History of Functional Analysis

Jean Dieudonné

1981

Contents

Introduction

One may give different definitions of "Functional Analysis". Its name name might suggest that it contains all parts of mathematics which deal with functions, but that would practically mean *all* mathematical Analysis. We shall adopt a narrower definition: for us, it will be the study of topological vector spaces and of mappings $u : \Omega \to F$ from a part of Ω of a topological vector space E into a topological vector space F, these mappings being assumed to satisfy various algebraic and topological conditions. A moment of reflection shows that this already covers a large part of modern Analysis, in particular the theory of partial differential equations.

Functional analysis thus appears as a rather complex blend of Algebra and Topology, and it should therefore surprise no one that the development of these two branches of mathematics had a strong influence on its own evolutions. As a matter of fact, it is almost impossible to dissociate the early history of General Topology (and even of the set-theoretic language) from the beginnings of Functional Analysis, since the sets and spaces which (after the subsets of $\mathbb{R}^n$) attracted most attention consisted of *functions*.

With regard to Algebra, as the most frequently studied mappings between topological vector spaces are *linear*, it is quite natural that linear Algebra should have greatly influenced Functional Analysis. In fact, at the end on the XIX$^{\text{th}}$ century, the old idea that infinitesimal Calculus was derived from the algebraic "Calculus of differences" by a "limit process" began to acquire a more precise and more influential form when Volterra applied a similar idea to an integral equation

$$\int_a^y \varphi(x)H(x,y)dx = f(y) \tag{1}$$

for an unknown function φ, the functions f and H being continuous in $[a,b]$ and $[a,b]\times[a,b]$ respectively, with $f(a)=0$. He divides $[a,b]$ into n subintervals by the points $y_k = a + k\frac{b-a}{n}$ $(1 \leq k \leq n)$, replaces y in (1) by these n values, and the integral by the corresponding Riemann sums, which gives him a system of n linear equations

$$\begin{cases} h_{11}z_1 & = b_1 \\ h_{21}z_1 + h_{22}z_2 & = b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots & \\ h_{n1}z_1 + h_{n2}z_2 + \dots + h_{nn}z_n & = b_n \end{cases} \tag{2}$$

with $j_{jk} = H(y_j, y_k), z_k = \varphi(y_k)$ and $b_k = f(y_k)$; the integral equation (1) was thus considered as obtained from systems (2) by a limit process when the number of unknowns became infinite.

Unfortunately, linear Algebra, as it was understood in the XIX$^{\text{th}}$ century (and even much later) did not readily lend itself to affording a good guidance to such generalizations. Its own evolution had been very slow and painful, stretching over 130 years, and in a succession of stages which, to our eyes, is exactly the *reverse* of the *logical* sequence of notions, namely

TODO: put diagram at p. 3.

In spite of the unsuccessful efforts of Grassmann and Peano, the intrinsic aspects and the geometric point of view in linear Algebra remained in the background until 1900; one would readily speak with Cayley (1843) of vectors and linear subspaces, but they were invariably considered as parts of some $\mathbb{R}^n$; in other words, everything in a vector space was always referred to a *fixed basis*, and linear maps were only handled through their *matrices* corresponding to these bases. The various "reduction" theorems were known in 1880, but only through complicated computations of determinants, and without any geometric interpretation. Furthermore, Frobenius, who had been the most influential mathematician in building up a synthesis of the linear Algebra of his time, had unfortunately taken a step backward (even with respect to Cayley) by electing to work systematically with bilinear forms $\sum_{p,q} a_{pq}x_p y_q$ instead of working with matrices (a_{pq}). Finally, before 1930 nobody had a correct conception of duality between finite dimensional *vector spaces*; even in van der Waerden's book (1931), such a vector space and its dual are still *identified*.

All this was to weigh heavily on the evolution of linear Functional Analysis; in particular it followed (over a shorted span of years) the same unfortunate succession of stages through which linear Algebra had to go; and it is only after it was realized that the current conception of vectors as "n-tuples" could not possibly be extended to infinite dimensional function spaces, that this conception was finally abandoned and that genuinely geometrical notions won the day.

The diagram at the end of this Introduction tries to depict graphically in some detail the successive stages of the history of Functional Analysis, by mentioning the actions and reactions of the various parts of mathematics which took part in it. If one were to reduce this complicated history to a few key words, I think the emphasis should fall on the evolution of two concepts: *spectral theory* and *duality*. Both of course stem from the very concrete problems encountered in the solution of linear equations (or systems of linear equations), where the unknowns are *functions*. The basic concepts of spectral theory: eigenvalues, eigenfunctions and expansions in series of such functions were already known at the beginning of the XIX$^{\text{th}}$ century, in the theory of Fourier series; they would form the model on which all further advances were patterned. But it took 60 years of strenuous efforts to extend the theory from the Sturm–Liouville problem in ordinary differential equations to the partial differential equation of the vibrating membrane. It was gradually realized that the heart of the matter lay, not in the differential (or partial differential) equations themselves, but in *integral equations* associated to them; at first they were not explicitly written down, so that one can only speak of "crypto-integral" equations, to designate the use of methods resting on evaluations of integrals, and which only later emerged as standard methods in the theory of integral equations.

The remarkable feature of this history is that, after such a slow incubation period, so to speak, spectral theory, in the span of a few years, reached complete maturity, giving birth in the process to the concept of linear duality, which began at last to be understood by analysts, before becoming later familiar to all mathematicians by a kind of backlash effect. What is interesting in this rapid advance is that it was accomplished in a series of what one may call discrete *jumps*, in each of which the decisive step was to ignore the special features of the problem under consideration, and to make it accessible by inserting it into a more general context.

The first of these "discontinuities" occurred in 1896–1900, when Le Roux, Volterra and Fredholm, instead of working on the *special* integral equations studied by their predecessors (Abel, Liouville, Beer–Neumann), elected to use *minimal* assumptions on the kernels, and in so doing discovered that the theory was far simpler that it was generally thought.

The second step was taken by Hilbert in his 1906 papers, subordinating the too special theory of symmetric integral equations to the much more general concept of infinite "bounded" quadratic forms, which turned out to provide the frame needed for all subsequenct progress in ordinary and partial differential equations.

The contemporary discovery of the Lebesgue integral, and the geometric and topological concepts introduced by Fréchet in Analysis immediately led Hilbert's successors to translate his results into the language of what we now call Hilbert space, linking the euclidean geometry to integration theory, and making possible the discussion of the most general systems of linear equations in such a space.

