

Chapter 10

Epilogue: Historical Notes on Functional Analysis

Differentiae et summae sibi reciprocae sunt, hoc est summa differentiarum seriei est seriei terminus, et differentia summarum seriei est ipse seriei terminus, quorum illud ita enuntio: $\int dx$ aequ. x ; hoc ita: $d \int x$ aequ. x .

G. Leibniz

10.1 Integral Calculus

In a concise description of mathematical methods, Henri Lebesgue underlined the importance of definitions and axioms (see [47]).

When a mathematician foresees, more or less clearly, a proposition, instead of having recourse to experiment like the physicist, he seeks a logical proof. For him, logical verification replaces experimental verification. In short, he does not seek to discover new materials but tries to become aware of the richness that he already unconsciously possesses, which is built in the definitions and axioms. Herein lies the supreme importance of these definitions and axioms, which are indeed subjected logically only to the condition that they be compatible, but which could lead only to a purely formal science, void of meaning, if they had no relationship to reality.

Leibniz conceived integration as the reciprocal of differentiation:

$$\int dx = d \int x = x.$$

The computation of the integral of f is reduced to the search for its *primitive*, solution of the differential equation

$$F' = f.$$

The textbooks by Cauchy, in particular the *Analyse algébrique* (1821) (see [7]) and the *Résumé des leçons données à l'Ecole Royale Polytechnique sur le calcul infinitésimal* (1823), opened a new area in analysis. Cauchy was the first to consider the problem of existence of primitives:

In integral calculus, it seemed necessary to me to demonstrate in general the existence of *integrals* or *primitive functions* before giving their various properties. In order to reach this, it was necessary to establish the notion of *integral between two given limits* or *definite integral*.

Cauchy defines and proves the existence of the integral of continuous functions:

According to the preceding lecture, if one divides $X - x_0$ into infinitesimal elements $x_1 - x_0, x_2 - x_1 \cdots X - x_{n-1}$, the sum

$$S = (x_1 - x_0)f(x_0) + (x_1 - x_2)f(x_1) + \cdots + (X - x_{n-1})f(x_{n-1})$$

will converge to a limit given by the definite integral

$$\int_{x_0}^X f(x)dx.$$

So Cauchy proved the existence of primitives of continuous functions using integral calculus.

Though every continuous function has a primitive, Weierstrass proved in 1872 the existence of continuous nowhere differentiable functions. In a short note [44], Lebesgue proved the existence of primitives of continuous functions without using integral calculus. His proof is clearly functional-analytic.

In 1881 ([37]), Camille Jordan defined the functional space of *functions of bounded variation*, which he called functions of *limited oscillation*. His goal was to linearize Dirichlet's condition for the convergence of Fourier series:

Let x_1, \dots, x_n be a series of values of x between 0 and ε , and y_1, \dots, y_n the corresponding values of $f(x)$. The points $x_1, y_1; \dots; x_n, y_n$ will form a polygon.

Consider the differences

$$y_2 - y_1, y_3 - y_2, \dots, y_n - y_{n-1}.$$

We will call the sum of the positive terms of this sequence the *positive oscillation* of the polygon; *negative oscillation* is the sum of the negative terms; *total oscillation* is the sum of those two partial oscillations in absolute value.

Let us vary the polygon; two cases may occur:

- 1° The polygon may be chosen so that its oscillations exceed every limit.
- 2° For every chosen polygon, its positive and negative oscillations will be less than some fixed limits P_ε and N_ε . We will say in that case that $F(x)$ is a function of *limited oscillation* in the interval from 0 to ε ; P_ε will be its *positive oscillation*; N_ε its *negative oscillation*; $P_\varepsilon + N_\varepsilon$ its *total oscillation*.

This case will necessarily occur if $F(x)$ is the difference of two finite functions $f(x) - \varphi(x)$, because it is clear that the positive oscillation of the polygon will be $\underset{<}{=} f(\varepsilon) - f(0)$, and its negative oscillation $\underset{<}{=} \varphi(\varepsilon) - \varphi(0)$.

