# The Development of the Modern Integration Theory from Cauchy to Lebesgue: a Historical and Epistemological Study with Didactical Implications

Georgeana Bobos-Kristof

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Master of Science (Mathematics) at
Concordia University
Montréal, Québec, Canada

August 2007

# ABSTRACT

## The development of the modern integration theory from Cauchy to Lebesgue: a historical and epistemological study with didactical implications

Georgeana Bobos-Kristof

In this thesis, we describe the historical development of the modern integration theory by presenting the key ideas and insights that led to its shaping, from the first rigorous definition of the definite integral given by Cauchy in 1823, to Lebesgue's theory as it first appeared in his doctoral thesis of 1902. In the final part we also present recent approaches in integration theory. We show that various problems motivated the enrichment of the notion of integral, while one in particular constituted the most important trigger of this development from the beginnings well into the 20th century: the search for a better understanding of Fourier series.

For our study, we principally look at original sources, and we provide detailed proofs for important results, often by elaborating on the authors' sketchy or heuristic arguments, making thus the original results accessible to the modern reader.

We then use this historical and epistemological analysis to raise some issues related to the teaching of integration at various levels at university, i.e., in calculus, analysis and measure theory courses. For this, we look at some typical textbooks and, inspired by the historical analysis, we give some suggestions for teaching. Our findings show that there might be some conceptual gaps between subsequent levels that are not necessarily insuperable, but require careful didactical analysis.

# ACKNOWLEDGEMENTS

I was lucky to have Dr. Galia Dafni as my first teacher in graduate mathematics and as my thesis advisor. She not only guided me wonderfully throughout the project, but also, as a teacher and mathematician, was great inspiration for my work in this thesis.

I wish to thank Dr. Anna Sierpinska for her insightful critique, which, as always, has made my work better. My gratitude goes to her also for the constant support she provided me with throughout my studies at Concordia.

I would like to thank Dr. Proppe for the time he took to review my thesis, and for the interest he showed in my work.

Thanks to my dear Norbert and Maya, for bearing with me so lovingly during the writing of this thesis.

I dedicate this thesis to my mother. Mama, thanks for being here for me.

# Contents

# List of Figures

# Chapter 1

## INTRODUCTION

### 1.1 Scope of the thesis

The foundations of the modern theory of integration were laid by the French mathematician Henri Lebesgue at the beginning of the 20th century. Around the 1850's, Riemann's interest in integrable functions was at the start of what was to become, with the work of Lebesgue about a half a century later, a greatly generalized and abstract theory of integrals. The major breakthrough of Riemann's theory was that, in variance with the earlier work in integration, he considered the class of functions to which the process of integration can be applied. In other words, he was the first mathematician to ask the question: *When can we integrate?* He used a definition of the integral that was very close to the earlier one of Cauchy, but, unlike Cauchy, he not only considered continuous functions, but also attempted – and succeeded – to establish a necessary and sufficient criterion for integrability, significantly weakening the assumptions on the function to be integrated. At the time, the class of functions that fell under his definition of integrable seemed to have comprised all known functions, and, indeed, seemed to assure the largest generality possible.

It was not until the full development of a measure theoretic point of view and by complete withdrawal from the classical definition of the integral, which involved the Cauchy sums, that a further generalization of the integral could be afforded, and ultimately realized by Lebesgue.

In Riemann's theory, the integral of a function $f$ on an interval $[a, b]$ can be approximated by sums of the form:

$$\sum_{i=1}^{n} f(x_i)\, l(I_i)$$

where $I_1, I_2, \ldots, I_n$ are intervals such that $\bigcup_{i=1}^{n} I_i = [a, b]$, $l(I_i)$ is the length of the interval $I_i$, and $x_i$ is an arbitrary value in $I_i$. Now, Lebesgue's theory generalizes the notion of length to that of <u>measure</u>, the class of sets $I_i$ to <u>measurable sets</u>, and the class of functions to which this process is applicable to the class of <u>measurable functions</u>. Countable unions and intersections of measurable sets are also measurable and a (positive) measure is countably additive, that is, if $\{I_n\}$ is a countable collection of pairwise disjoint measurable sets then:

$$l(\bigcup_n I_n) = \sum_n l(I_n)$$

The general aim of this project is to the describe the mathematical developments that took place historically to allow the generalization of the notion of definite integral. As is the case for other concepts and theories in mathematics, these developments were not, by far, as natural as it may appear from the above paragraphs. Integrals were known and used for centuries before the developments we have described. But it took Cauchy's preoccupation with rigor to provide a *definition* of the definite integral, Riemann's acceptance of the modern concept of function to introduce the concept of *integrability*, Cantor's new radical approach to lay convenient *conceptual foundations* for the definitions of measure related notions, a formalization of *geometric intuitions* of the integral undertaken by Peano and Jordan through the notions of content, and finally, Borel's insight to provide the desired *properties of a* *generalized measure*, before Lebesgue took the apparently distinct perspectives and put them together to create a new theory of integration.

These series of insights were not motivated solely by mathematicians' desire to create a more general theory of integration. The results mentioned above have often been obtained in the context of seeking solutions to very different problems, sometimes without realizing their usefulness for a general theory of integration. For example, commenting on his own notion of measure, Borel [5], goes so far as to state that it is totally unrelated to integration. We will show that various problems/interests, of philosophical, mathematical, or physical nature motivated the enrichment of the notion of integral, and one, in particular, constituted the most important trigger of this development: the search for a better understanding of Fourier series. Fourier's original work was fraught with contradictions, but it contained bold and broad insights whose exploration kept mathematicians busy for a century and more. This historical fact illustrates well the view, expressed in the following citation from Davis and Hersh's *The Mathematical Experience* [19] , "How completely inadequate it is to limit the history of mathematics to the history of what has been formalized and made rigorous. The unrigorous and contradictory play important parts in this history" (see also Byers, *How mathematicians think: using ambiguity, contradiction, and paradox to create mathematics* [7]).

Another aim of the study is to explain the significance of Lebesgue's work, beyond the elegant generalization it provided. As Lebesgue stated ([37], p. 194): "[...] a generalization made not for the vain pleasure of generalizing but in order to solve [...] problems is always a fruitful generalization". Our argument will be based on the application of the Lebesgue integral in Fourier analysis.

Finally, we attach a didactical value to our study. Integration is one topic that recurs at all levels in university mathematics: once taught in Calculus courses, it is then "re-taught" in undergraduate analysis, undergraduate and graduate measure theory courses, using different mathematical and didactic organizations of the content taught (see Chevallard [14]). In the thesis, we want to examine these organizations in more detail, and, by drawing inspiration from the gradual enrichment of the notion of integral in history, to call attention to the objectives that the teaching of integration has or should have and to the problems that may arise in this context at the different levels.

## 1.2 Context and originality of our study

In the previous section we presented an abridged "story" of the key developments that paved the way to Lebesgue's theory of integration. In the study, we will develop this story in more detail: the relevant mathematical results will be presented in full. However, in our approach we will try to bring out the *great ideas* that emerged from each author's work to come together in the new theory.

On the other hand, we will try not to give the impression that the development of the theory has been taking place in a straightforward, natural and cumulative way. In fact, Lebesgue's inspiration came from apparently incompatible perspectives, which he brought together to create a new, more powerful theory. We will therefore also draw attention to the motivations that led to the discovery of the presented mathematical results.

Thus we will focus more on the conceptual enrichment of the notion of integral, and on the problems of various nature that triggered it, and not so much on providing a chronicle of actual historical facts and events.

A similar approach is undertaken by Hawkins [27] in his book *Lebesgue's theory of integration. Its origins and development.* He underlines the fact that the results obtained by Lebesgue's predecessors had a great significance for the development of what he terms as "the first genuine theory of integration" ([27], p.ix.), because they provided Lebesgue both with a useful set-theoretic support and with unresolved problems that would *allow* and *motivate* the new theory. His historical account comprehensively traces these contributions and the problems that triggered them and, in that, it served as a guide for us. Our study is certainly of a narrower breadth in the sense that it is less concerned with the historical aspects largely portrayed by Hawkins. Given the scope of the thesis we cannot give justice to the many aspects that were part of the historical context, and which have certainly contributed to the development of integration theory as we know it today. We will focus on only a few ideas, that we have considered to be the key concepts in the advancement of the theory, but we will provide, for these results, a more detailed mathematical exposition, by principally looking at the original sources. In most instances, however, we will "enhance" the original proofs, which were often quite succinct and heuristic in nature, by providing more details and thus making them readable by the modern audience.

In that we are closer to Pesin's [42] approach, who comprises a higher level of mathematical detail, with complete definitions, theorems and proofs from original papers and treatises. His book, however, is closer to a survey of the body of mathematical results obtained historically, being less concerned with the various motivations that constituted the driving forces behind the mathematical results presented.

Van Dalen and Monna's study [38] gives an overview of the developments in set theory and integration in two separate sections of the book, also by giving a chronological presentation of the relevant mathematical results. To some extent, this approach does not illustrate the fact that the two are inextricably linked. Nevertheless, their book was an important source for our account of the development of the measure-theoretic support for integration theory.

In this context, with our approach, we endeavor to address a larger audience: our study is detailed enough mathematically to be interesting for mathematicians and mathematics students interested in the development of classical mathematical ideas, without, however, obscuring the reading with too many historical factual details. Besides, as we have mentioned earlier, we will also provide some didactical reflections, and thus appeal to readers interested in the teaching of the subject: mathematics educators or instructors of calculus or analysis.

## 1.3 Outline of the thesis

We begin the second chapter with a short overview of the existing knowledge in the domain of integration, with the advent of calculus by Newton and Leibniz. The main body of this chapter, however, is devoted to the contributions of Cauchy and Riemann, as we believe that the beginning of integration as a *theory* is marked by the definition of the definite integral introduced by Cauchy. An essential section in this chapter describes *the Fourier series problem*, that initiated Riemann's motivation to look more closely at the notion of integration. We also include here the convenient reformulation of Riemann's definition and criterion of integrability provided by Peano and Darboux. Another section is aimed at describing the importance of Riemann's insight in the subsequent development of the theory of integration. We conclude this chapter by reviewing the drawbacks of Riemann's theory that became evident by the end of the century.

The third chapter addresses the development of measure-theoretic notions and their relation to integration. Here, we highlight the insights provided by the works of Cantor, Peano, Jordan, and Borel, which made the generalizability of Riemann's definition of the integral more apparent.

In the fourth chapter we describe Lebesgue's notion of integral and measure, as introduced in his doctoral thesis. We will also include, in this chapter, a section that illustrates the significance of this new integration theory.

In the final chapter we look at mathematical and didactic organizations of integration theories in a few modern textbooks used in calculus, analysis and measure theory courses. Our critical reflection on these organizations is inspired by our study of the historical development of these theories. We conclude the thesis with some thoughts at the end of this journey through the insights that shaped the modern integration theory.

# Chapter 2

## THE CAUCHY AND RIEMANN INTEGRALS: THE BEGINNINGS OF AN INTEGRATION THEORY

In this chapter we look at the development of modern integration theory, the beginnings of which, we will show, were marked by the first rigorous definition of the definite integral, given by Cauchy at the beginning of the 19th century. We describe in detail the mathematical results obtained by Cauchy and Riemann on the topic of integration, focusing on the gradual enrichment of the concept of integral.

## 2.1 Introduction

Integration goes back far more than the above-mentioned 19th century mark. It first arose as a method for finding areas. Its roots can be found in the work of Archimedes (3rd century BC), when a given area was approximated by dividing it into rectangles or other shapes of known area, which were made smaller and smaller – as to approximate the desired area closely – through some sort of limiting process (see, for example, [38]).

In the 17th century, in *The Mathematical Principles of Natural Philosophy* [39], Newton studied the computation of areas through the problem of "quadrature", i.e. the problem of finding a square whose area is equal to a given area. He gave a

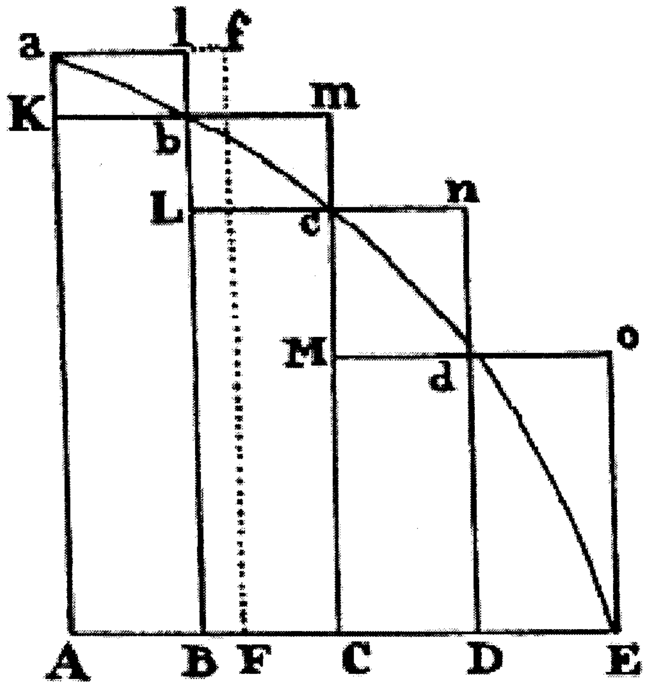


Figure 2.1: Area under a decreasing curve from Newton's *Principia*

procedure of calculating areas under curves, and justified it by an argument that may not satisfy the modern requirements of rigor, but is certainly sound. We summarize it below:

To approximate the area in the figure, Newton divides the domain into equal subintervals (see Figure 2.1). To each subinterval there correspond two rectangles: one whose height is the maximum value of the function on that interval and the other whose height is the minimum value of the function on the same interval. The desired area is between the sum of the areas of the circumscribed rectangles and the sum of the areas of the inscribed rectangles. The difference between these two areas is the sum of the areas of the rectangles *aKbl*, *bLcm*, *cMdn*, *dDEo*. Now, if these rectangles were all aligned under the upper most rectangle, *aKbl*, it's easy to see that the sum of their areas is equal to the change in height of the function multiplied by the length of any subinterval. Therefore, the difference in the areas can be made to approach zero, by taking smaller and smaller subintervals. According to Newton, all these areas approach the same value as the length of the subinterval approaches 0.

Newton also considers the case where the subintervals are of different length by drawing the dotted line $fF$ (see Figure 2.1). He notes that the sum of the differences of the areas is still less than the change in height multiplied by the length of the longest subinterval. This means that one will get the same limit for the ratio if the length of the longest subinterval approaches 0.

For Leibniz these small increments existed as a special kind of numbers: *infinitesimals* (see [15]). He came up with the notation $dx$ to stand for the difference between two infinitely close values of $x$, $dy$ for the difference of two infinitely close values of $y$, and $dy/dx$ for their ratio. He eventually used the following notation to denote the total area under the curve, obtained as a sum (an elongated $S$) of the increments in the area:

$$\int y\,dx$$

As Cooke [15] shows, this notation, although lacking a logical basis, was very productive at the time. The fundamental theorem of calculus followed somewhat more intuitively from his approach and notation: "Leibniz could argue that the ordinates to the points on a curve – *summing all the lines in the figure* – amounted to summing infinitesimal differences in area $dA$, which collapsed to give the total area. Since it was obvious that on the infinitesimal level $dA = ydx$, the fundamental theorem of calculus was an immediate consequence" ([15], p. 470).

The dual understanding of the integral – as a limit of sums of products and as the inverse of differentiation – that came about with Newton's and Leibniz's work, is what lies at the heart of calculus. Until the 19th century this understanding was certainly held and exploited by the scientific community, but a definition of the integral as the sum of products of $f(x)$ times the infinitesimal $dx$ was avoided, because this would have required a precise definition of the notion of "infinitesimal". Instead, it became customary to define integration as the inverse of differentiation, in other words, $\int$ applied to a function $f(x)$ is an operator that returns the function(s) whose derivative is $f$.

## 2.2 Problems with calculus

By the turn of the 18th century, much of what is today is called Calculus was already well known and included in the period's textbooks. The rules for differentiating and integrating elementary functions, the expansion of functions into power series or the solving of simple differential equations were used as powerful tools of Newton and Leibniz's calculus for a century thereafter. However, several pressing mathematical and foundational questions started to arise, in particular in relation to the notion of integral. The most obvious one was the problem of nonelementary integrals: one could differentiate functions by an algorithmic procedure, but that was not the case for integration [15]. Functions arising in fairly basic physical situations, such as $sinx/x$ or $e^{-x^2}$, turned out not to be the derivatives of elementary functions (an *elementary function* is defined as a function that can be obtained by addition, multiplication, division and composition from the rational functions, the trigonometric functions and their inverses, and the exponential and logarithmic functions). Another set of problems came up in relation to the subject of differential equations, where integration had become the main tool. Here, in more and more cases, again, modeling physical, and especially mechanical, phenomena, the equations were not reducible to a form where the solution could be obtained by integrating both sides to eliminate the derivatives. Finally, the foundational difficulties related to the introduction of the "infinite" in mathematics, specifically through the use of "infinitesimal" methods, first raised in 1734 by Berkley, were the source of many a debate during the eighteenth century [15].

## 2.3 Cauchy: the definition of the definite integral

Cauchy managed to address some aspects of these problems, in particular those regarding the foundational aspects of calculus. He is largely recognized to be the first mathematician who succeeded in providing the much needed rigor in calculus, by finding a way to avoid infinitesimals, and thus making the passage from *calculus* ("a technology" in terms of Chevallard's praxeology theory, [14]) to a theory of this calculus, now called *mathematical analysis.* Limiting processes had been used as a technique long before Cauchy, but he is generally credited with having constructed the notion of limit as a fundamental concept of analysis. In his *Cours d'Analyse* [11] he also gave a definition of continuity that was virtually the same as the one we are using today: "[...] la fonction $f(x)$ sera, entre les deux limites assignées à la variable $x$, fonction continue de cette variable, si, pour chaque valeur de $x$ intermédiaire entre ces limites, la valeur numérique de la différence $f(x+a) - f(x)$ décroît indéfiniment avec celle de $a$ ".

With Cauchy, integration knew an important theoretical advancement: in 1823, he defined the definite integral of a continuous function using finite approximating sums [12]:

Let $f : [a, b] \longmapsto \mathbb{R}$ be a continuous function. Let $P = \{x_i\}_{i=0}^{n}$ be a partition of $[a, b]$, i.e. $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$, and define the norm of the partition $P$ as the number

$$\Delta(P) \stackrel{def}{=} \max_{i=0}^{n-1}(x_{i+1} - x_i).$$

The Cauchy sum, associated to $f$ and a partition $P$, is defined by

$$S(P) = \sum_{i=1}^{n} (x_i - x_{i-1})\, f(x_{i-1}).$$

The sums $S$ posses a limit as the norm of $P$ approaches 0, which is called *the definite integral*, and denoted by $\int_a^b f(x)\, dx$.

Using various theorems of his *Cours d'Analyse* [11], Cauchy proved the existence of this limit, by showing that for any two partitions $P$, the corresponding sums approach the same limit, i.e.: the difference between them can be made arbitrarily small by restricting the length of the subintervals they form. Although his proof is somewhat vague by today's standards – for instance, Cauchy uses uniform continuity without distinguishing it from continuity – it contains the essential ideas of the proof we provide below:

Define, first, for every $\delta > 0$, the modulus of continuity (a term which was introduced later by Lebesgue [56]):

$$\omega_\delta(f) = \sup\{\, |f(x) - f(y)| \,:\, x, y \in [a,b]\,,\, |x - y| \leq \delta \,\} \tag{2.1}$$

and note that the following equality describes the property of uniform continuity of $f$:

$$\lim_{\delta \mapsto 0} \omega_\delta(f) = 0. \tag{2.2}$$

At the time, this fact was unknown – or at least it wasn't as clear as it is to us today – but Cauchy seemed to have intuited this property and he used it somehow between the lines.