This in turn led F. Riesz in 1910–1913 to introduce L^p and l^p spaces for any exponent p such that $1 < p < +\infty$, and to discover the natural duality between the *different* spaces L^p and L^q with $\frac{1}{p} + \frac{1}{q} = 1$, in sharp distinction from the muddleheaded ideas on the matter, which the accidental self-duality of Hilbert space had failed to dispel.

But although F. Riesz, in the treatment of systems of linear equations in l^p spaces, was the first to obtain a condition which later was seen to consist in a particular application of the Hahn–Banach theorem, he failed to visualize that condition as amounting to an extension property of a continuous linear form defined on a subspace. This fourth "jump" was only accomplished by Helly in 1921, again by generalizing the theory of systems of linear equations from the special l^p spaces to *any* normed subspace of $\mathbb{C}^{\mathbb{N}}$. After that, only two more steps were needed to reach the present status of the theory, with the passage to general normed spaces (together with the use of transfinite induction) by Hahn and Banach and a little later the extension of duality theory to locally convex spaces during the period 1935–1945.

This process of successive generalizations may thus have reached a point of diminishing returns around the middle of the century. Inasmuch as we are able to judge from events probably too recent to allow a proper perspective, the theory of topological vector spaces, after 1950, has stabilized as one of the standard tools of modern mathematics, together with linear and multilinear Algebra, General Topology and measure theory. The advances which have been achieved during the last 30 years mainly consist in new imaginative ways to use the fundamental tools of Functional Analysis, either in theories where they had noot been applied before, such as differential geometry and differential topology (K-theory, theory of the Atiyah–Singer index, foliations), or in the construction of more powerful methods to handle functional equations (distributions, Sobolev spaces, pseudo-differential operators and their generalizations)·

This volume grew out of a series of lectures which I gave in Rio de Janeiro in 1979, at the invitation of Prof. Jorge Alberto Barroso of the Universidade Federal de Rio de Janeiro, to whom go my most heartfelt thanks. I am also very grateful to him for the pains he took in supervising the preparation of the manuscript for publication.

Chapter I
Linear differential equations and the Sturm–Liouville problem

§1. Differential equations and partial differential equations in the XVIII$^{\text{th}}$ century.

Until around 1750, the notion of *function* of one variable was a very hazy one. The domain where it was defined was seldom described with precision; it was tacitly assumed that around each point x_0, the function was equal to a power series in $x - x_0$ and its derivatives were obtained by taking the derivatives of each term of the series. To solve a differential equation of order n

$$y^{(n)} = F(x, y, y', y'', \dots, y^{(n-1)}) \tag{1}$$

one would therefore substitute in (1) for y and its derivatives a power series $\sum_{k=0}^{\infty} c_k(x-x_0)^k$ and its derivatives, and identify the series on both sides, which would determine each c_k for $k \leq n$ as a function of $c_0, c_1, \dots, c_{k-1}$; the solution thus depended on n arbitrary parameters $c_0, c_1, \dots, c_{n-1}$. The very few cases in which it was possible to write explicitly the solution by means of primitives of known functions (such as the linear equation $y' = a(x)y + b(x)$ of order 1) were already known at the end of the XVII$^{\text{th}}$ century.

After 1760 began the first general study of linear equations of arbitrary order

$$L(y) \equiv y^{(n)} + a_1(x)y^{(n-1)} + \dots + a_n(x)y = b(x). \tag{2}$$

D'Alembert observed that the knowledge of a particular solution of the equation and of all solutions of the homogeneous equation $L(y) = 0$ yields by addition all solutions of (2). A little later, Lagrange [**135**] showed that the general solution of $L(y) = 0$ may be written $\sum_{k=1}^{n} c_k y_k$ where the c_k are arbitrary constants, and the $y_k (1 \leq k \leq n)$ particular solutions (which he tacitly assumed to be linearly independent). Then, by his famous method of "variation of constants" [**135**], he showed how to obtain also the solutions of (2) when the y_k were known: the solution is written in the form $y = \sum_{k=1}^{n} z_k y_k$, where the z_k are unknown functions, subject to $n-1$ linear relations

$$\sum_{k=1}^{n} z_k' y_k^{(\nu)} = 0 \quad (0 \leq \nu \leq n-2) \tag{3}$$

These conditions imply that $y^{(\nu)} = \sum_{k=1}^{n} z_k y_k^{(\nu)}$ for $0 \leq \nu \leq n-1$; replacing y by $\sum_{k=1}^{n} z_k y_k$ in (2) and using the fact that the y_k satisfy $L(y_k) = 0$, one obtains for the z_k' another linear equation

$$\sum_{k=1}^{n} z_k' y_k^{(n-1)} = b(x) \tag{4}$$

from which, by the Cramer formulas, one can compute the $z_k' (1 \leq k \leq n)$ and the problem is thus reduced to computing their primitives,

Lagrange also introduced [**135**] the notion of *adjoint* of a linear differential operator L, which was to acquire great importance later: he showed that there exists a linear differential operator M satisfying an identity

$$zL(y) - yM(z) = \frac{d}{dx}\left(B(y,z)\right) \tag{5}$$

where B is bilinear in $(y, y', \dots, y^{(n-1)})$ and $(z, z', \dots, z^{(n-1)})$, constituting a generalization of the classical "integration by parts"; he deduced from that formula that if a solution z of $M(z) = 0$ was known, solutions of $L(y) = 0$ could be obtained by solving an equation $B(y, z) = \text{Const}$, of order $n - 1$.

Partial differential equations were not considered until the middle of the XVIII$^{\text{th}}$ century, in connection with problems of Mechanics or Physics and then they were of order 2 at least (see §2). The study of partial differential equations of first order was only begun by Fuler and Lagrange after 1770. Euler was able to solve a few particular equations, and then Lagrange found general methods which enabled his followers, Charpit and Monge, to reduce the solution of a general equation of first order

$$F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) = 0 \tag{6}$$

to the solution of a system of ordinary differential equations, an idea which was developed later by Cauchy in his concept of "characteristic curves".

§2. Fourier expansions.

