

$$\{D^+(x) > C\} \subset \mathcal{F}_g \quad \text{where} \quad g(x) = f(x) - Cx$$

Proof. Suppose that $D^+(x) > C$ for some $a < x < b$. The definition of limes superior gives some $y > x$ such that

$$(i) \qquad \frac{f(y) - f(x)}{y - x} > C$$

Then $g(y) - g(x) = f(y) - f(x) - C(y-x) > 0$ and hence $x \in \mathcal{F}_g$.

1.8 Proposition *For every $C > 0$ the outer Lebesgue measure of the set $\{D^+ > C\}$ satisfies the inequality*

$$\big|\{D^+ > C\}\big|^* \leq \frac{f(b) - f(a)}{C}$$

Proof. With $g(x) = f(x) - Cx$ as above one has an interval decomposition $\mathcal{F}_g = \cup\,(\alpha_\nu, \beta_\nu)$ and the inclusion from Proposition 1.4 gives

$$|\{D^+ > C\}|^* \leq \sum (\beta_\nu - \alpha_\nu) \tag{1}$$

Apply Proposition 1.4 to the forward Riesz intervals of g. This gives for every ν:

$$f(\alpha_\nu) - C \cdot \alpha_\nu \leq f(\beta_\nu) - C \cdot \beta_\nu$$

Rewriting the last inequality we get

$$C(\beta_\nu - \alpha_\nu) \leq f(\beta_\nu) - f(\alpha_\nu)$$

Taking the sum over all ν we obtain

$$C \cdot \sum (\beta_\nu - \alpha_\nu) \leq \sum f(\beta_\nu) - f(\alpha_\nu) \leq f(b) - f(a) \tag{2}$$

where the last inequality holds since f is non-decreasing. Now (1) and (2) give the requested inequality in Proposition 1.8.

1.9 The d_+-function. To each $a < x < b$ we put

$$d_+(x) = \liminf_{k \to 0} \frac{f(x) - f(x-k)}{k}$$

Let $c > 0$ and set

$$h(x) = f(x) - c$$

A similar reasoning as in Proposition 1.4 gives the inclusion

$$\{d_+ < c\} \subset \mathcal{B}_h$$

where the right hand side is the backward Riesz set from 1.5.

1.10 Some inequalitites. Consider a pair $0 < c < C$ and the intersection

$$E = \{d_+ < c\} \cap \{D^+ > C\}$$

Now (1.9) gives the inclusion

$$E \subset \{D^+ > C\} \cap \mathcal{B}_h \tag{i}$$

Let $\{(\alpha_\nu, \beta_\nu)\}$ be the interval decomposition of the open set $\mathcal{B}_h$. For each ν we consider the restriction of $g(x) = f(x) - Cx$ to the interval (α_ν, β_ν) and Proposition 1.8 gives the inequality

$$|\{D^+ > C\} \cap (\alpha_\nu, \beta_\nu)\,|^* \leq \frac{f(\beta_\nu) - f(\alpha_\nu)}{C} \tag{ii}$$

Since (α_ν, β_ν) is a backward Riesz interval of $f(x) - c$ we have $f(\beta_\nu) - f(\alpha_\nu) \leq c(\beta_\nu - \alpha_\nu)$. Hence (i) gives:

$$|\{D^+ > C\} \cap (\alpha_\nu, \beta_\nu)\,|^* \leq \frac{c}{C} \cdot (\beta_\nu - \alpha_\nu) \tag{iii}$$

Since the backward Riesz intervals are disjoint a summation over ν and the inclusion (i) give:

$$|\{D^+ > C\} \cap \{d_+ < c\}|^* \leq \frac{c}{C}(b - a) \tag{*}$$

1.11 Proof of Lebesgue's theorem. The function f restricts to a non-decreasing function on an arbitrary open subinterval (a_*, b_*) of (a, b) and since both D^+ and d_+ are constructed by limits close to a point we get the same inequality as in (*) above, i.e. one has the inequality

$$|\{d_+ < c\} \cap \{D^+ > C\} \cap (a_*, b_*)|^* \leq \frac{c}{C} \cdot (b_* - a_*)$$

Now the criterion from §XX implies that $\{d_+ < c\} \cap \{D^+ > C\}$ is a null-set. Apply this for pairs $c = q < r = C$ where q, r are rational numbers. Since a denumerable union of null-sets is a null-set we conclude that the equality

$$\text{(i)} \qquad d_+(x) = D^+(x)$$

holds almost everywhere. In the same way one proves that the equality

$$\text{(i)} \qquad d^+(x) = D_+(x) \quad \text{holds almost everywhere}$$

Finally, it is obvious that when (i-ii) hold then f has an ordinary derivative which proves Lebegue's theorem that every monotone function has a derivative almost everywhere.

1.12 An extension of Lebesgue's theorem. Let f be a continuous function on the closed unit interval $[0, 1]$. Suppose that E is a measurable subset of $(0, 1)$ such that the restriction of f to E is non-decreasing. Removing an eventual zero set we also assume that $E = \mathcal{L}(E)$, i.e. every $x \in E$ is a point of density for E as explained in § XX. Using exactly the same methods as above it follows that there is a (possibly empty) null-set $S \subset E$, there exists a derivative at every $x \in E$ in the sense that

$$\text{(1)} \qquad \lim_{h+k \to 0} \frac{f(x+h) - f(x-k)}{h+k} = f'_E(x)$$

exists for each $x \in E \setminus S$ where the limit is restricted in the sense that $x + h$ and $x + k$ belong to E during the passage to $h + k \to 0$. But since x is a point of density (1) holds without this restriction, i.e. $f'_E(x)$ gives an ordinary derivative of f. Let us supply the details for this assertion. We may take $x = 0$ and replacing f by $f - f'_E(0)x - f(0)$ we can assume that $f'_E(0) = f(0) = 0$. Next, let $0 < \epsilon < 1/4$ which gives some $\delta > 0$ such that if $0 < x < \delta$ and $x \in E$ then

$$f(x) \leq \epsilon \cdot x$$

At the same time the density condition entails that if δ is small enough then

$$|E \cap (-x, x)| \geq 2x(1 - \epsilon) \quad : \quad 0 < x < \delta$$

If we consider some $0 < x < \delta/2$ we see that (xx) implies the interval $(x + 4\epsilon \cdot x, x)$ must intersect E and if $x^* \in E$ is in this interval we get

$$f(x) \leq f(x^*) \leq \epsilon \cdot x^* \leq \epsilon \cdot 2x$$

Since $\epsilon > 0$ this proves that $D^+(0) = 0$ and in the same way the reader can verify that the right derivative at $x = 0$ vanishes.

1.13 Proof of Theorem 1.1

For each non-negative integer $n = 0, 1, 2, \ldots$ and every rational number $r \in (a, b)$ we denote by $E_{n,r}$ the set of all $r < x < b$ such that

$$\frac{f(x) - f(\xi)}{x - \xi} > -n \quad : \quad r < \xi < x$$

Exercise. Show the set-theoretic inclusion

$$\{D_*(x) > -\infty\} \subset \bigcup E_{n,r}$$

where the union is taken over all $n \geq 0$ and every rational number $a < r < b$.

1.14 Proposition. *For each pair (n, r) the equality*

$$D^*(x) = D_*(x)$$

holds almost everywhere in the measurable set $E_{n,r}$.

Proof. Replacing the interval (a, b) by (r, b) and f by $f(x-r) + nx$ we can assume that $r = n = 0$ and now $E_{0,0} \subset (0, b - r)$ where the restriction of f to this measurable set is monotone, i.e.

$$0 < \xi < x \implies f(x) > f(\xi)$$

holds for every pair $\xi < x$ in $E_{0,0}$. To simplify notations we set $E = E_{0,0}$. Let E_* be the set of density for E as defined in XX and recall from XX that $E \setminus E_*$ is a null-set. Ignoring this null-set we consider the restriction of f to E_* which again is a non-decreasing function. The extended version from 1.12 of Lebegue's theorem applies and shows that after removing another null-set from E_* if necessary, then the limit below exists for each $x \in E_*$:

$$(*) \qquad D(x) = \lim_{h+k\to 0} \frac{f(x+h) - f(x-k)}{h+k}$$

In the same way one proves that if a null-set is removed from the set

$$D^*(x) = +\infty\}$$

then f has an ordinary derivative so that $D^*(x) = D_*(x)$. This finishes the proof of the Denjoy-Young theorem.

1.15 Examples and Exercises.

Above we have studied monotone continuous functions. There also exist non-decreasing jump functions which arise as follows: Let $\{\xi_n\}$ be a sequence of real numbers in $(0,1)$. They are not ordered and may give a dense set. For example, we can take some enumeration of all rational numbers in $(0,1)$. Next, let $\{\delta_n\}$ be a sequence of positive numbers such that $\sum \delta_n < \infty$. To each n we get the jump function $H_n(x)$ where

$$H_n(x) = 0 \quad : \quad x < \xi_n \quad \text{and} \quad H_n(x) = \delta_n \quad : \quad x \geq \xi_n$$

Now

$$s(x) = \sum H_n(x)$$

is a non-decreasing function which has jump-discontinuites at each ξ_n.

Exercise. Show that s is pointwise continuous at every x outside the set $\{\xi_n\}$, i.e. show that if $\epsilon > 0$ then there exists $\delta > 0$ such that

$$s(x+\delta) < s(x) + \epsilon \quad \text{and} \quad s(x-\delta) > s(x) - \epsilon$$

Less evident is the following:

1.16 Theorem. *$s(x)$ has an ordinary derivative which is equal to zero almost everywhere.*

Proof. Let $\alpha > 0$ and denote by E be the subset of $(0,1)$ which consists of numbers $0 < x < 1$ such that

$$\limsup_{h+k\to 0} \frac{s(x+h) - s(x-k)}{h+k} > \alpha$$

It suffices to show that E is a null-set. To prove this we consider some $\epsilon > 0$ and choose N so large that

$$\text{(i)} \qquad \sum_{n>N} \delta_n < \alpha \cdot \epsilon$$

Set $s_*(x) = s(x) - (H_1(x) + \ldots + H_N(x))$. If E_* is the corresponding set in (x) with s replaced by s_* then E and E_* only differ by the finite set $\xi_1, \ldots, \xi_N$ so the measures of e and E_* are the same. Now we apply Vitali's covering theorem using s_* and obtain a a sequence of disjoint intervals $\{a_n, b_n)\}$ which yields a Vitali covering of E_* and at the same time

$$\frac{s_*(b_\nu) - s_*(a_\nu)}{b_\nu - a_\nu} \geq \alpha$$

It follows that

$$\text{(ii)} \qquad s_*(1) - s_*(0) \geq \alpha \cdot \sum (b_\nu - a_\nu)$$

At the same time (i) entails that $s_*(1) - s_*(0) \leq \alpha \cdot \epsilon$ and hence we have

$$|E|^* = |E_*|^* \leq \sum (b_\nu - a_\nu) \leq \epsilon$$

Since ϵ was arbitrary we get $|E|^* = 0$ as requested.

1.17 Stieltjes' Moment problem.

For a full account of Stieltjes' results we refer to his collected work [Stieltjes]. Here we shall give some examples to illustrate the flavour of moment problems. We are given a real-valued and cointinuous function $f(x)$ defined on $x \geq 0$ whose absolute value gets so small as $x \to +\infty$ that the integrals

$$\int_0^\infty x^n \cdot |f(x)|\, dx < \infty$$

for all positive integers n. At first sight one may expect that if

$$\int_0^\infty x^n \cdot f(x)\, dx = 0$$

hold for all n, then f must be identically zero. However, this is not true. We shall give examples below. But first we insert the following

Exercise. Let $g(x)$ be an arbitrary continuous function on $[0, 1]$. Then the following two equalities hold:

$$\int_0^1 g^2(x)\, dx = \frac{2}{\pi} \lim_{R\to\infty} \int_0^R \big[\int_0^1 \cos(st) g(t)\, dt\big]^2\, ds$$

$$\int_0^1 g^2(x)\, dx = \frac{2}{\pi} \lim_{R\to\infty} \int_0^R \big[\int_0^1 \sin(st) g(t)\, dt\big]^2\, ds$$

The hint is to apply Parseval's formula for L^2-functions on the real line as explained in §§-distribution theory.

Next, consider a test-function $\phi(s)$ supported by $[0, 1]$. Now there exists the cosine integral

$$\Phi(x) = \int_0^1 \cos(sx) \cdot \phi(s)\, ds$$

This function satisfies the integrability condition (*). For example, if $m \geq 1$ we perform $2m$-many partial integrations and find that

$$x^{2m}\Phi(x) = (-1)^m \cdot \int_0^1 \cos(sx) \cdot \phi^{(2m)}(s)\, ds$$

Similarly we have the sine-transform:

$$\Psi(x) = \int_0^1 \sin(sx) \cdot \phi(s)\, ds$$

For each positive integer p a partial integration gives

$$\Phi(x) + i\Psi(x) = \frac{(-1)^p}{(is)^p} \int_0^1 e^{ixs} f^{(p)}(s)\, ds$$

Using this and the exercise above the reader may verify the equalities below for every non-negative integer p:

$$\frac{2}{\pi} \int_0^\infty x^{2p} \cdot \Phi(x)^2\, dx = \frac{2}{\pi} \int_0^\infty x^{2p} \cdot \Psi(x)^2\, dx = m_p^2$$

where we have put

$$m_p^2 = \int_0^1 [f^{(p)}(s)\,]^2\, ds$$

Now we consider the non-decreasing functions

$$G(x) = \frac{2}{\pi} \int_0^{\sqrt{x}} \Phi(s)^2\, ds \quad \text{and} \quad H(x) = \frac{2}{\pi} \int_0^{\sqrt{x}} \Psi(s)^2\, ds$$

From the above we obtain

$$m_p^2 = \int_0^\infty x^p \cdot G(x)\,dx = \int_0^\infty x^p \cdot H(x)\,dx$$

Since $G \neq H$ this gives an example where the moment problem has no unique solution with $c_p = m_p^2$.

Remark. The reason why uniqueness fails above is that the sequence $\{m_p\}$ increases rather rapidly. In fact, whenever ϕ is a test-function which is not identically zero then

$$\sum_{p=1}^\infty \frac{1}{m_p^{\frac{1}{p}}} = +\infty$$

This divergence is proved by analytic function theory in § XX. Stieltjes said that a sequence of non-negative numbers $\{c_n\}$ yields a *determined moment problem* if there exists a unique non-decreasing and continuous function $f(x)$ which satisfies (*)(and (**) above. We refer to the notes sbout functional analysis where the moment problem is discussed in more detail which includes conditions for the moment problem to be determined.

II. Abstract measure theory

Introduction.

To get some intuition we present the theory using concepts from probability theory. A sample space consists of a pair $(\Omega, \mathcal{B})$ where Ω is a set and $\mathcal{B}$ a Boolean σ-algebra of subsets. Thus, if $\{A_\nu\}$ is a denumarable family of sets in $\mathcal{B}$ indexed by non-negative integers, then $\cap A_\nu$ and $\cup A_\nu$ stay in $\mathcal{B}$.

0.1 Measurable functions. A real valued function f defined on Ω is measurable if the inverse sets

(0.1.1) $$f^{-1}(-\infty, a) \text{ and } f^{-1}(-\infty, a] \text{ both belong to } \mathcal{B} \quad : \quad a \text{ any real number}$$

The class of such functions is denoted by $\mathcal{M}$.

Exercise. Use that $\mathcal{B}$ is a σ-algebra to show that a real-valued function f is measurable if the sets in (0.1.1) belong to $\lfloor$ for every rational number a. Next, let f and g be a pair of measurable functions. Then $f + g$ is measurable. To see this the reader can check that

$$\{f + g < a\} = \cup_{q \in Q} \{f < q\} \cap \{g < a - q\}$$

where the union is taken over all rational numbers, and in a similar way one proves that $\{f+g \leq a\}$ are measurable sets for every real number a. Conclude that $\mathcal{M}$ is a vector space of the real number field.

0.2 Probability measures. A probability measure μ is a map

$$\mu\colon \mathcal{B} \to [0, 1]$$

which assigns a numer $0 \leq \mu(A) \leq 1$ for every set A in $\mathcal{B}$, and σ-additivity means that if $\{A_\nu\}$ is a sequence of pairwise disjoint sets in $\mathcal{B}$, then

$$\mu(\cup A\nu) = \sum \mu(A_\nu)$$

Finally, every probability measure is normalized, i.e. $\mu(\Omega) = 1$. The family of probablility measures is denoted by $P(\Omega, \mathcal{B})$.

0.3 Nullsets. A subset F of Ω is a null set with respect to a probability measure μ if there to each $\epsilon > 0$ exists $A \in \mathcal{B}$ such that

$$F \subset A \quad \text{and} \quad \mu(A) < \epsilon$$

The class of null-sets is denoted by $\mathcal{N}_\mu$. By σ-additivity $\mathcal{N}_\mu$ is stable under a denumarable union, i.e.

$$\{F_\nu\} \subset \mathcal{N}_\mu \implies \cup F_\nu \in \mathcal{N}_\mu$$

Notice that one does not require that a null-set belongs to $\mathcal{B}$.

0.4 μ-measurable sets. Let $B \subset \Omega$ be an arbitrary subset. Its outer - resp. inner μ-measure are defined by

$$\mu^*(B) = \min_{A \in \mathcal{B}} \mu(A) \quad : B \subset A \quad : \quad \mu_*(B) = \max_{A \in \mathcal{B}} \mu(A) \quad : \ A \subset B$$

If $\mu_*(B) = \mu^*(B)$ one says that B is μ-measurable and the common number is denoted by $\mu(B)$. Let $\mathfrak{M}_\mu$ be the family if all μ-measurable subsets of Ω. It is clear that we have the inclusion

$$\mathcal{B} \subset \mathfrak{M}_\mu$$

0.5 Exercise. Show that a set B is μ-measurable if and only if there exists a set $B_* \subset B$ such that $B \setminus B_* \in \mathcal{N}_\mu$ and B belongs to $\mathcal{B}$. Conclude that $\mathfrak{M}_\mu$ is a Boolean σ-algebra of subsets of Ω generated by $\mathcal{B}$ and $\mathcal{N}_\mu$. Show also that the map

$$B \mapsto \mu(B)$$

is σ-additive on $\mathfrak{M}_\mu$.

0.6 μ-measurable functions. A real-valued function f on Ω is μ-measurable if

$$f^{-1}(-\infty, a) \text{ and } f^{-1}(-\infty, a] \text{ both belong to } \mathfrak{M}_\mu \quad : \quad a \in \mathbf{R}$$

Denote this class by $\mathcal{M}_\mu$. Since $\mathcal{B} \subset \mathfrak{M}_\mu$ one has the inclusion

$$\mathcal{M} \subset \mathcal{M}_\mu$$

0.7 Teorem. *Let $f \in \mathcal{M}_\mu$. Then there exists a null set $F \in \mathcal{N}_\mu$ and some $f_* \in \mathcal{M}$ such that $f = f_*$ in $\Omega \setminus F$. Thus, after modifying a μ-measurable function on a null set with respect to μ it becomes "truly measurable".*

Proof. Let $\{q_\nu\}_1^\infty$ enumerate the set Q of rational numbers. The exercise in (0.5) to each q_ν a pair of disjoint sets $F_\nu \in \mathcal{B}$ and $G_\nu \in \mathcal{N}_\mu$ such that

$$f^{-1}(-\infty, q_\nu) = F_\nu \cup G_\nu$$

We can also regard the inverse image of the singleton sets and find for every q_ν a disjoint pair $H_\nu \in \mathcal{B}$ and $S_\nu \in \mathcal{N}_\mu$

$$f^{-1}(\{q_\nu\}) = H_\nu \cup S_\nu$$

Set

$$G^* = \cup G_\nu \bigcup \cup S_\nu$$

Here $G^* \in \mathcal{N}_\mu$ which gives the existence of a decreasing sequence of sets $\{W_\nu\}$ in $\mathcal{B}$ such that

$$\mu(W_\nu) < 2^{-\nu} \quad \& \quad G^* \subset \cap W_\nu$$

Now $W_* = \cap W_\nu$ belons to $\mathcal{B}$ and is also a null-set for μ. Finally, define the function f_* by

$$f_*(\omega) = f(\omega) \quad : \omega \in \Omega \setminus W_* \quad \text{and} \quad f_*|W_* = 0$$

By this construction $f_* \in \mathcal{M}_\mathcal{B}$ and f_* differs from f on a null set.

0.8 Equivalence classes of μ-measurable functions. A pair f, g in $\mathcal{M}_\mu$ are called equivalent if they are equal outside a null set. Theorem 0.7 shows that every equivalence class can be represented by a $\mathcal{B}$-measurable function.

0.9 Convergene in μ-measure. We say that a sequence $\{f_n\}$ in $\mathcal{M}_\mu$ converges in μ-measure to a limit function g in $\mathcal{M}_\mu$ if

$$\lim_{n\to\infty} \mu(\{|f_n - g| \geq \epsilon\}] = 0$$

hold for every $\epsilon > 0$. When it holds we write

$$f_n \xrightarrow{\mu} g \tag{0.9.1}$$

0.10 Convergence almost everywhere. A sequence of μ-measurable functions $\{f_\nu\}$ is said to converge μ-almost everywhere to a limit function g if there exists a null set $F \in \mathcal{N}_\mu$ such that every f_ν is defined in $\Omega \setminus F_\nu$ and one has a pointwise limit:

$$\lim_{\nu\to\infty} f_\nu(\omega) = g(\omega) : \ \omega \in \Omega \setminus F$$

When it holds we write

$$f_n \xrightarrow{a.e} g$$

0.11 Remark. Almost everywhere convergence means that pointwise convergence holds outside a null set F for μ. So the limit function g above is only defined in $\Omega \setminus F$. But when we regard μ-measurable functions one can ignore null sets for μ and the following hold:

0.12 Proposition. *If g is an almost everywhere limit of a sequence $\{f_\nu\}$ in $\mathcal{M}_\nu$ then $g \in \mathcal{M}_\mu$.*

Proof. For a real number a and a positive integer N we put

$$A_N(a) = \{w \in \Omega \setminus F : \max_{\nu \geq N} f_\nu(w) \leq a\}$$

If $\{g \leq a\} = \{w \in \Omega \setminus F : g(w) \leq a\}$ it is clear that

$$A_N(\epsilon) \subset \{g \leq a\}$$

hold for every N. Conversely, if δ is a positive number the pointwise convergence entails that

$$\{g \leq a\} \subset \bigcup_{n \geq 1} A_N(a+\delta)$$

Notice that the sets $A_N(a)$ increase with N and let $A^*(a)$ be the union. Then (i-ii) entail that the following inclusions hold for every positive integer m:

$$A^*(a) \subset \{g \leq a\} \subset A^*_N(a+1/m)$$

This implies that

$$\{g \leq a\} = \bigcap_{m \geq 1} A^*_N(a+1/m)$$

and the reader may verify that the right hand side is a μ-measurable set. Hence $\{g \leq a\} \in \mathcal{M}_\mu$. In a similar fashion one shows that $\{g \geq b\} \in \mathcal{M}_\mu$ hold for every real number and the reader can conclude that g is μ-measurable.