Assume that $P' = \{y_j\}_{j=0}^{m}$ is a partition of $[a,b]$, which is finer than $P$, i.e. $\{x_i\} \subset \{y_j\}$. Then, for two consecutive points of $P$, $a \leq x_{i-1} < x_i \leq b$, we have:

$$x_{i-1} = y_{i_k} < y_{i_k+1} < \cdots y_{i_k+p} = x_i,$$

where $y_{i_k}, \dots, y_{i_k+p}$ belong to $P'$. Then

$$(x_i - x_{i-1})\, f(x_{i-1}) = \sum_{s=1}^{p} (y_{i_k+s} - y_{i_k+s-1}) f(x_{i-1}),$$

hence:

$$\begin{aligned}
&\left| (x_i - x_{i-1})\, f(x_{i-1}) - \sum_{s=1}^{p} (y_{i_k+s} - y_{i_k+s-1})\, f(y_{i_k+s-1}) \right| \\
&= \left| \sum_{s=1}^{p} (y_{i_k+s} - y_{i_k+s-1}) f(x_{i-1}) - \sum_{s=1}^{p} (y_{i_k+s} - y_{i_k+s-1})\, f(y_{k_i+s-1}) \right| \\
&= \left| \sum_{s=1}^{p} (y_{i_k+s} - y_{i_k+s-1})\, (\, f(x_{i-1}) - f(y_{k_i+s-1})\,) \right| \\
&\leq \sum_{s=1}^{p} (y_{i_k+s} - y_{i_k+s-1})\, |\, f(x_{i-1}) - f(y_{k_i+s-1})\, | \\
&\leq \left( \sum_{s=1}^{p} (y_{i_k+s} - y_{i_k+s-1}) \right) \sup \{\, |f(x) - f(y)| \,:\, x, y \in [x_{i-1}, x_i] \,\} \\
&= (x_i - x_{i-1}) \sup \{\, |f(x) - f(y)| \,:\, x, y \in [x_{i-1}, x_i] \,\}.
\end{aligned}$$

The above supremum actually defines the oscillation of $f$ in $[x_{i-1}, x_i]$. But if we take into account the definition of the norm of $P$, and 2.1, we have

$$\sup \{\, |f(x) - f(y)| \,:\, x, y \in [x_{i-1}, x_i] \,\} \leq \omega_{\Delta(P)}(f)$$

and the above computation amounts to:

$$\left| (x_i - x_{i-1})\, f(x_{i-1}) - \sum_{x_{i-1} \leq y_{s-1} < x_i} (y_s - y_{s-1})\, f(y_{s-1}) \right| \leq (x_i - x_{i-1})\, \omega_{\Delta(P)}(f),$$

which gives:

$$\begin{aligned}
|S(P)-S(P')| &= \left|\sum_{i=1}^{n}(x_i-x_{i-1})f(x_{i-1})-\sum_{j=1}^{m}(y_j-y_{j-1})f(y_j)\right| \\
&= \left|\sum_{i=1}^{n}\left((x_i-x_{i-1})f(x_{i-1})-\sum_{x_{i-1}\le y_{s-1}<x_i}(y_s-y_{s-1})f(y_{s-1})\right)\right| \\
&\le \sum_{i=1}^{n}\left|(x_i-x_{i-1})f(x_{i-1})-\sum_{x_{i-1}\le y_{s-1}<x_i}(y_s-y_{s-1})f(y_{s-1})\right| \\
&\le \sum_{i=1}^{n}(x_i-x_{i-1})\,\omega_{\Delta(P)}(f)=\left(\sum_{i=1}^{n}(x_i-x_{i-1})\right)\omega_{\Delta(P)}(f) \\
&= (b-a)\,\omega_{\Delta(P)}(f)\,,
\end{aligned}$$

hence

$$|S(P)-S(P')|\le(b-a)\,\omega_{\Delta(P)}(f), \tag{2.3}$$

which holds for every $P'$ finer than $P$.

Let now $(P_n)_{n\ge1}$ be a sequence of partitions such that

$$P_{n+1} \quad \text{is finer than} \quad P_n \quad \text{and} \quad \Delta(P_n)\longmapsto 0. \tag{2.4}$$

Then (2.3) is equivalent to

$$|S(P_{n+p})-S(P_n)|\le(b-a)\,\omega_{\Delta(P_n)}(f) \quad \text{for every} \quad n,p\ge1,$$

therefore the sequence of reals $(S(P_n))_{n\ge1}$ is a Cauchy sequence, hence it has a limit, say $L$.

Now, let $(P'_n)_{n\ge1}$ be another sequence of partitions verifying (2.4), having the limit, say $L'$. To show that $L=L'$, let $Q_n=P_n\cup P'_n$. Then $Q_n$ is a new partition of

$[a, b]$, which is finer than both $P_n$ and $P'_n$, therefore, by (2.3),

$$|S(P_n) - S(Q_n)| \leq (b-a)\, \omega_{\Delta(P_n)}(f)\,,\, |S(P'_n) - S(Q_n)| \leq (b-a)\, \omega_{\Delta(P'_n)}(f),$$

and

$$\begin{aligned} |L - L'| &\leq |L - S(P_n)| + |S(P_n) - S(Q_n)| + |S(Q_n) - S(P'_n)| + |S(P'_n) - L'| \\ &\leq |L - S(P_n)| + (b-a)\, \omega_{\Delta(P_n)}(f) + (b-a)\, \omega_{\Delta(P'_n)}(f) + |S(P'_n) - L'| \end{aligned}$$

for every $n \geq 1$. Passing to the limit, as $n \mapsto +\infty$, gives $L = L'$.

If follows that $L$ depends only of $f$, and this number is Cauchy's definition for $\int_a^b f(x)\, dx$.

It should be noted that during Cauchy's time even the concept of real number was rather blurry and, in fact, Cauchy was the first to define a real number as a limit of sequences now called Cauchy sequences (see [31]).

Jourdain [31] points out that, by providing this definition for the definite integral, Cauchy reinstated the sum-conception of integral as the fundamental notion of integral calculus. While Leibniz conceived the integral calculus as *calculus summatorius,* in further development of the process of integration, in Euler's work for instance, the notion the integration as antidifferentiation became prevalent. In his *Institutionum calculi integralis* of 1768, Euler defined integral calculus as a method of finding, from a given relation of differentials, the relation of the quantities themselves (see [31]). On the other hand, the sum-conception was used for the approximate evaluation of integrals. Cauchy's essential contribution to integration theory is best captured in the following quote by Jourdain ([31], p. x): "[...] the existence of a limit – an arithmetical translation of the geometrical "area" – of certain sums formed with the aid of a continuous function $f(x)$ was proved to exist and called [...] *definite integral* [...]. Cauchy's manner of founding the integral calculus showed, by construction, the existence of the class of functions $F(x)$ which admit for derivative a given continuous function $f(x)$. Before Cauchy, such integrals were found, and thus shown to exist, for many $f(x)$'s; but Cauchy proved the *general* proposition."

We would like to add a final remark here: the original Cauchy sum

$$S(P) = \sum_{i=1}^{n}(x_i - x_{i-1})f(x_{i-1})$$

requires a particular intermediate point in every interval $[x_{i-1}, x_i]$, namely the left end-point $x_{i-1}$. Now, if we choose an arbitrary system of intermediate points $\xi_i \in [x_{i-1}, x_i]$, and set

$$R(P) \stackrel{def}{=} \sum_{i=1}^{n}(x_i - x_{i-1})f(\xi_i),$$

then for continuous functions:

$$|S(P) - R(P)| = \left|\sum_{i=1}^{n}(x_i - x_{i-1})(f(x_{i-1} - f(\xi_i))\right| \leq$$

$$\leq \sum_{i=1}^{n}(x_i - x_{i-1})\,|f(x_{i-1}) - f(\xi_i)| \leq (b-a)\omega_{\Delta(P)}(f).$$

It follows that, as $\Delta(P) \longmapsto 0$, $R(P)$ approaches the integral of $f$ on $[a, b]$, as defined by Cauchy. Riemann's step forward was to keep the $R(P)$ sums in his definition, but to drop the assumption of continuity.

## 2.4 The Fourier series problem

In 1822, in his paper on heat conduction, Fourier [24] studied the possibility of representing functions by trigonometric series. A trigonometric series is a series of the form:

$$\frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx), \tag{2.5}$$

where $a_0, a_1, \ldots, b_1, b_2, \ldots$ are real numbers.

One can easily observe that it is enough to study trigonometric series in an interval of length $2\pi$, since, if the above series does converge to a sum $S(x)$, then we have, for any natural number:

$$S(x + 2n\pi) = S(x)$$

The question that interested Fourier was whether a given arbitrary function can be represented by a trigonometric series. In other words, when can we have, for a function $f(x)$ defined, say, on $[-\pi, \pi]$:

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx)? \tag{2.6}$$

If this were true, i.e., if there exist sequences $(a_n)$ and $(b_n)$ of real numbers such that a series of the form (2.5) does converge to $f(x)$, we can use the properties of trigonometric functions to determine the coefficients $a_n$, $b_n$ in terms of $f(x)$. First, note the following obvious identities:

$$\begin{aligned}
\int_{-\pi}^{\pi} \sin mx \sin nx \, dx &= \pi\delta_{mn} \\
\int_{-\pi}^{\pi} \cos mx \cos nx \, dx &= \pi\delta_{mn} \\
\int_{-\pi}^{\pi} \sin mx \cos nx \, dx &= 0,
\end{aligned}$$

where:

$$\delta_{mn} = \begin{cases} 0 & \text{if} \quad m \neq n, \\ 1 & \text{if} \quad m = n. \end{cases}$$

Now, if we multiply both sides of the equations by $\cos nx$ and then by $\sin nx$, and assume that term-by-term integration is allowed, we get:

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx,$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx.$$

The coefficients $a_0, a_1, \ldots, b_1, b_2, \ldots$ defined by the above integral formulas are called the Fourier coefficients of $f$. The series on the right-hand side of the equation (2.6) is then called the Fourier series of $f$. This was, in essence, the argument that Fourier gave in his paper of 1822: if the equation (2.6) can be solved – i.e., if the values of the infinite number of unknowns $a_0, a_1, \ldots, b_1, b_2, \ldots$ can be determined – then the representation must be valid. He further assumed that term-by-term integration of an infinite series of functions is always possible and used this assumption to determine the coefficients. This claim, as we now know, is not always true. There are several issues with the above calculation. The most important question is related with the assumption on which it is based: when does a function admit a representation by Fourier series? To demonstrate the validity of the Fourier series expansion of $f$ one must start with the coefficients, defined as values of definite integrals, construct the infinite series shown on the right-hand side of equation (2.6), and then prove that this series converges to $f$ at each value of $x$ in $[-\pi, \pi]$, considering also the important question of the validity of term-by-term integration.

But the questions surrounding Fourier's proposition turned out to be very powerful triggers for the development of the modern theory of integration. That is because the formulation of the results obtained in the study of the possibility to represent a function by trigonometric series, as we will show, depends on what is meant by the representation of a function by a series, and how the integrals in the formulas of the Fourier coefficients are understood.

## 2.5 Riemann: integrable functions

When Riemann turned to the study of the Fourier series in his habilitation thesis he saw the need to first make some important clarifications about integration before addressing the main problem, that is, the representability of functions by trigonometric series [45]. He did this in a short section before the main body of the thesis, in just four pages. Here, he defined the definite integral and the improper integral, and gave necessary and sufficient conditions for integrability. He then gave an example of an integrable function that is discontinuous at every rational number with an even denominator, showing thus that while continuity is a sufficient condition for integrability, it is far from necessary. Riemann had studied with Dirichlet in Berlin, before going to Göttingen to complete his doctorate under Gauss' direction. This collaboration certainly influenced Riemann: it was Dirichlet who, in 1829, called attention to functions that were discontinuous on an infinite set of points in a finite interval and to the problem of extending the concept of integral to this kind of functions.

Riemann's definition of the integral [45] differs from Cauchy's in the calculation of the approximating sums: the value of the function is chosen in an arbitrary manner in the interval $[x_{i-1}, x_i]$. However, the essential step forward is that Riemann replaces the assumption of continuity by the much weaker one that the sums (Riemann sums) all converge to a unique limit:

Let $f$ be a bounded function on $[a, b]$. For partitions of this interval, $P = \{x_i\}_{i=0}^{n}$, consider the corresponding Riemann sums: $S(P, f) = \sum_{i=1}^{n}(x_i - x_{i-1})f(\xi_i)$, where $\xi_i \in [x_{i-1}, x_i]$.

If these sums approach a unique limit as the norm of the corresponding partitions approaches 0, that is, if $\lim_{\Delta(P)\mapsto 0} S(P, f)$ exists, then $f$ is said to be *Riemann integrable* on [a,b]. The limit $\lim_{\Delta(P)\mapsto 0} S(P, f)$ is called *the definite integral* of $f$ on $[a, b]$, and is denoted by $\int_a^b f(x)\,dx$

The statement $\lim_{\Delta(P)\mapsto 0} S(P, f) = L$, where $L$ is a real number, means that for every $\epsilon > 0$ there is a $\delta > 0$, such that, for any partition $P$ with $\Delta(P) < \delta$, and for any possible Riemann sum $S(P, f)$ relative to $P$, the inequality:

$$|S(P, f) - L| < \epsilon$$

is satisfied.

Next, Riemann asks the question: "[...] in what cases does a function admit integration and in what cases it does not?" ([45], p. 235), which, based on his understanding of integration meant: when do the Riemann sums approach a unique limit as the norm of the corresponding partitions approaches zero? He gives two integrability conditions. The first states that the function $f$ bounded on $[a, b]$ is integrable if and only if:

$$\lim_{\Delta(P)\mapsto 0}(\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n) = 0, \qquad (C1)$$

where $P = \{x_i\}_{i=0}^n$ is a partition that subdivides the interval $[a, b]$ into intervals of length $\delta_i$, while $D_i$ represents the oscillation of the function on the interval $[x_{i-1}, x_i]$.

Riemann did not prove $(C1)$; he used it as an assumption to prove the second condition for integrability $(C2)$, which is, in fact, known as his criterion of integrability. But, as we show below, $(C1)$ follows from Riemann's definition of integrability:

$$\begin{aligned}\sum_{i=1}^{n} \delta_i D_i &= \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) \\ &= \left( \sum_{i=1}^{n} M_i(x_i - x_{i-1}) \right) - \left( \sum_{i=1}^{n} m_i(x_i - x_{i-1}) \right) \\ &= \overline{S}(P, f) - \underline{S}(P, f),\end{aligned}$$

where:

$$M_i = \sup\{f(x) : \quad x_{i-1} \leq x \leq x_i\}, \quad m_i = \inf\{f(x) : \quad x_{i-1} \leq x \leq x_i\},$$

$\overline{S}(P, f)$ and $\underline{S}(P, f)$ are what we call, after a later development by Darboux, the upper and lower Darboux sums (see next section).

Obviously, for a given partition $P$ every Riemann sum corresponding to this partition lies between the upper Darboux sum and the lower Darboux sum. Note that $\overline{S}(P, f)$ and $\underline{S}(P, f)$ may not be Riemann sums since $m_i$ and $M_i$ may not be in the range of $f$ (unless $f$ is continuous); however, we can find Riemann sums that are arbitrarily close to the Darboux sums. Now, $f$ is Riemann integrable if and only if we can make all Riemann sums to be within $\epsilon$ of the specified value $L = \int_a^b f(x)\,dx$ by restricting the partitions to those with subintervals of length less than an appropriately chosen $\delta$ (by definition). This is possible if and only if the upper and lower Darboux sums for these partitions are also within $\epsilon$ of $L$. Then Riemann's $(C1)$ follows from the fact that $f$ is Riemann integrable if and only if we can make the difference between the upper and lower Darboux sums as small as we wish by restricting the lengths of the intervals in the partition:

$$((x_i - x_{i-1}) < \delta \quad \text{for all} \quad i \Rightarrow \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) < \epsilon) \Leftrightarrow (C1).$$

The statement and the proof of the second condition of integrability contain the real insight into Riemann's innovation of establishing the notion of integrable function independently of the continuity assumption:

($C2$): A necessary and sufficient condition for a function $f$ to be integrable on $[a, b]$ is that for any $\epsilon > 0$ and $\sigma > 0$, there exists $d > 0$ such that for any partition $P = \{x_i\}_{i=0}^n$ of $[a, b]$ with norm less than $d$, the total length of the subintervals $[x_{i-1}, x_i]$ for which the oscillation $D_i$ is greater than $\sigma$ is less than $\epsilon$.

In modern terms the above condition states that the function is Riemann integrable if and only if the set of points of discontinuity of the function has Lebesgue measure zero. Of course, Riemann did not regard it as such, since the measure-theoretic developments were only to be fully understood and used in the context of integration about half a century later. We give Riemann's proof of ($C2$) below:

Let $f$ be integrable and let $\Delta(d)$ be the least upper bound of the sums $\sum_{i=1}^n \delta_i D_i$ for all partitions of norm not larger than $d$; then $f$ is integrable if and only if $\Delta(d)$ converges to 0 with $d$ (by $C(1)$). Also, for each partition $P$, let $S_{\sigma,P} = \sum \delta_i$ be the total length of those intervals for which $D_i$ is greater than an arbitrary $\sigma > 0$. Then:

$$\sigma S_{\sigma,P} = \sigma \sum_{i:D_i \geq \sigma} \delta_i \leq \sum_{i=1}^{n} \delta_i D_i \leq \Delta(d).$$

Thus, $S_{\sigma,P} \leq \frac{\Delta(d)}{\sigma}$ and since $\Delta(d)$ converges to 0 with $d$, $S_{\sigma,P}$ also converges to 0 with $d$.

Conversely, assume that given $\epsilon > 0$, $\delta > 0$ , there exists $d > 0$ such that for any partition with norm less than $d$, the total length of the subintervals for which the oscillation $D_i$ is greater than $\sigma$ is less than $\epsilon$. Thus, for partitions with norm less than $d$, if $D$ is the oscillation of the function on $[a, b]$, we have:

$$\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n = \sum_{i:D_i \geq \sigma} \delta_i D_i + \sum_{i:D_i < \sigma} \delta_i D_i \leq S_{\sigma,P} D + \sigma(b-a) \leq D\epsilon + \sigma(b-a).$$

Since $\epsilon$ and $\sigma$ can be made arbitrarily small, the sum converges to 0 with $d$ and by $(C1)$, $f$ is integrable.

As an example of an integrable function having an infinite number of points of discontinuity in every interval, Riemann presents a function that has discontinuities at all rational numbers with even denominators:

$$f(x) = \sum_{n\geq 1} \frac{< nx >}{n^2},$$

where $< x >$ is defined as 0 if $x$ is of the form $k + \frac{1}{2}$, where $k$ is an integer, and as the difference between $x$ and the nearest integer otherwise. In other words,

$$< x >= \begin{cases} x - [x] & , \text{ if } \quad [x] \leq x < [x] + \frac{1}{2} \quad , \\ 0 & , \text{ if } \quad x = [x] + \frac{1}{2} \quad , \\ x - [x] - 1 & , \text{ if } \quad [x] + \frac{1}{2} < x < [x] + 1 \quad . \end{cases}$$

(Here, $[x]$ stands for the greatest integer function.)

The function $< x >$ has a discontinuity at every point that is half of an odd integer: the limit on the left is equal to $\frac{1}{2}$, the limit on the right is $\frac{-1}{2}$ , while at those points the value of the function is 0. On the other hand, $< x >$ is continuous at any other point. Now the function $\frac{<nx>}{n^2}$ carries the same properties: it is discontinuous whenever $nx$ is half of an odd integer for some $n$, and this happens for every $x$ that is a rational number with an even denominator, i.e., at those points $x_0 = \frac{p}{2q}$, where $p$ and $q$ are relatively prime and $n$ is an odd multiple of $q$, and it is continuous at all other points.

Let us now study the continuity of our original function: $f(x) = \sum_{n\geq 1} \frac{<nx>}{n^2}$.