In 1747, d'Alembert gave the first mathematical treatment of the general problem of the small vibrations of a string of length a, fixed at each extremity; the string moves in a plane where the axis $0x$ is along the position of the string at rest, the segment $0 \leq x \leq a$; if $y = u(x,t)$ is the equation of the string at time t, d'Alembert shows that, if $u(x,t)$ remains small, it satisfies the equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \tag{7}$$

where c is a known function of x alone, and is constant if the density of the string is constant. When c is constant, taking $X = x - ct$ and $Y = x + ct$ as new variables reduces the equation to $\frac{\partial^2 u}{\partial X \partial Y} = 0$, and d'Alembert concluded that the solution of (7) is given by

$$u(x,t) = f(x - ct) + g(x + ct) \tag{8}$$

where f and g are "arbitrary" functions. A year later, Euler interpreted this result as meaning that (for $c = 1$) $u(x,t)$ was known once the two functions of x,

$$u(x,0) = \varphi(x), \quad \frac{\partial u}{\partial t}(x,0) = \psi(x) \tag{9}$$

were prescribed, the value of $u(x,t)$ being explicitly given by

$$u(x,t) = \frac{1}{2}\left(\varphi(x-t) + \varphi(x+t)\right) + \frac{1}{2}\int_{-t}^{t} \psi(x-\xi)d\xi \tag{10}$$

(Euler only gives a geometric construction equivalent to this formula). Now it was well known experimentally that $\varphi(x)$ could be quite different from an analytic function, for ins tance it could have no derivative at some points, and this led Euler to introduce,

in addition to what he called "continuous" functions (i.e. analytic functions in our sense) more general ones which he baptized "mechanical" without giving their precise definition (from the context they seem to be piecewise twice differentiable functions in our terminology).

On the other hand, already in 1715, B. Taylor, by a direct argument which did not use equation (7), had concluded that (when c is constant) for any integer $n \geq 1$, the function

$$u_n(x,t) = \sin\frac{n\pi x}{a}\cos\frac{n\pi ct}{a} \tag{11}$$

represented vibrations of the string, namely for $n = 1$ the "fundamental" tone, and for $n = 2, 3, ...$, its "harmonics". As it was well known that the sound emitted by a vibrating string was in general a mixture of several "harmonics", Daniel Bernoulli, in 1750, proposed that the general solution (10) could also be written as a series

$$u(x,t) = \sum_{n=1}^{\infty} a_n \sin\frac{n\pi x}{a}\cos\frac{n\pi c}{a}(t - b_n) \tag{12}$$

for suitable values of the a_n and b_n. However, in 1753, Euler observed that this would imply that an arbitrary "mechanical" function defined in an interval $-a \leq x \leq a$ could be written as a series

$$\frac{a_0}{2} + a_1\cos\frac{\pi x}{a} b_1 \sin\frac{\pi x}{a} + a_2\cos\frac{2\pi x}{a} + b_2\sin\frac{2\pi x}{a} + ... \tag{13}$$

and he believed that such a series of analytic functions could only represent an analytic function. His opinion was shared (with some variations) by almost all other mathematicians of his time, and no progress was made on this question until the beginning of Fourier's work on the theory of heat (see [**065**]). Having to solve equations such as

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{14}$$

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y} = 0 \tag{15}$$

for various boundary conditions, he systematically looks for solutions of the form $u(x,y) = v(x)w(y)$ and, following D. Bernoulli, wants to obtain the most general solution as series whose terms are these particular ones. In so doing, he is brought back to the problem of expressing a function f as a series (13), but this time he adds to D. Bernoulli's argument the formulas giving actually the values of the coefficients a_n, b_n

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx dx, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx dx \tag{16}$$

(when $a = \pi$) which as a matter of fact had already been obtained by Clairaut and Euler, without realizing their interest. Using these formulas Fourier was able to show on many examples of non analytic functions that the corresponding Fourier series converged to $\frac{1}{2}(f(x_+) + f(x_-))$, and expressed his conviction that this was true for "arbitrary" functions, although his attempts and those of Cauchy to prove that result were unsuccessful and the first proof for a piecewise monotonic and piecewise continuous

function was only given by Dirichlet in 1829. One should also mention in that connection that in 1799, Parseval had given the formula

$$\frac{a_0^2}{2} + \sum_{n=1}^{\infty} (a_n^2 + b_n^2) = \frac{1}{\pi} \int_{-\pi}^{\pi} (f(t))^2 \, dt \tag{17}$$

by a purely formal computation, without any proof of convergence.

These results gave the impetus to the vast theory of *trigonometric series*, which was to be one of the main concerns of most analysts in the XIX$^{\text{th}}$ century, centered around the criteria of convergence of such series and the relations between its sum and its coefficients. The evolution of that theory was closely linked to a gradual precision and deepening of the notions of set of real numbers, of function and of integral. But before 1920 there was not much contact between that theory and the development of Functional Analysis as we understand it.

On the contrary, other results of Fourier in his *Theory of heat* triggered the birth of *spectral theory*. For instance [**067**] he shows that the "cooling off" problem for a solid sphere of radius r, when one assumes spherical symmetry for the problem, is governed by the partial differential equation

$$\frac{\partial u}{\partial t} = k \left(\frac{\partial^2 u}{\partial x^2} + \frac{2}{x} \frac{\partial u}{\partial x} \right) \tag{18}$$

with the "boundary conditions" that $u(x,t)$ must remain finite when x tends to 0, and satisfy the relation

$$\frac{\partial h}{\partial u} + hu = 0 \quad \text{for } x = r \text{ and all } t, \tag{19}$$

where h and k are constants. Using his favorite method of "separation of wvariables", Fourier obtains solutions

$$u(x,t) = \frac{1}{x} \exp(-k\lambda^2 t) \sin \lambda x \tag{20}$$

provided the parameter λ is a solution of the transcendental equation

$$\frac{\lambda r}{\operatorname{tg} \lambda r} = 1 - hr. \tag{21}$$

He easily proves that the equation has an infinity of real roots λ_n tending to $+\infty$. To obtain a solution of (18) with boundary condition (19) and such that $u(x,0)$ is a given function $f(x)$, he proceeds as before, writing $xf(x)$ as a series $\sum_{n=1}^{\infty} c_n \sin \lambda_n x$; he shows that one has again the "orthogonality" relations (of course he does not use that word)

$$\int_0^r \sin \lambda_n x \sin \lambda_m x dx = 0 \quad \text{for } m \neq n \tag{22}$$

and from them deduces the relations

$$c_n = \left(\int_0^r x f(x) \sin \lambda_n x dx \right) \Big/ \left(\int_0^r \sin^2 \lambda_n x dx \right) \tag{23}$$

without of course any rigorous justification, nor any proof of the fact that the series converges to $xf(x)$.

§3. The Sturm–Liouville theory.