0.13 A warning Convergence in measure need not imply convergence almost everywhere. But for suitable suitable subsequences one has the following:

0.14 Proposition. *Let $\{f_n\}$ be a sequence of measurable functions which converges in μ-measure to a function g. Then there exists a subsequence*

$$\{g_k = f_{\nu_k} \,:\, \nu_1 < \nu_2 < \ldots\} \quad \& \quad \phi_k \overset{a.e}{\to} g$$

Proof. Regarding the sequence $\{f_n - g\}$ it suffices to prove the result when $g = 0$. We first find an integer ν_1 such that

$$\nu \geq \nu_1 \implies \mu(\{|f_\nu \geq 2^{-1}\} \leq 2^{-1}$$

Next, we find $\nu_2 > \nu_1$ such that

$$\nu \geq \nu_2 \implies \mu(\{|f_\nu \geq 2^{-2}\} \leq 2^{-2}$$

Inductively we get a sequence $\nu_1 < \nu_2 \ldots$ where

$$\nu \geq \nu_k \implies \mu(\{|f_\nu \geq 2^{-k}\} \leq 2^{-k}$$

Then, the sequence $\{g_k = f_{\nu_k}\}$ converges almost everywhere to the zero function. To see this we set:

$$g^*_k = \max_{\nu \geq k} |g_\nu|$$

The inductive construction above gives

$$\text{(i)} \qquad \mu(\{g^*_k \geq 2^{-k}\} \leq \sum_{\nu \geq k} \mu(\{|g_\nu| \geq 2^{-k}\} \leq \sum_{\nu \geq k} 2^{-\nu} = 2^{-k+1}$$

Here $g^*_1 \geq g^*_2 \geq \ldots$ is a monotone sequence and (i) shows that

$$g^*_k \overset{a.e}{\to} 0$$

From this it is clear that $\{g_k\}$ also converges almost everywhere to zero.

0.15 Exercise Let $\{f_n\}$ be a sequence in $\mathcal{M}_\mu$ such that

$$\lim_{n,m} \mu(\{|f_n - f_m| \geq |\epsilon\}) = 0$$

hold for each $\epsilon > 0$, where (m, n) ted to ∞, i.e. to every $\epsilon > 0$ there exists some integer N such that for every pair $n \geq N$ and $m \geq N$ one has

$$\mu(\{|f_n - f_m| \geq |\epsilon\}) < \epsilon$$

When this holds we say that $\{f_n\}$ is a Cauchy-sequence with respect to the μ-measure Prove by a similar method as in Proposition 0.14 that there exists a subsequence $\{g_k = f_{n_K}\}$ which converges almost everywhere to a limit function f_*. Show also that f_* is unique in $\mathcal{M}_\mu$ since the almost everywhere convergence of the subsequence entails that $f_n \xrightarrow{\mu} f_*$.

1. Integrals

1.1 Elementary functions. If $B \in \mathcal{B}$ we get the characteristic function χ_B which is 1 on B and 0 in $\Omega \setminus B$. An elementary function is a finite linear sum

$$F = \sum a_k \cdot \chi_{B_k} \tag{*}$$

where $B_1, \ldots, B_N$ is a finite set of disjoint sets in $\mathcal{B}$ whose union is Ω and and $a_1, \ldots, a_N$ are real numbers. Notice that F is unchanged if we take a refined partition, i.e. if every B_k is a disjoint union of some $\mathcal{B}$-sets $\{C_{k\nu}\}$ then

$$F = \sum \sum a_k \cdot \chi_{C_{k\nu}}$$

Let $E(\mathcal{B})$ denote the family of all elementary functions. This is a a linear space of functions on Ω, and if F and G belong to $E(\mathcal{B})$ we can represent both as in (*) after a common refinement and the product $G \cdot F$ is the elementary function defined by

$$G \cdot F = \sum b_k \cdot a_k \cdot \chi_{B_K}$$

where $F = a_k$ and $G = b_k$ hold on B_k for every k. Next, let F be a bounded $\mathcal{B}$-measurable function i.e. there exists a constant K such that $-K \leq F(\omega) \leq K$ hold for all $\omega \in \Omega$. Then F can be *uniformly approximated* by elementary functions. Namely, if $N \geq 1$ we set

$$B_N(\nu) = \{-K + K \cdot \nu \cdot 2^{-N} \leq F < -K + K \cdot (\nu + 1) \cdot 2^{-N}\} \quad : 0 \leq \nu \leq 2^{2N} - 2$$

$$\text{and} \quad B_N(2^{2N} - 1) = \{K - 2^{-N} \leq F \leq K\}$$

We get the elementary function

$$F_* = \sum_{\nu=0}^{2^{2N}-1} (-K + \nu \cdot 2^{-2N}) \cdot \chi_{B_N(\nu)}$$

It is clear that

$$F_* \leq F \leq F_* + 2^{-N}$$

Since N can be arbitrary large this shows that F is uniformly approximated by elementary functions.

1.2 Integrals of elementary functions. Let μ be a probability measure. To each elementary function

$$F = \sum a_k \cdot \chi_{B_k}$$

we assign the μ-integral

$$\int F \cdot d\mu = \sum a_k \cdot \mu(B_k) \tag{1.2.1}$$

It is clear that (1.2.1) yields an additive map from $E(\mathcal{B})$ into the complex number where we regard complex-valued functions. The maximum-norm $||F||_\infty$ is the maximum of $\{|a_k|\}$. Since μ is a probability measure we get:

$$\Big|\int F \cdot d\mu\,\Big| \leq ||F||_\infty \tag{1.2.2}$$

1.3 Integrals of bounded functions. If $K > 0$ is a positive number we denote by $\mathcal{M}_\mu(K)$ the class of μ-measurable functions f such that the absolute value $|f|$ is $\leq K$ almost everywhere. So in the equivalence class we can take f to be a function in $\mathcal{M}_\mathcal{B}$ where this everywhere defined function has maximum norm K at most. By (1.1) f can be uniformly approximated by a sequence $\{F_n\}$ from $E(\mathcal{B})$ with maximum norms $\leq K$. For each pair n, m the triangle inequality gives

(1.3.1) $$\left|\int F_n \cdot d\mu - \int F_m \cdot d\mu\right| \leq \int |F_n - F_m| \cdot d\mu \leq ||F_n - F_m||_\infty$$

Since $||F_n - f||_\infty \to 0$ it follows that (1) tends to zero as n and m increase. Hence the evaluated integrals $\{\int F_n \cdot d\mu\}$ is a Cauchy sequence of complex numbers and there exists a limit

(1.3.2) $$\lim_{n\to\infty} \int F_n \cdot d\mu$$

Moreover, (1.2.1) implies that this limit is intrinsic, i.e. independent of the chosen sequence $\{F_n\}$ which approximates f uniformly. The limit (2) is called the μ-integral of f and is denoted by

(1.3.3) $$\int f \cdot d\mu$$

1.4 Absolutely integrable functions. For each positive real number K we define the the truncation operator T_K from $\mathcal{M}_\mu$ to $\mathcal{M}_\mu(K)$ by:

$$T_K(f)(x) = f(x) : |f(x)| \leq K \quad \text{and} \quad T_K(f)(x) = 0 : |f(x)| > K$$

When $f \in \mathcal{M}_\mu$ we get bounded functions $\{T_K(f) : K > 0\}$. We can also take absolute values and notice that

(1) $$K_1 < K_2 \implies |T_{K_1}(f)| \leq |T_{K_2}(f)|$$

From (1) the μ-integrals of the non-negative functions $\{|T_K(f)|\}$ increase with K, i.e.

(2) $$K_1 < K_2 \implies \int |T_{K_1}(f)| \cdot d\mu \leq \int |T_{K_2}(f)| \cdot d\mu$$

We can impose the condition that the non-decreasing sequence (2) is bounded. This leads to

1.5 Definition. *A function $f \in \mathcal{M}_\mu$ is said to be absolutely integrable if there exists a constant C such that*

(*) $$\int |T_{K_1}(f)| \cdot d\mu \leq C \quad \text{for all} \quad K > 0$$

The class if these functions is denoted $L^1(\mu)$.

1.6 Integrals of $L^1(\mu)$-functions. Let f belong to $L^1(\mu)$. From the construction of the T-operators it is clear that if $K_2 > K_1 > 0$ then

(i) $$\left|\int T_{K_2}(f) \cdot d\mu - \int T_{K_1}(f) \cdot d\mu\right| \leq \int |T_{K_2}(f)| \cdot d\mu - \int |T_{K_1}(f)| \cdot d\mu$$

Since $f \in L^1(\mu)$ the non-decreasing sequence $\{\int |T_{K_2}(f)| \cdot d\mu\}$ is bounded above and therefore convergent. Since a convergent sequence also is a Cauchy sequence, it follows from (i) that

$$\lim_{K_2, K_1} \int T_{K_2}(f) \cdot d\mu - \int T_{K_1}(f) \cdot d\mu = 0$$

as K_1 and K_2 tend to $+\infty$. Since Cauchy sequences of complex numbers converge there exists a limit

(*) $$\lim_{K\to\infty} \int T_K(f) \cdot d\mu$$

This limit is denoted by $\int f \cdot d\mu$ and is called the μ-integral of f.

1.7 Exercise. Show that the integral on $L^1(d\mu)$ is additive, i.e.

$$\int f d\mu + \int g d\mu = \int (f+g) d\mu \quad : \quad f, g \in L^1(d\mu)$$

Hence the μ-integral yields a linear functional on the space of absolutely integrable functions. Show also that if $f \in L^1(\mu)$ then the absolute value $|f|$ belongs to $L^1(\mu)$.

1.8 The L^1-norm. The linear space $L^1(d\mu)$ becomes a normed space when we set

$$||f||_1 = \int |f[d\mu \quad : \quad f \in L^1(d\mu$$

1.9 Exercise. Show that $E(\mathcal{B})$ is a dense subspace of $L^1(\mu)$.

1.10 Convergence in the L^1-norm.. By definition a sequence $\{f_n\}$ converges in the L^1-norm to a limit function g in $L^1(\mu)$ when

(i) $$\lim_{n\to\infty} ||f_n - g||_1 = 0$$

If $\epsilon > 0$ we notice that

(ii) $$\mu(\{|f_n - g| \geq \epsilon\} \leq \epsilon^{-1} \cdot ||f_n - g||_1$$

Hence (i) implies that

(3) $$\lim_{n\to\infty} \mu(\{|f_n - g| \geq \epsilon\} = 0$$

Here ϵ is arbitrary small which means that $f_n \xrightarrow{\mu} g$. Thus L^1-convergence implies convergence in μ-measure.

1.11 Example. The converse is not true. For example, let μ be the Lebesgue measure on $[0,1]$. To each $n \geq 1$ we define $f_n(x)$ to be n if $0 \leq x \leq \frac{1}{n}$ and otherwise zero. Then

$$\int_=^1 f_n(x) \cdot dx = 1$$

hold for all n. At the same time $f_n(x) \to 0$ for every $x > 0$, i.e. the sequence converges almost every where to zero and hence also in measure. But if we restrict the attention to bounded functions a converse result holds.

1.12 Proposition *Let $K > 0$ be fixed. Then a sequence $\{f_\nu\}$ in $\mathcal{M}_\mu(K)$ converges to a limit function g in the L^1-norm if and only if the sequence converges in measure to g.*

Proof. Consider a pair f, g in $\mathcal{M}_\mu(K)$. If $\epsilon > 0$ we get the measurable set

$$E_\epsilon(f,g) = \{|f - g| \geq \epsilon\}$$

Since the maximum norm $||f - g||_\infty \leq 2K$ it follows that

(i) $$||f - g||_1 = \int_{E_\epsilon(f,g)} |f - g| d\mu + \int_{\Omega \setminus E_\epsilon(f,g)} |f - g| d\mu \leq 2K \cdot \mu(E_\epsilon(f,g)) + \epsilon$$

If $f_\nu \xrightarrow{\mu} g$ where $\{f_\nu\}$ in $\mathcal{M}_\mu(K)$ one has

(ii) $$\lim_{\nu\to\infty} \mu(E_\epsilon(f_\nu, g)) = 0 \quad \text{hold for each} \quad \epsilon > 0$$

Hence (i) implies that $||f_n - g||_1||_1 \to 0$. Since we already proved that L^1-convergence entails convergence in measure we get the equivalence assertion in Proposition 1.12 follows.

1.13 Dominated convergence theorem

Let $\{f_n\}$ be a sequence in $L^1(\mu)$ which converges in μ-measure to a limit function g. The example in 1.11 shows that convergence need not hold in the L^1-norm. To compensate for this we impose a certain bound so that the situation is essentially the same as in Proposition 1.12

1.14 Theorem *Let $\{f_n\}$ be a sequence in $L^1(\mu)$ where $f_n \xrightarrow{\mu} g$ holds. Assume in addition that there exists a non-negative $\phi \in L^1(\mu)$ such that*

(1) $$|f_n| \leq \phi$$

hold almost everywhere for each n. Then the limit function g belongs to $L^1(\mu)$ and

(2) $$\lim_{\nu\to\infty} ||f_n - g||_1 = 0$$

1.15 Exercise Prove Theorem 1.14. The hint is to apply truncation operators to ϕ so that the L^1-norm of $\phi - T_K(\phi)$ are small and then use Theorem 1.13.

1.16 Remark. Recall that almost everywhere convergence implies convergence in measure. So if $\{f_n\}$ is some $L^1(\mu)$-sequence which converges almost everywhere to a function g and a dominating ϕ-function exists as above, then $||f_n - g||_1 \to 0$.

2. A completeness theorem.

Let $\{f_n\}$ be Cauchy sequence in the L^1-norm, i.e.

$$(*) \qquad \lim_{n,m\to\infty} \int |f_n - f_m| \cdot d\mu = 0$$

2.1 Theorem *When (*) holds there exists a unique L^1-function f_* such that*

$$\lim_{n,m\to\infty} \int |f_n - f_*| \cdot d\mu = 0$$

Hence the normed space $L^1(\mu)$ is complete, i.e. it is a Banach space.

Proof. From (*) it follows that there exists a subsequence $\{n_k\}$ such that if $g_k = f_{n_k}$ then

$$\text{(i)} \qquad ||g_{k+1} - g_k||_1 \leq 2^{-k}$$

Since $\sum 2^{-k} < \infty$ the non-negative function

$$G = |g_1| + \sum_{k=1}^{\infty} |g_{k+1} - g_k|$$

is integrable. It is clear that

$$\text{(ii)} \qquad |g_k| \leq G \quad : k = 1, 2, \dots$$

Together with (i) this entails that $\{g_k\}$ is a Cauchy sequence and Exercise 0.15 gives a function g where

$$g_k \xrightarrow{\mu} g$$

Next, (ii) enable us to apply Theorem 1.14 which gives

$$||g_k - g||_1 \to 0$$

Above $\{g_k\}$ is a subsequence of $\{f_n\}$ and the triangle inequality shows that $||f_n - g||_1 \to 0$ which shows that g is the requested limit in $L^1(\mu)$ of the given Cauchy-sequence $\{f_n\}$.

2.2 The space $L^2(\mu)$. On elementary functions we can introduce the L^2-norm, i.e. set

$$||f||_2 = \sqrt{|f|^2 \cdot d\mu}$$

Passing to the closure of elementary functions this yields a normed linear space denoted by $L^2(\mu)$ which again is complete and therefore becomes a Hilbert space. The details are left to the reader.

3. Signed measures and the Radon-Nikodym theorem.

Let $(\Omega, \mathcal{B})$ be a sample space. Consider an additive real-valued map $\mu\colon \mathcal{B} \to \mathbf{R}$ which may take negative values and for which there exists a constant C such that

$$(*) \qquad -C \leq \mu(A) \leq C \quad : \quad A \in \mathcal{B}$$

This uniform bound and additivity imply that if $\{A_\nu\}$ is a sequence of disjoint sets in $\mathcal{B}$ then

$$\sum |\mu(A_\nu)| < 2C$$

Hence the series $\sum \mu(A_\nu)$ is absolutely convergent. We say that μ is σ-additive if

$$\sum \mu(A_\nu) = \mu(\cup A_\nu)$$

hold for every sequence of disjoint sets in $\mathcal{B}$ and refer to μ as a *signed* measure. From now on all measures are σ-additive and bounded, i.e. (*) holds for some constant $C = C_\mu$. If $\mu(A) \geq 0$ for all $A \in \mathcal{B}$ we say that μ is a positive measure.

3.1 Definition *Two positive measures μ and ν are mutually singular if there exist nullsets $A \in \mathcal{N}_\mu$ and $B \in \mathcal{N}_\nu$ such that*

$$\mu(B) = \mu(\Omega) \quad : \quad \nu(A) = \nu(\Omega)$$

When this holds we write $\mu \perp \nu$.

3.2 Hahn's Theorem *Every signed measure μ has a unique decomposition*

$$\mu = \mu_+ - \mu_- \quad \text{where} \quad \mu_,\mu_- \quad \text{are both positive and} \quad \mu_+ \perp \mu_-$$

Proof. $A \in \mathcal{B}$ is called a μ-positive set if

$$\text{(i)} \qquad E \subset A \quad \text{and} \quad E \in \mathcal{B} \implies \mu(E) \geq 0$$

Denote this class by $P_+(\mu)$. Obviously the union of two μ-positive sets is again μ-positive and by σ-additivity there exists a $A^* \in P_+(\mu)$ such that

$$\text{(ii)} \qquad \mu(A^*) = \max_{A \in P_+(\mu)} \mu(A)$$

Sublemma. One has $\mu(B) \leq 0$ for each $B \subset \Omega \setminus A^*$.

Proof of Sublemma. We argue by contradiction. Suppose there is some $B_0 \subset \Omega \setminus A^*$ with $\mu(B_0) = \delta > 0$. The maximality of $\mu(A^*)$ implies that B_0 does not belong to $P_+(\mu)$ which gives some $\delta_1 > 0$ such that

$$\text{(i)} \qquad -\delta_1 = \min_{E \subset B_0} \mu(E)$$

Choose $E_1 \subset B_0$ with $\mu(E_1) \leq -\delta_1/2$ and set $B_1 = B \setminus E_1$. Now $\mu(B_1) \geq \delta + \delta_1/2$ and exactly as above we get a negative number

$$-\delta_2 = \min_{E \subset B_1} \mu(E)$$

Choose $E_2 \subset B_1$ with $\mu(E_2) \leq -\delta_2/2$. Inductively we get a decreasing sequence of sets

$$B_\nu = B_0 \setminus (E_1 \cup \ldots \cup E_\nu)$$

where $\{E_\nu\}$ are disjoint. Moreover, we have a sequence $\{\delta_\nu\}$ of positive numbers where

$$\text{(ii)} \qquad -\delta_\nu = \min_{E \subset B_\nu} \mu(E) \quad \text{and} \quad \mu(E_{\nu+1}) \leq -\delta_\nu/2$$

Since μ is a signed measure there is a constant A such that

$$\text{(iii)} \qquad \mu(E_1) + \ldots + \mu(E_N) \geq -A \quad \text{for all} \quad N \geq 1$$

It follows that

$$\text{(iv)} \qquad \delta_1 + \ldots + \delta_N \leq 2A \quad \text{for all} \quad N \geq 1$$

Hence the positive series $\sum \delta_\nu$ is convergent which gives $\delta_\nu \to 0$ as $\nu \to +\infty$. Put

(v) $$B_* = \cap B_\nu$$

For each $\nu \geq 1$ the inclusion $B_* \subset B_N$ and the definition of δ_ν from (ii) give:

(vi) $$\min_{E \subset B_*} \mu(E) \geq -\delta_\nu$$

This hold for every ν and since $\delta_\nu \to 0$ the minimum in (vi) is ≥ 0 which means that B_* belongs to $P_+(\mu)$. At the same time the construction above gives

(vii) $$\mu(B_\nu \geq \delta_0 + \frac{1}{2}(\delta_1 + \ldots + \delta_\nu) \geq \delta_0$$

for every ν. By σ-additivity we have

$$\mu(B_*) = \lim_{\nu \to \infty} \mu(B_\nu)$$

and then (vii) entails that $\mu(B_*) > 0$. This contradicts the (1) since A^* and B_* are disjoint and the Sublemma is proved.

Final part of the proof . The Sublemma gives positive measures μ_+ and μ_- defined by

$$\mu_+(E) = \mu(E \cap A^*) \quad : \quad \mu_-(E) = -\mu(E \cap (\Omega \setminus A^*)$$

We see that $\mu_+ \perp \mu_-$ and $\mu = \mu_+ - \mu_-$. This proves the existence of at least one Hahn-decompostion. The proof of *uniqueness* of such a decomposition is left to the reader.

3.3 Radon-Nikodym derivatives.

Let μ be a positive measure and consider a non-negative function $f \in L^1(d\mu)$. This gives a positive measure defined by the σ-additive map

$$E \mapsto \int_E f \cdot d\mu \quad : \ E \in \mathcal{B}$$

Denote this positive measure by $f \cdot \mu$. If $E \in \mathcal{N}_\mu$ the construction of μ-integrals implies that $\int_E f \cdot d\mu = 0$. Hence one has the inclusion

(3.1) $$\mathcal{N}_\mu \subset \mathcal{N}_{f \cdot \mu}$$

In general, a positive measure ν is called *absolutely continuous* with respect to μ if

$$\mathcal{N}_\mu \subset \mathcal{N}_\nu$$

It turns out that such positive measures are of the form $f \cdot \mu$ with $f \in L^1(\mu)$.