Note first that since $| < x > | \leq \frac{1}{2}$, the above series is uniformly convergent because

$$\left|\frac{< nx >}{n^2}\right| \leq \frac{1}{2n^2} \quad \text{and} \quad \sum_{n\geq 1} \frac{1}{2n^2} < +\infty. \tag{2.7}$$

(In Riemann's proof, the uniform convergence is not mentioned; he apparently assumes is throughout the paper)

As a consequence of the uniform convergence, the limits can be interchanged so $f$ has lateral limits at every point and:

$$f(x-0) = \sum_{n\geq 1} \frac{< nx - 0 >}{n^2}, \quad f(x+0) = \sum_{n\geq 1} \frac{< nx + 0 >}{n^2}.$$

Also because of the uniform convergence of the series, $f$ is continuous at those points at which all the terms of the series are continuous, i.e., at all points that are not of the form $x_0 = \frac{p}{2q}$ , where $p$ and $2q$ are relatively prime. For $x_0$ of this form, consider all $n$ which are odd multiples of $q$, i.e. let:

$$S_q = \{n : \quad n = (2m-1)q, \quad \text{for some} \quad m \in \mathbb{N}\}.$$

Then:

$$\begin{aligned} f(x_0 - 0) - f(x_0 + 0) &= \sum_{n\geq 1} \frac{< nx_0 - 0 >}{n^2} - \sum_{n\geq 1} \frac{< nx_0 + 0 >}{n^2} \\ &= \sum_{n\geq 1} \frac{< nx_0 - 0 > - < nx_0 + 0 >}{n^2} \\ &= \sum_{n\in S_q} \frac{< nx_0 - 0 > - < nx_0 + 0 >}{n^2} \\ &+ \sum_{n\notin S_q} \frac{< nx_0 - 0 > - < nx_0 + 0 >}{n^2}. \end{aligned}$$

Now,

if $n \in S_q$, then $< nx_0 - 0 > - < nx_0 + 0 >= 1$,

and

if $n \notin S_q$, then $< nx_0 - 0 > - < nx_0 + 0 >= 0$.

Therefore:

$$f(x_0 - 0) - f(x_0 + 0) = \sum_{n \in S_q} \frac{1}{n^2} > 0 = \sum_{m \geq 1} \frac{1}{(2m-1)^2 q^2} = \frac{\pi^2}{8q^2} > 0,$$

which shows that $f$ is not continuous at $x_0$. Thus $f$ is discontinuous on a dense subset of real numbers.

Yet, as Riemann shows next, $f$ is integrable on any interval, because for any $\sigma > 0$ , there are finitely many $q$'s such that $\frac{\pi^2}{8q^2} > \sigma$. So, in every interval there exist only a finite number of points $x_0 = \frac{p}{2q}$ at which the jump of the function is larger than $\sigma$. Hence for a sufficiently small $d$ (a bound for the norm of the partition), the sum $S$ (the total length of the intervals on which the oscillation is greater than $\sigma$) can be made arbitrarily small, that is, condition $(C2)$ is satisfied, and thus, $f$ is integrable.

## 2.6 Darboux and Peano: a convenient reformulation of Riemann's definition and criterion of integrability

In 1875, Darboux published his *Mémoire sur les fonctions discontinues* [18] where he proposed a way to bypass the difficulty caused by the variability of the values of $f$ due to the fact that $\xi_i$ can be any value in the interval $[x_{i-1}, x_i]$ by working with the least upper bound (or supremum) and the greatest lower bound (or infimum) of the set:

$$\{f(x) : \quad x_{i-1} \leq x \leq x_i\}.$$

For a bounded function $f$ on $[a, b]$ he associates with every partition $P$ of $[a, b]$, the upper and lower Darboux sums:

$$\overline{S}(P,f) = \sum_{i=1}^{n} M_i(x_i - x_{i-1}) \qquad M_i = \sup\{f(x): \quad x_{i-1} \leq x \leq x_i\},$$

$$\underline{S}(P,f) = \sum_{i=1}^{n} m_i(x_i - x_{i-1}) \qquad m_i = \inf\{f(x): \quad x_{i-1} \leq x \leq x_i\}.$$

If we take a finer partition we can easily see that the upper sum gets smaller, approaching the value of the Riemann integral. Similarly, the lower Darboux sums get bigger approaching the Riemann integral, if it exists. This suggests taking the infimum of the upper sums and the supremum of the lower sums. The greatest lower bound of the upper sums and the least upper bound of the lower sums (taken over all partitions) are called respectively the upper and the lower Darboux integrals, and are denoted as follows:

$$\overline{\int_a^b} f(x)\,dx = \inf \overline{S}(P,f)$$

$$\underline{\int_a^b} f(x)\,dx = \sup \underline{S}(P,f).$$

In fact, it was Peano [41] who, in 1883, used Darboux sums to define the so-called upper and lower integrals that exist for every function. If the integrals are nevertheless named after Darboux, it is probably because they are defined in terms of the sums introduced by Darboux. The function $f$ is said to be integrable if these integrals agree. We show below that this definition of integrability is equivalent to Riemann's definition.

Assume first that $f$ is Riemann integrable. Then, for any partition:

$$\overline{S}(P,f) < \underline{\int_a^b} f(x)\,dx < \overline{\int_a^b} f(x)\,dx < \overline{S}(P,f).$$

(It is easy to show that $\underline{\int_a^b} f(x)\,dx < \overline{\int_a^b} f(x)\,dx$. We'll omit the proof.)

By Riemann's $(C1)$, if $f$ is integrable, there is a partition $P$ for which $\overline{S}(P,f) - \underline{S}(P,f)$ can be made arbitrarily small. Thus, the difference $\overline{\int_a^b} f(x)\,dx - \underline{\int_a^b} f(x)\,dx$ can also be made less than any desired positive value, which means that the upper and lower Darboux integrals coincide.

Conversely, assume that the Darboux integrals are equal, and prove that $f$ is integrable.

Let $\underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx = L$. Then, since the upper Darboux integral is the infimum, over all partitions, of the upper Darboux sums, for any $\epsilon > 0$ there is a partition, say $P'$, such that $\underline{S}(P',f)$ is within $\frac{\epsilon}{2}$ of $L$. Similarly, we can find a partition, say $P''$, such that $\overline{S}(P'',f)$ is within $\frac{\epsilon}{2}$ of $L$. If we take a common refinement $P_3$ of $P'$ and $P''$, i.e., $P' \cup P''$, then every Riemann sum for the partition $P_3$, is within $\frac{\epsilon}{2}$ of $L$. While this is also true for any refinement of $P_3$, we still have to show that this holds in general: for any partition with sufficiently short intervals, regardless of whether they share points with $P_3$ or not, every Riemann sum is within $\epsilon$ of $L$.

By the same idea of Riemann's proof of the criterion of integrability, let $D$ denote the oscillation of the function on the interval $[a,b]$. Also, let $p$ denote the number of intervals in $P_3$. Now let $P$ be an arbitrary partition, such that $\Delta(P) < \frac{\epsilon}{2pD}$, and construct a partition $Q = P \cup P_3$, where $Q$ has at most $p-1$ points more than $P$. Consider an arbitrary Riemann sum for $P$, $S(P,f) = \sum_{i=1}^{n}(x_i - x_{i-1})f(\xi_i)$, and a particular Riemann sum for $Q$, namely one constructed as follows: for every $i$, $1 \le i \le n$, if $\xi_i$ belongs to an interval of $P$, then in the summation, the length of this interval is multiplied by $f(\xi_i)$; otherwise, i.e. for less than $p$ intervals, the point at which $f$ is evaluated is arbitrary. Then, we have:

$$|S(P,f) - S(Q,f)| < p \cdot D \cdot \frac{\epsilon}{2pD} = \frac{\epsilon}{2}.$$

On the other hand, since $Q$ was chosen to be a refinement of $P_3$, we have that:

$$|S(Q, f) - L| < \frac{\epsilon}{2}.$$

Thus, we have, for arbitrarily chosen $P$ (with norm less than a given value):

$$|S(P, f) - L| < \epsilon.$$

Hence, by definition, $f$ is integrable.

Riemann's criterion of integrability $(C1)$, in terms of Darboux sums is: $f$ is integrable if and only if there is a partition $P$ such that, for any $\epsilon > 0$, $\overline{S}(P, f) - \underline{S}(P, f) < \epsilon$

## 2.7 Riemann's insight

In order to appreciate Riemann's work on integrability, one has to take into account the fact that the basic concepts of analysis were not yet well established by mid-19th century. Although concepts of limit and continuity were fairly clear, uniform continuity and uniform convergence, as they are known today, were freshly intuited properties, only about to emerge.

Riemann was the first to study trigonometric series systematically without assuming that these series are Fourier series, i.e. series where the coefficients are given by Fourier's integral formulas. Instead, he started with arbitrary trigonometric series of the form (2.5) and asked what kind of properties such a function must possess. He formulated necessary and sufficient conditions for the representability of a function at a point by trigonometric series in terms of a continuous function $F$ which is the sum of the uniformly convergent series (again, Riemann did not explicitly distinguish between uniform and non-uniform convergence):

$$C + C'x + (\frac{a_0}{4})x^2 - \sum_{k=1}^{\infty}(\frac{a_k coskx + b_k sinkx}{k^2})$$

obtained after term-by-term integrating the series (2.5) twice. Riemann showed that the behavior of the series (2.5) at a point $x$ depends only on the behavior of $F$ in an arbitrarily small neighborhood of this point.

Riemann's major innovation in the in the construction of integration theory was that, in variance with the earlier work, he made the class of functions to which the process of integration can be applied an explicit object of his investigations. In other words, he was the first mathematician to ask the question: *When can we integrate?* in its most general form. Riemann posed the problem of integrability, thus giving meaning to Fourier coefficients. He used a definition of the integral that was very close to the earlier one of Cauchy, but, unlike Cauchy, he considered not only continuous functions, but also attempted – and succeeded – to establish a necessary and sufficient criterion for integrability, significantly weakening the assumptions on the function to be integrated.

While Riemann's contribution to integration theory was certainly triggered by the study of trigonometric series, another, perhaps less obvious, but equally powerful motivation lies at the heart of his work, namely his acceptance of the very general concept of function as any correspondence $x \mapsto f(x)$ between real numbers [27]. This must have been a bold decision because resistance against such generality was probably not uncommon among mathematicians of his time, since still in 1908, the great Henri Poincaré was writing: "Formerly, when a new function was invented, it was in view of some practical end. Today, they are invented on purpose to show our ancestors' at fault, and we shall never get anything more than that from them" [44].

Riemann started off with roughly the same concept of definite integral as

Cauchy, but then he asked the question: what are the conditions on a function $f(x)$ under which the Cauchy sums approach a unique limit as the norm of the corresponding partitions approaches zero? The elimination of the assumption of continuity would not have been meaningful, in Hawkins' view [27], if the possibility of arbitrary functions which are discontinuous in the modern sense at more than a finite number of points in a finite interval had not been considered seriously. Hawkins maintains that Cauchy, despite adopting the modern concept of continuity, fell short of actually acknowledging the logical possibility of existence of highly discontinuous functions. Although his definition of the definite integral did not involve an analytic expression of the function, thus applying to the modern conception of function as a correspondence, he did not appear to hold such a general notion of function. In fact, he was inclined to think of functions in terms of equations, and to restrict himself to "well-behaved" discontinuous functions, that only cease to be continuous for a finite number of values.

## 2.8 Some shortcommings of the Riemann integral

Riemann's integrability was successful in dealing with a large class of problems, and through the the uniform convergence property, in many cases, term by term integration became possible. However, this approach to integrability has some important weaknesses.

Firstly, the original definition of Riemann's integrability applies only to bounded functions, defined on bounded intervals. This difficulty is avoided by introducing the so-called improper Riemann integral as

$$\int_a^{b-0} f(t)\,dt = \lim_{x\mapsto b, x<b} \int_a^x f(t)\,dt,$$

provided that $f : [a,b) \mapsto R$ is Riemann integrable on every interval $[a,x]$ and the above limit actually exists. This approach is useful only for simple applications, and it quickly turns out to be rather complex and to need new concepts such as semi-convergent improper integrals or absolutely convergent improper integrals, almost useless from the point of view of the (future) Lebesgue integral.

Secondly, the condition of uniform convergence appears to be too restrictive a condition. If $\{f_n\}$ is a sequence of Riemann integrable functions such that $f_n(x)$ converges pointwise everywhere, not necessarily uniformly, an equality of the form

$$\lim_n \int_a^b f_n(t)\,dt = \int_a^b \lim_n f_n(t)\,dt$$

is not true in general, or impossible. Consider the following example (inspired from [13]):

Let $\mathbb{Q} \cap [0,1] = \{r_1, r_2, \ldots, r_n \ldots\}$ be an enumeration of the rational numbers in $[0,1]$ and define $f_n(x) = 1$, if $x = r_k$, where $1 \leq k \leq n$, and $f(x) = 0$ otherwise. Also, let $f(x) = 1$, if $x \in \mathbb{Q} \cap [0,1]$, or $f(x) = 0$ otherwise.

Then each $f_n$ is Riemann integrable, $\int_0^1 f_n(x)\,dx = 0$, but $f$ is not Riemann integrable (it is the Dirichlet function restricted to $[0,1]$).

Finally, it is possible to have a bounded derivative that cannot be integrated: in 1881, Volterra provided an example of a a bounded function on $(0,1)$, whose derivative exists everywhere, and is not Riemann integrable (see [43]). In his thesis [32], Lebesgue notes that this means that differentiation and integration cannot be considered as inverse operations for a sufficiently large class of functions.

# Chapter 3

## THE DEVELOPMENT OF THE MEASURE-THEORETIC SUPPORT FOR THE MODERN THEORY OF INTEGRATION

### 3.1 Introduction

If one looks back now at Riemann's definition of integration and his original formulation of the criterion of integrability, the notion of ordinary length appears to be central and a generalization of Riemann's integral seems to follow naturally by generalizing the notion of length. But some time would pass until the developments in set theory would be linked to the notion of definite integral.

### 3.2 Cantor's work in set theory and the first notion of measure

Cantor introduced the first notion of measure that was still, however, dissociated from the concept of definite integral. His groundbreaking work in set theory was in fact triggered by the same interest in trigonometric series. In 1872 he published a paper where he studied the uniqueness of the coefficients in the representation of a function by trigonometric series; in other words, he asked the question: could two different trigonometric series converge to the same function [8]?

Let us now go back to the problem of uniform convergence of series of functions that is linked to this question: uniqueness can be proved if uniform convergence is assumed, because uniform convergence, as shown by Weierstrass [42], assures term-by-term integration, and thus, Fourier's argument we have mentioned earlier stands (see section 2.4). However, the uniform convergence condition proved to be very restrictive: a series of continuous functions that converges uniformly converges to a continuous function. But then, interesting Fourier series that converged to discontinuous functions could not be uniformly convergent.

Heine addressed this problem in 1870 [28] and relaxed the sufficient condition for term-by-term integration by introducing the notion of *uniform convergence in general*: a series $\sum_n f_n(x)$ converges to $f(x)$ uniformly in general in $[a, b]$ if there exists a finite number of points $x_1, x_2, \ldots, x_n$ in $[a, b]$ such that the series converges uniformly to $f(x)$ on any sub interval of $[a, b]$ not containing the $x_i$.

He studied the question of uniqueness by considering the difference between two distinct trigonometric series that would converge to the same function. This difference would be then a trigonometric series with non-zero coefficients that converges to zero.

Heine thus reduced the problem of uniqueness to the following:

$$\textit{If} \quad \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + b_n \sin(nx) = 0 \quad (*)$$

*for all $x$ except those forming a set $P$ then are the coefficients $a_n$ and $b_n$ all equal to 0?*

He proved that, indeed, these coefficients are 0, if the convergence in (*) is uniform in general with respect to the finite set $P$. This amounts to saying that two distinct trigonometric series that are uniformly convergent in general cannot converge to the same function.

Cantor took up the uniqueness problem and eliminated the uniform convergence assumption. He improved on the uniqueness problem gradually (as shown in [27]): he first proved in 1870 that if a trigonometric series converges to 0 at all $x$, then all coefficients of the series must be 0. In 1871, he showed that the same proof works if the trigonometric series converges to 0 at all but finitely many points. He then went on 1872 to study the problem for infinitely many points. In this context, he began a careful investigation of infinite sets, and introduced the following definitions [8]:

> *Limit point*: Given a set $P$, a point $x$ is a limit point of $P$, if every neighborhood of $x$ contains infinitely many points of $P$. (The neighborhood of $x$ referred to an interval that had $x$ in its interior.)
>
> *Derived set, type 1 set* The set of limit points of $P$, denoted by $P'$ is called the derived set of $P$. A set is of type 1 if its derived set is finite.
>
> If $P'$ is not finite, then it has a derived set, and if this latter set is finite, then $P$ is of type 2. Similarly, sets of type $n$ are defined as follows: A set $P$ is of type $n$ if its $(n-1)$-th derived set is not finite but the $n$-th derived set is finite.

For example, Cantor shows that the set $\{1, 1/2, 1/3, \ldots, 1/n, \ldots\}$ has the set $\{0\}$ as its first derived set, and is thus of type 1. The first derived set of the set of all rational numbers between 0 and 1, on the other hand, contains all the real numbers between 0 and 1, and the derived set of this latter set is also made of all the real numbers between 0 and 1.

What Cantor accomplished in his 1872 paper was to prove that if a trigonometric series converges to 0 at all points except on a set of type $n$, then all the coefficients of the trigonometric series are 0.

With this paper, Cantor embarked on his fruitful work in the theory of infinite sets. It also became clear, in connection with his work, that the question of finding necessary and sufficient conditions under which term-by-term integration is possible is extremely difficult.

In 1874 Cantor introduced the notion of countability, and proved the surprising result that the set of algebraic numbers is countable, and started to study infinite sets and to classify them according to these new notions.

The first notion of measure appears in Cantor's 1884 papers [9], [10], where he introduces it for the study of properties of continua. What he termed *grandeur* (or *Inhalt* in his earlier German language paper) of a set is defined as follows [10]:

*For an arbitrary bounded set $P$ in $\mathbb{R}^n$, let $K(p,\rho)$ be the closed sphere centered at $p \in P$, of radius $\rho$. Take*

$$\Pi(P,\rho) = \bigcup_{p \in \bar{P}} K(p,\rho)$$

*and let $f(\rho)$ be its volume, which is a continuous function of $\rho$, that decreases monotonically with $\rho$. The volume $f(\rho)$ of $\Pi(P,\rho)$ is given by the multiple integral*

$$\int_{\Pi(P,\rho)} dx_1 dx_2 \cdots dx_n.$$

*The limit: $F(P) = \lim_{\rho \mapsto 0} f(\rho)$ is the measure of $P$.*

Hawkins [27] observes that Cantor's definition is based on two assumptions, which he treats "loosely": the first, is that if $P$ is composed of a finite number of n-dimensional regions, then the multiple integral $\int_P dx_1 \cdots dx_n$ is defined. This assured that the multiple integral $\int_{\Pi(P,\rho)} dx_1 \cdots dx_n$ is defined. But, as observed by both Hawkins [27] and Pesin [42], multiple integrals had yet to be studied with more rigor before such general assumptions could be made.

Secondly, he assumed that if $P$ is an arbitrary bounded set, then $\Pi(P, \rho)$ consists of a finite number of $n$-dimensional pieces, where $\bar{P}$ is $P$ together with its limit points. This can be justified by the fact a closed and bounded set is compact. But this property of sets hadn't been proved at the time either.

Cantor [9] observes that a set and its closure have the same measure. Consequently, this measure is not, in general, additive: if $P$ and $Q$ are two disjoint sets, the relation $F(P \cup Q) = F(P) + F(Q)$ does not necessarily hold if their closures are not disjoint also. If, for instance, $Q$ is the set of rational numbers in $I = [0,1]$, and $P$ is the set of irrational numbers in $[0,1]$, then $F(Q) = F(P) = 1$, and $F(Q \cup P) = 1$. Cantor mentioned this consequence in the following, more intuitive form: $F(P \cup Q) = F(P) + F(Q)$, if $P$ and $Q$ are in "completely separated" $n$-dimensional regions of the space.