The converse is easy to prove. Indeed, it is easy to verify that

- 1° The oscillation of a function from 0 to ε is equal to the sum of its oscillations from 0 to x and from x to ε , x being any quantity between 0 and ε .

2° We have that $F(x) = F(0) + P_x - N_x$, P_x and N_x denoting the positive and the negative oscillations from 0 to x . But $F(0) + P_x$ and N_x are finite functions nondecreasing from 0 to ε .

Hence Dirichlet's proof is applicable, without modification, to every function of bounded oscillation from $x = 0$ to $x = \varepsilon$, ε being any finite quantity.

The functions of limited oscillations constitute a well-defined class, whose study could be of some interest.

Functions of bounded variation will play a fundamental role in the following domains:

- (a) Convergence of Fourier series;
- (b) Rectification of curves;
- (c) Integration;
- (d) Duality.

Let $u : [0, 1] \rightarrow \mathbb{R}$ be a continuous function. The length of the graph of u is defined by

$$L(u) = \sup \left\{ \sum_{j=0}^k [(a_{j+1} - a_j)^2 + (u(a_{j+1}) - u(a_j))^2]^{1/2} : \right. \\ \left. k \in \mathbb{N}, 0 = a_0 < a_1 < \dots < a_{k+1} = 1 \right\}.$$

In 1887, in Volume III of the first edition of his *Cours d'Analyse* at the École Polytechnique, Jordan proved that $L(u)$ is finite if and only if u is of bounded variation. The case of surfaces is much more delicate (see Sect. 10.3).

In 1894 ([80]), Stieltjes defined a deep generalization of the integral associated to an increasing function φ :

More generally, let us consider the sum

$$f(\xi_1)[\varphi(x_1) - \varphi(x_0)] + f(\xi_2)[\varphi(x_2) - \varphi(x_1)] + \dots + f(\xi_n)[\varphi(x_n) - \varphi(x_{n-1})]. \quad (A)$$

It will still have a limit, which we shall denote by

$$\int_a^b f(u)d\varphi(u).$$

We will have only to consider some very simple cases like $f(u) = u^k$, $f(u) = \frac{1}{z+u}$, and there is no interest in giving to the function $f(u)$ its full generality. Thus it will suffice, as an example, to suppose the function $f(u)$ continuous, and then the proof presents no difficulty, and we have no need to develop it, since it is done as in the ordinary case of a definite integral.

It is easy to extend Stieltjes's definition to every function φ of bounded variation. Stieltjes breaks the reciprocity between integral and derivative.

In 1903 ([32]), J. Hadamard characterized the *continuous linear functionals* on $C([a, b])$:

It is easy to reach this, following Weierstrass and Kirchhoff, and introducing a function $F(x)$, with a finite number of maxima and minima and such that

$$\int_{-\infty}^{+\infty} F(x)dx = 1;$$

e.g., $F(x) = \frac{1}{\sqrt{\pi}}e^{-x^2}$.

Starting then from the well-known identity

$$\lim_{\mu=\pm\infty} \mu \int_b^a f(x)F[\mu(x-x_0)]dx = f(x_0), \quad a < x_0 < b,$$

and assuming (as the authors quoted before) the operation U to be continuous (in the sense of Bourlet), it will suffice to define

$$U[\mu F\mu(x-x_0)] = \Phi(x_0, \mu)$$

to show that our operation could be represented as

$$U[f(x)] = \lim_{\mu=\pm\infty} \int_a^b f(x)\Phi(x, \mu)dx.$$

In 1909 ([61]), F. Riesz discovered a representation depending on only one function:

In the present note, we shall develop a new analytic expression of the linear operation, containing only one generating function.

Given the linear operation $A[f(x)]$, we can determine a function of bounded variation $\alpha(x)$ such that for every continuous function $f(x)$, we have

$$A[f(x)] = \int_0^1 f(x)d\alpha(x).$$

Riesz's theorem asserts that every continuous linear functional on $C([0, 1])$ is representable by a Stieltjes integral.