3.4 Theorem. *Let μ be a positive measure. Then every positive measure ν which is absolutely continuous with respect to μ is of the form $f \cdot \mu$ for a unique non-negative function $f \in L^1(d\mu)$.*

Proof. For each pair (k, N) where N is a positive integer and k a non-negative integer we consider the following two signed measures

(i) $$\nu - k2^{-N} \cdot \mu \quad \text{and} \quad (k+1)2^{-N} \cdot \mu - \nu$$

The Hahn decomposition applied to $\nu - k2^{-N} \cdot \mu$ gives a maximal set

(ii) $$S_N(k) \in P_+(\nu - k2^{-N} \cdot \mu)$$

and similarly we find a maximal set

(iii) $$T_N(k) \in P_+((k+1)2^{-N} \cdot \mu - \nu)$$

If N is fixed the measure $(\nu - k2^{-N}) - (\nu - (k+1)2^{-N}) = 2^{-N} \cdot \mu \geq 0$. This implies that

(iv) $$S_N(k+1) \subset S_N(k) \quad \text{for all} \quad k = 0, 1, \ldots$$

Moreover, since ν is absolutely continuous with respect to μ it is clear that:

(v) $$\cap_{k\geq 1} S_N(k) = \emptyset$$

Finally the reader may observe that

(vi) $$S_N(k) \setminus S_N(k+1) = S_N(k) \cap T_N(k)$$

Let us put

(vii) $$W_N(k) = S_N(k) \cap T_N(k)$$

From (vi) it follows that $\{W_N(k)\}$ is a family of disjoint subsets of Ω and we notice that

(viii) $$k2^{-N}\mu(E) \leq \nu(E) \leq (k+1)2^{-N}\mu(E) \quad : \; E \subset W_N(k)$$

The reader may also verify the set-theoretic equality

(ix) $$W_N(k) = W_{N+1}(2k) \cup W_{N+1}(2k+1)$$

for all pairs k and N, Next we construct a sequence of functions by:

(x) $$f_N = \sum_{k=1}^{\infty} k2^{-N} \cdot \chi_{W_N(k)} \quad : \; N = 1, 2 \ldots$$

Using (ix) the reader may verify that $\{f_N\}$ increase with N and (viii) entails that

(xi) $$\int_E f_N \cdot d\mu \leq \nu(E) \leq \int_E f_N \cdot d\mu + 2^{-N} \cdot \mu(E) \quad : \; E \in \mathcal{B}$$

Now $\{f_N\}$ is a non-decreasing sequence and there exists a limit function

$$f_* = \lim_{N\to\infty} f_N$$

where the convergence holds almost everywhere and it is celar that (xi) implies that

$$\nu(E) = \lim_{N\to\infty} \int_E f_n \cdot d\mu = \int_E f_* \cdot d\mu \quad : \; E \in \mathcal{B}$$

This gives $\nu = f_* \cdot d\mu$ and Theorem 3.3, is proved.

3.4 A general decomposition.

Let μ be a positive measure. For any other positive measure ν there exists a unique decomposition of ν into a sum of one measure which is singular with respect to μ, while the other term is given by an $L^1(d\mu)$-function. More precisely one has:

3.5 Theorem *Given a positive measure μ every other positive measure ν is of the form*

$$\nu = \nu_s + f d\mu \quad : \quad \nu_s \perp \mu \quad \text{and} \quad f \in L^1(d\mu)$$

Proof To find the singular part ν_s we put

(1) $$M = \max \nu(A) \quad : \quad A \in \mathcal{N}_\mu$$

By σ-additivity there exists some $A_* \in \mathcal{N}_\mu$ such that $\nu(A_*) = M$. Define the measure ν_* by:

(2) $$\nu_*(E) = \nu(A_* \cap E)$$

Here $\nu_* \perp \mu$. Put $\gamma = \nu - \nu_*$. The construction of ν_* gives

(3) $$A \in \mathcal{N}_\mu \implies \gamma(A) = 0$$

Then Theorem 3.5 gives $\gamma = f \cdot \mu$ for some $f \in L^1(\mu)$ and Theorem 3.5 follows.

3.6 The Vitali-Hahn-Saks theorem.

Let $\Omega, \mathcal{B})$ be a probability space and μ a probability measuremwhich gives the σ-algebra $\mathcal{N}(\mu)$ of null-sets. A pair of μ-measurable sets are identified when their Boolean difference is a null-set. If $A \in \mathcal{B}$ then A_* denotes its equivalence class. Let X denote the space whole points are such equivalence classes of subdsets from $\mathcal{B}$. A metric is defined on X by

$$d(A_*, B_*) = \mu(A \setminus B) + \mu(B \setminus A)$$

3.6.1 Exercise. Show that X is a complete metric space.

Next, an additive real-valued function λ defined on $\mathcal{B}$ is called μ-continuous if

$$\lim_{\mu(A) \to 0} \lambda(A) = 0$$

3.6.2 Theorem. *Let $\{\lambda_n\}$ be a sequece of additive μ-continuus functions such that*

$$(*) \qquad \lim_{n \to \infty} \lambda_n(A)$$

exist for every $A \in \mathcal{B}$. Then, for each $\epsilon > 0$ there exists δ such that

$$(*) \qquad \mu(A) \leq \delta \implies \max_n |\lambda_n(A)| \leq \epsilon$$

Proof. For each $\epsilon > 0$ and every pair of integers n, m we get the closed sets in X:

$$\Sigma_{n,m} = \{A \in \mathcal{B} : \ |\lambda_n(A) - \lambda_m(A)| \leq \epsilon/2\}$$

Set

$$\Sigma_p^* = \bigcap_{n,m \geq p} \Sigma_{n,m}$$

Then (*) entails that $\cup \Sigma_p^* = X$. Exercise 3.6.1 and Baire's category theorem gives some p and $\delta > 0$ such that

$$\mu(A) < \delta \implies A \in \Sigma_p^*$$

Shrinking δ if necessary

$$\mu(A) < \delta \implies \max_{n \leq p_*} \lambda_n(A) \leq \epsilon/2$$

Now the reader can check that (*) holds.

3.6.3 Exercise. Show that if $\lambda_*(A)$ is the limit value in the theorem for each A, then λ_* is additive and μ-continuous.

III. Lebesgue Theory

Contents

Introduction.

In the euclidian space $\mathbf{R}^n$ it is natural to define the n-dimensional volume of a cube whose sides have length a to be a^n. After this one expects that it is possible to associate a non-negative number to every bounded open set and construct an additive measure which enjoys an invariance property, i.e. the n-dimensional measure of a bounded open set Ω should be the same after a translation and also under a rotation expressed by an orthogonal linear map. It turns out that this can be achieved in a unique way. But after one is confronted with a problem, i.e. how to extend the construction of n-dimensional volumes to a larger family than bounded open sets. More generally, one would like to find a Boolean σ-algebra of subsets of $\mathbf{R}^n$ which contain all bounded open sets on which there exists a σ-additive measure which may assign infinite volumes to "large unbounded sets".

The solution to all this was achieved by Henri Lebesgue. His famous book [Leb] from 1904 has found a wide range of applications and we shall expose the main results in Lebesgue's theory. From the abstract measure theory from Section II one can quickly attain the requested properties of the Lebesgue measure and after define the class of Lebesgue measurable functions and construct their integrals. However, the geometric picture is lost if one proceeds directly via the abstract notions. Instead we describe an effective procedure to construct the n-dimensional Lebesgue measure of a bounded open set Ω which will be denoted by $\mathrm{vol}_n(\Omega)$. The trick is to employ *dyadic grids*. To each positive integer N one has the family $\mathcal{D}_N$ of open cubes whose sides have length 2^{-N} and the corner points belong to the lattice $2^{-N} \cdot \mathbf{Z}^n$. In §1 we explain how dyadic grids enable us to construct $\mathrm{vol}_n(F)$ where F is a bounded open set or a compact set. After one defines the *inner* respectively the *outer* measure of an arbitrary subset E as follows:

$$\text{inner measure}(E) = \max_{K \subset E} \mathrm{vol}_n(K) \tag{1}$$

where the maximum is taken over compact subsets of E. Next

$$\text{outer measure}(E) = \min_{E \subset \Omega} \mathrm{vol}_n(\Omega) \tag{2}$$

where the minimum is taken over open sets which contain E. If (1) and (2) are equal one says that E is measurable in the sense of Lebesgue and the common number is its measure. It may occur that the common number is $+\infty$. But if E is relatively compact then its measure is always finite. The class of all measurable subsets of $\mathbf{R}^n$ is denoted by $\mathcal{L}(\mathbf{R}^n)$. It contain a subclass called *null sets*, where a set E is called a null-set if its outer measure is zero. The family of null sets is denoted by $\mathcal{N}(\mathbf{R}^n)$. One says that two measurable sets E and F are equivalent if their Boolean difference

$$(E \setminus F) \cup (F \setminus E) \in \mathcal{N}(\mathbf{R}^n) \tag{3}$$

0.1 Lebesgue points. Let E be measurable where $\text{vol}_n(E) > 0$. To each point $x_0 \in E$ we denote by $\mathcal{S}_\delta(x_0)$ the family of all open cubes $\square$ whose sides have length $\leq \delta$ and contain x_0. Here is not required that x_0 is the center of the cube or that the sides are parallell to the coordinate axis. Following Lebesgue we set

$$(*) \qquad \lambda_\delta^E(x_0) = \min_{\square \in \mathcal{S}_\delta(x_0)} \frac{\text{vol}_n(E \cap \square)}{\text{vol}_n(\square)}$$

Since the cube-family $\{\mathcal{S}_\delta(x_0)\}$ decreases as $\delta \to 0$, $\delta \mapsto \lambda_\delta^E(x_0)$ increases as δ shrinks to zero and we obtain a limit

$$(**) \qquad \lambda^E(x_0) = \lim_{\delta \to 0} \lambda_\delta^E(x_0)$$

It is clear that this λ-number is between 0 and 1. The set of Lebesgue points is defined by

$$(***) \qquad \mathcal{L}(E) = \{x \in E \quad : \lambda^E(x) = 1\}$$

It turns out that

$$(****) \qquad (E \setminus \mathcal{L}(E)) \in \mathcal{N}(\mathbf{R}^n)$$

0.2 Absence of phantoms. Let us remark that (****) already is non-trivial when F is a closed set, i.e. here (****) gives the null-set $F \setminus \mathcal{L}(F)$ which from a measure theoretic point means that F is like an open set around almost every point. Next, let Ω be a bounded open set and apply (****) to its compact boundary which gives a null-set S of $\partial\Omega$ such that

$$\lim_{\delta \to 0} \sup_{\square \in S_\delta(x)} \frac{\text{vol}_n(\Omega \cap \square)}{\text{vol}_n(\square)} = 0$$

hold when $x \in \partial\Omega \setminus S$. This means that Ω is thin around almost every boundary point.

0.3 Lebesgue measurable functions. A real-valued function $f(x)$ is measurable in the sense of Lebesgue if the inverse image sets $f^{-1}(-\infty, a)$ and $f^{-1}(-\infty, a]$ are measurable for every real number a. Functions whose absolute values attain large values too often are not so interesting, i.e. one studies foremost measurable functions which are locally integrable. In § XX we construct the Lebesgue integral which gives the space $L^1_{\text{loc}}(\mathbf{R}^n)$. Using (****) we shall prove that a function $f \in L^1_{\text{loc}}(\mathbf{R}^n$ has almost everywhere a Lebesgue value defined as

$$\mathcal{L}_f(x) = \lim \frac{1}{\text{vol}_n(\square)} \cdot \int_\square f(x) \cdot dx$$

where the limit is taken over open cubes $\square$ which contain x and their volumes tend to zero. Moreover, the limit value is equal to $f(x)$ almost everywhere. Thanks to this the notion of measurable L^1-functions becomes transparent. Namely, when $f \in L^1_{\text{loc}}(\mathbf{R}^n)$ and $\delta > 0$ and set:

$$F_\delta(x) = \frac{1}{\delta^n} \cdot \int_{\square_\delta(x)} f(\xi) \cdot d\xi$$

where $\square_\delta(x)$ is the cube centered at x and with sides of length δ. Lebegue's theorem implies that one has a pointwise convergence

$$\lim_{\delta \to 0} F_\delta(x) = f(x)$$

at every Lebesgue point of f. Identifying measurable functions which are equal almost everywhere we can agree that every L^1_{loc}-function is equal to its associated Lebesgue function $\mathcal{L}_f$, i.e one ignores "silly values". Concerning the F_δ-functions they are continuous functions. So outside a null set a given L^1_{loc}-function is the pointwise limit of a sequence of continuous functions.

1. Dyadic grids and Lebesgue points.

If N is a positive integer we obtain a family of pairwise disjoint open cubes of the form

$$\square(\nu_\bullet) = \{(x_1, \ldots, x_n) : \nu_k 2^{-N} < x_\nu < (\nu_k + 1)2^{-N} : 1 \leq k \leq n\}$$

where $\nu_\bullet = (\nu_1, \ldots, \nu_n)$ runs over all n-tuples of integers. This family of cubes is denoted by $\mathcal{D}_N$. Notice that each $\square \in \mathcal{D}_N$ contains 2^n many cubes in $\mathcal{D}_{N+1}$.

1.1 Dyadic exhaustion of open sets. Let Ω be a bounded open set. To each $N \geq 1$ we get the family $\mathcal{D}_N(\Omega)$ of cubes from $\mathcal{D}_N$ which are contained in Ω. Let $\rho_N(\Omega)$ be the number of such cubes. The volume of the union becomes

$$w_N(\Omega) = 2^{-nN} \cdot \rho_N(\Omega)$$

If $\square \in \mathcal{D}_N(\Omega)$, then the 2^n many dyadic cubes from $\mathcal{D}_{N+1}$ which are contained in $\square$ all belong to $\mathcal{D}_{N+1}(\Omega)$. It follows that

$$\rho_{N+1}(\Omega) \geq 2^n \cdot \rho_N(\Omega)$$

From this we conclude that $w_N(\Omega) \leq w_{N+1}(\Omega)$. Hence $\{w_N(\Omega)\}$ is increasing. Since Ω is bounded this sequence is bounded above. For example, if Ω is contained in the unit cube $\{0 \leq x_\nu \leq 1\}$ then $w_N(\Omega) \leq 1$ for every N. Since every non-decreasing sequence of non-negative real numbers which is bounded above has a limit we obtain the number

$$(*) \qquad \mathrm{vol}_n(\Omega) = \lim_{N\to\infty} w_N(\Omega)$$

The limit number is called the n-dimensional Lebesgue measure of the bounded open set.

1.2 The unique dyadic exhaustion. Let Ω be a bounded open set. We first seek cubes in $\mathcal{D}_1(\Omega)$. It may be empty but is otherwise a finite set of unit cubes. Then we consider the cubes in $\mathcal{D}_2(\Omega$ which are not contained in those from $\mathcal{D}_1(\Omega)$. We continue in this way, i.e. if $k \geq 3$ we take only those cubes in $\mathcal{D}_k(\Omega$ which are outside the union of all cubes from $\mathcal{D}_1(\Omega), \ldots, \mathcal{D}_{k-1}(\Omega)$. Let us denote this family by $\mathcal{D}^*_k(\Omega)$. In this way we obtain a *uniquely determined* sequence $\{\mathcal{D}^*_k(\Omega)\}$ where each non-empty family $\mathcal{D}^*_k(\Omega)$ is a finite subfamily of $\mathcal{D}_k$. Let $\rho^*_k(\Omega)$ be the number of cubes in $\mathcal{D}^*_k(\Omega)$. By this construction it is clear that

$$\Omega_N(\Omega) = \sum_{k=0}^{k=N} 2^{-nk} \cdot \rho^*_k(\Omega)$$

hold for every N. Passing to the limit we obtain

$$\mathrm{vol}_n(\Omega) = \sum_{k=0}^{\infty} 2^{-nk} \cdot \rho^*_k(\Omega)$$

1.3 Remark. If $\square$ is an arbitrary cube whose sides are parallell to the coordinate axis and of common length a then its Lebesgue measure is a^n. This equality is clear from the construction above if $2^N \cdot a$ is an integer for some N and the corner points belong to the lattice $2^{N} \cdot \mathbf{Z}^n$. The fact that $\mathrm{vol}_n(\Omega) = a^n$ for an arbitrary cube follows when the real number a is approximated by "dyadic integers" , i.e. for every $N \geq 1$ we find an integer $q_N$so that $q_N 2^{N} \leq a < (q_N + 1)2^{-N}$ and in a similar fashion we perform small translations of the corner points. Next, we may also consider open cubes whose sides are not parallel to the coordinate axis. Then we still have $\mathrm{vol}_n(\Omega) = a^n$ where a is the common length of the sides. To prove this one studies a general *linear transformation* on $\mathbf{R}^n$. Thus, let A be a real $n \times n$-matrix whose determinant is $\neq 0$. By $x \mapsto y = A(x)$ we get a bijective map from $\mathbf{R}^n$ onto itself. Then one has the equality

$$(*) \qquad \mathrm{vol}_n(A(\Omega)) = \big|\det(A)\big| \cdot \mathrm{vol}_n(\Omega)$$

for every bounded open set Ω.

Exercise. Prove (*). The hint is that we first have the equality in the special case when the A-matrix permutes the coordinates, i.e. when

$$A(x_1, \ldots, x_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

where $i \to \sigma(i)$ is a permutation. In addition to such linear maps we recall from Linear Algebra that every matrix with a non-zero determinant is the product of such permuting matrices and special matrices of the form

$$A(x_1,\ldots,x_n) = (y_1,x_2,\ldots,x_n)$$

$$\text{where} \quad y_1 = a_1x_1 + a_2x_2 + \ldots + a_nx_n \quad \text{and} \quad a_1 \neq 0$$

The proof that (*) holds for such linear transforms is left as an exercise.

1.4 Additivity. Let $\Omega_1,\ldots,\Omega_k$ be finite family of disjoint and bounded open sets. Since every open cube $\square$ is connected an inclusion $\square \subset \Omega_1 \cup \ldots \cup \Omega_k$ implies that $\square$ is contained in one of the cubes. From this we conclude that

$$\rho_N(\Omega_1 \cup \ldots \cup \Omega_2) = \rho_N(\Omega_1) + \ldots + \rho_N(\Omega_k)$$

hold for each N. Passing to the limit we get:

$$\text{vol}_n(\cup\,\Omega_\nu) = \sum_{\nu=1}^{\nu=k} \text{vol}_n(\Omega_\nu)$$

More generally, let $\{\Omega_\nu\}$ be a denumerable sequence of pairwise disjoint open sets such that the union is contained in a bounded set. Then we get convergent positive series and the reader can verify that

$$\text{vol}_n(\cup\,\Omega_\nu) = \sum_{\nu=1}^{\infty} \text{vol}_n(\Omega_\nu)$$

Thus, the volume is σ-additive on bounded open sets.

1.5 Interior approximation. Let Ω be a bounded open set. For each N we may also consider the family $\mathcal{D}_N^*$ of *closed* dyadic cubes. Let $\bar\rho_N(\Omega)$ be the number of closed cubes from $\mathcal{D}_N^*$ which are contained in Ω. Notice that every such closed cube appears as a *compact* subset of Ω and hence the finite union of these cubes is again compact subset of Ω. It is clear that the function

$$N \mapsto 2^{-nN} \cdot \bar\rho_N(\Omega)$$

is increasing. So the limit exists and we have

$$\text{(1)} \qquad \lim_{N\to\infty} 2^{-nN} \cdot \bar\rho_N(\Omega) \leq \text{vol}_n(\Omega)$$

It turns out that (1) is an equality. To prove this we notice that if $\square \in \mathcal{D}_N$ and k is another large positive integer, then almost every small cube from $\mathcal{D}_{N+k}$ which is contained in $\square$ has is closure contained in $\square$. More precisely, the reader may verify that if Ω is a bounded open set then the inequality below holds for for every pair of positive numbers N and k:

$$\text{(i)} \qquad \bar\rho_{N+k}(\Omega) \geq 2^{nk} \cdot \rho_N(\Omega) - 2n \cdot 2^{(n-1)k} \cdot \rho_N(\Omega) = 2^{nk}\rho_N(\Omega) \cdot (1 - 2n \cdot 2^{-k})$$

The point is that we can choose large k-numbers at the same time let $N \to \infty$. Namely, for each $\epsilon > 0$ we first choose k such that $2n \cdot 2^{-k} < \epsilon$. Then (i) and a passage to the limit obviously give:

$$\lim_{N\to\infty} 2^{-nN} \cdot \bar\rho_N(\Omega) \geq (1-\epsilon) \cdot \text{vol}_n(\Omega)$$

Since ϵ is arbitrary we get equality in (1).

1.6 The measure of compact sets. Let E be a compact subset of $\mathbf{R}^n$. For each N we denote by $\rho_N^*(E)$ the number of cubes $\square \in \mathcal{D}_N$ such that the closure $\bar\square$ has a non-empty intersection with E. Next, let Ω be some bounded open set which contains E. If $\square \in \mathcal{D}_N$ and $\bar\square \subset \Omega$ then we either have that $\bar\square \cap E \neq \emptyset$ or $\bar\square \subset \Omega \setminus E$. This gives the equality:

$$\bar\rho_N(\Omega) = \bar\rho_N(\Omega \setminus E) + \rho_N^*(E)$$

Passing to the limit where we apply the interior approximation from 1.5 to the open sets Ω and $\Omega \setminus E$, it follows that

$$\lim_{N\to\infty} 2^{-nN} \cdot \rho_N^*(E) = \mathrm{vol}_n(\Omega) - \mathrm{vol}_n(\Omega \setminus E)$$

In particular the limit in the left hand side exists and we use it as a definition of $\mathrm{vol}_n(E)$ for a compact set E. The reader may also verify that the limit exists directly. Namely, if $\square \in \mathcal{D}_N$ has a closure whose closure has a non-empty intersection wit E then it can only occur that some of the smaller closed cubes from $\mathcal{D}_{N+1}$ which are contained $\square$ have empty intersection with E. This entails that the function

$$N \mapsto 2^{-nN} \cdot \rho_N^*(E)$$

is increasing and hence has a limit which by the above is the measure of E.

1.7 An outer approximation. Let E be a compact set. If $\delta > 0$ we obtain the open set

$$E_\delta = \{x\colon\ \mathrm{dist}(x, E) < \delta\}$$

With this notation one has

(*)
$$\lim_{\delta\to 0} \mathrm{vol}_n(E_\delta) = \mathrm{vol}_n(E)$$

To prove this we take some $\epsilon > 0$ and find N so large that

$$2^{-nN} \cdot \rho_N^*(E) < \mathrm{vol}_n(E) + \epsilon$$

Next, with N fixed we notice that if $\square \in \mathcal{D}_N$ is such that $\bar{\square} \cap E = \emptyset$ then $\bar{\square} \cap E_\delta = \emptyset$ for a sufficiently small δ. Hence with N fixed we can find δ so small that

$$E_\delta \subset \cup \bar{\square}$$

where the union is taken over cubes in $\mathcal{D}_N$ whose closures intersect E. This gives

$$\mathrm{vol}_n(E_\delta) \leq 2^{-nN} \cdot \rho_N^*(E) < \mathrm{vol}_n(E) + \epsilon$$

Since ϵ is arbitrary we get (*).

1.8 Lebesgue points

Let E be a compact set. Examples show that E may have positive Lebesgue measure and yet its interior is empty. Suppose this holds and let $x_0 \in E$ be a given point. If $\square$ is an open cube which contains x_0 then $\square \setminus E$ is non-empty so we have the strict inequality

$$\mathrm{vol}_n(E \cap \square) = \mathrm{vol}_n(\square) = \mathrm{vol}_n(\square \setminus E)$$

In spite of this it turns out that

$$(*) \qquad \lim_{\square \to x_0} \frac{\mathrm{vol}_n(E \cap \square)}{\mathrm{vol}_n(\square)} = 1$$

can hold where the limit is taken over cubes which tend to the singleton set $\{x_0\}$. To be precise, (*) means that for every $\epsilon > 0$ there exists some integer M such that

$$x_0 \in \square \quad \text{and} \quad \mathrm{vol}_n(\square) < \frac{1}{M} \implies \mathrm{vol}_n(E \cap \square) > (1-\epsilon) \cdot \mathrm{vol}_n(\square)$$

where $\square$ are arbitrary open cubes. Notice that we do not require that x_0 is the center if $\square$ when we regard an inclusion $x_0 \in \square$.