Cantor's impact upon measure theory and, hence, upon integration, however, extends beyond the fact that he was among the first to introduce a notion of measure. It was the generality with which he treated sets that had the greatest influence. As Ulam [58] observes:

"Cantor's creation of set theory, where the notion of geometrical figure was generalized into that of an arbitrary subset of points of a given space, introduced also a need for an axiomatic investigation of the problem of measure of sets and, at the same time, made possible a logical analysis of the notion of measure in general." ([58], p. 597)

## 3.3 Peano: a rigorous definition of area.

The notions of measure introduced by Cantor , Stolz and Harnack were not linked to the concept of definite integral. What would inspire Lebesgue's approach to integration, according to Lebesgue himself [34] would be exactly that: the use of the notion of content to define area, in particular Jordan's work in this domain. The rigorous geometrical view of the definite integral of a function was first introduced by Peano in 1887. In his *Applicazioni geometriche del calcolo infinitesimale* [?] he introduced a notion of content that distinguished between inner and outer content. While his concept of outer content agrees with Cantor, Stolz, and Harnack's notions of measure, the introduction of an inner content was a new and important feature of Peano's work.

He gave separate definitions for content in $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$ (his terminology is actually inner and outer *length*, *area*, and *volume*, respectively). For $S$, a subset of the two-dimensional space, Peano defined the outer and inner content – $c_e(S)$ and $c_i(S)$ – as the infimum of the areas of a finite system of polygons containing the set, and correspondingly, the supremum of the areas of polygons contained in the set. $S$ is said to have content (area) if the two values agree, and this common value is defined as the area of $S$.

Peano noticed the following relationship between inner and outer content:

$$c_e(S) = c_i(S) + c_e(\partial S),$$

where $\partial S$, the boundary of $S$, is defined by Peano along with the notions of interior and exterior of a set as follows: a point $p$ is an *interior point* of a set $S$ if there exists a $\delta > 0$ such that every $x$ whose distance from $p$ is less than $\delta$ belongs to $S$; $p$ is an *exterior point* of $S$, if there exists $\delta > 0$, such that all points $x$ whose distance from $p$ is less than $\delta$ do not belong to $S$. Finally, $p$ is a *boundary point* if it is neither an interior nor an exterior point. In other words, the boundary of $S$, denoted by $\partial S$, consists of all points for which every neighborhood contains at least one point in $S$ and at least one point not in $S$. The consequence of this relationship is that the set $S$ has area if and only if $c_i(S) = c_e(S)$ .

Peano did not prove this relationship, but it follows quite easily from his definitions of interior, exterior and boundary points. Let us prove it for 2-dimensional sets:

Consider a grid of $\mathbb{R}^2$ with $2^{-k} \times 2^{-k}$ disjoint squares, and let $E_k$ the union of squares that intersect $S$ in at least one point, and $I_k$, the union of squares that are completely contained in $S$. Then $E_k$ is made of two kinds of squares: those that contain boundary points of S and those that do not, and thus are completely contained in S. In other words: $E_k = I_k \cup B_k$ , where $B_k$ is the union of those squares that contain boundary points, and

$$area(E_k) = area(I_k) + area(B_k) \quad .$$

Now, as $k$ increases, the area of the cover $E_k$ decreases and approaches the outer content of $S$, the area of $B_k$ also decreases and approaches the outer content of the boundary of $S$, while the area of $I_k$ increases and approaches the inner content of $S$. Then:

$$\begin{aligned} c_e(S) &= \lim_{k\mapsto\infty} area(E_k) = \lim_{k\mapsto\infty} area(I_k \cup B_k) \\ &= \lim_{k\mapsto\infty} (area(I_k) + area(B_k)) \\ &= \lim_{k\mapsto\infty} area(I_k) + \lim_{k\mapsto\infty} area(B_k) \\ &= c_i(S) + c_e(\partial S). \end{aligned}$$

The new notions of Peano, having provided a precise definition of area in terms of content, allowed for a new geometrical characterization of Riemann integrability. Peano had introduced in 1883 [41] the upper and lower Riemann integrals

$$\underline{\int_a^b} f(x)\,dx \quad \text{and} \quad \overline{\int_a^b} f(x)\,dx,$$

defined as the infimum and the supremum of the upper and lower Darboux sums of $f$, respectively. He came back to these definitions in the later work after introducing the notion of content, and reformulated Riemann's criterion of integrability ($C1$), giving it a geometric interpretation, as follows:

If $f$ is a non negative function defined on $[a,b]$ and $S$ denotes the region bounded by the graph of $f$ , i.e.:

$$S = \{\,(x,y)\,:\,a \leq x \leq b\,,\,0 \leq y \leq f(x)\,\},$$

then:

$$c_i(S) = \underline{\int_a^b} f(x)\,dx \quad \text{and} \quad c_e(S) = \overline{\int_a^b} f(x)\,dx.$$

It follows that $f$ is Riemann integrable if and only if $S$ has area in the sense that its inner and outer content are equal.

## 3.4 Jordan: measurability and finite additivity of measure.

Jordan made the concept of content more rigorous (Peano's work was thought to be lacking in this respect), but also more far reaching through his very influential *Cours d'analyse de l'École Polytechnique*, an analysis textbook published in 1893.

Based on the same concept of content as Peano, he introduced the notion of measurability, which despite being quite present between the lines in Peano's work, was never explicitly introduced [29]:

*A set $S$ is measurable if and only if the inner and outer content of the set are equal, $c_i(S) = c_e(S)$. The measure of $S$ is given by either the inner or outer content: $c(S) = c_i(S) = c_e(S)$.*

What motivated Jordan to focus on the notion of content was his interest in the theory of multi-dimensional integrals, which he managed to describe in a rigorous manner. The concept of measurability (denoted hereinafter as *Jordan measurability*) became useful when the domain of integration for a real valued function of two variables was an arbitrary set, which might be a very irregular region: in this case, the integral made sense if the inner and outer contents of the domain were equal. In this context, the finite additivity of the Jordan measure, that is, the property that the measure of the union of a finite collection of pairwise disjoint Jordan-measurable sets is equal to the sum of the measures, is critical.

Jordan noted the following properties of the measure, that we will describe, in what follows, using the letter $m$: $m_i(S)$ the inner measure,$m_e(S)$ the outer measure, and $m(S)$ the Jordan measure, respectively, of $S$:

(a) If $S_1 \subset S_2$, then $m_i(S) \leq m_i(S)$ and $m_e(S) \leq m_e(S)$ ;

(b) If $S_k$ are pairwise disjoint, then $m_e\left(\bigcup_{k=1}^{n} S_k\right) \leq \sum_{k=1}^{n} m_e(S_k)$;

(c) If $S_k$ are pairwise disjoint, then $m_i\left(\bigcup_{k=1}^{n} S_k\right) \geq \sum_{k=1}^{n} m_i(S_k)$.

Finite additivity follows:

$$\sum_{k=1}^{n} m_i(S_k) \leq m_i\left(\bigcup_{k=1}^{n} S_k\right) \leq m_e\left(\bigcup_{k=1}^{n} S_k\right) \leq \sum_{k=1}^{n} m_e(S_k).$$

Therefore, since each set is Jordan measurable,

$$m\left(\bigcup_{k=1}^{n} S_k\right) = \sum_{k=1}^{n} m(S_k).$$

Instead of defining upper and lower integrals as suprema and infima, he defined the "intégrale par excès" and "intégrale par défaut" in a slightly different way, but from the point of view of the generalizability of the definite integral, very fruitful. The most important difference was that he considered the upper and lower Darboux sums that correspond to partitions into arbitrary Jordan-measurable sets instead of simply intervals. The "intégrale par excès" and "intégrale par défaut" are then defined as limits (instead of supremum and infimum) of these sums as the dimensions of the sets constituting the partition approach 0.

Lebesgue, who studied Jordan's *Cours d'analyse* as an undergraduate at the École Normale Supérieure, later recounts how Jordan's work had paved the way for his integration theory: a more general notion of measure and measurability would allow him to create a more general notion of integral and integrability.

## 3.5 Borel: a generalized countably additive measure.

Borel took a different approach to defining measure. In his 1898 book *Leçons sur la Théorie des Fonctions* he defined measure in an axiomatic manner: measure and measurability were described in terms of the properties that they should posses. Cantor's set theory had provided the necessary background for such an approach. In particular, the main property he entailed for his measure was countable additivity [5]. This means that given an infinite collection of pairwise disjoint sets, $S_1, S_2, \ldots, S_n, \ldots$, the measure of their union is equal to the sum of their measures:

$$m\left(\bigcup_{n\geq 1} S_n\right) = \sum_{n\geq 1} m(S_n).$$

According to Borel, "When a set is formed of all the points comprised in a denumerable infinity of intervals which do not overlap and have total length $s$, we say that the set has measure $s$ ([5], p. 46). This means that the measure *assigned* by Borel to the class of sets that are unions of countable pairwise disjoint intervals is the sum of the measures of the covering intervals. Thus, Borel's concept of measure seems to be based on the property that any open set $S \subset \mathbb{R}$, is the union of countably many disjoint open intervals, but Borel doesn't use the terminology of open sets (an explicit definition of open sets was not introduced until 1899, by Baire, see [38]). At the time, the terminology of closed sets was rather used. In fact Borel points out that a closed set is measurable since it is the complement of a countable number of disjoint sets. We emphasize here that this is a characteristic of both Borel and Lebesgue's work: in their original arguments they use the terminology of *sets that are countable unions of nonoverlapping intervals*, rather then the modern terminology of *open sets.*

Borel also required that if a set $E$ has measure $s$ and has $E'$ as a subset, of measure $s'$, the set $E \backslash E'$ is said to have measure $s - s'$.

Thus, the following three assumptions define Borel's notion of measure in a unique manner:

(1) The measure of a bounded interval is its length;

(2) The measure of a countable union of pairwise disjoint sets is the sum of their measures.

(3) If $Q \subset S$, then $m(S \backslash Q) = m(S) - m(Q)$.

According to Borel, a set is said to be measurable if its measure "can be defined by virtue of the preceding definitions". In his treatise, Borel did not investigate in a rigorous manner the class of sets described by these properties, and the structure of *his* measurable sets is – using modern terminology – a $\sigma$-ring. Lebesgue [36] would later give a rigorous construction of the class of sets obtained by successive application of the union and complement operations starting with open sets. In the modern terminology they are called *Borel sets* or Borel measurable sets, they have a structure which by now is called $\sigma$-algebra of sets.

**Definition 3.5.1.** *For an arbitrary non empty set $T$, $\mathcal{F} \subset \mathcal{P}(T)$ is called a $\sigma$-algebra, if* *(a)* $\emptyset \in \mathcal{F}$ *and* $T \backslash A \in \mathcal{F}$ *whenever* $A \in \mathcal{F}$.

*(b) For every sequence* $A_n \in \mathcal{F}$,

$$\bigcup_{n \geq 1} A_n \in \mathcal{F}.$$

The collection of Borel sets, or Borel measurable sets, is the smallest $\sigma$-algebra that contains all the open sets.

Borel's motivation for the introduction of the new concept of measure came from his study of the size of the set of points on which certain infinite series converge, that had led earlier to the discovery of the Heine-Borel theorem. The Jordan measure turned out to be inappropriate for the study of this set, because it didn't have the property of countable additivity. But although he did find it useful to compare his definition with the earlier one of Jordan, Borel did not see the use of his idea of measure in the theory of integration. In fact, he went so far as to comment: "The problem we investigate here is [...] totally different from the one resolved by M. Jordan" ([5], p. 46).

Borel's measure is, in fact less general than the Jordan measure, in the sense that there are Jordan measurable sets which are not Borel measurable: the cardinality of the collection of all Borel sets, is $c$, the cardinality of the continuum, while the cardinality of the collection of Jordan measurable sets in $\mathbb{R}$ is $2^c$. This can be justified by the following argument: consider the well-known Cantor's ternary set, which, it is easy to see, has Jordan measure 0. Then any subset of this set will be Jordan measurable, and of measure 0. On the other side, the Cantor set $C$ is equal with the intersection of a decreasing sequence $(C_n)_n$, each $C_n$ being a finite union of intervals, hence having the cardinality $c$. Since the intersection of a decreasing sequence of sets with the cardinality $c$ has also the cardinality $c$, it follows that $C$ has the cardinality $c$, therefore the collection of its subsets has cardinality $2^c$. On the other hand, the cardinality of the set of all intervals is not larger than the cardinality of the set of pairs of real numbers, which is $c$. By taking complements and countable unions, the cardinality remains unchanged, thus the collection of all Borel sets has cardinality $c$, which is smaller than $2^c$, the cardinality of Jordan measurable sets.

Hence, Borel clearly did not intend for his measure to generalize the earlier work of Jordan, especially since his interests, as he has stated were totally different. Moreover, it is interesting that in the field of probability theory, as well, where Borel made important contributions, both in general theory and applications (such as the Borel-Cantelli lemma, the Strong Law of Large Numbers, or the Continued Fraction Theorem), he failed to see the natural application of his theory of measure. In his landmark paper in probability *Les probabilités dénombrables et leurs applications arithmétiques* of 1909 he used countable additivity very rarely and even avoided using it by offering alternate proofs. Barone and Novikoff [3] attribute this to that fact that Borel regarded rather the concept of countable independence as the essential ingredient of his theory (the property a countable collection $B_1, B_2, \ldots, B_n, \ldots$ has that $P(\cap_{n=1}^{\infty} B_n) = \Pi_{n=1}^{\infty} P(B_n)$). Bingham [4], on the other hand, looking at the development of the axiomatic theory of probability, asserts that Borel's failure to use his measure theory in probability, could be explained by the fact that his view of measure was still attached to the Euclidean and geometric settings. It would take some time, and some important contributions such as those of Fréchet and Levy (see [4]), for Borel's measure theory to be applied successfully by Kolmogorov, in his 1933 *Grundberigffe*, to develop an axiomatic theory of probability.

The usefulness of Borel's concept of measure in the theory of integration was to be revealed by Lebesgue's work, who attached a countably additive measure to a class of sets much more general than those that had been considered previously, and, most importantly, used this measure for integration.

# Chapter 4

## THE LEBESGUE MEASURE AND INTEGRAL

### 4.1 Lebesgue's measure: improving on Jordan and Borel

Henri Lebesgue is generally known as the creator of the modern theory of measure and integration. The first detailed exposition of the new notion of measure and integral appears in his doctoral thesis of 1902, *Intégrale, longueur, aire* [32].

In the previous chapter we have presented the gradual development of the measure theoretic support for the modern theory of integration. This development was far from being linear and every new idea was closely related to a particular problem which had to be solved with no general picture in sight.

Namely, Cantor (section 3.2), in his studies on Fourier series, became interested in finding a possibility to evaluate "how big" a particular set is, in other words, to find a measure for sets. But his approach was mainly topological – no geometrically intuitive lengths or areas were involved – and in the end he invented the measurement of sets in terms of cardinality.

Later, Peano (section 3.3) and Jordan (section 3.4) gave an intuitive definition of measure and measurability which applies well to sets which are close to our image of length and area, but countable unions of such sets failed to remain measurable (in fact these sets are measurable in our modern sense, it was Jordan's method of "measurement" that failed).

Borel's idea of measure already encompassed some very useful properties (section 3.5), but he did not seek to find the most general collection of sets for which such a measure could be defined.

In his thesis Lebesgue first takes what he calls a descriptive approach to defining his more general measure (today, we'd rather call it an "axiomatic approach", by postulating the conditions that this measure should possess. The *problem of measure* – as Lebesgue put it – was to find a measure satisfying these properties. Namely, $m(E)$, a non negative measure on some *bounded* subsets $E \subset \mathbb{R}$ has to satisfy the following ([32]):

$(m1)$ $\quad m(E) \neq 0 \quad \text{for some} \quad E$

$(m2)$ $\quad m(E+a) = m(E) \quad \text{for every real number} \quad a$

$(m3)$ If $E_n$ are pairwise disjoint for $n = 1, 2, 3, \ldots$ , then

$$m\left(\bigcup_{n\geq 1} E_n\right) = \sum_{n\geq 1} m(E_n).$$

$(m1)$ requires that $m$ is not identically zero, $(m2)$ is the translation invariance, while condition $(m3)$ is the countable additivity.

In the subsequent section, Lebesgue proposes to take an interval as "measurement unit", that is, to attribute to it the measure 1 by convention (assuming that the problem of measure has a solution). Then, any interval can be attached a measure, which is its length. That is, if, for instance, $m((0,1))$ is assumed to be 1, then from $(m2)$ and $(m3)$ it follows that every interval *should* be measurable and $m((a,b)) = b - a$ (in fact, in later expositions [36], Lebesgue replaced $(m1)$ by the requirement that $m((0,1)) = 1$). Moreover, any set that is the (bounded) union of a countable collection of pairwise disjoint intervals $(I_n)_n$ *should* have attached a measure and

$$m\left(\bigcup_{n\geq 1} I_n\right) = \sum_{n\geq 1} m(I_n).$$

Throughout his thesis, Lebesgue, like Borel, whom he often quotes, works with those sets that can be written as countable unions of pairwise disjoint intervals that are not necessarily open (as compared to the modern approach where Borel sets are "generated" by *open* sets), the measure of which, he assumes, in the sequel, to be the sum of the lengths of these intervals.

All the above suggested that if $m(E)$ is definable for some bounded set $E$, i.e., if the problem of measure had a solution, then its value should be less than $\sum_{n\geq 1} l(I_n)$, where $\left(\bigcup_{n\geq 1} I_n\right) \supset E$, and $l(I)$ denotes the length of the interval $I$. Thus Lebesgue proposed the following constructive definition of measure:

$$m_e(E) \stackrel{def}{=} \inf \left\{ \sum_{n\geq 1} l(I_n) \,:\, E \subset \bigcup_{n\geq 1} I_n, \quad I_n \quad \text{pairwise disjoint} \right\} \tag{4.1}$$

The above $m_e(E)$ is the outer measure of $E$.

In order to better understand Lebesgue's contribution, recall that for a bounded $E \subset \mathbb{R}$, Jordan defined the outer content of $E$ as

$$\bar{c}(E) = \inf \sum_{k=1}^{p} l(I_k),$$

where the infimum is taken over all *finite* collections of non-overlapping intervals $I_1, \ldots, I_p$ such that $E \subset I_1 \cup \cdots \cup I_p$, and the inner content as

$$\underline{c}(E) = \sup \sum_{k=1}^{p} l(I_k),$$

where the supremum is taken over all *finite* collections of non-overlapping intervals $I_1, \ldots, I_p$ such that $I_1 \cup \cdots \cup I_p \subset E$. We have $\underline{c}(E) \leq \bar{c}(E)$, and $E$ is said to be Jordan measurable if $\bar{c}(E) = \underline{c}(E)$, the common value being the Jordan measure of $E$. Now it is easy to see that the outer measure $m_e(E)$ is similar to Jordan's outer content, the change consisting of the fact that a *finite* number of intervals is replaced by sequences, i.e., by an infinite number of intervals. This apparently small change makes all the difference.

Namely, the major inconvenience of Jordan's measurability is the fact that, in general, countable unions of Jordan measurable sets fail to remain Jordan measurable. For example, every point on the real line is Jordan measurable but the set of rationals on $[0, 1]$ is countable but not Jordan measurable. On the other hand, Jordan measurability is strongly related to the Riemann integral and the fact that countable unions of Jordan measurable sets could easily fail to be Jordan measurable is "almost" equivalent to the fact that the limit of a sequence of Riemann integrable functions could fail to remain Riemann integrable. The example given in section 2.8 is a very simple illustration of this phenomenon. So, condition ($m3$) asks implicitly that countable unions of Lebesgue measurable sets remain Lebesgue measurable.

Inspired by Jordan's work, Lebesgue successfully modified Jordan's definition of outer measure, but his definition of the inner measure was different, it was not a direct generalization of Jordan's inner content. Lebesgue's idea for the inner measure was very intuitive: it was based on the same idea of generalizing the notion of length. He used the outer measure, but for the *complement* of $E$, relative to some interval $[a, b]$ such that $E \subset [a, b]$.

So, for every $E \subset [a, b]$, the inner measure $m_i(E)$ was defined as

$$m_i(E) = b - a - m_e([a, b] \backslash E). \tag{4.2}$$

Lebesgue noted that the number $m_i(E)$ does not depend on the choice of the interval $[a, b]$.