The results of Fourier on the theory of heat were continued and expanded by Poisson. Their work led Ch. Sturm in 1836 and J. Liouville one year later to build a general theory which would include all cases considered by Fourier and Poisson, without assuming the possibility of explicit integration. They consider a second order differential equation

$$y'' = q(x)y + \lambda y = 0 \tag{24}$$

where q is a real valued continuous function in a compact interval $[a, b]$ of $\mathbb{R}$, and λ a complex parameter. The first problem is to consider boundary conditions of the form

$$y(a)\cos\alpha - y'(a)\sin\alpha = 0, \quad y(b)\cos\beta - y'(b)\sin\beta = 0 \tag{25}$$

where *alpha* and β are two positive constants, and to determine for what values of λ the problem has a non trivial solution (an "eigenfunction" for the "eigenvalue" λ in our present day language).

A first remark, which had already essentially been made by Poisson, is that if λ, μ are two different eigenvalues, and u, v two corresponding "eigenfunctions", then from the relations

$$u'' - qu + \lambda u = 0, \quad v'' - qv + \mu v = 0,$$

one deduces

$$u''v - v''u + (\lambda - \mu)uv = 0$$

and as $\int_a^b (u''v - v''u)\,dx = (u'v - v'u)|_a^b = 0$ because of (25), one obtains

$$(\lambda - \mu)\int_a^b u(x)v(x)dx = 0. \tag{26}$$

A first consequence of this relation is that eigenvalues are necessarily *real* numbers, Indeed, if λ was not real, then $\overline{\lambda}$ would also be an eigenvalue with eigenfunction $\overline{u}$, and substituting $\overline{\lambda}$ and $\overline{u}$ for μ and v in (26), one obtains $\int_a^b |u(x)|^2 dx = 0$, contrary to assumption.

The main contribution of Sturm was the proof that there are infinitely many eigenvalues $\lambda_1 < \lambda_2 < \ldots < \lambda_n \ldots$, tending to $+\infty$. In his study of vibrating strings, d'Alembert had al ready considered an equation of the form $y'' - \lambda\varphi(x)y = 0$ where φ is not constant, and had tried to prove that there is a single value of λ for which there is a solution in $[a, b]$ vanishing at a and b and nowhere else; his idea was to study the corresponding Riccati equation for y'/y when λ varies [**065**]. Sturm elects a similar approach: he considers a solution $u(x, \lambda)$ of (24) satisfying the *first* condition (25), and fixed for instance by the condition $u(a, \lambda) = 1$ (or $u'(a, \lambda) = 1$ if $\alpha = 0$), and he studies the *variation* of $u(x, \lambda)$ as a function of λ; the λ_n are therefore the solutions of the equation $u(b, \lambda)\cos\beta - u'(b, \lambda)\sin\beta = 0$. He is thus led to *compare* solutions of two equations

$$y'' + q_1(x)y = 0, \quad y'' + q_2(x)y = 0 \tag{27}$$

when $q_1(x) < q_2(x)$, and discovers many remarkable such "comparison theorems", of which we will only quote the one which leads to the existence of the eigenvalues. Sturm's paper is rather long-winded and not very clear ([**209**],

TODO: add reference:, [S, p. 259–268]

) and there is a much simpler formulation of his result: an equation $y'' + q(x)y = 0$ is written as a system of two first order equations by the usual introduction of two functions $y_1 = y, y_2 = y'$, which gives $y_1' = y_2, y_2' = -q(x)y_1$, and then one takes as new unknowns two functions r, θ such that $y_1 = r\sin\theta, y_2 = r\cos\theta$, which leads to the system

$$r' = (1 - q(x))r\sin\theta\cos\theta \tag{28}$$

$$\theta' = \cos^2\theta + q(x)\sin^2\theta \tag{29}$$

where the second equation now is of the first order only(*). The comparison theorem which is needed is then the following one: consider solutions φ_0, φ_1 in $[a, b]$ of the two equations

$$\theta' = \cos^2\theta + q_1(x)\sin^2\theta, \quad \theta' = \cos^2\theta + q_2(x)\sin^2\theta \tag{30}$$

and suppose that $q_1(x) < q_2(x)$ in $[a, b]$. Then, if for a number $\alpha \in]a, b[$ one has $\varphi_1(\alpha) \leq \varphi_2(\alpha)$, one also has $\varphi_1(x) < \varphi_2(x)$ for $\alpha < x < b$. The proof is very simple and consists in computing the derivative of the function $w(x) = \varphi_2(X) - \varphi_1(x)$ and showing that there is a continuous function f in $[a, b]$ such that $w'(x) - f(x)w(x) \geq 0$, which implies that w cannot change sign,

If now we apply the preceding change of variable to (24), we get the equation

$$\theta' = \cos^2\theta + (\lambda - q(x))\sin^2\theta \tag{31}$$

and we consider the solution $\omega(x, \lambda)$ such that $\omega(a, \lambda) = \alpha$; the eigenvalues λ are the solutions of the equations

$$\omega(b, \lambda) = \beta + n\pi \quad \text{for } n \in \mathbb{Z} \tag{32}$$

Sturm's comparison theorem then shows that for each $x \in]a, b]$ the function $\lambda \mapsto \omega(t, \lambda)$ is *strictly increasing*, and in addition, from (31) it follows that if $\omega(x, \lambda) = k\pi$ for an integer k, then $\frac{\partial\omega}{\partial x}(x, \lambda) = 1$. From these facts it is easy to show that each equation (32) has one and only one solution λ_n for each $n \geq 1$ and no solution for $n \leq 0$; in addition, the corresponding eigenfunction u_n may be shown to have exactly n zeroes in the interval $]a, b[$ [**052**].