1.9 Definition. *A point $x \in E$ is called a Lebesgue point if (*) holds. The set of Lebesgue points in E is denoted by $\mathcal{L}(E)$.*

It turns out that $\mathcal{L}(E)$ is so large that the complement $E \setminus \mathcal{L}(E)$ is a null set, i.e. one has:

1.10 Theorem. *Almost every point in E is a Lebesgue point.*

The proof relies upon a covering Lemma due to Vitali. So we first expose this result and prove Theorem 1.10 in 1.13 below.

1.11 Vitali coverings. Let A be a subset of $\mathbf{R}^n$ and $\mathcal{V} = \{\square_\alpha\}$ a family of open cubes. It is called a Vitali covering of A if the following hold:

For each point $a \in A$ and each $\epsilon > 0$ there exists some $\square_\alpha \in \mathcal{V}$ such that $a \in \square_\alpha$ and vol$_n(\square_\alpha) < \epsilon$.

Remark. The $\mathcal{V}$ -cubes may consist of a non-denumarable family, i.e. the indices α can be taken from any set.

1.12 Vitali's Theorem *Let $\mathcal{V}$ be a Vitali covering of a bounded set A. Then there exists a sequence of pairwise disjoint cubes $\square_1, \square_2, \ldots$ in $\mathcal{V}$ such that for every N one has the inclusion*

$$A \subset \bar{\square}_1, \cup \ldots, \cup \bar{\square}_N \bigcup_{\nu > N} 3 \cdot \square_\nu$$

where $\bar{\square}_1, \ldots, \bar{\square}_N$ denote the closure of the first N cubes and if $\nu > N$ then $3 \cdot \square_\nu$ is the expanded cube whose sides are three times larger than those of $\square_\nu$.

Proof. Pick a cube $\square_1 \in \mathcal{V}$ such that

$$|\square_1|_n > \frac{2}{3} |\square_\alpha|_n \quad : \quad \forall\, \square_\alpha \in \mathcal{V}$$

Next, let $\mathcal{V}_1$ be the subclass of $\mathcal{V}$-cubes which have empty intersection with the closed cube $\bar{\square}_1$. Then we pick $\square_2 \in \mathcal{V}_1$ such that

$$|\square_2|_n > \frac{2}{3} |\square_\alpha|_n \quad : \quad \forall\, \square_\alpha \in \mathcal{V}_1$$

We continue in this way and obtain a sequence of pairwise disjoint cubes $\square_1, \square_2, \ldots$ where one for each $k \geq 2$ has

$$|\square_{k+1}|_n > \frac{2}{3} \cdot |\square_\alpha|_n \quad : \quad \forall\, \square_\alpha \in \mathcal{V}_k$$

and $\mathcal{V}_k$ is the family of cubes which have empy intersection with $\bar{\Box}_1 \cup \ldots \cup \bar{\Box}_k$. It remains to show that this sequence gives the covering lemma. First, since the $\mathcal{V}$-cubes all stay inside a bounded set we have

$$\sum_{\nu=1}^{\infty} |\Box_\nu|_n < \infty \tag{1}$$

Next, let $a \in A$ and let N be some positive integer. If a already belongs to $\bar{\Box}_1 \ \cup \ldots, \cup \ \bar{\Box}_N$ we are done. If the inclusion fails the Vitali covering gives some $\Box_\alpha \in \mathcal{V}$ which contains a and at the same time is so small that it has empty intersection with the union of the N first closed cubes. Now $|\Box_\alpha|_n > 0$ and we claim that $\Box_\alpha$ cannot be disjoint from $\bar{\Box}_{N+1} \cup \ldots, \cup \bar{\Box}_M$ for all $M \geq N+1$. For if it is disjoint from such a union up to some integer , the construction entails that

$$|\Box_\alpha|_n \leq \frac{2}{3} |\Box_M|_n$$

But this cannot hold when M is large since the convergence in (1) implies that $|\Box_M|_n \to 0$ as M increases. Hence we can find a *smallest* $M \geq N+1$ such that

$$\Box_\alpha \cap \bar{\Box}_M \neq \emptyset$$

Since M is minimal $\Box_\alpha$ is disjoint from the closed union of the first $M-1$-cubes and the construction of $\Box_M$ entails

$$|\Box_M|_n > \frac{2}{3} |\Box_\alpha|_n$$

This gives obviously the inclusion $\Box_\alpha \subset 3 \cdot \bar{\Box}_M$ and finishes the proof of Vitali's covering lemma.

1.13 Proof of Lebesgue's theorem

Let $\epsilon > 0$ and $0 < \rho < 1$ be kept fixed for a while. When N is a positive integer we define the subset $U_N(E)$ of E which consists of points $x \in E$ for which there exist a cube $\Box$ such that

$$\frac{\text{vol}_n(\Box \cap E)}{\text{vol}_n(\Box)} < \rho \quad : \quad \text{vol}_n(\Box) < \frac{1}{N}$$

It is obvious that $U_N(E)$ is an relatively open subset of E. Moreover these sets decrease as N increases. Put

$$A = \cap_{N \geq 1} U_N(E)$$

Next, by 1.5 above there exists $\delta > 0$ such that

$$|A_\delta \setminus A| < \epsilon$$

Next, let $\mathcal{V}(\rho)$ be the family of cubes $\Box$ for which

$$\frac{\text{vol}_n(\Box \cap E)}{\text{vol}_n(\Box)} < \rho \quad : \quad 3 \cdot \bar{\Box} \subset A_\delta$$

It is clear that $\mathcal{V}(\rho)$ is a Vitali covering of A. The Covering lemma gives a sequence of pairwise disjoint cubes $\Box_1, \Box_2, \ldots$. so that

$$A \subset \bar{\Box}_1 \cup \ldots \cup \bar{\Box}_N \bigcup 3 \cdot \Box_\nu \quad : \quad N \geq 1$$

Next, using the obvious the inequalities

$$\frac{\text{vol}_n(A \cap \Box_\nu)}{\text{vol}_n(\Box_\nu)} \leq \frac{\text{vol}_n(E \cap \Box_\nu)}{\text{vol}_n(\Box_\nu)} < \rho \quad : \quad 1 \leq \nu \leq N$$

it follows that

$$|A|_n < \rho \cdot \sum_{\nu=1}^{\nu=N} \text{vol}_n(\Box_\nu) + 3^n \cdot \sum_{\nu > N} \text{vol}_n(\Box_\nu)$$

Finally, since the cubes $\{\square_\nu\}$ are pairwise disjoint the series $\sum \text{vol}(\square_\nu)$ converges. So if N is sufficently large we have

$$3^n \cdot \sum_{\nu > N} \text{vol}_n(\square_\nu) < \epsilon$$

At the same time $\square_1, \ldots, \square_N$ are all contained in A_δ. Hence XX gives

$$|A|_n < \rho \cdot |A_\delta|_n + \epsilon < \rho \cdot |A|_n + \rho\epsilon + \epsilon = \leq \rho \cdot |A|_n + 2\epsilon$$

Here ϵ can be arbitrarily small and since $\rho < 1$ we conclude that A is a null set. Its construction dependson ρ so let us denote this null set with $A(\rho)$. As $\rho \to 1$ the condition in (x) becomes more relaxed so these null sets increase. We get in particular the null set

$$A^* = \cup_{\nu \geq 2} A(1 - \frac{1}{\nu})$$

At this stage the reader may verify that $E \setminus A^*$ is contained in $\mathcal{L}(E)$ and Lebesgue's theorem is proved.

Remark Above we found that $A(1 - \frac{1}{\nu})$ is a so called G_δ set, i.e. the intersection of a decreasing sequence of relatively open subsets of E. So A^* is the union of an increasing sequence of G_δ-subsets of E, i.e. it a set-theoretically "relatively nice" subset of E.

1.14 Example. Let $n = 1$ and consider a denumerable family of open intervals $\{(a_\nu, b_\nu)\}$ all of which are contained in $(0, 1)$. Suppose also that $\sum (b_\nu - a_\nu) = 1/2$. In addition these intervals can be chosen so that their union is dense in $(0, 1)$. We obtain the compact set

$$E = [0, 1] \setminus \cup (b_\nu - a_\nu)$$

It has no interior points and its Lebesgue measure is $1/2$. Suppose that $\xi \in \mathcal{L}(E)$ where $0 < \xi < 1$. Given a small $\delta > 0$ we put

$$\rho(\delta) = \sum_{\nu=1}^{\infty} |(b_\nu - a_\nu) \cap (\xi - \delta, \xi + \delta)|_n$$

By the definition of Lebesgue points it follows that

$$\lim_{\delta \to 0} \frac{\rho(\delta)}{\delta} = 0$$

The fact that this hold for all $\xi \in E$ outside a null set is remarkable since the chosen family of open intervals $\{(a_\nu, b_\nu)\}$ is quite general.

2. Measurable sets

Let Ω denote open sets and E compact sets. The outer - respectively the inner measure - of a set A are defined by

$$|A|^* = \min \text{vol}_n(\Omega) \quad : A \subset \Omega \quad : \quad |A|_* = \max \text{vol}_n(E) \quad : E \subset A$$

2.1 Definition. *A set A is measurable in the sense of Lebesgue if $|A|^* = |A|_*$. When it holds the common number is denoted by $vol_n(A)$ and called the Lebesgue measure of A. The class of measurable sets is denoted by $\mathfrak{Leb}(\mathbf{R^n})$.*

Remark. Let A be measurable. The definition of inner measure yields an increasing squence of compact subsets $\{F_\nu\}$ of A such that $\lim |F_\nu| \to |A|_*$. Since $|A| = |A|_*$ it follows that that if $F^* = \cup F_\nu$, then $A \setminus F^*$ is a null set. Similarly, using the equality $|A| = |A|^*$ we find that that there exists a descreasing sequence of open sets $\{\Omega_\nu\}$ containing A such that if $\Omega_* = \cap \Omega_\nu$, then $\Omega_* \setminus A$ is a null set. Thus, apart from a null set every measurable set can be taken as a G_δ-set, i.e. a denumerable intersection of open sets, or as a denumerable union of closed sets. In general, a pair of measurable sets A, B are called *equivalent* if the Boolean difference

$$(A \setminus B) \cup (B \setminus A)$$

is a null set. This is an equivalence relation on the class of measurable sets and whenever needed one may choose a special measurable set from such an equivalence class.

Lebesgue points of measurable sets

Let A be a measurable set. A point x_0 is called a Lebesgue point for A if

$$(*) \qquad \lim_{\square \to x_0} \frac{\text{vol}_n(A \cap \square)}{\text{vol}_n(\square)} = 1$$

holds where the limit is taken over cubes which tend to the singleton set $\{x_0\}$. The set of Lebesgue points is denoted by $\mathcal{L}(A)$. Notice that we do not require that x_0 belongs to A. However, if E is a closed set it is clear that a Lebesgue point of E must belong to E which means that our definition above extends that for closed sets in Definition 1.9. In particular we see that $\mathcal{L}(A)$ is contained in the clopsure of A. If Ω is an open set it is clear that $\mathcal{L}(\Omega) \subset \Omega$ and it may occur that the boundary $\partial\Omega$ contains some Lebesgue points, i.e. the set $\mathcal{L}(\Omega) \setminus \Omega$ can be non-empty.

2.2 Theorem. *For every measurable set A the Boolean difference of A and $\mathcal{L}(A)$ is a null set.*

Proof. First we prove that $A \setminus \mathcal{L}(A)$ is a null set. To show this we consider some $0 < \rho < 1$ and for each integer $N \geq 1$ we get the subset $U_N(A)$ which consists of all $x \in A$ for which there exists an open cube $\square$ such that

$$(1) \qquad \frac{\text{vol}_n(\square \cap A)}{\text{vol}_n(\square)} < \rho \quad : \quad \text{vol}_n(\square) < \frac{1}{N} \quad : x \in \square$$

It is clear that $U_N(A)$ is relatively open in A and hence measurable. Put

$$(2) \qquad A_*(\rho) = \cap U_N(A)$$

Then $A_*(\rho)$ is measurable and hence equal to its outer measure. So for any $\epsilon > 0$ there exists an *open set* Ω which contains $A_*(\rho)$ such that

$$(3) \qquad |\Omega \setminus A_*(\rho)| < \epsilon$$

Next, define the family $\mathcal{V}(\rho)$ of open cubes for which

$$(4) \qquad \frac{\text{vol}_n(\square \cap A)}{\text{vol}_n(\square)} < \rho \quad \text{and} \quad 3 \cdot \bar{\square} \subset \Omega$$

This yields a Vitali covering of $A_*(\rho)$ and we get a sequence of parirwise disjoint cubes $\square_1, \square_2, \ldots$ in $\mathcal{V}(\rho)$ such that

$$A_*(\rho) \subset \bar{\square}_1 \cup \ldots \cup \bar{\square}_N \bigcup 3 \cdot \square_\nu \quad : \quad N \geq 1$$

From this we conclude that

$$|A_*(\rho)| < \rho \sum_{\nu=1}^{\nu=N} \text{vol}_n(\square_\nu) + 3^n \cdot \sum_{\nu > N} \text{vol}_n(\square_\nu)$$

As in the proof for compact sets we choose N so that $3^n \cdot \sum_{\nu > N} \text{vol}_n(\square_\nu) < \epsilon$ and since each $\square_\nu \subset \Omega$ we obtain

$$|A_*(\rho)| < \rho \cdot |\Omega| + \epsilon < \rho \cdot |A_*(\rho)| + \rho\epsilon + \epsilon \leq \rho \cdot |A_*(\rho)| + 2\epsilon$$

where we used that $|\Omega \setminus A_*(\rho)| < \epsilon$. Hence

$$|A_*(\rho)| < \frac{2\epsilon}{1 - \rho}$$

Since $\epsilon > 0$ can be made arbitrary small we conclude that $A_*(\rho)$ is a null set. Now we choose $\rho = 1 - \frac{1}{k}$ with $k = 1, 2, \ldots$ and get the null set

$$A_* = \cup A_*(1 - \frac{1}{N})$$

Funally, from the construction of the sets $A_*(\rho)$ it is clear that one has the inclusion:

$$A \setminus A_* \subset \mathcal{L}(A)$$

Hence $A \setminus \mathcal{L}(A)$ is contained in the null set A_*. To show that $\mathcal{L}(A) \setminus A$ is a null set we can work locally and consider a bounded and relatively open subset B of $\mathbf{R}^n \setminus A$. Now $B \setminus \mathcal{L}(B)$ is a null set and at the same time it is obvious that

$$\mathcal{L}(A) \cap B \subset B \subset \mathcal{L}(B)$$

From this we conclude that $\mathcal{L}(A) \setminus A$ also is a null set and Theorem 2.2 follows.

2.3 Remark. The result above means that we can choose a canonical set in every equivalence class of measurable sets. Namely, if A and B are two measurable sets whose Boolean difference is a null set then $\text{vol}_n(A \cap \Box) = \text{vol}_n(B \cap \Box)$ hold for every cube. It follows that we have the equality $\mathcal{L}(A) = \mathcal{L}(B)$. So in the given equivalence class we can choose this common set denoted by A_* where we now have the equality

$$(*) \qquad A_* = \mathcal{L}(A_*)$$

2.4 A criterion for null sets Let A be a measurable set. Suppose there exists some $0 < \rho < 1$ such that

$$|A \cap \Box|_n \leq \rho \cdot |\Box|_n$$

hold for a family of open cubes which is a Vitali covering of A. Then we see that A cannot have any Lebesgue point and hence it must be a null set.

3. Measurable functions and their integrals

A real-valued function f on $\mathbf{R^n}$ is measurable if the sets

$$\{x\colon\, f(x) < a\} \quad : \quad \{x\colon\, f(x) \leq a\}$$

are measurable for every real number a. Among these occur finite $\mathbf{R}$-linear sums of characteristic functions. That is, let $A_1, \ldots, A_m$ be a finite family of pairwise disjoint and measurable sets. Then the functions

$$\sum c_\nu \cdot \chi_{A_\nu} \quad : \quad c_1, \ldots, c_m \text{ in } \mathbf{R}$$

are measurable where we for each set A denote by χ_A its characteristic function. Let us now study bounded and measurable functions which for simplicity takes values in the interval $[-1, 1]$. Given such a function f we can approximate it from below and from above. When $N \geq 1$ we get the pairwise disjoint and measurable sets

$$A_\nu(N) = \{\nu 2^{-N} \leq f < (\nu+1)2^{-N}\}$$

Put

$$S_*(N) = \sum_{\nu=-2^N}^{\nu=2^N-1} \nu 2^{-N} \chi_{A_\nu(N)} \quad \text{and} \quad S_N^* = \sum_{\nu=-2^N}^{\nu=2^N-1} (\nu+1)2^{-N} \chi_{A_\nu(N)}$$

It follows that

$$S_*(N) \leq f \leq S^*(N) \quad : \quad S^*(N) - S_*(N) \leq 2^{-N}$$

Remark. Notice that $\{S_*(N)$ is an increasing sequence of functions, while $\{S^*(N)\}$ is decreasing. Hence we can approximate the bounded function f *uniformly* by characteristic functions from below or from above.

3.1 The Lebesgue integral. Let f be bounded and measurable. Consider also some bounded measurable set E which is kept fixed. To each $N \geq 1$ we define the integrals

$$\int_E S_*(N)(x)dx = \sum_{\nu=-2^N}^{\nu=2^N-1} \nu 2^{-N} \mathrm{vol}_n(A_\nu(N) \cap E)$$

$$\int_E S^*(N)(x)dx = \sum_{\nu=-2^N}^{\nu=2^N-1} (\nu+1) 2^{-N} \mathrm{vol}_n(A_\nu(N) \cap E)$$

It is clear that their difference is $\leq (b-a)2^{-N}\mathrm{vol}_n(E)$ and their exists the common limit

$$\lim_{N\to\infty} \int_E S_*(N)(x)dx = \int_E S^*(N)(x)dx$$

It is denoted by $\int_E f(x)dx$ and called the Lebesgue integral of f over the measurable set E.

Exercise Show that the integral is approximated by small partitions of the range of f. More precisely, let $-1 = a_0 < a_1 \ldots < a_N = 1$ and put

$$I(a_\bullet) = \sum a_\nu \cdot \chi_{\{a_\nu \leq f < a_{\nu+1}\}}$$

Then one has the inequality

$$|\int_E f(x)dx - I(a_\bullet)| \leq \mathrm{vol}_n(E) \cdot \max\,(a_{\nu+1} - a_\nu)$$

Above the rate of convergence these $I(a_\bullet)$-sums to the Lebesgue integral of f is *independent* of f as long as its range is confined to $[-1,1]$. This means that the construction of the Lebegue inegral is "robust".

Exercise Show that the Lebesgue integral is additive in the sense that

$$\int_{E_1\cup E_2} f(x)dx = \int_{E_1} f(x)dx + \int_{E_2} f(x)dx \quad : E_1, E_2 \text{ disjoint bounded measurable sets}$$

Show also additivity with respect to f, i.e. that

$$\int_E (f_1+f_2)dx = \int_E f_1 dx = \int_E f_2 dx$$

for a pair of bounded and measurable functions.

3.2 Lebesgue points. Let f be a bounded measurable function. A point x_0 is called a Lebesgue point of f if

$$(*) \qquad \lim_{\delta\to 0} \frac{1}{|\square|_n} \int_\square f(x) \cdot dx = f(x_0)$$

hold where the limit it is taken over cubes $\square$ which contain x_0 with measure $< \delta^n$. Thus, if $f = \chi_A$ for some measurable set A, then we encounter the previous notion of Lebesgue points for sets.

3.3 Theorem *Outside a null set a bounded measurable function has Lebesgue points.*

Proof To prove this we employ the functions $S^*(N)$ and $S_*(N)$ above. Each S-function is finite linear sum of characteristic functions and by Theorem 2.2 measurable sets have Lebesgue points almost every where. Since the union of a denumerable family of null sets is again a null set we see that there exists a null set F such that the following hold for every point $x_0 \in \mathbf{R}^n \setminus F$ and every $N \geq 1$:

$$\lim_{\delta\to 0} \frac{1}{|\square|_n} \int_\square S^*(N)(x)dx = S^*(N)(x_0) \quad \text{and} \quad \lim_{\delta\to 0} \frac{1}{|\square|_n} \int_\square S_*(N)(x)dx = S_*(N)(x_0)$$

Since $f \leq S^*(N)$ for all N the first limit formula entails that

$$\text{Lim.sup}_{\,\delta\to 0}\,\frac{1}{|\Box|_n}\int_\Box f(x)dx \leq f(x_0)$$

Similarly, the second limit formula gives

$$\text{Lim.inf}_{\,\delta\to 0}\,\frac{1}{|\Box|_n}\int_\Box f(x)dx \geq f(x_0)$$

Hence the equality (*) holds almost everywhere.

3.4 Approximating bounded measurable functions. Let f be a bounded measurable function. To each $\delta > 0$ we set

$$F_\delta(x) = \delta^{-n}\cdot\int_{\Box_\delta(x)} f(x)dx$$

where $\Box_\delta(x)$ is the square centered at x with volume δ^n. The definition of Lebesgue points give

$$(*) \qquad \lim_{\delta\to 0} F_\delta(x) = f(x) \quad : \quad x_0 \in \mathcal{L}(f)$$

Concerning the F_δ-functions one has:

3.5 Proposition *There exists a constant C_n which depends on n only such that when$|f|\leq 1$ and $0<\delta\leq 1$ then*

$$|F_\delta(x) - F_\delta(y)| \leq C_n\cdot\frac{|x-y|}{\delta}: \quad \text{for any pair } x, y$$

3.6 Remark The convergence in (*) is *pointwise*. Thus, every bounded measurable function outside a null set is a pointwise limit of Lipschitz continuous functions. Identifying measurable functions which are equal almost everywhere we may assume from the start that f is equal to its Lebesgue value outside a nullset $\mathcal{N}_f$. Then the pointwise limit in (*) holds at all points outside $\mathcal{N}_f$.

3.7 Another approximation of measurable functions. Set

$$(1) \qquad F^*_\delta(x) = \max_{0<\rho\leq\delta} F_\rho(x)$$

Here the maximum is taken over a family of Lipschitz continuous functions which implies that F^*_δ is upper semi-continuous for every $\delta > 0$. As δ decreases the maximum is taken over smaller families of functions. So $\{F^*_\delta\}$ is a non-increasing sequence of functions which converge pointwise to f outside $\mathcal{N}_f$.