Since for every outer measure $m_e(A \cup B) \leq m_e(A) + m_e(B)$, we have

$$b - a = m_e([a,b]) = m_e(E \cup ([a,b] \backslash E)) \leq m_e(E) + m_e([a,b] \backslash E),$$

and therefore $m_i(E) \leq m_e(E)$. (Lebesgue used the terms *exterior measure* and *interior measure* – whence the choice of notation $m_e$ and $m_i$; we are using the modern terminology). The set $E$ is said to be *measurable* if $m_i(E) = m_e(E)$, the common value being the measure of $E$.

With this definition, Lebesgue was able to prove that every Borel set is measurable (see section 3.5 for the definition of a Borel set), that every union of a countable collection of measurable sets is also measurable, and the measure is countably additive. He only gives the idea of the proof, which is based on an equivalent definition of measurable sets. We present below a detailed proof, based on Lebesgue's ideas.

Let us first fix an interval $[a, b]$ and denote by $\mathcal{G}$ the set of all $A \subset [a, b]$ which are unions of a countable collection of pairwise disjoint intervals (not necessarily open). Lebesgue uses implicitly the following properties of such sets (that had been shown by Borel): if $A$ and $B$ are in $\mathcal{G}$, then so is $A \cap B$ and if $A \subset B$ then $B \backslash A$ has a measure, and $m(B \backslash A) = m(B) - m(A)$. Most importantly, the measure of any set $A \subset \mathcal{G}$ is equal to the sum of the lengths of the intervals that cover it.

Now, the definition of $m_e(E)$, where $E \subset A$, from (4.1) translates into:

$$m_e(E) = \inf\{\, m(A) \;:\; A \in \mathcal{G} \,\}. \tag{4.3}$$

Lebesgue begins with the remark that $E \subset [a, b]$ *is measurable if and only if for every* $\epsilon > 0$ *there exist* $A, B \in \mathcal{G}$ *such that*

$$E \subset A, \quad [a,b]\backslash E \subset B \quad and \quad m(A \cap B) < \epsilon. \tag{4.4}$$

He didn't prove this claim, but it is not difficult to show the equivalence of the two definitions.

Indeed, let $E \subset [a,b]$ be fixed. Then, for every $A, B \in \mathcal{G}$ such that $E \subset A$ and $[a,b]\backslash E \subset B$, since $[a,b] = E \cup ([a,b]\backslash E) \subset A \cup B \subset [a,b]$, we have

$$[a,b] = A \cup B = (A\backslash(A \cap B)) \cup (B\backslash(A \cap B)) \cup (A \cap B),$$

hence

$$\begin{aligned} b - a &= m([a,b]) = m(A\backslash(A \cap B)) + m(B\backslash(A \cap B)) + m(A \cap B) \\ &= m(A) - m(A \cap B) + m(B) - m(A \cap B) + m(A \cap B) \\ &= m(A) + m(B) - m(A \cap B), \end{aligned}$$

therefore:

$$b - a + m(A \cap B) = m(A) + m(B). \tag{4.5}$$

Assume that for an arbitrary $\epsilon > 0$, there are $A, B$ such that (4.4) holds. Then, by (4.3) we have $m_e(E) \leq m(A)$, $m_e([a,b]\backslash E) \leq m(B)$, hence

$$m_i(E) = b - a - m_e([a,b]\backslash E) \geq b - a - m(B),$$

and

$$0 \leq m_e(E) - m_i(E) \leq m(A) + m(B) - (b - a) \overset{from\ (4.5)}{=} m(A \cap B) \leq \epsilon.$$

It follows that $m_e(E) = m_i(E)$ because $\epsilon$ was arbitrary, i.e., $E$ is measurable. Conversely, assume that $E$ is measurable $(m_e(E) = m_i(E))$ and fix $\epsilon > 0$. By (4.3) there are $A, B \in \mathcal{G}$ such that

$$m_e(E) \leq m(A) \leq m_e(E) + \epsilon, \quad m_e([a,b]\backslash E) \leq m(B) \leq m_e([a,b]\backslash E) + \epsilon$$

and from (4.5):

$$\begin{aligned} m(A \cap B) &= m(A) + m(B) - (b-a) \\ &\leq (m_e(E) + \epsilon) + (m_e([a,b]\backslash E) + \epsilon) - (b-a) \\ &= 2\epsilon + m_e(E) - (b-a) - m_e([a,b]\backslash E) \\ &= 2\epsilon + m_e(E) - m_i(E) \\ &= 2\epsilon, \end{aligned}$$

We now prove that every union of a countable collection of measurable sets is measurable.

Let $E_1, E_2, \ldots, E_n \ldots$ be a sequence of measurable sets in $[a,b]$. Fix an arbitrary $\epsilon > 0$ and set

$$E = \bigcup_{n \geq 1} E_n.$$

For every $n \geq 1$, since $E_n$ are measurable, there are $A_n, B_n \in \mathcal{G}$ such that

$$E_n \subset A_n \quad , \quad [a,b]\backslash E_n \subset B_n \quad \text{and} \quad m(A_n \cap B_n) \leq \frac{\epsilon}{2^n}.$$

Now set $A'_2 = A_2 \cap B_1$, $B'_2 = B_2 \cap B_1$, $A'_3 = A_3 \cap B'_2$, $B'_3 = B_3 \cap B'_2 \ldots$, or,

$$A'_{p+1} = A_{p+1} \cap (B_1 \cap B_2 \cap \cdots \cap B_p) \quad \text{and} \quad B'_{p+1} = B_{p+1} \cap (B_1 \cap B_2 \cap \cdots \cap B_p)$$

for every $p \geq 1$. We claim that

$$E \subset A \overset{def}{=} A_1 \cup \left( \bigcup_{p \geq 2} A'_p \right) \quad \text{and} \quad [a,b] \backslash E \subset B,$$

where $B$ can any of the sets $B_1, B'_2, \ldots, B'_p, \ldots$.

Let $x \in E$. If $x \in E_1 \subset A_1$, then it is obvious that $x \in A$. If $x \notin A_1$ then there is a first $p \geq 1$ such that

$$x \in E_{p+1} \subset A_{p+1} \quad \text{and} \quad x \notin E_1 \cup \cdots \cup E_p,$$

$$\text{hence} \quad x \in ([a,b] \backslash E_1) \cap \cdots \cap ([a,b] \backslash E_p) \subset B_1 \cap \cdots \cap B_p,$$

therefore $x \in A'_{p+1} \subset A$. For the second inclusion it is sufficient to note that $x \in [a,b] \backslash E$ implies that

$$x \in ([a,b] \backslash E_1) \cap \cdots \cap ([a,b] \backslash E_p) \subset B_1 \cap \cdots \cap B_p \quad \text{for every} \quad p \geq 1.$$

It follows that $A, B \in \mathcal{G}$ and $A \cap B \subset \bigcup_{p \geq 1} (A_p \cap B_p)$ , hence

$$m(A \cap B) \leq \sum_{p \geq 1} m(A_p \cap B_p) \leq \sum_{p \geq 1} \frac{\epsilon}{2^p} = \epsilon,$$

therefore $E$ is measurable.

Now, if $E_n$ are pairwise disjoint measurable sets, then, for any $\epsilon > 0$ and for every $n \geq 1$, $E_n \subset A'_n$, where $A'_n$ is defined as above ($A'_1 = A_1$), and we have:

$$m(A'_n) - m(E_n) \leq m(A_n) - m(E_n) = m(A_n \backslash E_n) \leq m(A_n \cap B_n) \leq \frac{\epsilon}{2^n}, \quad n \geq 1.$$

On the other hand, we have shown that $E = \bigcup_{n\geq 1} E_n \subset A = \bigcup_{n\geq 1} A'_n$, thus:

$$m(A) - m(E) = m(A \backslash E) \leq m(A \cap B) \leq 2\epsilon.$$

It follows that:

$$\sum_{n\geq 1} m(E_n) - m(E) = m(A) - m(E) - \sum_{n\geq 1} m(A'_n) + \sum_{n\geq 1} m(E_n) \leq 2\epsilon - \sum_{n\geq 1} \frac{\epsilon}{2^n} = \epsilon.$$

(Note that $m(A) = \sum_{n\geq 1} m(A'_n)$, since $A \in \mathcal{G}$)

Thus, countable additivity holds:

$$m\left(\bigcup_{n\geq 1} E_n\right) = \sum_{n\geq 1} m(E_n).$$

One can see in this construction the significance of Borel's results. Lebesgue did state his purpose to improve on the clarity and rigor of Borel's theory of measure ([32], p. 2), but he entailed – and achieved – more than that, by applying his improved theory of measure to integration.

## 4.2 Lebesgue's integral: partitioning of the range

Next, Lebesgue turns to the problem of integration. Through the concept of measurable functions, his theory of measure proves to be more than a simple generalization of Jordan's and Borel's measures, and finds its great application to integration.

He starts out with the following geometric characterization of the integral in terms of measure of sets: if $E$ is measurable,

$$\int_a^b f(x)\,dx = m(E^+) - m(E^-), \tag{4.6}$$

where $E^+$ and $E^-$ are the planar sets bounded by the graph of $f$ and lying above and below the $x$-axis, respectively, and $m$ denotes their 2-dimensional measure. This geometrical definition of the integral had been proposed before, by Peano, but for the Riemann integral and in terms of the Peano (Jordan) content [41].

But Lebesgue's way of constructing the integral analytically provided its power: he abandons his predecessors' approach of partitioning the $x$-axis, because this approach, he shows in a conference talk of 1926 (see [13]), is to some extent limited: roughly speaking, by partitioning the domain $(a, b)$ into smaller and smaller intervals $(x_k, x_{k+1})$, one gets smaller and smaller differences between $\overline{f}_k$ and $\underline{f}_k$ (the supremum and the infimum, respectively, of $f$ on $(x_k, x_{k+1})$), which makes the difference between the upper and lower Darboux sums tend to 0, if $f$ is continuous, or has only a few points of discontinuity. However, this process may not work for functions that are highly discontinuous, and he proposes a procedure for integration that starts out with the goal to be attained: namely to "collect" values of $f(x)$ which differ by very small amounts. For this, he partitions the range of the function, instead of its domain.

That is, if $f \geq 0$ is defined on $[a, b]$, $\underline{f} = \inf f(x)$ and $\overline{f} = \sup f(x)$, let

$$P : \underline{f} = y_0 < y_1 < \cdots y_{n-1} < y_n = \overline{f}$$

be an arbitrary partition of the interval $[\underline{f}, \overline{f}]$ and consider the sets

$$E_k = \{t \in [a, b] \ : \ y_k \leq f(t) < y_{k+1}\}$$

Each of these sets is made up of a union of intervals, and they partition the domain $[a, b]$. Then, if $E$ is the region bounded by the graph of $f$, the (planar) measure of $E$, $m(E)$ lies between the the sums: $\sum_{k=0}^{n-1} y_k m(E_k)$ and $\sum_{k=0}^{n-1} y_{k+1} m(E_k)$, i.e., $\int_a^b f = m(E)$ lies between these two quantities. These sums have a common limit as the norm of $P$ approaches 0 and this number is taken by Lebesgue as the definition for the integral.

$$\int_a^b f = \lim_{\Delta(P)\mapsto 0} \sum_{k=0}^{n-1} y_{k+1} m(E_k) = \lim_{\Delta(P)\mapsto 0} \sum_{k=0}^{n-1} y_k m(E_k).$$

(The argument that justifies the existence of the common limit is typical. For any two partitions the difference between the two sums can be made arbitrarily small by restricting the lengths of the subintervals formed by the two partitions: one proceeds by first taking a partition $Q$ that's finer than $P$ and then, one that contains the points of both $P$ and $Q$.)

The sums formed in this manner, however, are only defined for functions $f(x)$ for which the sets $E_k$ are measurable for all $y_k$ – Lebesgue calls these functions summable, and in later presentations *measurable.* More precisely, measurable functions are those functions $f : [a, b] \longmapsto \mathbb{R}$ for which the sets

$$\{\, t \in [a, b] \;:\; c \leq f(t) \leq d \,\}$$

are measurable for all real numbers $c, d$ [32].

Equipped with this definition, Lebesgue then shows the regular properties of the integral and of the measurable functions. The most important property of measurable functions followed easily from the properties of measurable sets: if a function $f$ is the limit of a sequence $f_n$ of measurable functions, then $f$ is measurable.

Finally, the most useful property of the integral – allowing for term-by-term integration – was revealed. In his thesis, Lebesgue only proved the following result that is sometimes called the bounded convergence theorem (the dominated convergence theorem only appears in a paper of Lebesgue later, in 1908):

**Theorem 4.2.1.** *If $f_n$ is a sequence of measurable functions defined on some measurable set $E$, such that $\lim_{n\mapsto\infty} f_n = f$, and $|f - f_n| \leq M$ for all $n$, then:*

$$\int_E f = \lim_{n\mapsto\infty} \int_E f_n$$

The proof is again very neat, typical for the kind of techniques used in arguments involving measurable sets and, in particular, limiting processes on sequences of sets.

For a fixed $\epsilon \geq 0$, consider the set:

$$E_n = \{x \in E: \quad |f(x) - f_{n+p}(x)| \geq \epsilon, \quad \text{for some} \quad p \geq 0\}.$$

Then $E_n$ is measurable, $E_{n+1} \subset E_n$, and, since $\lim_{n \mapsto \infty} f_n = f$, it follows that $\bigcap_{n \geq 1} E_n = \emptyset$, which gives $\lim_{n \mapsto \infty} m(E_n) = 0$.

<u>Note</u>: Lebesgue had proven the following two useful properties of measurable sets:

$m\left(\bigcup_{n \geq 1} E_n\right) = \lim_{n \mapsto \infty} m(E_n)$, where $E_n \subset E_{n+1}$;
$m\left(\bigcap_{n \geq 1} E_n\right) = \lim_{n \mapsto \infty} m(E_n)$, where $E_{n+1} \subset E_n$, provided that, for some $n$, $m(E_n) \neq \infty$.

Now, by the following inequalities:

$$\begin{aligned} \left|\int_E f - \int_E f_n\right| &= \left|\int_E (f - f_n)\right| \leq \int_E |f - f_n| = \int_{E_n} |f - f_n| + \int_{E \setminus E_n} |f - f_n| \\ &\leq \int_{E_n} M + \int_{E \setminus E_n} \epsilon = Mm(E_n) + \epsilon(m(E) - m(E_n)), \end{aligned}$$

which approaches 0 as $n \mapsto \infty$, and thus

$$\int_E f = \lim_{n \mapsto \infty} \int_E f_n.$$

◇

It must be noted that Lebesgue didn't limit his theory to bounded sets or bounded functions, as we have sketched in the above presentation. In fact, Lebesgue's thesis was the starting point of a lifetime's work in this field in which many – if not almost all – results bear his name or can be traced back to his ideas: he generalized his theory for functions of more variables and to certain classes of unbounded functions, and he also showed the power of these generalizations in important applications.

It is also important to point out that Lebesgue was the first to prove that there exists a unique measure on the Borel sets in $\mathbb{R}$ such that the measure of every interval is its length.

## 4.3 Lebesgue on Fourier

Fourier had argued in 1822 [24] that if if a bounded function can be represented by a trigonometric series, i.e.:

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx), \tag{4.7}$$

then this is the Fourier series of $f$, i.e., the coefficients $a_n$ and $b_n$ are given by the integral formulas:

$$a_n = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos nx\, dx,$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin nx\, dx.$$

His argument involved term-by-term integration, which he had taken for granted. The question about the conditions under which term-by-term integration is permissible arose. Secondly, what is understood by integration had to be clarified, in particular the issue of *integrability*, if the coefficients were to have any meaning. We described the remarkable insights that the search for answers to these questions has led to.

After publishing his thesis, Lebesgue declared his objective to demonstrate the usefulness of his newly introduced concept of integral ([33], p.453), and the first domain where he would do this was in the study of trigonometric series.

His first application was proving that Fourier's assertion is true in the context of the Lebesgue integral, where the convergence in (4.7) is understood in the weak sense of conditional convergence, and there are no assumptions made about the function $f$. His proof was based on the bounded convergence theorem, that he had proved in his thesis, which allowed term-by-term integration, and the Riemann function corresponding to the series above (see section 2.7). We present below the main ideas of Lebesgue's proof:

Cantor showed that $a_n$ and $b_n$ approach 0 as $n$ approaches $\infty$, so it is possible to define a continuous function $F(x)$,

$$F(x) = \frac{a_0}{4}x^2 - \sum_{n=1}^{\infty}\left(\frac{a_n cosnx + b_n sinnx}{n^2}\right),$$

obtained by integrating the series in (4.7) term-by-term twice. Riemann showed that $F$ is related to $f$ through the following relation:

$$\lim_{\alpha\mapsto 0}\frac{\Delta^2 F(x)}{\alpha^2} = \lim_{\alpha\mapsto 0}\frac{F(x+\alpha)+F(x-\alpha)-2F(x)}{\alpha^2} = f(x). \tag{4.8}$$

Now Lebesgue considers the function:

$$f(x,r) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} r^n(a_n \cos nx + b_n \sin nx) \quad (0 \leq r < 1). \tag{4.9}$$

By a theorem due to Abel, the convergence of the series in (4.7) implies:

$$f(x) = \lim_{r\mapsto 1} f(x,r) \quad \text{for every fixed} \quad x \in [0, 2\pi]. \tag{4.10}$$

Two results, concerning $F(x)$ and $f(x,r)$, were used in the proof. The first one, about $F(x)$, established a property analogous to the intermediate value theorem:

**Lemma 4.3.1.** $\frac{\Delta^2 F(x)}{\alpha^2}$ *lies between the infimum and the supremum of* $f(x)$, *for* $x \in$

$[x - \alpha, x + \alpha]$.

Secondly, based on Riemann's theorem (4.8), Lemma 4.3.1, and the properties of harmonic functions, Lebesgue proved the following property of $f(x, r)$:

**Lemma 4.3.2.** *If $M$ and $m$ are the supremum and the infimum of $f(x)$, then they are also the infimum and the supremum of $f(x, r)$.*

Now, note that the convergence of $\{a_n\}$ and $\{b_n\}$ implies their boundedness, and thus for any $r < 1$ the series on the right hand side of (4.9) is uniformly convergent by the Weierstrass M-test, since:

$$\sum_{n=1}^{\infty} r^n |a_n \cos nx + b_n \sin nx| \leq \sum_{n=1}^{\infty} r^n. \tag{4.11}$$

It follows that $f(x, r)$ is continuous in $x$ for every fixed $r < 1$ and, by uniform convergence, the equality (4.9) can be integrated term-by-term, which gives

$$r^n a_n = \frac{1}{\pi} \int_0^{2\pi} f(x, r) \cos nx \, dx \quad , \quad r^n b_n = \frac{1}{\pi} \int_0^{2\pi} f(x, r) \sin nx \, dx,$$

and if we choose an arbitrary sequence $r_k \uparrow 1$,

$$a_n = \lim_{k \mapsto \infty} \frac{1}{\pi} \int_0^{2\pi} f(x, r_k) \cos nx \, dx \quad , \quad b_n = \lim_{k \mapsto \infty} \frac{1}{\pi} \int_0^{2\pi} f(x, r_k) \sin nx \, dx.$$

Now, since every $f(x, r_k)$ is continuous in $x$ and $f(x, r_k) \longmapsto f(x, 1) = f(x)$, it follows that $f$ is measurable, hence integrable because it is bounded.

Finally, the uniform boundedness of the sequence $(f(x, r_k))_{k \geq 1}$ , insured by Lemma 4.3.2, allows for the application of Lebesgue's bounded convergence theorem (4.2.1), hence term-by-term integration yields the Fourier coefficients:

$$a_n = \lim_{k \mapsto \infty} \frac{1}{\pi} \int_0^{2\pi} f(x, r_k) \cos nx \, dx = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx \, dx,$$

$$b_n = \lim_{k \mapsto \infty} \frac{1}{\pi} \int_0^{2\pi} f(x, r_k) \sin nx \, dx = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin nx \, dx.$$

However, this first result of Lebesgue in the domain only gave a partial answer to the main problem of Fourier analysis. The other side of the problem can be formulated as follows:

*Given an integrable function $f$, and its Fourier series, i.e., the trigonometric series where $a_n$, $b_n$ are given by Fourier's integral formulas, when does it represent the function $f$, in other words, when does this hold:*

$$f = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx)?$$

The results depend on what the meaning given to the "=" in the relation above: in what sense does the series converge and how does it represent the function?