Building on these results of Sturm, Liouville then proceeds to give a general formulation to the expansions of Fourier and Poisson. From relation (26) where λ and μ are replaced by λ_n and λ_m it follows that

$$\int_a^b u_m(x)u_n(x)dx = 0 \quad \text{for } m \neq n. \tag{33}$$

To each function f, defined and continuous in $[a, b]$, Liouville associates its "generalized Fourier coefficients"

$$c_n = \left(\int_a^b f(x)u_n(x)dx\right) \Big/ \left(\int_a^b u_n^2(x)dx\right) \tag{34}$$

(*) This device seems to have first been introduced by H. Prüfer [**180**].

and considers the "generalized Fourier series" $\sum_{n=1}^{\infty} c_n u_n(x)$. In order to study its convergence, he needs more information on the behavior of λ_n and u_n when n tends to $+\infty$. He observes that, if $\lambda = \rho^2 > 0$, any solution of (24) satisfies a relation of the form

$$y(x) = A\cos\rho x + B\sin\rho x + \frac{1}{\rho}\int_a^x q(t)y(t)\sin\rho(x-t)dt \tag{35}$$

(which can be deduced from Lagrange's "variation of constants" method, by writing (24) as $y + \rho^2 y = q(x)y$, although this is not the way Liouville proves (35)). Applying this to $y = u_n$, so that ρ is replaced by $\lambda_n^{\frac{1}{2}}$, he gives a sketchy proof that $\rho_n = \frac{(n-1)\pi}{b-a} + O(1/n)$ and (if $\cos\alpha \neq 0$) $u_n(x) = \sqrt{\frac{2}{b-a}}\cos\rho_n x + O(1/n)$ (when u_n is normalized by the condition $\int_a^b u_n^2(x)dx = 1$). This allows him to prove that the series $\sum_{n=1}^{\infty} c_n u_n(x)$ converges, provided the usual Fourier series of f converges. He still has to show that, if f is continuous, the function $F(x) = \sum_{n=1}^{\infty} c_n u_n(x)$ is equal to $f(x)$; he assumes (without proof) that F is continuous and that $c_n = \int_a^b F(x)u_n(x)dx$, and is reduced to proving that the relations $\int_a^b (F(x) - f(x))u_n(x)dx = 0$ for all n imply $F = f$ (first appearance of the property of "completeness" of an orthonormal system); but this he can only do under the additional assumption that $F - f$ has only a finite number of zeroes in $[a, b]$. The complete proof of the relation $f(x) = \sum_{n=1}^{\infty} c_n u_n(x)$ was only given (for f piecewise C^2) at the end of the XIX$^{\text{th}}$ century, as well as the relation $\sum_{n=1}^{\infty} c_n^2 = \int_a^b f^2(x)dx$; Liouville had only proved the corresponding inequality $c_1^2 + \ldots + c_N^2 \leq \int_a^b f^2(x)dx$ for all N (named after Bessel, who had proved it for the trigonometric system) ([**151**]

TODO: add reference: [S, p.268–281]

).

These remarkable results were to form the pattern of *spectral theory*, the main efforts of analysts in that direction being directed to a generalization of the Sturm–Liouville theory to some types of partial differential equations; but in the first half of the XIX$^{\text{th}}$ century, the theory of these equations was far less advanced than the theory of ordinary differential equations, and it is only after 1880 that progress became possible (see Chapter III).

Chapter II
The "crypto-integral" equations

§1. The method of successive approximations.

The study of celestial mechanics during the XVIII$^{\text{th}}$ century by the method of perturbations consisted, for the theory of the movements of planets, to first neglect their mutual attraction, which gave for each planet a Keplerian orbit around the sun, and then to find the deviations of the actual orbits from the Keplerian ones by taking into account the attraction of other planets; due to the fact that the masses of the planets are much smaller than the mass of the sun, these deviatims were expected to be small. Translated into mathematical terms, this amounted, in the simplest cases, to find good approximatims for the solutions of a system of differential equations

$$y_i' = \varepsilon f_{1i}(x, y_1, \dots, y_n) + \varepsilon^2 f_{2i}(x, y_1, \dots, y_n) + \dots \quad (1 \leq i \leq n) \tag{1}$$

where the parameter ε on the right-hand sides is "small". The general conception of function in XVIII$^{\text{th}}$ century mathematics naturally led to try to express the y_i as a power series in ε

$$y_i = a_i + \varepsilon y_{1i} + \varepsilon^2 y_{2i} + \dots \quad (1 \leq i \leq n), \tag{2}$$

to substitute these expressions in and identify the coefficients of the successive powers of ε on both sides. This led to a succession of equations

TODO: the line of dots at the end is not visible

$$\begin{aligned} y_{1i}' &= f_{1i}(x, a_1, \dots, a_n) \\ y_{2i}' &= F_{2i}(x, y_{11}, \dots, y_{1n}) \\ y_{3i}' &= F_{3i}(x, y_{11}, \dots, y_{1n}, y_{21}, \dots, y_{2n}) \end{aligned}$$

all of which had right-hand sides which were known functions, hence were reduced to mere "quadratures". No attempt was made to justify mathematically those procedures; the goal of these computations was merely to obtain a satisfactory agreement with observations.

It is well-known that Cauchy was the first mathematician who proved existence theorems for *general* types of differential equations, for which no explicit solution is available. His strategy was to consider the various methods introduced earlier for the purpose of numerical computations, and to show that, under certain conditions, these methods actually gave *convergent* approximation processes having a solution as limit. In particular, in a paper published in 1835 in Prag ([**040**], (2), vol. XI, p. 399–465), he takes up the method outlined above, not for an ordinary differential equation, but for a linear partial differential equation of first order (which was known to be equivalent to a system of ordinary differential equations)

$$\frac{\partial U}{\partial t} = \sum_{i=1}^{p} A_i(t, x_1, \dots, x_p) \frac{\partial U}{\partial x_i}; \tag{3}$$

the problem is to find a solution which for $t = 0$ reduces to a given function $u(x_1, \dots, x_n)$, and Cauchy transforms (3) into the equivalent "integro-differential" equation by considering $x_1, \dots, x_p$ as parameters:

$$U(t, x_1, \dots, x_p) = u(x_1, \dots, x_p) + \int_0^t \left(\sum_{i=1}^{p} A_i(s, x_1, \dots, x_p) \right) \frac{\partial U}{\partial x_i} ds \tag{4}$$

which he solves by successive approximations, starting with $U_0 = u$, and defining

$$U_n(t, x_1, \dots, x_p) = u(x_1, \dots, x_p) + \int_0^t \left(\sum_{i=1} A_i(s, x_1, \dots, x_p) \right) \frac{\partial U_{n-1}}{\partial x_i} ds$$

by induction; but he is only able to prove convergence towards a solution when the A_i are analytic functions.