3.8 Egoroff's Theorem.

Let f be a bounded measurable function defined on a compact set, say the unit cube $\Box$ in $\mathbf{R}^n$. If $\epsilon > 0$ we can choose a compact set $E \subset \mathcal{L}(f)$ such that $|\Box\setminus E|_n < \epsilon$. With $\{g_n = F_{1/n}\}$ we have a sequence of continuous functions on E which converges pointwise to f. Now we construct a subset of E where the pointwise convergence is uniform. To each pair of positive integers ν, N we set

$$G_N(\nu) = \bigcup_{k,m>N}\{x\in E:\ |g_k(x) - g_m(x)| \geq \frac{1}{\nu}\}$$

Since $\{g_k\}$ converges pointwise on E, the decreasing sequence $N \mapsto G_N(\nu)$ tends to the empty set for each fixed ν. So if $\epsilon > 0$ there exists for every $\nu \geq 1$ some $N(\nu)$ such that

$$\text{vol}_n(|G_{N(\nu)}(\nu)) < 2^{-\nu}\cdot\epsilon$$

Next, since $G_{N(\nu)}(\nu)$ is measurable we find an open set Ω_ν such that

$$G_{N(\nu)}(\nu) \subset \Omega_\nu \quad : \; |\Omega_\nu| < 2^{-\nu} \cdot \epsilon$$

Put $E_* = E \setminus \cup\Omega_\nu$. Then $|\square \setminus E_*| < 2\epsilon$ and from the construction of the G-sets we see that the sequence $\{g_k\}$ converges *uniformly* on E_* which implies that the limit function f restricts to be continuos on E_*. Hence we have proved:

3.9 Theorem. *To every $\epsilon > 0$ there exists a compact set $E \subset \square$ such that the measure of $\square \setminus E$ is $< \epsilon$ and the restriction of f to E is a continuous function.*

4. Fubini's Theorem.

We study images of measurable sets under projections. Let n and m be positive integers and denote points in $\mathbf{R}^{n+m}$ by (x, y). We have the projection $\pi(x, y) = x$ onto the n-dimensional x-space. Let A be a bounded measurable set in $\mathbf{R}^{n+m}$. To each x we get the slice

(1) $$A(x) = \{y : \; (x, y) \in A\}$$

4.1 Theorem. *There exists a null set $\mathcal{N}$ in the x-space such that the slices $A(x)$ are measurable for all x outside $\mathcal{N}$ and the function*

$$x \mapsto |A(x)|_m$$

is measurable. Moreover one has the equation

(*) $$|A|_{m+n} = \int |A(x)|_m \, dx$$

Proof. Consider first the case when E is a compact set. in $\mathbf{R}^{n+m}$. Then every slice $E(x)$ is compact and hence measurable and the reader may verify that the function

$$x \mapsto |E(x)|_m$$

is upper semi-continouos and in particular measurable. Next, if Ω is a bounded open set the slices are open and this time we get a lower semi-continuous function

$$x \mapsto |\Omega(x)|_m$$

Taking the dyadic decomposition of Ω from § xx the reader may verify that (*) holds for bounded open sets and expressing compact sets is complents of open subsets of large cubes we also have (*) for compact sets. Next, if A is a bounded measurable set we choose an increasing sequence of compact subsets $\{E_k|\}$ where $|A \setminus E_k|_{m+n} \to 0$. Similarly $\{\Omega_k|\}$ is an increasing sequence of open sets containing A such that $|\Omega_k \setminus A|_{m+n} \to 0$. Set

$$g_k(x) = |\Omega_k(x)|_m - |E_k(x)|_m$$

Then $\{g_k\}$ is a non-increasing sequence of functions and since (*) hold for open and compact sets while $|\Omega_k \setminus E_k|_{m+n} \to 0$ it follows that

$$\lim_{k \to \infty} \int g_k(x) \, dx = 0$$

By the remark in § xx this entails that $\{g_k\}$ converge poinitwise to zero almost everywhere. So we find a null set $\mathcal{N}$ in $\mathbf{R}^n$ and $g_k(x) \to 0$ hold when x is outside $\mathcal{N}$. For a single slice it means that the outer and the inner measure of $A(x)$ are equal. Moreover, the funciton defined outside $\mathcal{N}$ by

$$x \mapsto |A(x)|_m$$

is equal to the function

$$\lim_{k \to \infty} |E_k(x)|_m$$

Here (ii) is the pointwise limit of an increasing sequece of upper semi-continuous functions and therefore measurable. hence (i) is measurable and at this stage the reader can also verify the equation (*) in Theorem xx.

Fubini's theorem for functions. Let $f(x,y)$ be a bounded measurable function with compact support. We agree to identify f with its Lebegue function and set

$$G_k = F^*_{1/k} \quad : k = 1, 2, \ldots$$

here $\{G_k(x,y)\}$ are upper semi-conionuous functions and the same hold for the functions

$$g_k(x) = \int G_k(x,y)\,dy$$

Now $\{g_k\}$ is a non-increasing sequence of functions and there exists the limit

$$\lim_{k\to\infty} \int g_k(x)\,dx$$

By § xx it follows that $\{g_k\}$ converge almost everywhere, i.e. outside a null set $\mathcal{N}$ in the x-space there exists a limit function

$$g_*(x) = \lim_{k\to\infty} \int g_k(x)$$

For each fixed x outside $\mathcal{N}$ it follows that there exists the limit

$$\lim_{k\to\infty} \int G_k(x,y)\,dy$$

Since the sequence of functions $\{y \mapsto G_k(x,y)\}$ is non-increasing we get a limit function

$$G_*(x,y) = \lim_{k\to\infty} \int G_k(x,y)\,dy$$

So here G_* is defined outside the sliced null set $\mathcal{N} \times \mathbf{R}^m$. Using Theorem xx and approximating f unifornly by elementary functions the reader may verify the equality

$$\iint f(x,y)|, dxdy = \iint G_*(x,y)\,dxdy = \int g_*(x)\,dx$$

Remark. We refer to $G_*(x,y)$ as the Fubini's slice function under the projection $\pi(x,y) = x$. Returning to the function f which is identified eith its Lebesgue function we leave the following as an exercise to the reader:

Exercise. Show that for every x outside $\mathcal{N}$ one has the equality

$$G_*(x, y = f(x,y)$$

for all y such that (x,y) is a Lebegue point of f and this set of y-values excludes at most a null set in $\mathbf{R}^m$ which may depend upon x. It means that the bounded measurable function $f(x,y)$ has an almost everywhere defined slice fun ction $y \mapsto f(x,y)$ for every x outside the null set $\mathcal{N}$ which by the above can be identified with the everywhere defined slice function $y \mapsto G_*(x,y)$.

IV. Riesz measures

Contents

Introduction

The Lebesgue measure is invariant under translations as well as rotations in $\mathbf{R}^n$. Thus, the volume of a cube depends only upon the length of its sides and not upon its position. We shall construct measures which do not enjoy this invariance. If f is a locally integrable function in $\mathbf{R}^n$ in the sense of Lebesgue then we obtain a measure which to every bounded Lebesgue measurable set E assigns the mass

$$\int_E f \cdot dx$$

We say that $f \cdot dx$ is an absolutely continuous measure with respect to dx. The new class consists of *singular* Riesz measures μ whose mass is concentrated to a null set in the sense of Lebesgue. Among these occur discrete measures which attach masses at some sequence of points. Less evident is the description of singular Riesz measures whose mass of every singleton set is zero. If $n = 1$ we described such measures in § I via non-decreasing and continuos functions $f(x)$ on the real x-line whose derivative is zero outside a set with Lebegue measure zero. When $n \geq 2$ we have no clear "geometric picture" of a singular Riesz measure. They are of a more implicit character and in general one can just say that a Riesz measure μ represents a continuous linear form on the space of continuous functions. This follows from a result due to Friedrich Riesz which asserts that if K is a compact subset of $\mathbf{R}^n$ and $C^0(K)$ the Banach space of continuous and real-valued functions on K, then its dual space consists of singular Riesz measures supported by K. The *Riesz representation formula* will be established in § 5.

In § 1 we construct Riesz measures in $\mathbf{R}^n$ using dyadic grids which is similar to the construction of the Lebesgue measure, except that one now arrives at a family of σ-additive measures which may assign positive mass to null sets. In $\mathbf{R}^n$ there exists the algebra $\mathcal{B}(\mathbf{R}^n)$ of *Borel measurable functions.*. Using results from the section about from abstract measure theory we will show that if f is a bounded Borel function and μ some compactly supported Riesz measure then there exists an integral

$$(*) \qquad \int f \cdot d\mu$$

Their existence and properties constitute a major part in the theory about Riesz measures in $\mathbf{R}^n$.

1. Dyadic grids.

If $N \geq 0$ we obtain the family $\mathcal{D}_N$ of half-open cubes where we for each n-tuple $\nu_\bullet \in \mathbf{Z}^n$ assign the cube

$$\square_N(\nu_\bullet) = \{x\colon \ \nu 2^{-N} \leq x_\nu < (\nu+1)2^{-N}\}$$

It is clear that $\{\square_N(\nu_\bullet)\}$ is a disjoint covering of $\mathbf{R}^n$ for each fixed N. Next, let Ω be a bounded open set. For each $x \in \Omega$ there exists a unique smallest integer $N(x)$ such that x belongs to a cube $\square \in \mathcal{D}_{N(x)}$ with $\square \subset \Omega$. This cube is denoted by $\square(x)$. The reader may verify the implication

$$y \in \square(x) \cap \Omega \implies \square(y) = \square(x)$$

Hence Ω is a unique union of disjoint half-open dyadic cubes. In other words, there exists a denumerable sequence $\{x_\nu\}$ in Ω such that the cubes $\{\square(x_\nu)\}$ are disjoint and

$$(*) \qquad \Omega = \bigcup \square(x_\nu)$$

We shall use this unique decomposition of open sets to construct a family of additive functions. Let $\mathcal{D}_*$ be the family of subsets of $\mathbf{R}^n$ given by a finite and disjoint union of cubes $\{\square_\alpha\}$ where each $\square_\alpha \in \mathcal{D}_{N(\alpha)}$ for some integer $N(\alpha) \geq 0$. Suppose that

$$\mu\colon \mathcal{D}_* \mapsto \mathbf{R}^+$$

is a non-negative and additive function which means that

$$\mu(\cup\square_\alpha) = \sum \mu(\square_\alpha)$$

hold for every finite and disjoint family of dyadic half-open cubes.

1.1 The μ-measure on open sets. Let Ω be a bounded open set. The unique decomposition (*) defines the number:

$$(**) \qquad \mu(\Omega) = \sum \mu(\square(x_\nu))$$

The right hand side is a positive series. Since Ω is bounded the series is convergent. To see this it suffices to observe that Ω is contained in some finite union of K many cubes from the family $\mathcal{D}_0$. Since μ is non-negative and additive, each finite partial sum of the series in (**) is bounded by the sum of μ-masses over this K-tuple of cubes in $\mathcal{D}_0$. which entails that (**) is convergent. Next, let $\{\Omega_\nu\}$ be a sequence of pairwise disjoint open sets whose union is bounded. The uniqueness of dyadic decompositions implies that

$$(1) \qquad \mu(\cup\,\Omega_\nu) = \sum \mu(\Omega_\nu)$$

1.2 The μ-measure on compact sets. Let E be a compact subset of $\mathbf{R}^n$. If $N \geq 0$ we set

$$\rho_N(E) = \sum \mu(\square) \quad : \quad \square \in \mathcal{D}_N \quad \text{and} \quad \square \cap E \neq \emptyset$$

If $N \geq 1$ and some $\square \in \mathcal{D}_N$ has a non-empty intersection with E it may occcur that some of the half-open cubes in $\mathcal{D}_{N+1}$ which are contained in $\square$ have empty intersection with E. This observation shows that

$$N \mapsto \rho_N(E) \quad \text{decreases with} \quad N$$

Hence there exists the limit

$$\lim_{N\to\infty} \rho_N(E)$$

The limit is denoted by $\mu(E)$ and is called the μ-measure of the compact set.

1.3 Additivity. Let E be a compact set and Ω some bounded open set which contains E. Consider some cube $\square(x_\nu) \subset \Omega$ in the decomposition of Ω. Here two cases can occur : Either $\square(x_\nu) \cap E \neq \emptyset$ or else $\square(x_\nu) \subset \Omega \setminus E$. From this we conclude that the following equality holds for every integer $N \geq 0$:

$$\rho_N(\Omega) = \rho_N(\Omega \setminus E) + \rho_N(E)$$

where we have set

$$\rho_N(\Omega) = \sum \mu(\square) \quad : \quad \square \in \mathcal{D}_N \quad \text{and} \quad \square \subset \Omega$$

and similarly we define $\rho_N(\Omega \setminus E)$. Since the positive series (**) in 1.1 is convergent for open sets, a passage to the limit as $N \to \infty$ gives:

1.4 Proposition. *If E is a compact set and if $E \subset \Omega$ for some bounded open set, then*

$$\mu(E) + \mu(\Omega \setminus E) = \mu(\Omega)$$

1.5 Exercise. Let $\mathfrak{B}_*$ denote the Boolean algebra of subsets of $\mathbf{R}^n$ generated by all open and all compact subsets. Let μ be a non-negative and additive function on $\mathcal{D}_*$ which has a bounded support in the sense that $\mu(\square) \neq 0$ only occurs for a finite set of cubes in $\mathcal{D}_0$. Use the results above to show that μ extends to an additive map on $\mathfrak{B}_*$, i.e for each finite family of disjoint sets $\{F_k\}$ in $\mathfrak{B}_*$ it holds that

$$\mu(\cup F_\nu) = \sum \mu(F_\nu)$$

whenever F and G are disjoint and bounded sets in $\mathfrak{B}_*$.

1.6 σ-additivity.

Above we found an additive measure defined on bounded sets in $\mathfrak{B}_*$. Now there exists the σ-algebra $\mathfrak{B}$ of Borel sets which by definition this is the smallest σ-algebra containing $\mathfrak{B}_*$. In order that μ can be extended to a σ-additive measure defined on $\mathfrak{B}$ we must impose additional conditions. If $\square \in \mathcal{D}_N$ is a half-open dyadic cube and $0 < a < 1$ we obtain the closed cube $a\bar{\square}$ defined by

$$\nu_k 2^{-N} \leq x_k \leq (\nu_k + a)2^N \quad : 1 \leq k \leq n\}$$

By the construction in § 1.2 we can assign the μ-measure to each of these compact cubes and since μ is non-negative the function

$$\text{(i)} \qquad a \mapsto \mu(a \cdot \bar{\square})$$

is increasing. Moreover we have

$$\text{(ii)} \qquad \mu(a \cdot \bar{\square}) \leq \mu(\square) \quad : \text{for each} \quad a < 1$$

Hence (i-i) suggest the following:

1.7 Definition. *We say that the additive measure μ is regular if*

$$\lim_{a \to 1} \mu(a\bar{\square}) = \mu(\square)$$

holds for every half-open dyadic cube.

1.8 Remark. This regularity condition is necessary in order that a σ-additive extension. In fact, this follows since one has the empty intersection:

$$\bigcap_{a<1} (\square \setminus a \cdot \bar{\square}) = \emptyset$$

The case when μ is regular. Let μ satisfy the condition from Definition 1.7. If Ω is a bounded open set and $\epsilon > 0$ we find N such that

$$\text{(i)} \qquad \mu(\Omega) < \rho_N(\Omega) + \epsilon$$

Here $\rho_N(\Omega) = \sum \mu(\square_\nu)$ where the sum extends over a finite set of cubes in $\mathcal{D}_N$. The regularity condition gives some $a < 1$ such that

$$\text{(ii)} \qquad \sum \mu(\square_\nu) < \sum \mu(a \cdot \bar{\square}_\nu) + \epsilon$$

With $E = \cup\, a \cdot \bar{\square}_\nu$ it follows that

$$\text{(iii)} \qquad \mu(\Omega) < 2\epsilon + \mu(E)$$

Notice that E is a *compact* subset of Ω. Since $\epsilon > 0$ can be arbitrary small we have therefore proved:

1.9 Proposition. *Let μ be regular, Then*

$$\mu(\Omega) = \max_K \mu(K)$$

hold for every bounded open set Ω where the maximum is taken over compact subsets of Ω.

1.10 Exercise. Let μ be regular and E is a compact set. If $\delta > 0$ we set $E_\delta = \{x\colon\ \text{dist}(x, E) < \delta\}$. Show that

$$\lim_{\delta \to 0} \mu(E_\delta) = \mu(E)$$

1.11 Extension to the Borel algebra. Denote by $\mathfrak{B}$ the σ-algebra generated open and compact subsets of $\mathbf{R}^n$. Let μ be a regular and bounded measure. Using the results above it follows that μ extends n a unique fashion to a σ-additive measure defined on all sets in $\mathfrak{B}$. So here

$$\mu(\cup F_\nu) = \sum \mu(F_\nu)$$

hold for every denumerable disjoint union of Borel sets.

1.12 Non-negative Riesz measures. This class consists of non-negative and regular measures μ defined as above. We have seen that the σ-additivity follows from the regularity. Moreover the resulting σ-additive function on bounded Borel sets is fully determined by the values assigned to all dyadic cubes.

$$(*) \qquad \square_N(\nu\cdot) = \{x = (x_1, \ldots, x_n)\colon\ \nu_j 2^{-N} \le x_j < (\nu_j + 1)2^{-N}\}$$

Let us restrict the attention to the cube

$$\square_* = \{0 \le x_j < 1\colon\ 1 \le j \le n\}$$

Then $\mathcal{D}_N$ contains 2^{Nn} many pairwise disjoint and half-open cubes whose union cover $\square_*$.

By an induction over N we assign a non-negative mass to these cubes in $\mathcal{D}_N$. The constraint is to preserve additivity. First $\mathcal{D}_0$ is a single cube to which we assign the unit mass. Next, in $\mathcal{D}_1$ we have 2^n cubes and we assign a non-negative mass to each of them so that the sum is one. Next, let $N \ge 1$ and suppose that a mass $m_N(\nu\cdot)$ is assigned to every cube $\square_N(\nu\cdot) \in \mathcal{D}_N$. Notice that a given cube $\square_N(\nu\cdot) \in \mathcal{D}_N$ is the union of 2^n many cubes from $\mathcal{D}_{N+1}$. To each of these we assign a mass so that the sum of these 2^n-many masses is equal to $m_N(\nu\cdot)$.

This inductive construction works for all N. Suppose we have performed such a construction over all positive integers N. Hence a mass $m_N(\nu\cdot)$ is assigned to $\square_N(\nu\cdot)$ for every N and each of the 2^{Nn} many n-tuples $\nu\cdot = (\nu_1, \ldots, \nu_n)$.

The regularity condition. Above we have seen how an inductive construction yields an additive measure μ. But it does not follow that it is regular, i.e. the condition from Definition 1.7 is not automatic. So here one encounters a set-theoretic problem, i.e. to describe conditions during the inductive construction of the assigned masses $\{m_N(\nu\cdot)\}$ in order that μ is regular. A necessary and sufficient criterium for this regularity is not known, i.e. this would mean that there exists a fully *constructive* method to achieve all Riesz measures. So the reader should be aware of the fact that the space of Riesz measures quite abstract and we remark that it contains measures whose existence rely upon the Axiom of Choice.

2. Singular Riesz measures

By the Radon Nikodym theorem every Riesz measure μ whose total mass is finite is the unique sum of a singular part and an absolutely continuous measures defined as a density by some $L^1(\mathbf{R}^n)$-function. Consider for simplicity a measure μ supported by the unit cube, i.e. it is constructed via the dyadic grid on $\square$. The result below gives the condition in order that μ is singular, i.e. its total mass taken over a set of Lebesgue measure.

2.1 Theorem *The measure μ is singular if and only if the following hold: For each $\epsilon > 0$ there exists an integer N such that we can find a subfamily $\mathcal{F}_N = \{\square_N(\nu\cdot)\}$ of $\mathcal{D}_N$ where the number of cubes in $\mathcal{F}_N$ is $< \epsilon \cdot 2^{Nn}$ while*

$$\sum \mu(\square_N(\nu\cdot)) > 1 - \epsilon \quad : \quad \sum \text{ taken over cubes } \square_N(\nu\cdot) \in \mathcal{F}_N$$

Exercise. Prove this result.

3. Signed measures

.

A signed measure μ supported by the unit cube arises by the inductive construction above when we allow that some of the numbers $m_N(\nu\cdot)$ are negative. In addition to additivity we impose the condition that

$$\sum_\nu |m_N(\nu\cdot)[= 1$$

hold for each N. In this case μ is again a σ-additive measure defined on the Borel algebra $\mathcal{B}$. Moreover, by the Hahn decomposition theorem from § 2: XX we can write μ as difference of two non-negative Riesz measures ν_1, ν_2 which are $\perp$ to each other.

4. μ-integrals.

Let μ be a non-negative measure on the unit cube constructed as above. Let $f(x)$ be a continuous function defined on the compact unit cube. Let $M = |f|_\square$ be the maximum norm of $|f|$ over $\square$. The integral $\int f d\mu$ is defined as the limit of the increasing sequence

$$\rho_N = \sum_{\nu=0}^{\nu=2^N-1} M \cdot \nu \cdot 2^{-N} \cdot \mu(\,\nu 2^{-N} \leq f < M \cdot (\nu+1) 2^{-N}\,)$$

The convergence of $\{\rho_N\}$ is robust exactly as in the abstract theory from XX. In this way we obtain an additive map

$$f \mapsto \int f d\mu \tag{1}$$

With the terminology of functional analysis this means that the μ-integral yields a continuous linear form on the normed space of continuous functions on $\square$. The *converse* is fundamental, i.e. one has

5. Riesz' representation theorem.

Let $\mathcal{L}$ be a continuous linear functional on the Banach space $C^0(\square)$ of real-valued continuos functions on $\square$. Using a similar procedure as in the proof of the Hahn decomposition theorem for signed measures one shows that $\mathcal{L}$ can be written in a unique way as the difference of two non-negative linear forms. Let us then assume that that $\mathcal{L}(f) \geq 0$ for every non-negative continuous function. If 1 is the identity number we have the positive number $\mathcal{L}(1)$ and without loss of generality we assume that $\mathcal{L}(1) = 1$. Now we construct a regular measure μ as follows: Let $\square$ be a dyadic cube in $\mathcal{D}_N$ for some $N \geq 1$. To each $0 < a < 1$ we get the compact cube $a \cdot \square$ and set

$$\text{(i)} \qquad \mu_a(\square) = \min_f \mathcal{L}(f): \quad \text{where} \quad f|a \cdot \square = 1 \quad \text{and} \quad f \geq 0$$

It is clear that $a \mapsto \mu_a(\square)$ increases with a and we set

$$\text{(ii)} \qquad \mu(\square) = \lim_{a \to 1} \mu_a(\square)$$

If $\square_1$ and $\square_2$ are two different dyadic cubes in $\mathcal{D}_N$ we notice that the compact cubes $a \cdot \square_1$ and $a \cdot \square_2$ are disjoint. This observation and the possibility to construct continuous functions in (i) which vanish outside arbitrary small open neighborhoods of $a \cdot \square$, imply that (i) yields an additive function on the family of all dyadic cubes. Moreover, by the construction via (ii) it is automatically regular. and hence μ is a Riesz measure. Next, let E be a closed subset of $\square$. One says that E is a null set with respect to $\mathcal{L}$ if

$$\min_f \mathcal{L}(f) = 0 \quad : \quad f|E = 1 \quad : f \geq 0$$

This makes sense because for every $\delta > 0$ one constructs a continuous function f which is 1 on E and has compact support in its δ-open neighborhood E_δ. Let us now *assume* that the boundaries of all dyadic cubes are null sets for $\mathcal{L}$. This hypothesis is not restrictive since we otherwise can use a translation of $\mathbf{R}^n$ by an n-tuple of numbers from the non-denumarable set of real numbers, i.e. after such a translation we get null sets and place the dyadic grid in the corresponding unit cube to construct measures as in 0.X. Under this assumption we obtain a non-negative Riesz measure by the following procedure: If $N \geq 1$ and $\square_N(\nu \cdot)$ is a dyadic cube we define:

$$\rho_N(\nu) = \max_f \mathcal{L}(f) \quad : 0 \leq f \leq 1 \text{ and has compact support in } \square_N(\nu \cdot)$$

One verifies easily that these ρ-numbers satisfy the additivity during the inductive construction as N increases and hence we obtain a Riesz measure μ. Moreover, it is clear from the whole construction that

$$\mathcal{L}(f) = \int f d\mu \quad : \ f \in C^0(\square)$$

Notice that this gives a canonical procedure to find μ when $\mathcal{L}$ is given.