In 1905 Lebesgue provided his answer to this more general question, too [35]: if $f$ is Lebesgue integrable, then the Fourier series of $f$ converges to $f$ almost everywhere, in a weaker sense, namely, in the Cesaro sense.

The sequence $(a_k)$ is said to have Cesaro limit $l$ if:

$$\lim_{n \mapsto \infty} \frac{a_1 + a_2 + \cdots + a_n}{n} = l.$$

But Lebesgue's notion of measure and integral also had a deeper influence on the subsequent development of analysis: it allowed for the question to be formulated in different terms, giving new meanings to the word "represent" in the above question, and thus opening the avenue for much more powerful investigations through functional analysis.

Namely, the notion of sets of measure zero allowed a line of thinking where functions can be thought of as equal, when they are equal almost everywhere. One can think then in terms of equivalence classes of functions where $f \sim g$ if $f = g$ a.e. or

$\int_a^b |f(x)-g(x)|\,dx = 0$. This integral naturally defines a distance $D(f,g) = \int_a^b |f(x)-g(x)|\,dx$ between Lebesgue integrable functions defined on $[a,b]$. We end up with a space where each point is an equivalence class of functions, which turns out to be a Banach space (complete normed vector space). The most important contribution in this domain, which actually placed the Lebesgue integral in the context of functional analysis, belonged to Riesz, who in 1910 defined $L^p$ spaces as follows [47]:

**Definition 4.3.3.** *The space $L^p$, $p \geq 1$, consists of all the functions such that $|f|^p$ is integrable over $[a,b]$, together with the norm defined as:*

$$\|f\|_p = \left(\int_a^b |f(x)|^p\,dx\right)^{\frac{1}{p}},$$

*and the distance:*

$$d_p(f,g) = \|f-g\|_p$$

Two functions are considered equal (i.e., belonging to the same equivalence class) if the distance between them is 0. The most natural space to work in is the $L^2$ space, in which we have orthogonality. This was generalized to the notion of Hilbert spaces, in which Fourier series expansions can be viewed in terms of orthonormal bases.

Then the partial sums of the Fourier series of $\widetilde{f}$ (equivalence class of $f$ – the integral is the same for functions that are equal a.e., their Fourier coefficients are thus the same and so is then their Fourier series) can be seen as points in the space and new, or rather more powerful questions can be asked: do these points converge to $\widetilde{f}$? Is there a unique trigonometric series converging to an equivalence class? These questions open the way to more powerful mathematics through functional analysis.

As late as 1966, a stronger result was obtained by Carleson by including a stronger assumption: namely, he proved *pointwise convergence* of the Fourier series of a function $f$ by requiring that $f$ be square integrable; in 1970, Hunt proved the same result for $L^p$ spaces [30].

# Chapter 5

## EPILOGUE: A LOOK AT MODERN TEXTBOOK APPROACHES TO INTEGRATION AND DIDACTICAL REFLECTIONS

We have concluded our historical account of the development of integration theory with the first instances where Lebesgue's theory of integration proved its usefulness, but we also touched upon the more far reaching effects of the new theory on the development of analysis in the context of the early 20th century. Analysis would take a new course, shaped by the general context of the beginning of the century, when the axiomatic method became the landmark of a new way of doing mathematics; by 1910 it had become obvious that working with functions as points in an abstract space was much more fruitful. For integration, the notion of mapping was starting to become central, but this is a whole new – indeed, truly exciting – period in the evolution of mathematical thought that we do not aim to dwell upon in much depth in this study. In this chapter, however, in the context of our didactical analysis, we will have a look at some analysis textbooks that introduce the Lebesgue integral, whose approaches also encompass modern developments in integration theory.

In the introductory section of this chapter we give indications about our objectives in doing this analysis, our method of approaching the teaching issues related to integration, and the limitations of our approach. We then proceed to examine certain aspects of the teaching of integration at different levels at university.

## 5.1 Some methodological considerations

In North-American universities integration is taught on three levels:

- in a freshman or pre-university calculus course, the Riemann integral is used, but most of the definitions and proofs are omitted;
- in a more advanced undergraduate analysis course, Riemann integration is presented, including proofs of the results that were used in calculus;
- finally, in an undergraduate course or a beginning graduate course, typically in measure theory, where the Lebesgue integral is introduced.

Our aim is to describe and examine this organization in more detail, and to call attention to the objectives that the teaching of integration has or should have and to the problems that may arise in this context at the different levels. The teaching and learning of calculus has been extensively researched, and that is why we do not endeavor to delve very deeply into identifying problems that could arise or ways to address them in the teaching of integration at this level. The more advanced level of mathematics taught in analysis or measure theory courses has received much less attention in mathematics education research, and thus we believe that it is rather here that our study is of interest.

For the analysis we will look at modern textbooks, in particular at those used in a large, urban North American university (hereafter called "The University"), and we will draw inspiration from our account of the gradual enrichment of the notion of integral in history to raise some didactical issues. The sequencing of the analysis courses we have mentioned above reflects, to a certain extent, the historical succession of events in that it presents Riemann's theory as preceding that of Lebesgue

– of course, with significant variations, across various textbooks, with regards to the organization of the material. However, at the outset, we want to express the following standpoint: through our teaching recommendations we do not naively aim at "re-creating the past"; an organization of the material different from the historical evolution of the discipline may well be justified by expository and didactical reasons. Still, history can be interesting and informative for pedagogical purposes, especially if looked at, as we did, with an interest in the problems, conceptual contexts, and difficulties that marked the development of the subject.

Our analysis is theoretical, and lacking empirical confirmation, it will necessarily result in conjectures needing further investigation. We believe, however, that our reflections on the teaching of integration will be a welcomed contribution to a domain of mathematics education that has received little attention in research.

## 5.2 Integration in calculus courses

Typically, pre-university or freshman calculus courses cover the following topics in integration: the Riemann integral, the fundamental theorem of calculus, some applications of integration (to volume, work, lengths of curves, etc.), and various integration techniques.

In particular, at The University, the calculus course is based on the popular textbook of Stewart [55], where integration is introduced through the section titled *The Area problem.* In this section the area of a region $S$ bounded by the graph of a continuous function $f$ (with $f(x) \geq 0$), the vertical lines $x = a$ and $x = b$ , and the $x$-axis, is found by approximating rectangles; better and better numerical approximations are obtained by increasing the number of rectangles. The area of a region under a curve is then *defined* as the limit of the sums of the approximating rectangles. The second section begins with the definition of the definite integral, with the use of Riemann sums (i.e., Riemann's definition, presented in section 2.5), which are interpreted geometrically by reference to the previous section. A paragraph follows on the topic of "evaluating integrals" using the above definition, with the help of some formulas for sums of powers of positive integers. Finally, the properties of the definite integral are given. Antiderivatives are the subject of the following section (according to the course outline; in the textbook, antiderivatives are treated in an earlier chapter, the one on differentiation), and a table of antidifferentiation formulas is given for some basic functions. A section is then devoted to the fundamental theorem of calculus, while the rest of the course material on integration deals with applications and techniques of integration.

We have stressed in Chapter 2 that the insight at the heart of calculus lies in the dual understanding of the integral as the limit of a sum and as the inverse of differentiation, and it appears that, theoretically, the course, through the sequencing of the topics, aims at conveying such an understanding: the integral is first defined as the limit of a sum, then an operation that is inverse to differentiation is defined – antidifferentiation, and finally they are brought together through a section on the fundamental theorem of calculus. However, the exercises proposed in the textbook are, in their majority, heavily procedural. For instance, they involve calculating estimates for a given number of intervals for areas under given graphs using left endpoints or finding an expression for the area under the graph of a given function as a limit; this requires only a substitution in the definition of area. Nothing seems to bring to together the two very different interpretations of the integral.

That calculus courses generally fail to achieve conceptual understanding in students has been well known for a long time (for a review of students' difficulties with calculus, see, for instance [57]). The fast pace of the course is certainly an important contributing factor to this situation (see [54]), and the fact that two class sessions (each lasting approximately two hours and a half) are devoted to covering the material we described above (Areas, Definite Integral, Antiderivatives, and the Fundamental Theorem of Calculus), while the rest of the class time allotted for the topic of integration, consisting of five sessions, deals with techniques of integration and with applications, shows that the course is geared more towards developing students' practical abilities with integration rather then a more subtle conception of integration.

*But is this really a problem?* Historically, as we have seen, defining integration as the inverse of differentiation was prevalent and useful for a long time after Calculus was "invented" by Newton and Leibniz, until the re-instatement of the sum-conception of the integral by Cauchy. Certainly, the *informal* idea of these operations as being only two aspects of a single process is not meaningless. As Byers [7] states: "The Fundamental Theorem says that these processes [differentiation and integration] are inverses of one other (when the functions are "reasonable"). Now it may be possible to start with integration and then develop differentiation or vice versa, but the theorem says that, for functions of one variable, neither process is the more fundamental. Actually, the theorem says that there is in fact one process in calculus that is integration when it is looked at in one way and differentiation when it is looked at in another. Another way of putting this is that without the Fundamental Theorem there would be two subjects: differential calculus and integral calculus. With it there is just the calculus, albeit with a multiple perspective. This multiple perspective is essential to an understanding of calculus." ([7], p.50)

The answer to the above question, we believe, depends on the objectives that one or another calculus course has. The course we have looked at is a prerequisite for the science stream, its clientele consists of students enrolled or wishing to enroll in B.Sc.'s in Mathematics, Physics, Chemistry or Biology, or for Engineering and Computer Science students. We believe that at least students majoring in mathematics should come out of this course with the understanding of the definite integral as the limit of a sum, as Cauchy did, and although lacking the empirical confirmation of classroom research, based on our theoretical analysis, we believe that the calculus course may well fail to achieve this goal. Future engineers, on the other hand, may do without the more subtle understanding of the integral that came about with Cauchy's definition of the integral. Students in the engineering or computer science stream may benefit more from a course centered upon applications of calculus. This does not mean that conceptual understanding should be sacrificed for this audience, but introducing the formal definition of the definite integral as the limit of a sum, and then not reinforcing this understanding of integration through appropriate theoretical tasks is useless.

We talked about conceptual understanding, about teaching calculus with applications or about using tasks that reinforce understanding the theory. Research shows that these goals are not easy to achieve. Underlying calculus is the notion of limit, which, research shows, is the source of many difficulties (see [51], [52], [16]). The topic of integration, in particular the understanding of the integral as the limit of a sum, replicates the problems students have with understanding limits. Schneider [50] shows how some students, in considering upper and lower sums for the function $f(x)$ from 0 to 1 by taking more and more rectangles, think that "as long as the rectangles have thickness, they do not fill up the surface under the curve, and when they become reduced to lines, their areas are equal to 0 and cannot be added" ([50], p.33). In his study about college students' understanding of integration, Orton [40] found that students were fairly proficient in applying basic techniques of integration, but among the tasks she gave them during interviews, a core of a few items, concerned with the understanding of the integral as the limit of a sum, constituted the main stumbling block. He indicates that the procedure of breaking up an area or volume, making use of a limiting process, and providing the reasons why such a method works were not part of the students' understanding of the integral. Orton argues for laying a satisfactory groundwork in understanding the notion of limit before introducing more advanced topics in calculus, but also draws the attention to the fact that students' difficulties with algebra may also obscure their attempts to deal with calculus concepts. For instance, in any approach to integration which can be described as an attempt at a formal proof, such as the method based on summing rectangular areas, certain formulae for the summation of series are required, that for some students may be a first encounter. For the teaching of integration at these early stages, Orton proposes that calculus be linked with appropriate illustrations, such as exploring polygons with an increasing number of sides, investigating elementary number series, studying areas of irregular shapes, etc., as well as numerical explorations using calculators.

## 5.3 Integration in analysis courses

Mathematics majors' second encounter with integration is in analysis courses. A first analysis course at The University starts with a preparatory section in *Set Theory and Elements of Proofs*, and covers the following topics over a one-semester span: Sequences, Limits of Functions and Continuity, and Derivatives. The sequel analysis course covers Riemann Integration, Series of Real Numbers, Sequences and Series of Functions, Taylor's Theorem, Power Series and the Derivative of Inverse Functions. The textbook for this course, *Introductory Real Analysis*, by F. D'Angello and M. Seyfried [17], devotes a chapter to Riemann integration, and, according to the course outline, the topic is covered over four weeks of classes. In the first section, Riemann integrability and the Riemann integral of a bounded function on an interval are defined in terms of upper and lower sums, and necessary and sufficient conditions for Riemann integrability are then obtained. Modern textbooks thus rely mostly on Darboux and Peano's reformulations of Riemann's work on integration (see section 2.6). In the subsequent section, however, Riemann's original definition of the integral, as a limit of Riemann sums is given, and the equivalence of the two definitions is proven.

The properties of the Riemann integral are presented next, and then continuous and monotone functions are shown to be integrable. A section is devoted to the Fundamental Theorem of Calculus, and finally, improper integrals are described and comparison tests are established for deciding when an improper integral converges or diverges.

The change of scene from calculus to analysis courses is abrupt. For students, the proofs "migrate" from the the theoretical realm to the practical one: for instance, students are now required to prove that a function is integrable. Their tools are epsilons and deltas, convergent sequences, suprema and infima, etc. The first section on Riemann integration is titled *Existence of the Riemann integral*, and is characteristic for this shift from the symbolic manipulations students have grown accustomed to in calculus to the formalism of existence theorems of analysis. The Riemann integral of a function was defined in the calculus course as the limit of Riemann sums, under the very much implicit assumption of continuity. In the analysis course, on the other hand, it is shown to exist if a certain condition is satisfied, and this condition – i.e., that upper and lower Riemann integrals coincide – is carefully examined. For instance, the author of the textbook and the teacher would go through great "trouble" to prove what a student may well see as an obvious statement: $\underline{\int_a^b} f = \overline{\int_a^b} f$, and then $f$ is said to posses a Riemann integral, if the two agree.

Clearly, the objective of teaching integration in the undergraduate analysis course seems to be the understanding of the notion of integrable function, in the sense that Riemann first did, by asking the question: in what cases does a function admit integration and in what cases it does not? We see two main obstacles to reaching this objective. Firstly, students' encounter with formal proofs and definitions makes the passage from calculus to analysis very difficult. Tall [57] mentions this problem when he categorizes the representations available in the teaching of calculus. He asserts that enactive representations may act as an intuition basis for calculus built with numeric, symbolic and visual representations, but that the formal definitions and theorems of analysis require different cognitive qualities, and significant constructions and reconstructions of knowledge. He argues that the conceptual gap between the practical calculations and symbolic manipulations in calculus and the theoretical proofs and definitions of analysis is so wide that this causes a severe chasm in courses. He gives the example of the mean value theorem that "inhabits the realms of existence theorems in analysis and sits uncomfortably in the computations of elementary calculus" ([57], p. 23) and we would add that the same is true for the definite integral: it is the subject of computations in calculus (in using techniques of integration, or in applications to areas, volumes, etc.), while in analysis one questions its very existence for a given function.

However, besides the problem of formalism, specific to analysis courses in general, which is also implicitly recognized by the introductory section on set theory and proofs of the first analysis course, we want to draw attention to another obstacle that may become important in the teaching of integration in analysis courses. We discussed in Chapter 2 the fact that it was Riemann's acceptance of the modern concept of function that underlay his innovative approach to integration. Research in mathematics education documents how students' problems with the concept of function hinder their learning of calculus concepts. Dreyfus and Eisenberg, [21] cited in [57], found that students hold the belief that a relation is a function when it can be represented by only one formula, and that students view algebraic and graphical data as separate, with a graphical representation with no formula having no meaning for most students. Bakar and Tall [2] also found that students had certain misconceptions about the notion of function: that it has to be given by a formula that should always include $x$, that the graph should have a recognizable shape (e.g. polynomial, trigonometric or exponential), and that it has to have certain "continuous" properties. Another type of finding is that of Graham [23], who showed that students tend to believe that given a function, they are expected to do something with it, such as substitute a value (see also [6], [53]).

It is less likely, however, that the more mathematically sophisticated clientele of analysis courses – consisting of students majoring in mathematics, statistics or actuarial studies – would have such problems with functions. We believe the main problem for these students, that may act as an obstacle for the understanding of the notion of integrable function, is of a "higher order" type: it has to do with their conceiving of the notion of function in the sense that Cauchy did, as mainly continuous (at most, with a finite number of discontinuities) and given by an analytic expression. In precalculus courses students would gradually move from linear to quadratic, then to polynomial and rational functions, and later to trigonometric, exponential and logarithmic functions. In calculus courses, one also deals only with elementary functions. So, although, in theory, students would be given a definition stating that a function $f$ is a rule that assigns to each element in a set $X$ a unique element in a set $Y$, in practice, they would only see functions given by formulas. The set-theoretic definition proved to be very powerful for the development of mathematics and Riemann's approach to the problem of integral is only one example that illustrates it. However, mathematicians and/or teachers, we believe, do not rely entirely on such formal concepts. For them, too, more intuitive ideas of functions as change, or even "images" of functions as formulas coexist with the more formal notion of ordered pairs, but most probably in a more fluid and effective way. We believe this kind of composite of ideas regarding functions should be conveyed in teaching, as opposed to an approach where classes of elementary functions are presented and the ability to perform algebraic manipulations is developed. This is possible, we believe, in analysis courses, where sequences and series of functions are introduced, thus, an opportunity for dealing with not so well behaved functions exists. Then, possibly, students conceiving of functions as more than analytic expressions and continuous, would better grasp the understanding of integral that came about with Riemann's work.

## 5.4 Integration in measure theory courses

While the content in calculus and analysis courses on the topic of integration is generally similar in various textbooks, when it comes to measure theory courses, the situation is different: the way the Lebesgue measure and integration is introduced varies significantly across different textbooks. Therefore, we will first provide the background for our analysis by producing a more detailed overview of a few ways to present the Lebesgue measure and integration. In this overview, we will trace, at the same time, the origins of the main ideas of these modern approaches, concluding, in this way, our historical account of the development of the modern integration theory.

### 5.4.1 The Carathéodory method: generating measures through an outer measure

Constantin Carathéodory did pioneering work in abstract measure theory (see [38]). In 1918, he proposed a general method to produce an abstract measure with the help of a set function which is called outer measure. The novelty of Carathéodory's construction is the fact that it makes no use of the inner measure, so there is no need to restrict the argument to bounded sets, as a preliminary step. There are sets that are not bounded, and which we would expect to have finite measure: the most notable example is the set of rational numbers, $\mathbb{Q}$, which is countable, and has measure zero.

The ideas in the following presentation are due to Carathéodory and can be found – with variations – in many textbooks, a classic being the book of Paul Halmos, *Measure theory* [26], and one that is also extensively used being Royden's *Real Analysis* [48].

**Definition 5.4.1.** *For a fixed non empty set $T$, an* outer measure $\lambda$ *is a set function defined on $\mathcal{P}(T)$ (the power set of $T$, i.e., the set of all subsets of $T$) such that $\lambda(E) \geq 0$ for any $E \subset T$, and the following conditions are satisfied:*

*(a) $\lambda(\emptyset) = 0$ and $\lambda(E) \leq \lambda(F)$ whenever $E \subseteq F$;*

*(b) For every sequence $(A_n)_{n\geq 1}$,*

$$\lambda\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} \lambda(A_n).$$

*( The above property is usually called countable subadditivity.)*

Given an outer measure, Carathéodory defines a notion of measurability with respect to it:

**Definition 5.4.2.** *With $\lambda$ as in the previous definition, we say that $E \subset T$ is $\lambda$-measurable if*

$$\lambda(M) = \lambda(M \cap E) + \lambda(M \cap (T \backslash E)) \quad \text{for every} \quad M \subset T. \tag{5.1}$$

*We shall denote by $\mathcal{F}_\lambda$ the set of all $\lambda$-measurable sets in $T$.*

Carathéodory's condition requires that a measurable set can be used to cut any other set and the outer measure of the set that is cut will be equal to the sum of the outer measures of the two pieces.