10.2 Measure and Integral

Les notions introduites sont exigées par la solution d'un problème, et, en vertu de la seule présence parmi les notions antérieures, elles posent à leur tour de nouveaux problèmes.

Jean Cavaillès

In 1898, Emile Borel defined the measure of sets in his *Leçons sur la théorie des fonctions*:

The procedure that we have employed actually amounts to this: we have recognized that a definition of measure could be useful only if it had certain fundamental properties: we have

stated these properties a priori, and we have used them to define the class of sets that we consider measurable.

Those essential properties that we summarize here, since we shall use them, are the following: The measure of a sum of a denumerable infinity of sets is equal to the sum of their measures; the measure of the difference of two sets is equal to the difference of their measures; *the measure is never negative; every set with a nonzero measure is not denumerable*. It is mainly this last property that we shall use. Besides, it is explicitly understood that we speak of measures only for those sets that we called *measurable*.

Of course, when we speak of the sum of several sets, we assume that every pair them have no common points, and when we speak of their difference, we assume that one set contains all the points of the other.

Following Lebesgue,

The descriptive definition of measure stated by M. Borel is without doubt the first clear example of the use of actual infinity in mathematics.

However, Borel does not prove the existence of the measure!

The Lebesgue integral first appeared on the 29 April 1901. In the note [42], Lebesgue proved the existence of the Borel measure as a restriction of the Lebesgue measure.

In the introduction of his thesis [43], Lebesgue stated his program:

In this work, I try to give definitions as general and precise as possible of some of the numbers considered in Analysis: definite integral, length of a curve, area of a surface.

He formulated the *problem of the measure of sets*:

We intend to assign to every bounded set a positive or zero number called its measure and satisfying the following conditions:

1. There exist sets with nonzero measure.
2. Two equal sets have equal measures.
3. The measure of the sum of a finite number or of a countable infinity of sets, without common points, is the sum of the measures of those sets.

We will solve this *problem of measure* only for the sets that we will call measurable.

In his *Leçons sur l'intégration et la recherche des fonctions primitives* of 1904, see [45], Lebesgue formulated the *problem of integration*.

We intend to assign to every bounded function $f(x)$ defined on a finite interval (a, b) , positive, negative, or zero, a finite number $\int_a^b f(x)dx$, which we call the integral of $f(x)$ in (a, b) and which satisfies the following conditions:

1. For every a, b, h , we have

$$\int_a^b f(x)dx = \int_{a+h}^{b+h} f(x-h)dx.$$

2. For every a, b, c , we have

$$\int_a^b f(x)dx + \int_b^c f(x)dx + \int_c^a f(x)dx = 0.$$

3.

$$\int_a^b [f(x) + \varphi(x)]dx = \int_a^b f(x)dx + \int_a^b \varphi(x)dx.$$

4. If we have $f \geq 0$ and $b > a$, we also have

$$\int_a^b f(x)dx \geq 0.$$

5. We have

$$\int_0^1 1 \times dx = 1.$$

6. If $f_n(x)$ increases and converges to $f(x)$, then the integral of $f_n(x)$ converges to the integral of $f(x)$.

Formulating the six conditions of the integration problem, we define the integral. This definition belongs to the class of those that could be called *descriptive*; in those definitions, we state the characteristic properties of the object we want to define. In the *constructive* definitions, we state which operations are to be done in order to obtain the object we want to define. Constructive definitions are more often used in Analysis; however, we use sometimes descriptive definitions; the definition of the integral, following Riemann, is constructive; the definition of primitive functions is descriptive.

In 1906, in his thesis [23], Maurice Fréchet tried to extend the fundamental notions of analysis to abstract sets.

In this Mémoire we will use an absolutely general point of view that encompass these different cases.

To this end, we shall say that a *functional operation* U is defined on a set E of elements of every kind (numbers, curves, points, etc.) when to every element A of E there corresponds a determined numerical value of $U : U(A)$. The search for properties of those operations constitutes the object of the *Functional Calculus*.