In his 1837 papers on the Sturm–Liouville problem Liouville independently applied a similar method to the linear differential equation $y'' = f(x)y$, for which he wants to find a solution in $[a, b]$ satisfying the boundary condition $y'(a) - hy(a) = 0$. He starts from the function $y_0(x) = 1 + h(x - a)$ satisfying that condition, and considers the series

$$y = y_0 + y_1 + \dots + y_n + \dots \tag{5}$$

where the y_n are determined for $n > 0$ by the recursive equations

$$y_{n+1}(x) = \int_a^x dt \int_a^t f(s) y_n(s) ds.$$

It must be remembered that at that time the concept of uniform convergence had not yet been formulated, and no justifi cation had been given for asserting the continuity of a con vergent series of continuous functions, or differentiating or integrating such a series termwise, Liouville proves very easily that there is a constant C such that

$$|y_n(x)| \leq C^n (x - a)^{2n} / (2n)!$$

from which he concludes that the series (5) giving $y(x)$ converges for every x; but he tacitly takes for granted that y is a C^2 function and a solution of his problem.

In addition, Liouville makes the interesting remark that the function y can also be defined by the relation

$$y = y_0 + \int_a^x dt \int_a^t f(s) y(s) ds \tag{6}$$

(which he could also have writtem $y = y_0 + \int_a^x (x - t) f(t) y(t) dt$), thus giving what is probably the first example of what will be called later a "Volterra integral equation of the second kind" (see chap. IV); if one writes $z_n = y_0 + y_1 + \dots + y_n$, Liouville observes that the z_n are given by $z_0 = y_0$, and the recursive equations

$$z_{n+1}(x) = y_0 + \int_a^x dt \int_a^t f(s) z_n(s) ds \tag{7}$$

which is the standard process of "successive approximations" for these equations ([**151**]

TODO: add reference: [S, p. 268–281]

).

We have already seen that a little later in his papers of 1837, Liouville gives another "integral equation" equivalent to an equation $y' = f(x)y$ (chap. I, §3, equation (35)). This exemplifies a general idea: if a linear differential operator P is such that the equation $P \cdot u = f$ can be solved by a formula $u = y_0 + G \cdot f$, where G is a linear operator, then the equation $P \cdot u + Q \cdot u = 0$, where Q is an operator, is equivalent to $u - G \cdot (Q \cdot u) = y_0$; in the case of Liouville, $P \cdot u = u'' + \rho^2 u$ and $Q \cdot u = -qu$, and G is an integral operator (cf. chap. IX, §5).

The simplest application of this idea is to the proof of Cauchy's existence and uniqueness theorem for an ordinary differential equation $y' = f(x, y)$, which, with the initial con dition $y(x_0) = y_0$, is equivalent to $y = y_0 + \int_{x_0}^{x} f(t, y)dt$. In this general form it is given by E. Picard in his 1890 paper on successive approximations [**172**], where it comes as an afterthought, the bulk of the paper being concerned with applications of the method to partial differential equations. However, in these applications, Picard is directly influenced by the fundamental earlier works of C. Neumann on the Laplace equation and of H.A. Schwarz on the equation of vibrating membranes, which are the direct forerunners of the theory of integral equations; we will describe in detail C. Neumann's results in §4 of this chapter, and H.A. Schwarz's paper in chap. III, §1.

§2. Partial differential equations in the XIX$^{\text{th}}$ century.

During the whole XIX$^{\text{th}}$ century, the theory of partial dif ferential equations (in contrast with the theory of ordinary differential equations) has remained in an embryonic stage. The only general theorem, patterned after the Cauchy theorem on local existence and uniqueness of solutions of ordinary differential equations, is the Cauchy–Kowalewska theorem: suppose we have a system of r equations in r unknown real functions $v_1, \dots, v_r$ of $p+1$ real variables $x_1, \dots, x_p$, of type

$$\frac{\partial v_j}{\partial x_{p+1}} = \tag{8}$$
$$= H_j\left(x_1, \dots, x_{p+1}, v_1, \dots, v_r, \frac{\partial v_1}{\partial x_1}, \frac{\partial v_1}{\partial x_2}, \dots, \frac{\partial v_r}{\partial x_{p-1}}, \frac{\partial v_r}{\partial x_p}\right) (1 \le j \le r)$$

where the right hand sides do not contain any derivative with respect to x_{p+1} and are supposed to be real and *analytic* with respect to their $p + 1 + r + rp$ variables, in a neighborhood V_0 of 0 in $\mathbb{R}^{p+1+r+rp}$; then there is a small neighborhood V of 0 in $\mathbb{R}^{p+1}$ such that (8) has in V a unique solution $(v_1, \dots, v_r)$ consisting of *analytic* functions in V, such that $v_j(x_1, \dots, x_p, 0) = 0$ in $V \cap \mathbb{R}^p$ for $1 \le j \le r$.

The tendency (inherited from the XVIII$^{\text{th}}$ century) to consider that the most interesting functions were analytic was still very strong during the whole XIX$^{\text{th}}$ century, and therefore at first the analyticity restrictions of the Cauchy–Kowalewska theorem did not worry mathematicians very much, However, as it was known that some special types of partial differential equations, such as the scalar equation of first order and some types of second order equations, had solutions under much less stringent restrictions, people began to wonder if some other method than Cauchy's "method of majorants" (which could only be applied to analytic functions) would not yield a generalization of the Cauchy–Kowalewska theorem, at least for C^∞ functions. The question remained unanswered until 1956, when H. Lewy gave the surprising example of a system of two linear equations in 3 variables, with C^∞ coefficients

$$\begin{cases} \dfrac{\partial v_1}{\partial x_1} = \dfrac{\partial v_2}{\partial x_2} - 2x_2\dfrac{\partial v_1}{\partial x_3} - 2x_1\dfrac{\partial v_2}{\partial x_3} - f(x_3) \\ \dfrac{\partial v_2}{\partial x_1} = -\dfrac{\partial v_1}{\partial x_2} + 2x_1\dfrac{\partial v_1}{\partial x_3} - 2x_2\dfrac{\partial v_2}{\partial x_3} \end{cases}$$

which, for a suitable choice of the real C^∞ function f, has *no solution* whatsoever around any point (even if one al lows solutions which are distributions).

We shall not discuss the numerous local studies of analytic systems of partial differential equations (not necessarily reducible to the form (8)) which followed the Cauchy–Kowalewska theorem, since they had no influence on the development of Functional Analysis as we understand it.

The remainder of the theory of partial differential equations until 1890 was limited to very special scalar equations (mostly linear equations or order 2) generally derived from physical problems(*), such as the equation of vibrating strings and its generalizations to 3 and 4 variables (the "wave equations"), the Laplace equation $\Delta u = 0$ in 2 and 3 variables, the heat equation in 2, 3 and 4 variables. For these equations, the techniques of "separation of variables" or of Fourier transforms (see chapter VII, §6) gave special solutions or solutions depending on "arbitrary" functions. But until 1825 the determination of solutions by boundary condi tions (of which we have seen a few examples in Chapter I) was always restricted to *explicitly described and particular* such conditions.