The main result is that the given outer measure "becomes" a measure, in particular, it acquires the property of countable additivity, if restricted to measurable sets defined according to Carathéodory's condition. In other words, this result holds:

**Theorem 5.4.3 (Carathéodory).** *$\mathcal{F}_\lambda$ is a $\sigma$-algebra and the restriction of $\lambda$ to $\mathcal{F}_\lambda$ is a measure,*

where the notion of measure is defined as follows:

**Definition 5.4.4.** *If $T$ is a non-empty set and $\mathcal{F}$ is a $\sigma$-algebra of subsets of $T$, then $\mu : \mathcal{F} \longmapsto \overline{\mathbb{R}}_+$ is called a* measure *on $\mathcal{F}$ if*

$$\mu(\emptyset) = 0 \quad and \quad \mu\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mu(A_n)$$

*for every pairwise disjoint sequence of sets $A_n \in \mathcal{F}$.*

Note: One can take $\mathcal{F}_0 = \{\emptyset, T\}$ and $\mu(\emptyset) = 0$ and $\mu(T) = +\infty$. Then $\mathcal{F}_0$ is in fact a $\sigma$-algebra and $\mu$ a measure on $\mathcal{F}_0$. Some authors want to avoid this trivial case and use a small modification of the above definition (see next sections).

**Remark 5.4.5.** It follows from the definition that $A \in \mathcal{F}_\lambda$ whenever $\lambda(A) = 0$, therefore the measure given by Carathéodory's theorem is also a complete measure.

One can arguably find the origin of Carathéodory's condition in Lebesgue's notion of measure, at least in a logical sense. From the definition of the inner measure:

$$m_i(E) = b - a - m_e([a, b] \backslash E),$$

it can be observed that a set $E \subset [a, b]$ is measurable according to Lebesgue's definition ($m_i(E) = m_e(E)$), if and only if:

$$b - a = m([a, b]) = m_e(E) + m_e([a, b] \backslash E).$$

Now $m_e$ has all the properties of an outer measure and the bounded set $E$ is measurable, in the original Lebesgue construction, if and only if

$$b - a = m([a, b]) = m_e([a, b] \cap E) + m_e([a, b] \backslash E) \quad \text{for every} \quad a < b,$$

not necessarily with the condition $E \subset [a, b]$. But $m_e(E)$ can be computed for every subset $E \subset \mathbb{R}$, not necessarily bounded, and the particular case of the interval $[a, b]$

can be dropped. Then, one can *define* as "measurable" every subset $E \subset \mathbb{R}$ with the property that

$$m_e(M) = m_e(M \cap E) + m_e(M \backslash E) \quad \text{for every} \quad M \subset \mathbb{R}.$$

◇

Carathéodory's theorem is by now a starting point of many measure theory textbooks. Through particular choices of the "generating" outer measure one can obtain different measures. In fact, most textbooks use this method in such a concrete way with the purpose of proving the existence of the Lebesgue measure on $\mathbb{R}$ (see [48]). Namely, the Lebesgue measure is obtained by means of the following outer measure:

Let $\mathcal{C}$ be the set of all intervals of the form $I = [a, b)$, and set $l(I) = b - a$, where $l$ stands for the length of the interval $I$. Define, for every $A \subset \mathbb{R}$,

$$m^*(A) = \inf \left\{ \sum_{k \geq 1} l(I_k) \; : \; A \subset \bigcup_{k \geq 1} I_k, \quad I_k \in \mathcal{C} \right\}.$$

Once it is established that $m^*$ is an outer measure, let $\mathcal{L}$ be the set of all $m^*$ measurable sets as defined in 5.4.2. Carathéodory's theorem ensures that $\mathcal{L}$ is a $\sigma$-algebra and the restriction of $m^*$ to $\mathcal{L}$ is a measure, and it can be easily shown that for every $I = [a, b)$,

$$m^*(I) = l(I) = b - a.$$

Finally, the following result holds:

**Proposition 5.4.6.** *Every interval $I = [a, b) \in \mathcal{C}$ is $m^*$-measurable.*

*Proof.* The proof is a relatively easy consequence of the following obvious properties of the intervals in the class we have denoted by $\mathcal{C}$ :

(a) $\emptyset = [a, a) \in \mathcal{C}$ and $I \cap J \in \mathcal{C}$ whenever $I, J \in \mathcal{C}$.

(b) If $I, J \in \mathcal{C}$, then either $I \backslash J \in \mathcal{C}$ or $I \backslash J = J_1 \cup J_2$, where $J_1, J_2$ are disjoint elements of $\mathcal{C}$.

(c) If $I, J \in \mathcal{C}$, then

$$l(I \cap J) + l(I \backslash J) = l(I),$$

where, if $I \backslash J \notin \mathcal{C}$, we take $l(I \backslash J) = l(J_1) + l(J_2)$ if $I \backslash J = J_1 \cup J_2$.

Now, fix an arbitrary interval $I = [a, b)$, an arbitrary $M \subset \mathbb{R}$ and a sequence $I_n \in \mathcal{C}$ such that

$$M \subset \bigcup_{n \geq 1} I_n.$$

From our observation (a), $I_n \cap I \in \mathcal{C}$ ,

$$M \cap I \subset \bigcup_{n \geq 1} (I_n \cap I), \quad \text{hence} \quad m^*(M \cap I) \leq \sum_{n \geq 1} l(I_n \cap I),$$

$$M \cap (\mathbb{R} \backslash I) \subset \bigcup_{n \geq 1} (I_n \backslash I),$$

and, by our observation (b), $I_n \backslash I$ $(n \geq 1)$ can be rewritten as a sequence of intervals of $\mathcal{C}$ which also covers $M \cap (\mathbb{R} \backslash I))$, therefore

$$m^*(M \cap (\mathbb{R} \backslash I) \leq \sum_{n \geq 1} l(I_n \backslash I).$$

By our observation (c),

$$\sum_{n \geq 1} l(I_n) = \sum_{n \geq 1} (\, l(I_n \cap I) + l(I_n \backslash I)\,) =$$

$$= \sum_{n \geq 1} l(I_n \cap I) + \sum_{n \geq 1} l(I_n \backslash I) \geq m^*(M \cap I) + m^*(M \cap (\mathbb{R} \backslash I)).$$

Since $I_n \in \mathcal{C}$ was an arbitrary sequence which covers $M$, it follows

$$m^*(M) \geq m^*(M \cap I) + m^*(M \cap (\mathbb{R} \backslash I)),$$

therefore $I = [a, b)$ is $m^*$ measurable (the opposite direction inequality is assured by countable subadditivity). □

◇

We have reached the end of our argument. Since $\mathcal{L}$ is a $\sigma$-algebra and contains every interval $I = [a, b)$ by Proposition 5.4.6, it follows that it contains the $\sigma$-algebra generated in $\mathbb{R}$ by the set of *all* such intervals which is the $\sigma$-algebra of all Borel subsets of $\mathbb{R}$, usually denoted by the symbol $\mathcal{B}$. If we denote by $m$ the restriction of $m^*$ to $\mathcal{B}$, then $m$ is a measure and the measure of an interval, whether open, closed or half-open is, as we had expected, its length.

Assuming the completeness of such a measure, uniqueness is immediate.

The elements of $\mathcal{L}$ are the usual Lebesgue measurable sets and, by the definition of $m^*$ it is easy to check that:

> $A \subset \mathbb{R}$ is Lebesgue measurable if and only if $A = B \cup C$, where $B$ is a Borel set and $m^*(C) = 0$, therefore – see 5.4.5 – m(A)=m(B).

Thus, there exists a unique measure $m$ on the Borel sets of $\mathbb{R}$ such that the measure of an interval is its length.

This is in fact Lebesgue's discovery around 1902. It is a fundamental property of $\mathbb{R}$, it is the natural and the unique translation invariant measure on $\mathbb{R}$.

### 5.4.2 Riesz's method: the integral before the measure

Another approach found in textbooks introducing Lebesgue integration puts integration rather than measure to the fore and constructs Lebesgue integrable functions as differences of limits of increasing sequences of step functions with bounded integrals. This was originally done by the Hungarian mathematician F. Riesz in 1912 [46], and adopted by other writers in analysis (such as Apostol in [1], Weir in [60], or Chae in [13]).

Riesz's method is to arrive at the Lebesgue integral by a step by step construction, without using the measure. His motivation appeared to be the preoccupation with making Lebesgue's integration theory more accessible to mathematics students (see [42]), and the main difficulty of this new theory – he believed – lay in the prerequisite study of measure theory.

In what follows we present a scheme of this method, which starts with an independent definition of sets of measure zero, the only instance where the notion of measure appears:

**Definition 5.4.7.** *We say that $A \subset \mathbb{R}$ has measure zero if for every $\epsilon > 0$, there exists a sequence of bounded open intervals $I_n = (a_n, b_n)$, such that*

$$A \subset \bigcup_{n \geq 1} I_n \quad \text{and} \quad \sum_{n \geq 1} l(I_n) = \sum_{n \geq 1} (b_n - a_n) \leq \epsilon.$$

An immediate consequence of this definition is that any (at most) countable set has measure zero and a countable union of such sets sets is again of measure zero.

These sets are mainly used for the following definition:

**Definition 5.4.8.** *If $G \subset \mathbb{R}$, $f_n : G \longmapsto \mathbb{R}$ $(n \geq 1)$ and $f : G \longmapsto \mathbb{R}$ are arbitrary, we say that $(f_n)$ converges to $f$ almost everywhere, in short*

$$f_n(t) \longmapsto f(t) \quad a.e.,$$

*if the set of $t$ such that $f_n(t)$ does not converge to $f(t)$ has measure zero.*

Recall also that the *characteristic function* of a set $E$ is

$$\chi_E(t) = \begin{cases} 1 & \text{for} \quad t \in E, \\ 0 & \text{for} \quad t \notin E. \end{cases}$$

Then step functions and their integrals are introduced.

**Definition 5.4.9.** *A function $\varphi : [a,b] \longmapsto \mathbb{R}$ is called a* step function *if there is a partition of the interval $[a,b]$*

$$a = x_0 < x_1 < \cdots < x_n = b$$

*such that $\varphi$ is constant on every open interval $I_k = (x_{k-1}, x_k)$, i.e., $\varphi(x) = a_k$ for $x \in I_k$, $k = 1, 2, \ldots, n$.*

It follows from the definition that every step function can be written as

$$\varphi(x) = \sum_{k=1}^{n} a_k \chi_{I_k}(x) \quad a.e.$$

and the class of all step functions behaves well with respect to elementary algebraic operations, i.e. sums, products, etc. It is obvious that a step function has points of discontinuity only at the points of the partition and is a very simple Riemann integrable function.

**Definition 5.4.10.** *Let*

$$\int_a^b \varphi(x)\, dx = \sum_{k=1}^{n} a_k (x_k - x_{k-1})$$

*be the integral of the step function $\varphi(x) = \sum_{k=1}^{n} a_k \chi_{I_k}(x)$.*

The integral of a step function is its Riemann integral and, consequently, has all the common properties of an integral, such as linearity, monotonicity, and so on. Then the integral is gradually extended to larger classes of functions. To this end, the following two lemmas are essential:

**Lemma 5.4.11.** *Let $\varphi_n$ be a monotonically decreasing sequence of positive step functions defined on $[a,b]$. Then $\varphi_n(x) \longmapsto 0$ almost everywhere in $[a,b]$ if and only if*

$$\int_a^b \varphi_n(x)\,dx \longmapsto 0.$$

**Lemma 5.4.12.** *Let $\varphi_n$ be a monotonically increasing sequence of step functions on $[a,b]$, and assume that there is a real number $A$ such that*

$$\int_a^b \varphi_n(x)\,dx \leq A \quad \textit{for every} \quad n \geq 1 \quad .$$

*Then $\varphi_n(x)$ converges almost everywhere.*

The above two lemmas are the main tools in the construction of the Lebesgue integral: more general functions are approximated by step functions. First, the class $L^+$ of functions which are almost everywhere (a.e.) limits of increasing sequences of step functions with bounded integrals, is introduced and their integral is defined:

**Definition 5.4.13.** *Let $L^+$ be the class of all functions $f : [a,b] \longmapsto \overline{\mathbb{R}}$ (the extended real numbers set) such that*

*(a) There exists an increasing sequence $\varphi_n$ of step functions on $[a,b]$ with*

$$\varphi_n(x) \longmapsto f(x) \quad a.e..$$

*(b) The sequence of integrals*

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

*is bounded .*

*For every $f \in L^+$, let*

$$\int_a^b f(x)\,dx = \lim_n \int_a^b \varphi(x)\,dx. \tag{5.2}$$

Finally, the class of Lebesgue integrable functions is defined as the set of differences $\{\, g - h \,:\, g, h \in L^+\}$:

**Definition 5.4.14.** *Let $L$ be the class of all those $f : [a, b] \longmapsto \overline{\mathbb{R}}$ which can be written as $f = g - h$, where $f, g \in L^+$.*

*$L$ is the set of Lebesgue integrable functions and*

$$\int_a^b f(x)\,dx \stackrel{def}{=} \int_a^b g(x)\,dx - \int_a^b h(x)\,dx \tag{5.3}$$

*is the integral of $f$.*

The above definitions hide in their few lines a lot of work to be done. First of all, it has to be shown that $\int_a^b f(x)\,dx$ from (5.2) do not depend on the particular sequence of step functions and $\int_a^b f(x)\,dx$ from (5.3) do not depend on the functions $g, h \in L^+$ such that $f = g - h$.

Everything is achieved in the Riesz approach to the Lebesgue integral through a sequence of simple arguments, the most difficult being the above two lemmas. The main properties of the Lebesgue integral are proved without any mention of a measure, and, once all the tools are ready, the Lebesgue measure on $\mathbb{R}$ is defined in terms of the integral. First, measurable functions are defined as functions that can be represented almost everywhere as a limit of a sequence of step functions, then measurable sets are said to be the sets such that their characteristic function is measurable, and finally, the measure is defined by the integral of the characteristic function of that set.

### 5.4.3 Radon, Fréchet and Riesz: abstract integration and measure

Without claiming to have covered in an extensive or systematic manner the large number of approaches to Lebesgue integration and measure, we now turn to one of the most recent approaches to introducing the Lebesgue integral and measure.

Walter Rudin wrote a very popular text book in real analysis in which the Lebesgue integral is presented in an abstract context [49]. The ideas on which this exposition is based can be traced back to Radon, who, in 1913, introduced an integral in the $n$-dimensional Euclidean space, that was based on an abstract measure – a completely additive set function – rather than on the particular measure of Lebesgue (see [38]). Lebesgue's and Stieltjes integrals (the latter not discussed in our study) were special cases of his definition. Even further abstraction was achieved through the work of Fréchet in 1915, where the integral was defined for real functions defined on completely abstract sets, freed of any Euclidean ties.

Fréchet had also introduced, in his thesis of 1906, the notion of distance on an abstract set, the basis of today's notion of metric space, which seems to have inspired Riesz to introduce the notion of $L^p$ spaces. In Riesz's work one can also find the origins of the modern view of integrals as linear functionals, namely in his representation theorem. In its original form, *the theorem* stated that all continuous linear functionals are integrals [47]:

> *Ètant donée l'opération $A[f(x)]$, on peut déterminer la fonction à variation bornée $\alpha(x)$, telle que, quelle que soit la fonction continue $f(x)$, on ait:*
>
> $$A[f(x)] = \int_0^1 f(x)\, d\alpha(x).$$

Integrals came to be later *defined* as continuous linear functionals (for a very interesting account of the metamorphosis of this theorem from its beginnings to the present times, see a paper by J.D. Gray [25]). Not surprisingly, Riesz is considered to have the most important role in popularizing the work of Lebesgue, and in giving it its important place in the theory of function spaces.

Rudin's exposition of an abstract theory of measure and integral starts by giving legitimisation to such an approach. If one looks at Riemann's definition of the integral for some $f : [a, b] \longmapsto \mathbb{R}$, then $\int_a^b f(x)\,dx$ appears as the limit of sequences of the form

$$R_n = \sum_{k=0}^{n-1} c_k(x_{k+1} - x_k),$$

where $a = x_0 < x_1 < \cdots < x_n = b$ are partitions of $[a, b]$, with norm approaching 0. On the other hand, if we think of $m$ as the "natural" measure on $\mathbb{R}$, that is, if we let:

$$m([x_k, x_{k+1}]) = x_{k+1} - x_k \quad (0 \le k \le n-1),$$

and if we put (see *Definition* 5.4.9 and *Definition* 5.4.10):

$$\varphi(x) = \sum_{k=0}^{n-1} c_k \chi_{[x_k, x_{k+1}]}(x),$$

then the above $R_n$ can be re-written as

$$R_n = \sum_{k=0}^{n-1} c_k(x_{k+1} - x_k) = \int_a^b \varphi(x)\,dx = \sum_{k=0}^{k-1} c_k\, m([x_k, x_{k+1}]) \quad .$$

Now, a natural question arises: why should we limit the study of integration to the special case of intervals, instead of working with arbitrary measurable sets, and why should we consider only step functions [*Definition* 5.4.8] and not the more general

$$s = \sum_k c_k \chi_{A_k},$$

where $A_k$ are measurable sets, with the integral of $s$, suggested by 5.4.10, as

$$\int s(x)\,dx = \sum_{k=0}^{n-1} c_k\, m(A_k)?$$

Thus, the geometry of the underlying space is "removed" and this makes integration become a tool of much wider applicability. We will use, in the sequel, Rudin's notation.

A measurable space is an ordered pair $(X.\mathfrak{M})$, where $X$ is a non empty set and $\mathfrak{M}$ is a $\sigma$-algebra of subsets of $X$. (The definition of a $\sigma$-algebra can be found in 3.5.1).

**Definition 5.4.15.** *If $f : X \longmapsto \overline{\mathbb{R}}$, we say that $f$ is* measurable *if $f^{-1}(D) \in \mathfrak{M}$ for every open set $D \subset \overline{\mathbb{R}}$.*

*$s : X \longmapsto \mathbb{R}$ is said to be* simple *if it can be written as*

$$s = \sum_{k=1}^{n} c_k\, \chi_{A_k},$$

*where $c_k$ are real numbers and $A_k \in \mathfrak{M}$. (Note: the function $s$ could have several representations in the form of the above sum.)*

A fundamental property is the following

**Theorem 5.4.16.** *$f : X \longmapsto \overline{\mathbb{R}}$ is measurable if and only if there is a sequence $s_n$ of simple functions such that $s_n(x) \longmapsto f(x)$ for almost every $x \in X$. Moreover, if $f$ is positive, then $s_n$ can be found such that*

$$0 \leq s_n(x) \leq f(x) \quad \textit{for almost every} \quad x \in X.$$

Measurable functions have the regular algebraic properties, i.e. sums, products of measurable functions remain measurable, and if $f : X \longmapsto \overline{\mathbb{R}}$ is measurable, then

$|f|$ is measurable and

$$f^+(x) = \max\{\, f(x), 0\,\} \quad \text{and} \quad f^-(x) = -\min\{\, f(x), 0\,\}$$

are also measurable. Note that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Finally, every simple function is measurable because $\chi_A$ is measurable if and only if $A \in \mathfrak{M}$.