Fréchet defined *distance* which he called, in French, *écart*:

We can associate to every pair of elements A, B a number $(A, B) \geq 0$, which we will call *the distance of the two elements* and which satisfies the following properties: (a) The distance (A, B) is zero only if A and B are identical. (b) If A, B, C are three arbitrary elements, we always have $(A, B) \leq (A, C) + (C, B)$.

In [24], Fréchet defined *additive families of sets* and *additive functions of sets*:

An *additive family of sets* is a collection of sets such that:

1. If E_1, E_2 are two sets of this family, the set $E_1 - E_2$ of elements of E_1 , if they exist and that are not in E_2 , belongs also to the family.
2. If E_1, E_2, \dots is a denumerable sequence of sets of this family, their sum, i.e., the set $E_1 + E_2 + \dots$ of elements belonging at least to one set of the sequence, belongs also to the family.

A set function $f(E)$ defined on an additive family of sets \mathcal{F} is *additive* on \mathcal{F} if E_1, E_2, \dots being a denumerable sequence of sets of \mathcal{F} and *disjoint*, i.e., without pairwise common elements, we have

$$f(E_1 + E_2 + \dots) = f(E_1) + f(E_2) + \dots$$

When the sequence is infinite, the second member has obviously to converge regardless of the order of the terms. Hence the series in the second member has to converge absolutely.

Fréchet defined the integral without using topology. *Additive functions of sets* will be called *measures*.

In [12], Daniell chose a different method. He introduced a space \mathcal{L} of *elementary functions* and an *elementary integral*

$$\mathcal{L} \rightarrow \mathbb{R} : u \mapsto \int u \, d\mu$$

satisfying the axioms of linearity, positivity, and monotone convergence.

The two axiomatics are equivalent if to Daniell's axioms we add *Stone's axiom* (1948):

$$\text{for every } u \in \mathcal{L}, \min(u, 1) \in \mathcal{L},$$

or the axiom

$$\text{for every } u, v \in \mathcal{L}, uv \in \mathcal{L}.$$

The choice of primitive notions and axioms is rather arbitrary. There are no absolutely undefinable notions or unprovable propositions.

The axiomatization of integration by Fréchet opened the way to the axiomatization of probability by Kolmogorov in 1933. The unification of measure, integral, and probability was one the greatest scientific achievements of the twentieth century.

In his thesis [5], Banach defined the *complete normed spaces*:

There exists an operation, called norm (we shall denote it by the symbol $\|X\|$), defined in the field E , having as an image the set of real numbers and satisfying the following conditions:

$$\begin{aligned} \|X\| &\geq 0, \\ \|X\| &= 0 \text{ if and only if } X = \theta, \\ \|a \cdot X\| &= |a| \cdot \|X\|, \\ \|X + Y\| &\leq \|X\| + \|Y\|. \end{aligned}$$

If 1. $\{X_n\}$ is a sequence of elements of E , 2. $\lim_{p \rightarrow \infty} \|X_r - X_p\| = 0$, there exists an element X such that

$$\lim_{n \rightarrow \infty} \|X - X_n\| = 0.$$

Banach emphasized the efficiency of the axiomatic method:

The present work intends to prove theorems valid for different functional fields, which I will specify in the sequel. However, in order not to be forced to prove them individually for

every particular field, a tedious task, I chose a different way: I consider in some general way sets of elements with some axiomatic properties, I deduce theorems, and I prove afterward that the axioms are valid for every specific functional field.

The fundamental book of Banach ([6]), *Théorie des opérations linéaires*, was published in 1932. Banach deduces Riesz's representation theorem from the Hahn–Banach theorem.