A first attempt of classification of second order equations in 2 variables had been made by Laplace [**137**] He considered "quasi-linear" equations, i.e. those of the form

$$\begin{aligned} &A(x,y)\frac{\partial^2 z}{\partial x^2} + B(x,y)\frac{\partial^2 z}{\partial x \partial y} + \\ &+C(x,y)\frac{\partial^2 z}{\partial y^2} + F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) = 0 \end{aligned} \tag{9}$$

linear in the second order derivatives. As he did not have a clear idea of the distinction between real and complex vari ables, and therefore did not hesitate to give complex values to x and y, he asserted that a suitable change of variables could reduce the terms of (9) containing second order derivatives either to $\frac{\partial^2 z}{\partial x \partial y}$ or to $\frac{\partial^2 z}{\partial x^2}$ when A, B, C are not all identically zero! With the development of the theory of functions of one complex variable, it was soon realized that, for real variables x, y, equations (9) where the second order derivatives enter by $\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}$ (called *elliptic* equations) had to be sharply distinguished from those (called *hyperbolic* equations) where the second order derivatives enter by $\frac{\partial^2 z}{\partial x \partial y}$ or $\frac{\partial^2 z}{\partial x^2} - \frac{\partial^2 z}{\partial y^2}$. The study of general boundary conditions for hyperbolic equations only begins around 1860 and will have little contact with Functional Analysis until around 1925 (see chapter IX, §5). On the contrary, the various problems connected with the Laplace equation in 2 or 3 wvariables will be one of the main concerns of analysts from 1828 onwards, and will become the impetus leading to the theory of integral equations, and thence to our modern Functional Analysis.

§3. The beginnings of potential theory.

In 1748, D. Bernoulli had introduced in the theory of newtonian attraction the function $\Omega(M) = \sum_i (m_i \mu / r_i)$ for a point M of mass μ attracted by a finite number of punctual masses m_i, where r_i, is the distance of M to the mass m_i; and in 1773 Lagrange observed that the knowledge of that function immediately gave the components of the attraction exerted on M, by taking the derivatives of Ω with respect to the coordinates x, y, z of M. When the finite number of masses is replaced by a solid V of density ρ and

(*) See the interesting description of these problems given by Poincaré in the Introduction of his 1890 paper on the equations of mathematical physics ([**177**], vol. IX, p. 28–32)

the point M is outside V, the function Ω becomes

$$\Omega(x,y,z) = \mu \iiint_V \frac{\rho(\xi,\eta,\zeta)d\xi d\eta d\zeta}{r(x,y,z,\xi,\eta,\zeta)} \tag{10}$$

with $r(x,y,z,\xi,\eta,\zeta) = \left((x-\xi)^2+(y-\eta)^2+(z-\zeta)^2\right)^{\frac{1}{2}}$ [**135**].

In 1782 and 1785, Laplace showed that outside of V the function Ω satisfied the equation

$$\Delta\Omega \equiv \frac{\partial^2\Omega}{\partial x^2} + \frac{\partial^2\Omega}{\partial y^2} + \frac{\partial^2\Omega}{\partial z^2} = 0 \tag{11}$$

[**137**], and in 1813 Poisson completed that result by showing that if ρ discon tinuous in V, the integral (10) is still meaningful inside V, and Ω satisfies the "Poisson equation"

$$\Delta\Omega + 4\pi\rho = 0 \tag{12}$$

([**178**],

TODO: add reference: S, p. 342–346

). His idea is to consider the value of Ω at a point M in V as the sum of the corresponding functions Ω_1, Ω_2 relative to a small ball V_1 of center M and to the complement V_2 of V_1 in V; one has then $\Delta\Omega_2 = 0$, and when the radius of V_1 tends to 0 Poisson shows that $\Delta\Omega_1$ tends to $-4\pi\rho(M)$. (In fact his argument is not rigorous when one only assumes the continuity of ρ, and the existence of $\Delta\Omega$ is only guaranteed when ρ satisfies a Hölder condition; when ρ is merely continuous, equation (12) is valid only if the second order derivatives are taken in the sense of the theory of distributions (chap. VIII, §3)).

After the discovery of Coulomb's laws (1785) the Laplace equation became of central importance in electrostatics; it also was found to govern "stationary" phenomena in hydrodynamics and the theory of heat. Finally the so-called "Cauchy–Riemann" equations for real functions P, Q of x, y such that $P + iQ$ is an analytic function of $x + iy$, were known since the middle of the XVIII$^{\text{th}}$ century, and they implied that P and Q were solutions of the Laplace equation in 2 variables. Very early in the XIX$^{\text{th}}$ century, Gauss was well aware of this connection and of the fact that one obtained solutions of the Laplace equation in 2 variables by replacing the function (10) by

$$\Omega(x,y) = \iint_D \rho(\xi,\eta)\log\frac{1}{r(x,y,\xi,\eta)}d\xi d\eta \tag{13}$$

for a bounded domain D in the plane. The development by Cauchy of the theory of holomorphic functions of a complex variable could thus be used to yield properties of harmonic functions of 2 variables, such as for instance the non existence of relative extrema for such a function in its domain of definition; it was then natural to conjecture that similar properties were also valid for harmonic functions of 3 (and later for $n \geq 4$) variables, although they had to be proved by other means.

The first paper dealing with *general* boundary conditions for a partial differential equation was written in 1828 by George Green, a self-taught English mathematician (1793–1841); it is concerned with electrostatics and the general study in that theory of what Green for the first time calls *potential functions*. By that he not only means the functions of the form (10), but also what will later be called *simple layer potentials*, namely functions of the type

$$\Omega(M) = \iint_\Sigma \frac{\rho(P)}{MP}d\sigma(P) \tag{14}$$

where Σ is a smooth surface, ρ (the "density") a continuous function on Σ and $d\sigma$ the element of area on Σ; he was naturally led to such functions by the known experimental fact that on conductors the electric charges are concentrated on their surface.