Now assume that $\mu : \mathfrak{M} \longmapsto \overline{\mathbb{R}}_+$, is a fixed measure, defined as follows:

**Definition 5.4.17.** *A (positive) measure on the $\sigma$-algebra $\mathfrak{M}$ is a function*

$$\mu : \mathfrak{M} \longmapsto \overline{\mathbb{R}}_+$$

*such that*

$$\mu\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mu(A_n)$$

*for every sequence of pairwise disjoint sets $A_n \in \mathfrak{M}$, and there is at least one set $A \in \mathfrak{M}$ such that $\mu(A) < +\infty$.*

The integral is defined next, first for simple functions and then for positive measurable functions:

**Definition 5.4.18.** *For every $E \in \mathfrak{M}$ and every simple function*

$$s = \sum_{k=1}^{n} c_k \, \chi_{A_k}, \quad A_k \in \mathfrak{M}, \tag{5.4}$$

*define the* integral of $s$ over $E$ *as*

$$\int_E s \, d\mu \stackrel{def}{=} \sum_{k=1}^{n} c_k \, \mu(A_k \cap E),$$

*with the convention $0 \cdot \infty = 0$, meaning that we take $c_k \, \mu(A_k \cap E) = 0$, whenever $c_k = 0$ and $\mu(A_k \cap E) = +\infty$. (It is shown that $\int_E s \, d\mu$ does not depend on the particular* *representation in (5.4).)*

It is assumed here that $\mu\{x: \ s(x) \neq 0\} < \infty$; this avoids the $\infty - \infty$ situation.

**Definition 5.4.19.** *For every positive measurable function $f : X \longmapsto \overline{\mathbb{R}}_+$ and $E \in \mathfrak{M}$, set*

$$\int_E f \, d\mu = \sup \left\{ \int_E s \, d\mu \ : \ s \ \textit{simple}, \ 0 \leq s \leq f \right\}.$$

*$\int_E f \, d\mu$ is the* integral of $f$ over $E$ *or simply the* integral of $f$ *in the $E = X$ case. (Note that $\int_E f \, d\mu = +\infty$ is a possibility.)*

Finally, everything is ready for the following fundamental definition of *integrability*:

**Definition 5.4.20.** *If $f : X \longmapsto \overline{\mathbb{R}}$ is an arbitrary measurable function, and $E \in X$, then $f$* is integrable over $E$ *if*

$$\int_E |f| \, d\mu < +\infty.$$

*In this case*

$$\int_E f^+ \leq \int_E |f| \, d\mu < +\infty \quad \textit{and} \quad \int_E f^- \leq \int_E |f| \, d\mu < +\infty,$$

*and* the integral of $f$ over $E$ *is defined as*

$$\int_E f \, d\mu = \int_E f^+ \, d\mu - \int_E f^- \, d\mu.$$

Thus, given a measure, an integral on measurable functions can be associated with it. The properties of this integral can be proved using only elementary arguments, indeed in an abstract setting, but without making the theory any more difficult than the special case of the real line, Rudin asserts. Besides, there is no need for bounded intervals, bounded functions or partitions of intervals.

Conversely, when it comes to "producing" a measure Rudin's approach is also abstract. After laying down some topological preliminaries, he uses the Riesz representation theorem to prove the existence of a measure (actually a class of measures) with certain properties. Namely, any positive linear functional defined on a space satisfying certain topological properties "arises" as the integral with respect to a measure.

Then, defining the Lebesgue measure on $\mathbb{R}^k$ – a locally compact topological space – becomes equivalent to the task of defining a suitable positive linear functional on $C_c(\mathbb{R}^k)$ (the space of continuous functions with compact support defined on $\mathbb{R}^k$).

Keeping with the modern approach the subsequent chapter deals with classes of integrable functions, $L^p$ spaces. It is obvious that, in Rudin's book, integration and measure are presented rather as prerequisites, or tools, for the powerful applications that follow.

### 5.4.4 Discussion of teaching issues

We saw that the Lebesgue measure and integration can be introduced in at least three different ways; the obvious – and somewhat simplistic – question is *which one is better*?

Again, we believe that the answers greatly depend on the objectives of the teaching of Lebesgue integration and measure. That is why we will give our answer to the above question and, in general, reflect on the teaching of Lebesgue's integration theory through two more specific questions: *should Lebesgue integration be taught exclusively*? and, *should Lebesgue integration be taught at all?*

We started our research with the belief that not introducing the Riemann integration in undergraduate analysis and starting directly with the Lebesgue integration might be an interesting option. That is because Riemann integration is not logically necessary for the introduction of the modern, more general, theory of integration.

The only reason, we believed, for introducing the Riemann integral in a first stage, would be that is is the only option for a population of students lacking mathematical maturity. We have acquired a different point of view after undertaking the historical research: Riemann integration might well require a great deal of mathematical maturity and sophistication, too. Riemann's definition of the integral, and his contribution to integration theory, consisted of more than a slight generalization of Cauchy's definition of the integral as the limit of approximating sums. With his notion of integral, Riemann actually opened a new line of questioning: how discontinuous can a function be while still being integrable? This is an important question that also provides the opportunity for developing in students a more sophisticated notion of function, and Riemann integration is, we believe, the right context to start asking it. That is why, provided an effort is made to convey Riemann's insight in teaching, our opinion is that Riemann integration should not be replaced, in undergraduate mathematics, by Lebesgue integration.

Now, at the other end of the spectrum is the question about whether Lebesgue integration should be introduced at all at university. Besides the obvious time constraint – a first acquaintance with Lebesgue's theory requires significantly more time than the presentation of the Riemann integral – the conceptual difficulties of the former are certainly the main concern. Norbert Wiener, who attributes the early achievements of his career to his acquaintance with the Lebesgue integral, trying to suggest the main idea of the Lebesgue integration, expresses this concern: "It is easy enough to measure the length of an interval along a line or the area inside a circle or other smooth, closed curve. Yet when one tries to measure sets of points which are scattered over an infinity of segments or curve-bound areas, or sets of points so irregularly distributed that even this complicated description is not adequate for them, the very simplest notions of area and volume demand high-grade thinking for their definition. The Lebesgue integral is a tool for measuring such complex phenomena.

The measurement of highly irregular regions is indispensable to the theories of probability and statistics; and these closely related theories seemed to me to be on the point of taking over large areas of physics." ([59], p. 23).

At The University, the existence of two versions of the measure theory course, appears to address this concern. The undergraduate course, most probably assuming a less mathematically mature clientele uses Royden's textbook (see section 5.4.1), where the Lebesgue measure and integral are introduced through Carathéodory's method in a less abstract setting with the generating measure chosen at the outset to be the Lebesgue outer measure. The graduate course, on the other hand, is based on Rudin's book (section 5.4.3), where abstract integration with respect to any given measure is presented, and measures are produced through the Riesz Representation Theorem.

While this organization into two courses may succeed in providing the two different audiences with appropriately tailored "doses" of abstraction, the question still remains about whether students come out of this course with a deep understanding of the significance of Lebesgue's theory in analysis, or in mathematics, in general. We have seen that this theory was not the result of a search for a more general theory of integration, but of a search for answers to problems for which the Riemann integral was inadequate; and the main problem was certainly not the inability to integrate Dirichlet's function – an impression that students may come out of these courses with. Mathematicians developed a new theory of integration sometimes because they needed more powerful analytical tools in their search for answers to quite practical problems. The notable example we have looked at closely in our research, the questions related to Fourier series, was at the origin of many important discoveries that not only translated into direct answers to these questions, but also ended up having a larger influence on mathematics. We have talked about the deeper significance of the new integration theory in the emergence of functional analysis, and about how, in a way, it made analysis simpler. Therefore, we believe that the challenge in the teaching of advanced integration theory lies not so much in the logical difficulties of the theory, but in getting students to have a feel of what motivates the introduction of Lebesgue integration, and of its analytical power in applications.

One way to do this is by introducing "meta" aspects into teaching, that is, by maintaining an explicit discourse in class where information or knowledge *about* mathematics would be embedded into the teaching of it. The effectiveness of such an approach in university level teaching is documented by Dorier et al.[20] in their research on the use of the "meta-lever" in the teaching and learning of linear algebra. Of course, it is neither necessary nor possible to take students through all the intricacies of the development of the modern integration theory, but some indications as to what motivated this development can be given, and the discussion about Fourier series can be carried out in class without particular difficulties.

Still, if we are to be realistic, this may be not enough, especially in the undergraduate version of the measure theory course. We have argued that Riemann integration already requires a certain level of mathematical maturity, and although approaches such as Royden's, that builds a measure starting with a concrete outer measure, or Chae's, that bypasses abstract measure theory, may overcome the logical difficulties, it would be even harder to convey to students lacking mathematical sophistication the insights of the Lebesgue integration theory. Taking this into account, a solution would be to delay the introduction of Lebesgue integration for later, in a graduate course, and in that context, we believe a better choice would be Rudin's textbook. We have described the approach to integration and measure in this textbook as highly abstract. However, even if the methods, the proofs, and the suggested exercises are more difficult, the notion of abstract integral that can be defined for measurable functions with respect to any given measure appears like a very intuitively acceptable generalization, obtained by removing any Euclidean "ties". On the contrary, there seems to be nothing intuitive in Carathéodory's condition (5.4.2), and certainly not in the proof of what may seem to an undergraduate student as an obvious fact: the measure of an interval $[a, b)$ is its length. On the other hand, Rudin's textbook could also succeed in conveying the tool-like power of the Lebesgue integral, since it contains many applications, treated in the modern setting of $L^p$ spaces.

Finally, one more issue we want to touch upon, is related to a domain of application of the Lebesgue measure and integration to which we have not given much attention in our research. Probability and statistics emerged, around the 1920's, as domains where the Lebesgue integral became a tool for "measuring" phenomena. Here highly irregular regions occur and the Lebesgue integral becomes indispensable. For instance, Lebesgue's notion of measure allowed Wiener in the 1920's to provide an elegant mathematical model of the Brownian motion. He looked at the random irregular movement of the molecules of a fluid as describing not only curves, but statistical assemblages of curves, i.e., spaces of curves, and proved that, except for a set of cases of probability 0 (i.e. of probability measure 0), all the Brownian motions were continuous non-differentiable curves. As a final thought, we would like to add that the teaching of the Lebesgue integral in such bordering areas, where analysis and probability, or more generally, mathematics and physics meet would be an interesting option to consider, especially for students majoring in probability and statistics.

## 5.5 Final remarks

In this research we described the development of the modern integration theory from Cauchy to Lebesgue, and concluded with some of the modern developments following Lebesgue at the beginning of the 20th century. We have followed a somewhat linear strand in that we ignored the connections with some neighboring domains such as complex analysis, partial differential equations, multi-variable analysis, or the emergent probability theory.

Nevertheless, by considering some questions related to Fourier series, we aimed at describing the driving forces behind some of the important insights that shaped the modern concepts. On the other hand, we could not give justice to all the complex aspects that surrounded this development, and thus chose to ignore other, perhaps equally important, triggers of this evolution. In particular, we did not look at Lebesgue's and his predecessors' work on the Fundamental Theorem of Calculus, in research that aimed, roughly speaking, at making differentiation and integration inverse operations for a larger class of functions. But it was not our intention to provide the most exhaustive account of the development of analysis in the last half of the 19th century. Our hope is that, besides capturing the key ideas that led to Lebesgue's integration theory, we managed to tell a "story" that is illustrative for the entire mathematical genre. We tried to depict the emergence of a unified theory with the contribution of many mathematicians, not all aiming to solve the same problems, on the contrary, having sometimes quite different questions, motivated by either physical, mathematical or philosophical concerns.

Finally, we employed this historical and epistemological analysis to raise some issues related to the teaching of integration at various levels at university. We have found that there might be some conceptual gaps between subsequent levels, that are not necessarily insuperable, but require careful didactical analysis. We looked at some typical textbooks and, inspired by the historical analysis, we gave some suggestions for teaching. A future avenue for research would be the investigation of the advanced hypotheses through empirical research.

# Bibliography


[1] T. M. Apostol, *Mathematical analysis*, 2nd ed., Addison-Wesley, 1974.

[2] M. Bakar and D. O. Tall, *Students mental prototypes for functions and graphs*, International Journal for Mathematics Education in Science and Technology ,23 (1), 1992, 39-50.

[3] J. Barone and A. Novikoff, *A history of the axiomatic formulation of probability theory from Borel to Kolmogorov: Part I*, Arch. Hist. Exact Sci., 18(2000), p. 123-190.

[4] N.H. Bingham, *Studies in the history of probability and statistics XLVI. Measure into probability: From Lebesgue to Kolmogorov*, Biometrika, 87 (2000), p. 145-146.

[5] E. Borel, *Leçons sur la théorie des fonctions*, Paris, 1898.

[6] D. Breidenbach, E. Dubinsky, J. Hawks, D. Nichols, *Development of the process conception of function*, Educational Studies in Mathematics 23/3, 1992, p. 247-285.

[7] B. Byers, *How mathematicians think: using ambiguity, contradiction, and paradox to create mathematics*, Princeton University Press, Princeton and Oxford, 2007.

[8] G. Cantor, *Ueber die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen*, Math. Ann., 5 (1872), p. 92-101.

[9] G. Cantor, *Ueber unendliche, lineare Punktmannichfaltigkeiten*, Pt 5. Math. Ann., 23 (1884), p. 453-488.

[10] G. Cantor, *De la puissance des ensembles parfaits de points: Extrait d'une lettre adressé à l'editeur*, Acta Math., 4(1884), p.381-392.

[11] A.L. Cauchy, *Cours d'analyse de l'Ecole royale polytechnique: 1. T., Analyse algebrique* , Wissenschaftliche Buchgesellschaft, Darmstadt, 1968.

[12] A.L. Cauchy, *Œuvres complètes d'Augustin Cauchy*, v.4, Gauthier-Villars, Paris, 1882-1938.

[13] S. B. Chae, *Lebesgue Integration*, Marcel Dekker, Inc., New York, 1970.

[14] Y. Chevallard, *L'analyse des pratiques enseignants en théorie anthropologique du didactique*, Recherches en Didactique des Mathématiques, 19(2), 1999, p. 221-266.

[15] R. Cooke, *The history of mathematics. A brief course* , 2nd ed., Wiley-Interscience, Hoboken, New Jersey, 2005.

[16] B. Cornu, *Limits*, in *Advanced Mathematical Thinking*, D. O. Tall, eds., Kluwer, Dordrecht, 1991, p.153- 166.

[17] F. D'Angello and M. Seyfried, *Introductory Real Analysis*, Houghton Mifflin Company, Boston, New York, 2000.

[18] G. Darboux, *Mémoire sur les fonctions discontinues*, Annales scientifiques de l'École Normale Supérieure Sér. 2, 4 (1875), p. 57-112.

[19] P. J. Davis and R. Hersh, *The mathematical Experience*, Birkhäuser, Boston, 1981.

[20] J.L. Dorier, *On the Teaching of Linear Algebra*, Kluwer Academic Publishers, Dordrecht, 2000.

[21] T. Dreyfus and S. Vinner, *Some aspects of the function concept in college students and junior high school teachers*, in *Proceeding of the Sixth International Conference for the Psychology of Mathematics Education*, 1982, Antwerp, Belgium.

[22] J. Ferrini-Mundy and K.-G. Graham, *An overview of the curriculum reform effort: issues for teaching, learning, and curriculum development*, Amm. Math. Monthly, 98(1991), 7, p.627-635.

[23] K.-G. Graham and J. Ferrini-Mundy, *An exploration of student understanding of central concepts in calculus*, Ann. Mtg. of the American Ed. Res. Assoc., 1989.

[24] I. Grattan-Guinness and J. R. Ravetz, *Joseph Fourier, 1768-1830; a survey of his life and work, based on a critical edition of his monograph on the propagation of heat, presented to the Institut de France in 1807*, MIT Press, Cambridge, 1972.

[25] J.D. Gray, *The shaping of the Riesz representation theorem: A chapter in the history of analysis*, Arch. Hist. Exact Sci., 31 (1983), p.127-187.

[26] P.R. Halmos, *Measure theory*, Springer-Verlag, New York, 1974.

[27] T. Hawkins, *Lebesgue's theory of integration. Its origins and development*, 2nd ed., Chelsea Pub. Co., New York, 1975.

[28] H. Heine, *Ueber trigonometrishce Reihen*, Creele, Jl. Math., 71(1870), p.353-365.

[29] C. Jordan, *Cours d'analyse de l'École Polytechnique*, 3rd ed., Gauthier-Villars, Paris, 1959.

[30] O.G. Jørsboe, *The Carleson-Hunt theorem on Fourier series*, Springer-Verlag, Berlin, 1982.

[31] P.E.B Jourdain, *The origin of Cauchy's conception of a definite integral and of the continuity of a function*, ISIS, I(1913), 661-703.

[32] H. Lebesgue, *Intégrale, longueur, aire.*, Ann. Mat. Pura ed Appl. 7, 3(1902), p. 231-259.

[33] H. Lebesgue, *Sur les séries trigonométriques*, Paris, Ann. École Normale, 3(1903), 20, p. 453-485.

[34] H. Lebesgue, *Leçons sur l'intégration et la recherche des fonctions primitives* Gauthier-Villars, Paris, 1904.

[35] H. Lebesgue, *Recherches sur la convergence des séries de Fourier*, Math. Ann., 61(1905), 251-280.

[36] H. Lebesgue, *Leçons sur l'intégration et la recherche des fonctions primitives*, 2nd ed., Gauthier-Villars, Paris, 1928.

[37] H. Lebesgue, *Measure and the integral*, Holden-Day, San Francisco, 1966.

[38] D. Van Dalen and A.F. Monna, *Sets and Integration. An outline of the development*, Wolters-Noordhoff Pub., Groningen, 1972.

[39] I. Newton, *The mathematical principles of natural philosophy*, Philosophical Library, New York, 1964.

[40] A. Orton, *Students' Understanding of Integration*, Educational Studies in Mathematics, 14 (1983), 1.

[41] G. Peano, *Selected works of Giuseppe Peano*, University of Toronto Press, Toronto, 1973.

[42] I. N. Pesin, *Classical and modern integration theories*, Academic Press, New York and London, 1970.

[43] J.-P. Pier, *Development of Mathematics: 1900-1950*, Birkhauser Verlag, Berlin, 1994.

[44] H. Poincaré, *Science and Method*, Dover, New York, 1908/1952.

[45] B. Riemann, *Gesammelte mathematische Werke und wissenschaftlicher Nachlass*, 2nd ed, Dover Publications, New York, 1953.

[46] F. Riesz and S. Béla, *Functional Analysis*, Ungar, New York, 1955.

[47] F. Riesz, *Riesz Frigyes összegyujtött munkái*, Akadémiai Kiadó, Budapest, 1960.

[48] H.L. Royden, *Real Analysis*, New York: Macmillan; London: Collier Macmillan, 1988.

[49] W. Rudin, *Real and complex analysis*, 3rd ed., McGraw-Hill, New York, 1987.

[50] M. Schneider, *A way to analyse several difficulties the pupils meet with in calculus*, in *Proceedings of Working Group 3 on Students Difficulties in Calculus*, ICME-7M, Artigue and G. Ervynck, eds., Québec, Canada, 1993, p. 31-34.

[51] A. Sierpinska, *Obstacles epistémologiques relatifs à la notion de limite*, Recherches en Didactique des Mathématiques, 6 (1), 1985, p. 5-67.

[52] A. Sierpinska, *Humanities students and epistemological obstacles related to limits*, Educational Studies in Mathematics 18, 1987, p. 37187.

[53] A. Sierpinska, *On understanding the notion of function*, in *The concept of function: Elements of Pedagogy and Epistemology. Notes and Reports Series of the Mathematical Association of America*, E. Dubinsky an G. Harel, eds., , Vol. 25, 1992, p. 25-58.

[54] A. Sierpinska, G. Bobos, C. Knipping, *Sources of students' frustration in pre-university level, prerequisite mathematics courses*, Instructional Science, 2007, in press.

[55] J. Stewart, *Calculus: early transcendentals*, 5th ed., Brooks/Cole Pub. Co., 2002.

[56] K.-G. Steffens, *The history of approximation theory*, Birkhäuser, Boston, 2006.

[57] D. O. Tall, *Functions and Calculus*, in *International Handbook in Mathematics Education*, A. J. Bishop et al, eds., Kluwer, Dordrecht, 1997, p. 289-325.

[58] S.M. Ulam, *What is measure?*, The American Mathematical Monthly, 50(1943), 10, p. 597-602.

[59] N. Wiener, *I am a mathematician, the later life of a prodigy; an autobiographical account of the mature years and career of Norbert Wiener and a continuation of the account of his childhood in Ex-prodigy*, The M.I.T. Press, Cambridge, 1956.

[60] A. J. Weir, *Lebesgue integration and measure*, Cambridge, 1973.