The original proof of the Hahn–Banach theorem holds in every real vector space. Let $F : X \rightarrow \mathbb{R}$ be a positively homogeneous convex function and let $f : Z \rightarrow \mathbb{R}$ be a linear function such that $f \leq F$ on the subspace Z of X . By the well-ordering theorem, the set $X \setminus Z$ can be so ordered that each nonempty subset has a least element. It follows then, from Lemma 4.1.3, by transfinite induction, that there exists $g : X \rightarrow \mathbb{R}$ such that $g \leq F$ on X and $g|_Z = f$.

Let us recall the *principle of transfinite induction* (see [72]). Let \mathcal{B} be a subset of a well-ordered set \mathcal{A} such that

$$\{y \in \mathcal{A} : y < x\} \subset \mathcal{B} \Rightarrow x \in \mathcal{B}.$$

Then $\mathcal{B} = \mathcal{A}$.

In set theory, the well-ordering theorem is equivalent to the axiom of choice and to Zorn's lemma. In 1905, Vitali proved the existence of a subset of the real line that is not Lebesgue measurable. His proof depends on the axiom of choice.

10.3 Differential Calculus

L'activité des mathématiciens est une activité expérimentale.

Jean Cavaillès

Whereas the integral calculus transforms itself into an *axiomatic theory*, the differential calculus fits into the *general theory* of distributions.

The fundamental notions are

- Weak solutions;
- Weak derivatives;
- Functions of bounded variation;
- Distributions.

In [60], Poincaré defined the notion of *weak solution* of a boundary value problem:

Let u be a function satisfying the following conditions:

$$\frac{du}{dn} + hu = \varphi, \tag{3}$$

$$\Delta u + f = 0. \quad (4)$$

Now let v be an arbitrary function, which I assume only continuous, together with a first-order derivative. We shall have

$$\int \left(v \frac{du}{dn} - u \frac{dv}{dn} \right) d\omega = \int (v\Delta u - u\Delta v) d\tau,$$

so that

$$\int v f d\tau + \int u\Delta v d\tau + \int v\varphi d\omega = \int u \left(h v + \frac{dv}{dn} \right) d\omega. \quad (5)$$

Condition (5) is thus a consequence of condition (3).

Conversely, if condition (5) is satisfied for every function v , condition (3) will be also satisfied, *provided that u and $\frac{du}{dn}$ are finite, well-defined, and continuous functions.*

But it can happen that in some cases, we are unaware that $\frac{du}{dn}$ is a well-defined and continuous function; we cannot assert then that condition (5) entails condition (3), and it is even possible that condition (3) is meaningless.

Poincaré named condition (5) a *modified condition* and asserted (p. 121),

It is obviously equivalent to condition (3) from the physical point of view.

This Mémoire of Poincaré contains (p. 70) the first example of an integral inequality between a function and its derivatives:

Let V be an arbitrary function of x, y, z ; define:

$$A = \int V^2 d\tau, \quad B = \int \left[\left(\frac{dV}{dx} \right)^2 + \left(\frac{dV}{dy} \right)^2 + \left(\frac{dV}{dz} \right)^2 \right] d\tau.$$

I will write to shorten:

$$B = \int \sum \left(\frac{dV}{dx} \right)^2 d\tau.$$

I assume first that V satisfies the condition:

$$\int V d\tau = 0$$

and I intend to estimate the lower limit of the quotient $\frac{B}{A}$.

The maximum principle is stated on p. 92. Poincaré's principle appears in [59] for the formal construction of the eigenvalues and eigenfunctions of the Laplacian. In [60], Poincaré proved the existence of eigenvalues (for Dirichlet boundary conditions) using the theory of meromorphic functions (see [50]).