The conclusion from this is the following result from the original work by F. Riesz:

Theorem. *Let Ω be an open set in $\mathbf{R}^n$ and $\mathcal{L}$ is a linear form on the space $C_0^0(\Omega)$ of continuous functions with compact support in Ω. Assume also that to every compact subset K of Ω there exists a constant C_K such that*

$$|\mathcal{L}(f)| \leq C_K \cdot |f|_K \quad : \quad \text{Supp}(f) \subset K$$

Then there exists a unique - in general signed - Riesz measure μ in Ω such that

$$\mathcal{L}(f) = \int f \cdot d\mu \quad : \quad f \in C_0^0(\Omega)$$

6. Borel measurable functions.

In many applications one needs to integrate functions which are not necessarily continuous. To construct μ-integrals with discontinuous functions as integrands we regard the unique smallest σ-algebra of subsets of $\mathbf{R}^n$ generated by open sets. It is denoted by $\mathcal{B}$ and sets in this σ-algebra are called Borel sets. Since a Riesz measure μ is σ-additive and hence we can assign a measure $\mu(B)$ to every Borel set B. Next, we consider $\mathcal{B}$-measurable functions f. It means that a real-valued function f is Borel measurable if the sets

$$(*) \qquad \{x \in \mathbf{R}^n \quad : \ f(x) < t\} \in \mathcal{B} \quad : \text{for every real number } t$$

Using (*) one constructs integrals $\int f \cdot d\mu$ exactly as in the abstract theory from Section § XX.

7. Weak convergence.

Consider a sequence of non-negative measures $\{\mu_\nu\}$ where each measure has total mass one on $\square$. One says that the sequence converges weakly to a limit measure μ^* if

$$(*) \qquad \lim_{\nu\to\infty} \int f d\mu_\nu = \int f d\mu^* \quad : \ f \in C^0(\square)$$

7.1 Remark. Suppose that the standard dyadic grid is a null set for all measures -$\{\mu_\nu\}$ and also for the limit measure μ^*. Then the *uniform* continuity of each $f \in C^0(\square)$ shows that weak convergence holds if and only if

$$\lim_{\nu\to\infty} \mu_\nu(\square_N(k\cdot)) = \mu^*(\square_N(k\cdot)) \quad : \ N \geq 1 \quad : \text{for every } k(\cdot) - \text{tuple}$$

Thus, weak convergence means precisely that the mass distributions during the inductive partition of $\square$ into smaller dyadic cubes give convergent sequences of the associated real 2^N-tuples for every $N \geq 1$. Next, recall Bolzano's theorem, i.e. that every infinite sequence of real numbers on a bounded interval has a convergent subsequence. Using this and the "diagonal procedure" the reader may verify the following fundamental fact:

7.2 Theorem. *Let $\{\mu_\nu\}$ be a sequence of non-negative measures on $\square$ - each with total mass one. Then there exists at least one subsequence which converges weakly.*

7.3 Weak density of discrete measures. Among Riesz measures occur finite sums of point-masses. Consider a sequence

$$\mu_j = \sum c_{j,k} \cdot x_k \quad : \ j = 1, 2, \ldots$$

where $\{x_1, x_2, \ldots\}$ is a denumarable sequence of points in the unit cube $\square$. At the same time each μ_j is a probability measure, i.e.

$$\sum_j c_{j,k} = 1 \quad : \ k = 1, 2, \ldots$$

The condition for the sequence to converge weakly is now easy to check. Namely, to every $N \geq 1$ and every cube $\square_N(\dot{\nu})$ in $\mathcal{D}_N$ we have

$$\mu_j(\square_N(\nu\cdot)) = \sum_* c_{j,k} \quad : \text{sum taken over those } k \text{ with } x_k \in \square_N(\nu\cdot)$$

So above weak convergence holds if and only if these numbers have a limit as $j \to \infty$ for every $N \geq 1$ and every cube in $\mathcal{D}_N$. Expanding each x_k in the binary system one can proceed to get an understanding when such a weak convergence takes place. Notice that if the sequence $\{x_k\}$ is given from the start then there may exist many doubly indexed c-sequences where we get a weak limit and moreover one may attain different limit measures depending on the chosen c-sequence. This occurs in particular when the sequence $\{x_k\}$ is everywhere dense. For example, if one has enumarated points in the unit cube whose coordinates are all rational numbers. In this situation every probability measure on $\square$ is the weak limit of a sequence $\{\mu_j\}$.

7.4 Exercise. Show that if μ is a Riesz measure with compact support then it can be approximaed weakly by a sequence of discrete measures.

8. Products and convolution of Riesz measures

Let n and m be two positive integers and μ and ν are Riesz measures in $\mathbf{R}^n$ and $\mathbf{R}^m$ respectively. The family of dyadic cubes in $\mathbf{R}^{n+m}$ is obtained by products of cubes from grids in $\mathbf{R}^n$ and $\mathbf{R}^m$. This gives a unique product measure $\mu \times \nu$ which satisfies

$$(\mu \times \nu)(\Omega \times U) = \mu(\Omega) \cdot \nu(U)$$

where $\Omega \subset \mathbf{R}^n$ and $U \subset \mathbf{R}^n$ are open sets.

8.1 Projections of measures. Let γ be a measure in $\mathbf{R}^{n+m}$ where the coordinates are (x, y) with $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$. We have the projection from $\mathbf{R}^{n+m}$ onto $\mathbf{R}^n$ defined by

$$\pi(x, y) = x$$

Assume that γ has a finite total mass which gives a measure $\pi_*(\gamma)$ in $\mathbf{R}^n$ defined on Borel sets F by

(*) $$\pi_*(\gamma)(F) = \gamma(F \times \mathbf{R}^m)$$

One refers to γ_* as the direct image measure.

8.2 Convolution. Let μ, ν be a pair of measures in $\mathbf{R}^n$. We get the product measure $\mu \times \nu$ in $\mathbf{R}^{2n}$ whose points are denoted by (x, y) where we regard μ as a measure in the n-dimensional x-space and ν as a measure in the y-space. In $\mathbf{R}^{2n}$ we take coordinates (t, s) where

$$t_\nu = x_\nu + y_\nu \quad : \quad s_\nu = x_\nu - y_\nu$$

Now $\mu \times \nu$ is a measure in the (t, s) space and the projection $\pi(t, s) = t$ gives the measure

$$\pi_*(\mu \times \nu)$$

It is denoted by $\mu * \nu$ and called the convolution of μ and ν. If $f(t)$ is a continuous function with compact support in the t-space the constructions of $\pi_*(\mu \times \nu)$ and $\mu \times \nu$ give

(*) $$\int f(t) \cdot d(\mu * \nu)(t) = \iint f(x + y) d\mu(x) d\nu(y)$$

Remark. The last integral is symmetric with respect to μ and ν. Hence the convolution satisfies the commutative law:

$$\mu * \nu = \nu * \mu$$

One can continue and construct the convolution of a triple of measures, and more generally the convolution of any number of measures in $\mathbf{R}^n$. For example, if μ, ν, γ is a triple of measures in $\mathbf{R}^n$ then we get the measure defined by

$$\int_{\mathbf{R}^n} f(t) \cdot d(\mu * \nu * \gamma) = \iiint f(x + y + z) \cdot d\mu(x) d\nu(y) d\gamma(z)$$

8.3 Exercise. Show that the convolution satisfies the associative law. So if $M(\mathbf{R}^n)$ denotes the space of Riesz measures with finite total variation, then convolution equips it with a commutative product, i.e. $M(\mathbf{R}^n)$ becomes a commutative algebra. The multiplicative unit is the point mass at the origin. One has also the inequality for norms:

$$||\mu * \nu|| \leq ||\mu|| \cdot ||\nu||$$

i.e. the total mass of a convolution is at most the product of the individual total masses.

8.4 Example. Let E and F be two compact sets in $\mathbf{R}^n$, each with a positive Lebesgue measure. We get the new compact set defined by

$$E + F = \{x + y \quad : \quad x \in E \quad : y \in F\}$$

Next, the two characteristic functions χ_E and χ_F define measures where

$$\int f \cdot d\chi_E = \int_E f \cdot dx \quad : \quad \int f \cdot d\chi_F = \int_F f \cdot dx$$

The construction of the convolution shows that

$$\chi_E * \chi_F = \chi_{E+F}$$

8.5 Exercise. Let μ be the discrete measure on the real line which has mass $1/2$ at $x = 0$ and at $x = 1$. Let $N \geq 2$ be an integer and consider the N-fold convolution μ^N. Show that this is a discrete measure supported by the integers $0, 1, \ldots, N$ and the mass at $0 \leq k \leq N$ is given by $2^{-N} \cdot \binom{N}{k}$.

8.6 Exercise. Let $n = 2$ and μ is the Riesz measure supported by the unit circle $\{x^2 + y^2 = 1\}$ where it is given by the constant angular measure, i.e. if $f(x, y)$ is a continuous function then

$$\int f \cdot d\mu = \frac{1}{2\pi} \cdot \int_0^{2\pi} f(\cos\theta, \sin\theta) \cdot d\theta$$

Consider the convolution of μ with itself, i.e. the Riesz measure $\nu = \mu * \mu$. By (8.5) the support is $T + T$, i.e. the set of points (x, y) of the form

$$x = \cos\theta + \cos\phi \quad : \; y = \sin\theta + \sin\phi$$

The reader is invited to find the picture of this set and try to investigate the resulting measure. One gateway for this is to employ Fourier transforms. First one has

$$\widehat{\mu}(\xi, \eta) = \int_0^{2\pi} e^{-i\xi\cos\theta - i\eta\sin\theta}\, d\theta$$

We use polar coordinates in the (ξ, η)-space. With $\xi = r\cos\phi$ and $\eta = r\sin\phi$ the addition formula for the cosine function shows that the integral becomes

$$\text{(i)} \qquad \int_0^{2\pi} e^{-ir\cdot\cos(\theta-\phi)}\, d\theta$$

Since the cosine function is periodic (i) depends on r only and expressed by the function

$$\text{(ii)} \qquad r \mapsto \int_0^{2\pi} e^{-ir\cdot\cos\theta}\, d\theta$$

As explained in the chapter devoted to distribution theory the Fourier transform of $\mu * \mu$ is $\widehat{\mu}^2$ which entails that it is a function of r only. Via Fourier's inversion formula it follows that the measure $\mu * \mu$ is invariant under rotations. By (xx) it is supported by a disc of radius 2 centered at the origin and determined by the function

$$\text{(iii)} \qquad \phi(s) = (\mu * \mu)(\{x^2 + y^2 \leq s^2\}) \quad : 0 \leq s \leq 2$$

The precise expression of this ϕ-function is not so easy. See § xx for a further account which involves the construction of certain Bessel functions. This iilustrates that the passage to convolutions can lead to quite cumbersome computations.

5. Stokes Theorem

Consider a bounded and connected open set Ω in $\mathbf{R^n}$. Its boundary $\partial\Omega$ is compact but we do no assume that $\partial\Omega$ is connected. A boundary point p is *regular* if there exists some open ball B centered at p and a real-valued C^1-function ϕ in B such that

$$\Omega \cap B = \{\phi < 0\} \quad : \partial\Omega \cap B = \{\phi = 0\} \quad : \nabla(\phi) \neq 0$$

It is clear that the set of regular boundary points is a relatively open subset of $\partial\Omega$. Denote this set by $\mathfrak{reg}(\partial\Omega)$. Put

$$K = \partial\Omega \setminus \mathfrak{reg}(\partial\Omega) \tag{1}$$

We impose the condition that K has Hausdorff measure zero in dimension $n-1$. Next there exists the area measure on $\mathfrak{reg}(\partial\Omega)$ to be denoted by dS which is found locally as follows: Let p be a regular boundary point and assume that the partial derivative $\frac{\partial\phi}{\partial x_n} \neq 0$ at p. The implicit function theorem gives an open neighborhood U of p where

$$\phi(p+x) = (x_n - a(x')) \cdot \rho(x)$$

for a pair of real-valued c^1-functions a and ρ. Here $\rho > 0$ in U and

$$\Omega \cap U = \{x_n < a(x')\}$$

where a only depends on $x' = (x_1, \ldots, x_{n-1})$ and $a =$ when $x' = 0$. So locally one has

$$\partial\Omega \cap U = \{x_n = a(x')\} \tag{i}$$

We can choose U so that (i) holds while x' varies in a small open cube $\square$ centered at the origin in this $(n-1)$-dimensional space. By a wellknown result in calculus which follows from the theorem of Pythagoras, the $(n-1)$-dimensional area of $\partial\Omega \cap U$ is given by the integral

$$\int_\square \sqrt{1 + |\nabla(a)(x')|^2}\, dx' \tag{ii}$$

where

$$|\nabla(a)(x')|^2 = \sum_{j=1}^{j=n-1} (\frac{\partial a}{\partial x_j})^2$$

The reader may verify that (ii) can be written as

$$\int_\square \frac{1}{|\frac{\partial\phi}{\partial x_n}|} \cdot |\nabla(\phi)|\, dx' \tag{xx}$$

where the partial derivatives of ϕ are evaluated at $x = (x', a(x')$ during the integration over $\square$.

5.0 Definition. *The domain Ω is of Federer type if the Hausdorff measure $|K|_{n-1} = 0$ and the the total area*

$$\int_{\mathrm{reg}(\partial\Omega)} dS < \infty \tag{*}$$

5.1 Theorem. *Let Ω be a domain of the Federer type and $f(x)$ a function of class C^1 defined in some open neighborhood of the closure $\bar{\Omega}$. Then*

$$\int_\Omega \partial f/\partial x_j \cdot dx = \int_{\mathrm{reg}(\partial\Omega)} f \cdot \mathfrak{n}_j \cdot dS \quad : \quad 1 \leq j \leq n$$

where $\mathfrak{n}_j$ is the x_j component of the outer normal $\mathfrak{n}$.

The proof requires several steps and we begin with a case where the "ugly set" K is empty.

5.2 The case of graphic domains. Consider an open cube of the form

$$\square = \{x\colon\ -A \leq x_\nu < A\colon\ \nu = 1, \ldots, n\}$$

Put $x' = (x_1, \ldots, x_{n-1}$ and let $\phi(x')$ be a C^1 function defined on the cube

$$\Box_* = \{-A < x_\nu < A \colon \nu = 1, \ldots, n-1\}$$

Here $-A < \phi(x') < A$ is assumed which gives the graphic domain

$$\Omega = \{(x', x_n) \colon -A < x_n < \phi(x') \quad : x' \in \Box_*\} \tag{1}$$

Next, let $f(x)$ be a C^1-function with compact support in $\Box$. Then we shall prove that

$$\int_\Omega \partial f/\partial x_j \, dx = \int_{\partial\Omega)} f \cdot \mathfrak{n}_j \cdot dS \quad : \quad 1 \leq j \leq n \tag{*}$$

If $j = n$ the volume integral is equal to the repeated integral

$$\int_{\Box_*} \big[\int_{-A}^{\phi(x')} \partial f(x', x_n)/\partial x_n \cdot dx_n\big] \, dx' = \int_{\Box_*} f(x', \phi(x') \, dx' \tag{1}$$

Now $dx' = \mathfrak{n}_\mathfrak{n} \cdot dS$ on the part of $\partial\Omega$ defined by $x_n = \phi(x')$. Moreover, since f has compact support in $\Box$ it vanishes on the remaining boundary of $\partial\Omega$ where at least some $|x_j| = A$. From this it is clear that (*) holds if $j = n$

The case $1 \leq j \leq n-1$. We may assume that $j = 1$ to simplify the notations. Set

$$\psi(x') = \int_{-A}^{\phi(x')} f(x', s) ds$$

Partial integration with respect to x_1 gives

$$\partial\psi/\partial x_1(x') = f(x', \phi(x')) \cdot \partial\phi/\partial x_1(x') + \int_{-A}^{\phi(x')} \partial f/\partial x_1(x', s) \, ds \tag{i}$$

Now we integrate over $\Box_*$. Since the ψ-function is zero when $|x_1| = A$ we have trivially

$$\int_{\Box_*} \partial\psi/\partial x_1(x') dx' = 0 \tag{ii}$$

In the right hand side of (i) the second term yields the repeated integral

$$\int_{\Box_*} \big[\int_{-A}^{\phi(x')} \partial f/\partial x_1(x', s) \cdot ds\big] \cdot dx' = \int_\Omega \partial f/\partial x_1(x) \, dx$$

So the vanisihing of (ii) entails that this volume integral is equal to

$$-\int_{\Box_*} f(x', \phi(x')) \cdot \partial\phi/\partial x_1(x') \cdot dx'$$

Next, on the part of $\partial\Omega$ defined by $\{x_n - \phi(x') = 0\}$ one has the equality

$$\mathfrak{n}_1 \cdot dS = -\partial\phi/\partial x_1(x') \cdot dx'$$

Hence (*) holds with $j = 1$ and the case $2 \leq j \leq n-1$ is treated in the same way which finishes the proof of (*).

5.3. Using partition of the unity Consider the situation in Theorem 1 and impose the extra assumption that $f = 0$ in an open neighborhood U of the ugly set K. In this case the result in 5.2 gives Theorem 5.1 after a partition of the unity. Namely, the compact set $W = \partial\Omega \setminus U$ is contained in $\mathfrak{reg}(\partial\Omega)$ and for each $p \in W$ there exists a cube $\Box$ centered at p such that $\Box \cap \Omega$ is a graphic domain. By Heine-Borel's Lemma we can cover W by a finite set of such cubes, say $\Box_1, \ldots, \Box_M$. Finally we can cover the compact set $\Omega \setminus U \cup \Box_1 \cup \ldots \cup \Box_M$ by a finite set of cubes $\Box_{M+1}, \ldots, \Box_{M+N}$ where the closure of these cubes are contained in Ω. Then we construct a C^∞-partition of the unity, i.e. we find a family $\{\phi_\nu \in C_0^\infty(\Box_\nu)\}$ so that $\sum \phi_\nu = 1$ holds in a neighborhood of the support of f. Now Stokes formula from (1.2) hold for every function $f \cdot \phi_\nu$ and by adding the result we get Stokes formula for f in Theorem 5.1.

5.4 How to avoid K

Consider the general case with no assumption on the support of f. To prove the integral formula in Theorem 5.1 for each x_j we can without loss of generality take $j = n$. Let $\pi(x', x_n) = x'$ be the projection to the $n-1$-dimensional x'-space. The assumption that $|K|_{n-1} = 0$ entails that the image set $\pi(K)$ is a null set in the $(n-1)$-dimensional x'-space. Let $\phi_1(x'), \phi_2(x'), \ldots$ be a sequence of test-functions in $\mathbf{R}^{n-1}$ which converge almost everywhere to 1 outside $\pi(K)$, while each of them has compact support in $\mathbf{R}^{n-1} \setminus K$ and $0 \leq \phi_\nu \leq 1$ hold for every ν.

By (5.3) Stokes Formula holds for every function $\phi_\nu f$. Since the functions $\{\phi_\nu(x')\}$ are independent of x_n we get

$$\text{(5.4.1)} \qquad \int_\Omega \phi_\nu \cdot \partial_n(f) \cdot dx = \int_{\mathfrak{reg}(\partial\Omega)} \phi_\nu \cdot f \cdot \mathfrak{n}_n \cdot dS$$

Since $\phi_\nu \to 1$ holds almost everywhere in the x'-space, Lebesgue's dominated convergence theorem implies that the left hand side in (5.4.1) tends to the volume integral of $\partial_n(f)$. There remains to prove that the right hand side tends to the integral of $f \cdot \mathfrak{n}_n$ taken over the regular boundary. Since f is a bounded function and the total area is finite the requested limit this follows if we have proved:

$$\text{(5.4.2)} \qquad \lim_{\nu \to \infty} \int_{\mathfrak{reg}\partial\Omega)} (1 - \phi_\nu) \cdot \mathfrak{n}_n \cdot dS = 0$$

To prove (5.4.2) we let $\epsilon > 0$ and put

$$\Gamma_\epsilon = \{|\mathfrak{n}_n| \geq \epsilon\} \cap \mathfrak{reg}(\partial\Omega)$$

Set $M = \text{Area}\,[\mathfrak{reg}(\partial\Omega)]$. Without loss of generality we can assume that the maximum norm of f is ≤ 1. Then (5.4.2) is majorized by

$$\text{(5.4.3)} \qquad M \cdot \epsilon + \int_{\Gamma_\epsilon} (1 - \phi_\nu) \cdot \mathfrak{n}_n \cdot dS = 0$$

Next, on Γ_ϵ the area measure dS is majorized by $\frac{1}{\epsilon} \cdot dx'$. Hence the sequence $\{g_\nu = 1 - \phi_\nu\}$ regarded as functions on the measure space $(\Gamma_\epsilon\,;\,dS)$ converge almost everywhere to zero which gives

$$\text{(5.4.4)} \qquad \lim_{\nu \to \infty} \int_{\Gamma_\epsilon} (1 - \phi_\nu) \cdot \mathfrak{n}_n \cdot dS = 0$$

In particular we can take ν so large that (5.4.4) is $< \epsilon$. Since ϵ was arbitrary the limit formula (5.4.2) holds and the proof of Theorem 5.1 is finished.