Let us recall that we denote by $L(u)$ the length of the graph of the continuous function $u : [0, 1] \rightarrow \mathbb{R}$. Following Jordan, $L(u) < \infty$ if and only if u is of bounded variation. It follows then from a theorem due to Lebesgue that u is almost everywhere differentiable on $[0, 1]$. In [82], Tonelli proved a theorem equivalent to

$$L(u) = \int_0^1 \sqrt{1 + (u'(x))^2} dx \iff u \in W^{1,1}([0, 1]).$$

A counterexample due to Schwarz, published in 1882 in the *Cours d'Analyse* of Hermite, shows that it is not possible to extend the definition of length due to Jordan to surfaces. Let $z = u(x, y)$ be a nonparametric surface, with u continuous on $[0, 1] \times [0, 1]$. Let $\Omega =]0, 1[\times]0, 1[$ and define, on $X = C(\overline{\Omega})$, the distance

$$d(u, v) = \max\{|u(x, y) - v(x, y)| : (x, y) \in \overline{\Omega}\}.$$

The space of *quasilinear functions* on $\overline{\Omega}$ is defined by

$$Y = \{u \in X : \text{there exists a triangulation } \tau \text{ of } \Omega \\ \text{such that, for every } T \in \tau, u|_T \text{ is affine}\}.$$

The graph of $u \in Y$ consists of triangles. The sum of the areas of those triangles is called the *elementary area* of the graph of u and is denoted by $B(u)$.

The *Lebesgue area* of the graph of u is defined by

$$A(u) = \inf \left\{ \lim_{n \rightarrow \infty} B(u_n) : (u_n) \subset Y \text{ and } d(u_n, u) \rightarrow 0, \quad n \rightarrow \infty \right\}. \quad (*)$$

In [83] (see also [53]), Tonelli stated two theorems equivalent to

$$A(u) < \infty \iff \|Du\|_{\Omega} < \infty,$$

$$A(u) = \int_{\Omega} \sqrt{1 + \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2} dx dy \iff u \in W^{1,1}(\Omega).$$

Lebesgue area is a lower semicontinuous function on X . It extends the elementary area: for every $u \in Y$, $A(u) = B(u)$.

In [25], Fréchet observed that Lebesgue's definition allows one to extend lower semicontinuous functions. Let Y be a dense subset of a metric space X and let $B : Y \rightarrow [0, +\infty]$ be an l.s.c. function. The function A defined by (*) is an l.s.c. extension of B on X such that for every l.s.c. extension C of B on X and for every $u \in X$, $C(u) \leq A(u)$.

In [48], Leray defined the weak derivatives of L^2 functions, and called them *quasi-dérivées*.

In [75], announced in [74] and translated in [78], Sobolev defined the distributions of finite order on \mathbb{R}^N , which he called *fonctionnelles*. (A distribution f on \mathbb{R}^N is of *order* k if for every sequence $(u_n) \subset \mathcal{D}(\mathbb{R}^N)$ such that the supports of u_n are contained in some compact set and such that $\sup_{|\alpha| \leq k} \|\partial^\alpha u_n\|_{\infty} \rightarrow 0, n \rightarrow \infty$, we have

$\langle f, u_n \rangle \rightarrow 0, n \rightarrow \infty$.) Sobolev defined the derivative of a *fonctionnelle* by duality and associated a *fonctionnelle* to every locally integrable function on \mathbb{R}^N .

Without reference to his theory of *fonctionnelles*, Sobolev defined in [77] the weak derivatives of integrable functions. Regularization by convolution is due to Leray for L^2 functions (see [48]) and to Sobolev for L^p functions (see [77]).

In [69], Laurent Schwartz defined general distributions. In [70], he defined the tempered distributions and their Fourier transform. The treatise [71] is a masterful exposition of distribution theory.

Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a function of bounded variation on every bounded interval. The formula of integration by parts shows that for every $u \in \mathcal{D}(\mathbb{R})$,

$$\int_{\mathbb{R}} u \, d g = - \int_{\mathbb{R}} u' g \, dx.$$

The Stieltjes integral with respect to g is nothing but the derivative of g in the sense of distributions! Riesz's representation theorem asserts that every continuous linear functional on $C([0, 1])$ is the derivative in the sense of distributions of a function of bounded variation.

10.4 Comments

Some general historical references are [15, 19, 29]. We recommend also [46] on Jordan, [52] on Hadamard, [81] on Fréchet, and [38] on Banach.