5.5 Remarks about Federer domains. Suppose that f is a non-negative C^2-function and the bounded domain Ω is defined by $\{f < a\}$ for a pair of real constants $a > 0$. To each $1 \leq j \leq n$ we set $\partial_j(f) = \frac{\partial f}{\partial x_j}$. On the regular part of $\partial\Omega$ the outer normal

$$\mathbf{n}_j = \frac{\partial_j(f)}{|\nabla(f)|}$$

Let us assume that $\nabla(f) \neq 0$ in Ω. So for ech $1 \leq j \leq n$ there exists the C^1-function in Ω defined by

$$g_j = \frac{\partial_j(f)}{|\nabla(f)|}$$

If $\delta > 0$ is small we have the domain $\Omega(\delta) = \{f < a - \delta\}$ and Stokes theorem gives

$$\int_{\Omega(\delta)} \partial_j(g_j)\, dx = \int_{\partial\Omega(\delta)} g_j \cdot \mathbf{n}_j \cdot dS \quad : 1 \leq j \leq n$$

By (xx) it follows that

$$\sum_{j=1]}^{j=n}\int_{\Omega(\delta)}\partial_j(g_j)\,dx=\int_{\partial\Omega(\delta)}dS$$

If Ω is a Federer domain one expects that there is a limit formula

$$\lim_{\delta\to 0}\int_{\partial\Omega(\delta)}dS=\int_{\mathrm{reg}(\partial\Omega)}dS$$

The passage to the limit can be understood via distribution theory. For each $\delta>0$ we denote by χ_δ the characteristic function of the open det $\Omega(\delta)$. At the same time $\{g_j\}$ are bounded funtions in Ω and the defintion of distribution derivatives entail that

$$\frac{\partial\chi_\delta}{\partial x_j}(g_j)=-\int_{\Omega(\delta)}\partial_j(g_j)\,dx$$

So (xx) means that

$$\sum\frac{\partial\chi_\delta}{\partial x_j}(g_j)=\mathrm{vol}_{n-1}(\partial\Omega(\delta))$$

At the same time the distributions $\{\chi_\delta\}$ converge to the distribution defined by the characteristic function of Ω. Indeed, this is so because the n-dimensional Lebesgue measure of $\Omega\setminus\Omega(\delta)$ tends to zero. Passing to distribution derivatives it follows that

$$\lim_{\delta\to 0}\frac{\partial\chi_\delta}{\partial x_j}=\frac{\partial\chi_\Omega}{\partial x_j}$$

But this weak limit in the space of distributions need not impliy that evaluations on the bounded g-functions give a limit since they need not extend so nicely to the closure of Ω where the gradient vector of f may vanish at some points. This illustrates why the conditions in Theorem 5.2 are essential.

The polynomial case. A case of special interest occurs when $f(x)$ is a real-valued polynomial. By Sard's Lemma the set of critical values is sparse and situations as above arise when we in addition assume that the polynomial is hyper-elliptic, i.e when

$$\lim_{|x|\to+\infty}|f(x)|=+\infty$$

In this situation the boundsries of $\{f<a\}$ are smooth excpet for a finite set of positive numbers and it turnd out that they become Federer domains. However, this cannot be deduced by pure measure theoretic arguments. One must use a more refined analysis to profit upon the hypothesis that f is a polynomial. It would lead us too far to discuss this in more detail. Let us only mention that the assertion that we get Federer domains rely upon the study of semi-algebraic sets, i.e. here one empoys the Seidenberg-Tarski theorem and also certain straitfications and a geometric integration theory developed in Whitney's very interesting text-book *Geometric integration theory.* So even if Theorem 5.1 is quite general the reader should be aware of the fact that extra work often is needed to verify that a specific domain is of the Federer type. More generally domains arising as real-analytic polyhedra satisfy the Federer condition and this leads to general Stokes formulas while currents are integrated on so called semi-analytic sets in arbitrary real-analytic manifolds. Here inequalities due to Lojasievicz are used. But it would bring us to far to discuss this in more detail.

6. The Hardy-Littlewood maximal function

.

Let f be a non-negative L^1-function on the real line supported by $x \geq 0$. Its forward maximal function is defined by

$$f^*(x) = \max_{h>0} \frac{1}{h} \int_x^{x+h} f(t)dt$$

Above it suffices to seek the maximum when h runs over positive rational numbers which we enumerate as $\{q_\nu\}$. For a fixed ν we have a continuous function

$$g_\nu(x) = \frac{1}{q_\nu} \int_x^{x+q_\nu} f(t)dt$$

Hence $f^*(x)$ is the maximum of $\{g_\nu\}$ which implies that f^* is *lower semi-continuous*. So if $\lambda > 0$ we get the open set:

$$E(\lambda) = \{x : \, f^*(x) > \lambda\}$$

6.1. Theorem *One has the inequality*

$$\text{meas}\, E(\lambda) \leq \frac{1}{\lambda} \cdot \int_{E(\lambda)} f(x)dx$$

Proof. With $g(x) = \frac{f(x)}{\lambda}$ we put $E_g = \{g^* > 1\}$ and there remains to show that :

(1) $$\text{meas}(E_g) \leq \int_{E_g} g(x)dx$$

To get (1) we consider the primitive function

(2) $$G(x) = \int_0^x g(t)dt$$

It is non-decreasing and continuous. Let $\{(a_\nu, b_\nu)\}$ be the disjoint intervals of E_g.

Sublemma. For each ν one has

$$G(b_\nu) - G(a_\nu) \geq b_\nu - a_\nu$$

Proof. Suppose strict inequality holds. By continuity we find $x > a_\nu$ with $x - a_\nu$ small so that

$$G(b_\nu) - G(x) < b_\nu - x$$

Next, since $x \in E_g$ there exists $\xi > x$ such that

$$G(\xi) - G(x) > \xi - x$$

Then $b_\nu < \xi$ must hold. On the other hand, since b_ν is a boundary point of E_g we also have

$$G(\xi) - G(b_\nu) \leq \xi - b_\nu$$

Now we get a contradiction since it would follows that

$$G(\xi) - G(x) = G(\xi) - G(b_\nu) + G(b_\nu) - G(x) < \xi - b_\nu + b_\nu - x = \xi - x$$

Proof continued. Since the intervals (a_ν, b_ν) are disjoint we get

$$\int_{E_g} gdt = \sum \int_{a_\nu}^{b_\nu} gdt = \sum G(b_\nu) - G(a_\nu)$$

By the Sublemma the last sum is $\geq \sum (b_\nu - a_\nu)$ which is equal to $\text{meas}(E_g)$ and Theorem 6.1 is proved.

An L^2-inequality. If x is a Lebesgue point of f we have the inequality $f^*(x) \geq f(x)$. Using Theorem 1 we shall now establish an inequality for L^2-norms.

6.2. Theorem *For every square integrable function $f(x)$ supported $x \geq 0$ one has the inequality*

$$||f^*||_2 \leq 2 \cdot ||f||_2$$

where $||\cdot||_2$ denotes the L^2-norm.

Proof. Since we claim an a priori inequality it suffices to prove the result when $f(x)$ is bounded and has compact support. Let N be a positive integer which eventually will be very large. Keeping N fixed we set

$$E(\nu) = \{\frac{\nu}{N} \leq f^* < \frac{\nu+1}{N}\} \quad : \quad \nu = 0, 1, \ldots$$

Set

$$f^*_- = \sum \frac{\nu}{N} \cdot \chi_{E(\nu)} \quad : \quad f^*_+ = \sum \frac{\nu+1}{N} \cdot \chi_{E(\nu)}$$

Sublemma For each $N \geq 1$ we have

$$\int_0^\infty (f^*_-(x))^2 dx \leq 2 \cdot \int_0^\infty f^*_+(x) \cdot f(x)\, dx$$

Proof of Sublemma Set $E_\nu = \{f^* > \frac{\nu}{N}\}$. The left hand side is equal to

$$\sum_{\nu \geq 0} \frac{\nu^2}{N^2} \cdot (\,|E_\nu| - |E_{\nu+1}|\,) = \sum_{\nu \geq 0} \frac{1}{N^2} \cdot |E_\nu| \cdot (\nu^2 - (\nu-1)^2) \leq \sum_{\nu \geq 0} \frac{2\nu}{N^2} \cdot |E_\nu|$$

By Theorem 6.1 the last sum is majorized by

$$\sum_{\nu \geq 0} \frac{2}{N} \int_{E_\nu} f dt = \sum_{\nu \geq 0} \frac{2(\nu+1)}{N} \int_{E(\nu)} f dt$$

By the construction of f^*_+ the last sum is $2 \int f^*_+(x) \cdot f dx$ and the Sublemma is proved.

Proof continued. The Cauchy-Schwarz inequality for L^2-norms gives:

$$\int f^*_+(x) \cdot f\, dx \leq ||f^*_+||_2 \cdot ||f||_2$$

Hence the Sublemma gives

$$||f^*_-||_2^2 \leq 2 \cdot ||f^*_+||_2 \cdot ||f||_2$$

Theorem 1 follows since the L^2-norms of f^*_+ and f^*_- both tend to $||f^*||_2$ when N increases.

Remark. The proof of the L^2-inequality extends easily to L^p inequalities when $1 < p < 1$ where the constant 2 is replaced by $\frac{p}{p-1}$. In Theorem 1 we regarded the forward maximum function. One can also define the backward maximum function and without any restriction on the support of an L^2-function f on the x-line its full maximal function

$$f^{**}(x) = \max_{\xi,\eta} \frac{1}{\xi+\eta} \int_{x-\eta}^{x+\xi} f(t)dt$$

with the maximum taken over all pairs of positive numbers ξ, η. Then Theorem 6.2 gives:

$$||f^{**}||_2 \leq 4 \cdot ||f||_2$$

7. Rademacher functions

For each positive integer N the interval $[0,1)$ can be decomposed into half-open intervals

$$\Delta_N(\nu) = [\nu 2^{-N}, \nu+1)2^{-N}) \quad : \quad 0 \le \nu \le 2^N - 1$$

Define the function $R_N(x)$ by

$$R_N(x) = (-1)^\nu \quad : \quad x \in \Delta_N(\nu)$$

This construction applies to each $N \ge 1$ and it is clear that

$$\int_{\Delta_N(\nu)} R_M(x)dx = 0 \quad : \quad 0 \le \nu < 2^N - 1 \quad : M > N$$

It follows easily that

$$\int_0^1 R_N(x) \cdot R_M(x)dx = 0 \quad : \quad M \ne N$$

Next, let $\{\alpha_\nu\}$ be a sequence in ℓ^2, i.e. $\sum |\alpha_\nu|^2 < \infty$ and consider the partial sum functions

$$S_N(x) = \sum_{\nu=1}^{\nu=N} \alpha_\nu \cdot R_\nu(x)$$

We shall analyze the limit as $N \to \infty$ using expansions of real numbers in the binary system. That is, when $0 < x < 1$ we have a series expansion

$$x = \epsilon_1(x)2^{-1} + \epsilon_2(x)2^{-2} + \ldots \quad : \quad \epsilon_\nu(x) = 1 \text{ or } 0.$$

The expansion is unique unless $2^N x$ is an integer for some N. Ignoring this denumerable set we define for each x the partial sum of its binary expansion

$$\xi_N(x) = \sum_{\nu=1}^{\nu=N} \epsilon_\nu(x) \cdot 2^{-\nu}$$

So here $\xi_N(x) \le x < \xi_N(x) + 2^{-N}$ and with these notations the following hold for the given ℓ^2-sequence above: '

7.1. Proposition *For each x and $N \ge 1$ one has the equality*

$$S_N(x) = \sum_{\nu=1}^{\nu=N} (2\epsilon_\nu(x) - 1) \cdot \alpha_\nu$$

The straightforward proof is left to the reader. Now we announce

7.2. Theorem *For each sequence $\{\alpha_\nu\}$ in ℓ^2 there exists a null set E in $\{0 \le x \le 1\}$ such that one has a pointwise limit when x is outside E:*

$$\lim_{N\to\infty} S_N(x) = \sum_{\nu=1}^{\infty} (2 \cdot \epsilon_\nu(x) - 1) \cdot \alpha_\nu$$

Proof. Since the R-functions is an orthonormal family in $L^2[0,1]$ we have

$$\int_0^1 S_N(x)^2 dx = \sum_{\nu=1}^{\nu=N} \alpha_\nu^2 \quad : \quad N \ge 1$$

It follows that $\{S_N\}$ is a Cauchy sequence in $L^2[0,1]$ and hence it converges in the L^2-norm to a limit function $S(x)$. Let $\mathcal{L}(S)$ be the set of Lebesgue points for S. If $x \in \mathcal{L}(x)$ we have

$$S(x) = \lim_{N\to\infty} \int_{\xi_N(x)}^{\xi_N(x)+2^{-N}} S(t)dt$$

From the construction of the partial sum functions it follows that $\lim S_N(x) = S(x)$, i.e. the pointwise limit exists at each Lebesgue point of S. Theorem 7.2. follows since $(0,1) \setminus \mathcal{L}(S)$ is a null set.

8. The Brownian motion

Introduction. Inspired by P.J. Daniell's work [DA] which fascilated previous results about infinite products of measure spaces, Norbert Wiener gave a rigorous construction of a stochastic process $\mathcal{W}$ in continuous time with the following properties: For each $t > 0$ one has a normally distributed random variable W_t whose variance is t and $\{W_t\}$ have independent increments, i.e if $s < t$ then the random variables $W_t - W_s$ and W_s are independent. Wiener employed Fourier series to construct such a process whose sample space is a *denumerable* product of $\mathbf{R}$. The book [Pa-Wi] contains an interesting historic account about the role of the Browninan motion outside the "abstract world of mathematics". Here we add some further comments.

Albert Einstein's first scientific paper was published in december 1900. It contains a statistical analysis of experimental data from phenomenon of capillarity and shows that Einstein already during the early period of his career was familiar with statistical arguments. Five years later it led to his articles [XX] which justified the existence of atoms and made it possible to determine the Avogrado number. After Einstein's work many skilled experimental scientists used stochastic analysis based upon the Brownian motion. Among those who applied Einstein's recipe was Theo Svedberg who received the Nobel Prize in chemistry in 1925 (6 ?). The interested reader may consult the book *Atoms* by Jean Perrin [Per] which gives a fascinating description of the nature of Brownian motion and describes experiments which predict that individual paths are non-differentiable.

8.0 Wiener's construction. The first rigouros proof that almost every Brownian path is continuous and even has a certain Hölder continuity is due to Wiener who used random Fourier series expansions of sine-functions. In general, let $\{a_k\}$ be sequence of real numbers such that $\sum a_k^2 < \infty$. Next, $\{\rho_1, \rho_2, \ldots\}$ is a sequence of independent normally distributed variables, i.e. each ρ_k has mean-value zero and variance one. With t regarded as a time parameter we get a stochastic process:

$$\text{(i)} \qquad t \mapsto W(t) = \sum_{k=1}^{\infty} a_k \cdot \sin kt \cdot \rho_k \quad : 0 \leq t \leq \pi/2$$

It is obvious that the mean value of each $W(t)$ is zero and the variance becomes

$$\text{(ii)} \qquad \sum_{k=1}^{\infty} a_k^2 \cdot \sin^2 kt$$

Next, for a pair $0 < s, t \leq \pi/2$ the mean value

$$\text{(iii)} \qquad E(W(t) \cdot W(s)) = \sum_{k=1}^{\infty} a_k^2 \cdot \sin kt \cdot \sin ks = \frac{1}{2} \sum_{k=1}^{\infty} a_k^2 \cdot (\cos k(t+s) - \cos k(t-s))$$

At this stage Winer employed a special a-sequence. Expand the continuous function t on $[0, \pi]$ in a cosine series

$$\text{(iv)} \qquad t = \sum b_k \cdot \cos(kt)$$

Here

$$\frac{\pi}{2} \cdot b_k = \int_0^{\pi} t \cdot \cos kt \, dt = \frac{1}{k} \cdot \int_0^{\pi} \sin kt \, dt$$

The right hand side is $\frac{2}{k^2}$ if k is odd and otherwise zero. So if k is odd one has

$$b_k = \frac{4}{\pi \cdot k^2}$$

Set

$$\text{(v)} \qquad a_{2k+1} = \sqrt{b_{2k+1}} \quad : \ k = 0, 2, \ldots$$

Then (iii-iv) entail that

$$(*) \qquad E(W_\epsilon(t)\cdot W_\epsilon(s)) = \frac{1}{2}(t+s-|t-s|) = \min(t,s)$$

Recall from basic probability theory that a pair of non-correlated normal distributions are independent. From (*) it follows that the stochastic process $\{W(t)\}$ has independent increment, i.e. if $0 < s < t \leq \pi$ then $W(s)$ and $W(t)-W(s)$ are independent normally distributed variables with variance s and $t-s$. This means that $\{W(t)\}$ yields a Brownian motion.

By Wiener's construction the sample space for the whole stochastic process in the continuous time parameter is given by a denumerable product of independent normal distributions on the real line. Starting from the construction above, Wiener also analyzed the continuity of individual paths. For a given sequence of values of the indepedent ρ-variables we have a function of t with the sine-series expansion (i). From the above

$$a_k = \frac{2}{\sqrt{\pi}\cdot k}$$

when k is odd while $a_k = 0$ when k is even. Since the series $\sum k^{-1}$ diverges we do not get an absolutely convergent series even if the set of ρ-values is bounded. So the continuity of $W(t)$ is not automatic. The reader can consult [Paley-Wiener page xx-xx] for a proof that if $\delta > 0$ then $W_\epsilon(t)$ is Hölder continuous of order $1/2-\delta$ for almost every outcome of the ρ-sequence. In § 8.1 we give another construction of the Brownian motion and establish a precise result about the continuity for individual Browninan paths.

Parabolic PDE-equations. Einstein's work has shown that a probabilistic interpretation to solutions of partial differential equations which appear in diffusion processes of colloidal material is very useful. This inspired later work dealing with the Brownian motion and the heat-equation such as in the joint article [KKP] by Khintchine, Kolmogorv and Petrowsky from 1930. Here one finds precise results about continuity properties of individual Brownian paths based upon studies of boundary value problems for the heat equation. Later Paul Levy found a more direct proof of the continuity properties of individual Brownian paths which we give in § 8.2. From a historic point of view considerable credit must be given to Fourier and Laplace. Even though the interplay between PDE-equations of parabolic type for heat conduction and probability theory was not fully explicit in Laplace's pioneeringt work in probability theory from 1820, the mathematical frame was paved via Fourier-Laplace transforms. In § 8.2 we follow Fourier and Laplace to construct parabolic PDE-equations associated to stochastic processes in continuous time with independent increments.

Bachelier's work. To the historic account one must add the pioneering work by X.X Bachelier whose article *La Bourse* from 1900 calculates probability densities attached to the Brownian motion with barriers. For example, given some $a > 0$ we can ask for the time when a Brownian first hits the barrier $x = a$. This yields a probability distribution which depends on a and was determined explicitly by Bachelier. His text-book on probability theory from 1910 is the first modern account dealing with stochastic processes on an advanced level. The Brownian motion can be used to investigate various probability distributions, apart from the "hitting probability" on a given arc on the boundary of a domain in $\mathbf{C}$ which correspond to a harmonic measure. At several occasions in these notes we expose this "tautology" between probabilistic interpretations via the Brownian motion and solutions to the heat equation and other boundary value problems which are obtained via solutions to the Dirichlet problem.

8.1 The Levy-Löf construction.

The subsequent material is inspired from notes by Anders Martin-Löf who attributes the main constructions to P. Levy. We seek a stochastic process $\{X(t)\colon\ t \geq 0\}$ such that for every pair $0 \leq s < t$, $X(t) - X(s)$ is a normally distributed random variable with variance $t - s$ and independent of $X(s)$. In particular each $X(t)$ is normally distributed with variance t. Recall that two normally distributed variables are independent if and only if they are uncorrelated. Hence, if E denotes expected value, the condition that the process has independent increments is equivalent to the condition that

$$E(X(t) \cdot X(s)) = \min(s,t) \quad : \text{for every pair } s,t$$

Before we give the construction of a stochastic process satisfying the conditions above we make some observations if such a process is given.

1. Conditioned mean-values. Let $s < t < u$ and consider the mean value of $X(t)$ when $X(u)$ and $X(s)$ are given, i.e. we seek

$$E((X(t) \mid X(s), X(u)$$

Since the stochastic variables $X(s), X(t), X(u)$ are normally distributed it is wellknown that this conditioned mean value is a linear function of $X(s)$ and $X(u)$, i.e. there exists a pair of real numbers (a,b) such that:

$$\bar{X}(t) = E((X(t) \mid X(s), X(u)) = a \cdot X(s) + b \cdot X(u))$$

Since we are assuming independent increments, $X(t) - \bar{X}(t)$ is independent of both $X(s)$ and $X(u)$. A calculation which is left to the reader gives

(*) $$a = \frac{u-t}{u-s} \quad : \quad b = \frac{t-s}{u-s}$$

If $t - s \to 0$ while u stays fixed, then $a \to 1$ which reflects that the conditioned mean value is close to $X(s)$. Inserting the values of a, b we now have a linear function

$$\bar{X}(t) = \frac{(u-s)X(s) + (t-s)X(u)}{u-s} \quad : \quad s < t < u$$

which coincides with $X(s)$, respectively $X(u)$ at the end points of the interval $[s,u]$.

2. The conditioned variance. With a and b chosen as in (*) we get:

$$V(t) = E((X(t) - \bar{X}(t))^2 \mid X(s), X(u)) =$$

$$E[(a(X(t) - X(s)) + b(X(t) - X(u))^2] = a^2(t-s) + b^2(u-t) = \frac{(u-s)(t-s)}{u-s}$$

Hence the graph of $V(t)$ is a parabola whose values at the end points s and u are zero. The maximum is attained at the mid-point which reflects that we obtain a maximal variance when random values taken by $X(s)$ and $X(u)$ affect $X(t)$ in equal parts.

3. An inductive construction. We shall construct the Brownian motion over the unit interval $0 \leq t \leq 1$. After one can of course continue to any time interval. Let $N \geq 1$ and put

$$t_\nu(N) = \nu \cdot 2^{-N} \quad : \quad 0 \leq \nu \leq 2^N$$

From previous remarks we obtain a stochastic process as follows: Let $\chi_1, \ldots, \chi_{2^N}$ be independent and normally distributed variables, each with mean value zero and variance 2^{-N}. Consider a time value

$$t = k2^{-N} + \delta \quad : \quad 0 \leq k \leq 2^N - 1 \quad : 0 \leq \delta \leq 2^{-N}$$

At this moment of time we define the random variable

(i) $$X_N(t) = \sum_{\nu=0}^{\nu=k-1} \chi_\nu + (1-\delta)\chi_k + \delta\chi_{k+1}$$

With a sample space of dimension 2^N over the real line, the outcome consists of values taken by the independent and normally distributed χ-variables. This sample determines X_t for every time value $0 \leq t \leq 1$. Hence the outcome of each sample point consists of a piecewise linear curve $t \mapsto \gamma(t)$ defined on $[0,1]$ with eventual corner points at when $t = k2^{-N}$, where $\gamma(k2^{-N})$ is the sum of the first k many χ-variables.

4. Passage to the limit. We have constructed a process $\{X_N(t)\}$ whose sample space is of dimension 2^N. For the inductive construction we pass from stage N to $N+1$ allowing "white noise" during time intervals $[k2^{-N},(k+1)2^{-N}]$. This means that we introduce 2^N many new random variables $g_1,\ldots,g_{2^N}$ which are independent of those which were used to get $\{X_N(t)\}$. Moreover, every g_k is normally distributed with variance 2^{-N-1}. Thus, at stage $N+1$ we have a process $t\mapsto X_{N+1}(t)$ where

$$X_{N+1}(k2^{-N}+\delta 2^{-N-1})=X_N(k2^{-N})+\delta g_k\colon\ 0\leq\delta\leq 1$$

$$X_{N+1}((k+1/2)2^{-N}+\delta 2^{-N-1})=X_N(k2^{-N})+g_k+\delta g_{k+1}\colon\ 0\leq\delta\leq 1$$

Inductively we get a sequence of processes $\{X_N(t)\}_{N\geq 1}$ whose sample spaces increase. Thus, at each stage "white noise" appears, expressed by a block of 2^N many g-variables. The whole sample space is a denumerable product of the real line on which we define the usual product of the Lebesgue measure. The outcome from a sample point is a sequence of polygons $\gamma_1,\gamma_2,\ldots$. At a dyadic time value $t=k2^{-N}$ the polygons $\{\gamma_\nu\}$ pass the same point, There remains to investigate when the sequence of these polygons converge to a continuous limit curve. It turns out that the convergence holds for almost every sample point which by the above is taken from the denumerable product of copies of the real line $\mathbf{R}$.

5. Continuity of Brownian paths. We shall study the effect of white noise during the passage from stage $N-1$ to the stage N. For this purpose we consider a family $\chi_1,\ldots,\chi_{2^N}$ of independent normally distributed with mean value zero and variance one. Define for each $\alpha>0$:

$$\Pi_N(\alpha)=\text{Prob}\{\max:\ 2^{-N/2}|\chi_k|\geq\alpha : 1\leq k\leq 2^N\}$$

Thus, if $g_k=2^{-N/2}\chi_k$ are the random variables which cause white noise, then $\Pi_N(\alpha)$ is the probability that at least some g_k has absolute value $\geq\alpha$. By the previous discussion this means that we measure the probability for the *maximal distance* between the curves γ_{N-1} and γ_N over the whole time interval. So we want are upper bounds for the Π-numbers with α small as N increases. Here is the crucial inequality:

6. Proposition. *For all pairs N,α one has the inequality*

$$\Pi_N(\alpha)\leq\sqrt{\frac{2}{\pi}}\cdot\frac{2^{N/2}}{\alpha}\cdot e^{-2^N\cdot\alpha^2/2}$$

Proof. Consider first a single normal distribution with variance one. If $A>0$ we have the tail probability

$$\frac{1}{\sqrt{2\pi}}\int_A^\infty e^{-x^2/2}dx\leq\frac{1}{\sqrt{2\pi}}\int_A^\infty\frac{x}{A}\cdot e^{-x^2/2}dx=\frac{1}{\sqrt{2\pi}}\cdot\frac{e^{-A^2/2}}{A}$$

Put

$$A=2^{N/2}\cdot\alpha\quad:\quad\xi=\sqrt{\frac{2}{\pi}}\cdot\frac{1}{A}\cdot e^{-A^2/2}$$

Notice that $1-\Pi_N(\alpha)$ is the probability that $|\chi_k|\leq A$ for every k. The tail inequality applied to every χ_k gives

$$\text{Prob}\{|\chi_k|\geq A\}\geq\sqrt{\frac{2}{\pi}}\cdot\frac{1}{A}\cdot e^{-A^2/2}=\xi$$

The elementary inequality $(1-t)^M\leq Mt$ for every $0<t<1$ therefore gives

$$\Pi_N(\alpha)\leq(1-\xi)^{2^N}\leq 2^N\cdot\xi=2^N\cdot\sqrt{\frac{2}{\pi}}\cdot\frac{1}{A}\cdot e^{-A^2/2}$$

Since $A=2^{N/2}\alpha$ the last inequality proves Proposition 6.

7. Specific choice of α. With $\alpha = 2^{-N/2}\beta$ we have

$$\Pi_N(\alpha) \leq \sqrt{\frac{2}{\pi}} \cdot 2^N \cdot \frac{1}{\beta} \cdot e^{-\beta^2/2} = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\beta} \cdot e^{\,\mathrm{Log}(2)\cdot N - \beta^2/2} \tag{7}$$

Let us choose

$$\beta = [2N \cdot \log 2\,]^{\frac{1}{2}} \quad : \quad \theta > 1$$

Then the last term in (7) becomes

$$\frac{1}{\sqrt{\pi \cdot \log 2}} \cdot 2^{N(1-\theta^2)}$$

So with this specific choice of α we have established the inequality

$$\Pi_N(\alpha) \leq \frac{1}{\sqrt{\pi \cdot \log 2}} \cdot 2^{N(1-\theta^2)} \quad : \quad \alpha = [2N\mathrm{Log}(2)\,]^{\frac{1}{2}} \cdot 2^{-N/2} \tag{8}$$

8. Remark. Keeping $\theta > 1$ fixed and choosing $\alpha(N) = [2N\mathrm{Log}(2)\,]^{\frac{1}{2}} \cdot 2^{-N/2}$ for each N, we get the convergence of the series

$$\sum_{N=1}^{\infty} \Pi_N(\alpha(N)) \tag{9}$$

In the construction of the Brownian motion the "white noise" as we pass from period N to $N+1$ arises from 2^N many independent normally distributed variables, each with variance 2^{-N}. So $\Pi_N(\alpha(N))$ measures the maximal deviation over the whole time interval $[0,1]$ from a polygon at period N to the new polygon at period $N+1$.

The convergent series (9) implies that a single Brownian path is almost surely continuous and the choice of $\{\alpha(N)\}$ entails that the resulting limit curves are almost surely Hölder continuous of order $\sqrt{t}| \cdot \mathrm{Log}(\frac{1}{t})$. Summing up we have established the following:

9. Theorem. *A Brownian path is almost surely Hölder continuous of order* $\sqrt{t} \cdot \log(\frac{1}{t})$.

10. A result by Khintchine. Theorem 9 asserts that almost every Brownian path is Hölder continuous of order $\sqrt{t} \cdot \mathrm{Log}(\frac{1}{t})$ for *every* value of t. A relaxed condition would be request that Hölder continuity which holds for almost all time values. This leads to a result due to Khintchine which asserts that almost every Brownian path is Hölder continuous of order

$$\sqrt{t} \cdot \mathrm{Log}(\mathrm{Log}(\frac{1}{t})) \tag{*}$$

for almost every time value. Notice that since we have taken a double Log-function we come closer to Hölder continuity of order $\frac{1}{2}$ in (*). The proof of (*) uses Khintchine's *Law of the iterated logarithm.* For the proof of (*) we refer to [Khi] and the interested reader may also consult his excellent text-book on statistical mechanics in [Khi] which offers a good introduction for students of mathematics without prerequisite knowledge about physical laws.

8.2 Parabolic PDE-equations and stochastic processes.

Following original work by Laplace and Fourier we derive the parabolic PDE-equation which is used to express the joint probability distribution of a vector-valued stochastic process in continuous time. Let $n \geq 1$ and consider a process $X(t) = (X_1(t), \ldots X_n(t)$ which is governed by infinitesmal deterministic drift together with a Brownian motion:

$$dX_k(t) = b_k(X(t),t) \cdot dt + \sigma_k(X(t),t) \cdot dW_k(t) \quad : 1 \leq k \leq n \tag{*}$$

Above $\{b_k(x_1,\ldots,x_n,t)\}$ and $\{\sigma_k(x_1,\ldots,x_n,t)\}$ are real-valued functions defined for $x \in \mathbf{R}^n$ and $t \geq 0$ and $W(t) = (W_1(t),\ldots,W_n(t)$ is a vector-valued Gaussian distribution expressed by a density function $g(\omega_1,\ldots,\omega_n;t)$ where no drift occurs, i.e.

$$\text{(i)} \qquad \int_{\mathbf{R}^n} \omega_k \cdot g(\omega;t)\, d\omega = 0 \quad : 1 \leq k \leq n$$

We have also covariance functions

$$\text{(ii)} \qquad B_{jk}(t) = \int_{\mathbf{R}^n} \omega_j \omega_k \cdot g(\omega;t)\, d\omega \quad : 1 \leq j,k \leq n$$

At time $t = 0$ the random variable $X(0)$ has a joint density distribution $f_0(x) = f_0(x_1,\ldots,x_n)$. If $t > 0$ we denote by $f(x;t)$ the density of $X(t)$ and $\{\partial_j = \frac{\partial}{\partial x_j}\}$ are derivation operators with respect to the x-variables.

8.2.1. Theorem. *The function $f(x,t)$ satisfies the PDE-equation:*

$$\text{(8.2.1)} \qquad \frac{\partial f}{\partial t} = \sum_{k=1}^{k=n} \partial_k(b_k \cdot f) + \frac{1}{2} \sum\sum B_{j,k}(t) \cdot \partial_j \partial_k(\sigma_j \sigma_k \cdot f)$$

Proof. Introduce the characteristic function

$$\phi(\xi;t) = \int_{\mathbf{R}^n} e^{i\langle \xi,x\rangle}\, f(x;t)\, dx$$

By definition (*) means that if Δ is a small positive number then

$$\phi(\xi;t+\Delta) = \iint_{\mathbf{R}^n \times \mathbf{R}^n} \exp\Big[\sum_{k=1}^{k=n} i\xi_k(x_k + b_k(x;t)\cdot\Delta + \sigma_k(x;t)\omega_k \cdot \sqrt{\Delta}\Big] \cdot f(x;t) g(\omega;t)\, dx d\omega$$

Then (i) and (ii) together with a Taylor expansion up to order two gives:

$$\lim_{\Delta \to 0} \frac{\phi(\xi;t+\Delta) - \phi(x;t)}{\Delta} =$$

$$\sum_{k=1}^{k=n} i\xi_k \cdot \int_{\mathbf{R}^n} e^{i\langle \xi,x\rangle} b_k(x,t) \cdot f(x;t)\, dx - \frac{1}{2} \sum\sum B_{j,k}(t) \cdot \xi_j \xi_k \cdot \int_{\mathbf{R}^n} e^{i\langle \xi,x\rangle} \sigma_j(x,t) \cdot \sigma_k(x,t) \cdot f(x;t)\, dx$$

Now (8.2.1) follows from Fourier's inversion formula.

8.2.2 Remark. In order that Fourier's inversion formula applies one should assume that the functions $\{b_k\}$ are in L^1 and the σ-functions are in L^2. In practice more regularity is imposed, i.e. the σ-functions are of class C^2 and the b-functions of class C^1 in which case the parabolic PDE-equation (8.2.1) is expressed without distribution derivatives. One refers to (8.2.1) as a diffusion equation and in the context of PDE-theory the equation in Theorem 8.2.1 is said to be parabolic. Some text-books prove Theorem 8.2.1 via the "Ito Calculus". Personally I prefer the original proof by Laplace and Fourier above where the parabolic PDE-equation is established under relaxed conditions on the b- and the σ-functions. Moreover, Parseval's equation gives a precise insight about L^2-properties.

8.2.3 Uniqueness results. Already the case $n = 1$ leads to subtle problems. In fact, it was discovered by Gevrey and Hadamard that solutions to the basic heat equation

$$\frac{\partial f}{\partial t}(x,t) = f''_{xx}(x,t)$$

in general are not unique, i.e. non-trvial solutions can exist even if $x \mapsto f(x,t_0)$ vanishes identically for some time value t_0. It is ony in the analytic case where solutions expressed are expressed by the integral formula

$$f(x,t) = \frac{1}{\pi \cdot \sqrt{t}} \int_{-\infty}^{\infty} e^{(x-\xi)^2/4t}\, f(\xi,0)\, d\xi$$

However, this integral formula is not valid if $x \mapsto f(x,0)$ increases too fast. Here an involved analysis is needed which goes beyond elementary studies of stochastic PDE-equations. For results concerned with the non-analytic case we refer to work by Beurling and Holmgren, which is exposed in § xx.

9. Riesz' representation theorem.

Let S be a compact Hausdorff space and $\mathcal{B}_*(S)$ the Boooelan algebra generated by closed sets. Let μ be a non-negative and additive function on $\mathcal{B}_*(S)$. So in particular $\mu(S)$ is finite. Moreover, the additivity entails that

$$\big|\sum a_k \cdot \mu(E_k)\big| \leq \max\{|a_k|\} \cdot \mu(S) \tag{1}$$

for every finite family of disjoint sets $\{Ek\}$ and real a-numbers. Next, define a function μ^* on $\mathcal{B}_*(S)$ by

$$\mu^*(A) = \min_{A \subset U} \mu(U) \tag{2}$$

where the mimium is taken over open sets U which contain A.

9.1 Exercise. Show that μ^* is additive. The hint is that $\mu^*(U) = \mu(U)$ for open sets and in the next step one verifies that $\mu(F_1 \cup F_2) + \mu(F_1) + \mu(F_2)$ hold for each pair of disjoint closed sets.

9.2. Riesz' equality. As already explained in § 0.1 from the introduction the two additive measures μ and μ^* yield integrals with respect to continuos functions on S. Now we show the equality below for every $f \in C^0(S)$:

$$\int f \cdot d\mu^* = \int f \cdot d\mu \tag{9.2.1}$$

Proof. Since every continuous function is the difference of two non-negative functions it suffices to prove (*) when $f \geq 0$ and after scaling that $0 \leq f \leq 1$. We may also assume that $\mu(S) = 1$. Given a positive integer N we set for each $0 \leq k \leq N-1$:

$$a_k = \max_{k/N \leq x \leq (k+1)/N} f(x) : \; b_k = \min_{k/N \leq x \leq (k+2)/N} f(x)$$

$$A_k = \{k/N < f \leq (k+1)/N\}$$

Since μ^* is non-negative one has

$$\int f \cdot d\mu^* \leq \sum a_k \cdot \mu^*(A_k) \tag{i}$$

Next, if $\epsilon > 0$ the construction of μ^* gives an N-tuple of open sets $\{U_k\}$ where U_k contains A_k and

$$\mu^*(A_k) > \mu(U_k) - \epsilon/N \tag{ii}$$

Above we can shrink U_k and assume that

$$U_k \subset \{k/N < f(x) < (k+2)/N\} \tag{iii}$$

From (ii) a summation gives

$$\epsilon + \int f \cdot d\mu^* \leq \sum a_k \cdot \mu(U_k) \tag{iv}$$

Set

$$\omega_f(N) = \max\{a_k - b_k\}$$

The normalisation $\mu(S) = 1$ gives

$$\omega_f(N) + \epsilon + \int f \cdot d\mu^* \leq \sum b_k \cdot \mu(U_k) \tag{v}$$

Next, let $\{\chi_k\}$ be a partition of the unity with respect to $\{U_k\}$, i.e. the χ-functions are continuous and each χ_k has compact support in U_k while their sum is identically one on S. The definition of the b-numbers entail that $f \geq \sum b_k \cdot \chi_{U_k}$ and (v) gives

$$\omega_f(N) + \epsilon + \int f \cdot d\mu^* \leq \int f \cdot d\mu \tag{vi}$$

Above N is arbitrary large and by the uniform continuity of f we have $\omega_f(N) \to 0$ as $N \to \infty$. Since ϵ can be arbitrary small we conclude that

$$\int f \cdot d\mu^* \leq \int f \cdot d\mu$$

In the same way we get

$$\int (1-f) \cdot d\mu^* \leq \int (1-f) \cdot d\mu$$

Since $\mu^*(S) = \mu(S)$ we conclude that (9.2.1) holds.

9.3 The σ-additivity of μ^*.

To simplify notations we set $\mu^* = \mu$. If $A \in \mathcal{B}_*(S)$ the construction from (2) applied to its complement gives

$$\mu(A) = \max_{F \subset A} \mu(F) \tag{9.3.1}$$

with the maximum taken over closed subsets. Let us then take a denumerable sequence $\{A_n\}$ of pairwise disjoint sets in $\mathcal{B}_*(S)$ and set $A^* = \cup A_n$. Define the outer measure by

$$\widehat{\mu}(A^*) = \min_{A^* \subset U} \mu(U)$$

with the minimum taken over open sets which contain A^*. Given $\epsilon > 0$ we choose U so that

$$\mu(U) < \widehat{\mu}(A^*) + \epsilon$$

Next, apply (9.3.1) to A_n which gives an closed set $F_n \subset A_n$ such that

$$\mu(A_n) \leq \mu(F_n) + \epsilon \cdot 2^{-n}$$

Since μ is non-negative we have

$$\sum_{n=1}^{n=N} \mu(F_n) \leq \mu(U)$$

This hold for every N and since $\epsilon > 0$ was arbitrary we conclude that

$$\sum \mu(A_n) \leq \widehat{\mu}(A^*) + \epsilon$$

Above ϵ was arbitrary and hence

$$\sum \mu(A_n) \leq \widehat{\mu}(A^*) \tag{*}$$

To prove that equality holds in (*) we use (2) and find open sets $\{U_n\}$ with $A_n \subset U_n$ such that

$$\mu(U_n) \leq \mu(A_n) + \epsilon \cdot 2^{-n} \tag{i}$$

Since $\cup U_n$ is an open set which contains A^* we have

$$\widehat{\mu}(A^*) \leq \mu(\cup U_n) \tag{ii}$$

Next, apply (9.4.1) to $\cup U_n$ which gives a compact subset F such that

$$\mu(\cup U_n) \leq \epsilon + \mu(F) \tag{iii}$$

Heine-Borel's Lemma gives an integer N such that F is contained in the finite union $U_1 \cup \ldots \cup U_N$. Hence (i-iii) give

$$\widehat{\mu}(A^*) \leq \epsilon + \sum_{n=1}^{n=N} \mu(U_n) \tag{iv}$$

Finally, by (i) the last term is majorized by $\epsilon + \mu(A_1) + \ldots + \mu(A_n)$ which entails that

$$\widehat{\mu}(A^*) \leq 2\epsilon + \sum \mu(A_n)$$

Hence equality holds in (*). One can continue by an induction in the transfinite construction of Borel's σ-algebra and conclude that μ has a unique extension to a σ-additive measure to the Borel algebra $\mathfrak{B}(S)$.

Remark. The reader may consult [Kolmogorov] for a further account about consistency during the construction of σ-additive measures.

9.4 The case of signed measures.

In general, let μ be a signed measure, i.e. an additvie function on $\mathcal{B}_*(S)$ where its is assumed that there exists a constant C such that $|\mu(A)| \leq C$ for ever $A \in \mathcal{B}_*(S)$. Set

$$\mu_+(A) = \max_{F \subset A} \mu(F)$$

where the maximum is taken over subsets of A in $\mathcal{B}_*(S)$. The reader can check that μ_+ is additive. We also define

$$\mu_(A) = \min_{F \subset A} \mu(F)$$

This yields another additive function and the reader can verify the equality

$$\mu(A) = \mu_+(A) + \mu_(A)$$

From the construction μ_+ is non-negative and we get μ_+^*. Similarly, μ_- is non-negative and we get the non-negative measure $(-\mu_-)^*$. Set

$$\text{(xxx)} \qquad \mu^* = \mu_+^* - (-\mu_-)^*$$

So this is the difference of two non-negative and measures which both extend to σ-additive measures on $\mathfrak{B}(S)$.

9.4.1 Exercise. Show that

$$\int f \cdot d\mu^* = \int f \cdot d\mu$$

hold for continuous functions. Moreover, show that (xxx) corresponds to the Hahn-decomposition of the signed σ-additive measure μ^*as in § xx.

Riesz representation formula. Let S be a compact topolgical space. We assume that is is Hausdorff and recall from a general result in topology that S is normal, i.e. for each pair of disjoint closed sets F and G there exists a continuous function f whose range is the closed unit interval while $f = 1$ on F and zero on G. Now there exists the Banach space $C^0(S)$ of real-valued continuous functions equipped with the maximum norm. Next, we have the Boolean algebra $\mathcal{B}_*(S)$ generated by closed subsets of S. By set-theory one constructs the unique smallest σ-algebra which contains $\mathcal{B}_*(S)$. It is denoted by $\mathcal{B}(S)$ and called the Borel algebra of subsets in S. We remark that it can be found via a certain transfinite induction which was carried out in great detail by Baire. Now there exists the algebra $\mathfrak{B}^\infty(S)$ of bounded real-valued Borel functions. Such a function f has a bounded range and for every pair of real numbers $a < b$, the set $\{a \leq f < b\}$ belongs to $\mathcal{B}(S)$. Under the maxium norm it is clear that $\mathfrak{B}^\infty(S)$ is a Banach space. Since uniformly convergent sequences of continuous functions have a continuous limit, it follows that $C^0(S)$ is a closed subspace of $\mathfrak{B}^\infty(S)$. Next, consider the dual space of $\mathfrak{B}^\infty(S)$. Let μ be a continuous linear functional. To each Borel set E in S we associate the characteristic function χ_E and get the real number

$$\mu(E) = \mu(\chi_E)$$

It is clear that

$$E \mapsto \mu(E)$$

is an additive map on $\mathcal{B}(S)$. Moreover, if $||\mu||$ is the norm of the functional μ one has the inequality

(*) $$\left|\sum a_k \cdot \mu(E_k)\right| \leq \max |a_k|$$

for every finite and disjoint family of Borel sets and real a-numbers. Conversely, the reader may verify that if μ is an additive function on $\mathcal{B}(S)$ for which (*) holds, then it gives a continuous linear functional on $\mathfrak{B}^\infty(S)$. Thus, the dual space of $\mathfrak{B}^\infty(S)$ can be identified with additive μ-functions satisfying (*). Next, for each such μ, Borel's construction give integrals

(1) $$\int_S f \dot{d}\mu$$

when $f \in C^0(S)$. This corresponds to the restriction map from the dual of $\mathfrak{B}^\infty(S)$ to the dual of $C^0(S)$. The Hahn-Banach theorem shows that this map is surjective and hence every vector $\gamma \in C^0(S)^*$ is represented by some μ as above.

The Riesz' Theorem. The inclusion $C^0(S) \subset \mathfrak{B}^\infty(S)$ is strict so via the Hahn-Banach theorem there exist certain μ for which the integrals in (1) are zero for every continuous function f. So if we start with a vector $\gamma \in C^0(S)^*$ the question arises if there exists a specific μ which represents γ in the sense that

(2) $$\gamma(f) = \int_S f \cdot d\mu$$

hold for every continuous function. Riesz proved that when γ is given, then there exists a unique μ such that (2) holds, where μ in addition is σ-additive. This condition can be expressed by saying that

(3) $$\lim_{n\to\infty} \mu(E_n) = 0$$

for every decreasing sequence $\{E_n\}$ in $\mathcal{B}(S)$ for which $\cap E_n = \emptyset$. The class of those μ for which both (*) and (3) hold are called Riesz measures on S, and from the above this class is isomorphic to the dual of $C^0(S)$. We shall give a more detailed exposition of this important result in § 9 after we have become familiar with the material in § 1. The case when S is replaced by an euclidian space is treated in § 4 where the constructions are self-contained since one can profit upon the geometry in $\mathbf{R}^n$.