Green was interested in the relations between the surface density ρ and the potential it defines. He first establishes the famous theorem which, for the operator Δ, generalizes to 3 dimensions the relation between a differential operator and its adjoint (Chapter I, formula (5)):

$$\iiint_V (u\Delta v - v\Delta u)\, d\omega = \iint_\Sigma \left(v\frac{\partial u}{\partial n} - u\frac{\partial v}{\partial n}\right) d\sigma \tag{15}$$

where Σ is a smooth surface limiting a bounded volume V, u and v are C^2 in a neighborhood of $\overline{V}$, $\frac{\partial u}{\partial n}$ is the derivative of u along the exterior normal of Σ (*). He then has the original idea (*) of considering a function u which, still C^2 for all points different from a point M in V, becomes infinite at M in such a way that the difference $u(P) - (1/MP)$ is bounded when P tends to M; he applies (15) to the volume V from which a small ball of center M has been excised, and by letting the radius of the ball tend to 0, he obtains the formula

$$4\pi v(M) + \iiint_V (u\Delta v - v\Delta u)\, d\omega = \iint_\Sigma \left(v\frac{\partial u}{\partial n} - u\frac{\partial v}{\partial n}\right) d\sigma \tag{16}$$

provided of course the triple integral exists. Taking in particular $u(P) = 1/MP$ would give for a solution v of $\Delta v = 0$

$$4\pi v(M) = \iint_\Sigma \left(v\frac{\partial\left(\frac{1}{r}\right)}{\partial n} - \frac{1}{r}\frac{\partial v}{\partial n}\right) d\sigma \qquad (\text{with } r(P) = MP) \tag{17}$$

in other words, an integral formula which would solve the Laplace equation when v *and* $\frac{\partial v}{\partial n}$ were known on Σ. This was in agreement with what was known at the time for partial differential equations of the second order, such as the equation of vibrating strings (Chapter I, §2). However, experi ments showed that v was entirely determined by its values on Σ, and therefore it was not possible to take for both v and $\frac{\partial v}{\partial n}$ on Σ *arbitrary* continuous functions, so that the situation appeared quite different from the boundary conditions for hyperbolic equations. Furthermore, there was at least one case when an explicit formula gave v inside V by an integral extended to Σ, mnamely the *Poisson formula* for a ball V of center 0 and radius a, published in 1820:

$$v(M) = \frac{1}{4\pi}\iint_\Sigma \frac{a^2 - \rho^2}{ar^3} v(P)d\sigma \tag{18}$$

with $\rho = OM$. Green observed that one would have a similar formula for general domains V:

$$v(M) = \frac{1}{4\pi}\iint_\Sigma v(P)\frac{\partial G}{\partial n}(M,P)d\sigma \tag{19}$$

(*) Lagrange [**135**] and Gauss [**82**] had already obtained more particular relations of that kind between volume and surface integrals.

(*) It is of course the same idea which leads to the Cauchy formula giving the value of a holomorphic function inside a domain D when it is known on the boundary of D. However, it is unlikely that Green knew Cauchy's papers

by substituting in his formula (16) for u a function $G(M,P)$ such that: 1° in $V \times V$, G is C^2 provided $M \neq P$ and $\frac{\partial G}{\partial n}$ exists on Σ; 2° $G(M,P) - 1/MP$ remains bounded when P tends to M; 3° $G(M,P) = 0$ when M is in V and P on Σ; 4° when M is fixed in V, G, as a function of P, satisfies the Laplace equation in V. He could not prove the existence of such a "Green function", but made it plausible by an appeal to experimental facts: when the surface Σ is connected to the ground, and an electric charge $+1$ is put at the point M, it "induces" an electric charge on Σ such that the total potential of that charge and the punctual charge at M is 0 on Σ; that potential should be the function $G(M,P)$ ([**090**],

TODO: add reference: [S, p. 347–358]

).

Finally, by an ingenious use of his formula (15), Green could prove that in $V \times V$, one had $G(P,M) = G(M,P)$ for $M \neq P$.

§4. The Dirichlet principle.

Gauss had very early been interested in the Laplace equation, both in 2 variables in connection with his work on complex numbers, and in 3 variables in relation with his astronomical studies, and we have seen that in his 1813 paper on the attraction of spheroids, he had proved particular cases of the Green formula (15). After 1830, he devoted much of his time to the study of magnetism, both experimentally and theoretically, and thus was led to new research on potential theory, which he published in 1840 [**090**]. In that paper, he quotes no other work on the subject, and it is very unlikely that he ever heard of Green (whose work was not widely known, even in England)(*); he expands his 1813 formulas and obtains in this way some new particular cases of Green's formula (15), although he does not seem to have thought of formula (16). The closest approach to the latter is his famous "mean value formula"

$$v(0) = \frac{1}{4\pi} \iint_{\Sigma} v(P) d\sigma \tag{20}$$

for a harmonic function v in a sphere Σ of center 0, for which it is quite surprising that he should not have observed that it was a special case of Poisson's formula (18) which he cannot have failed to know.

As Green had done, Gauss was particularly interested in the behavior of simple layer potentials (14) when M tends to a point on the surface Σ; by a careful study, he shows that the potential Ω is coutinuous everywhere, and that the normal derivatives at a point M_0 of Σ exist on both sides of the surface, but have *different values*, their difference being $4\pi\rho(M_0)$; all this had been taken for granted without proof by Green.

Gauss attacked several problems related to potential theory, some of which were to become the focus of active research after 1930. One was the *equilibrium problem*: find a distribution of electric charges on a closed surface Σ giving a potential which is *constant* on Σ; another consisted in replacing charges inside Σ by charges on Σ in such a way that the potential outside Σ remains the same (what would later be called a "sweeping-out" process), and Gauss showed that it could be solved if the equilibrium problem had a solution.

Regarding the latter, Gauss introduced a new idea which was to become quite central in potential theory: he observed that if the potential Ω is given by (14) with $rho \geq 0$,

(*) The fact that Gauss also uses the word "potential" with the same meaning may be attributed to the fact that the word (in its Latin form) was commonly used in the XVIII$^{\text{th}}$ century by "natural philosophers".

and U is any continuous function on Σ, then if ρ is chosen such that the integral $\iint_\Sigma (\Omega - 2U)\rho d\sigma$ takes the *smallest* possible value among all possible choices of ρ, then $\Omega - U$ is constant on Σ, and he added that the existence of such a density ρ was obvious.

By adding to Ω a suitable constant, this method of Gauss solved the problem of finding a harmonic function u in the volume V, continuous in $\overline{V} = V \cup \Sigma$, and equal on Σ to a given function U [(*)]. The same problem was considered a little later by W. Thompson (the future Lord Kelvin) in 1847 and by Dirichlet around the same time in his lectures (published long afterwards)

TODO: add reference: [S, p. 380–387]

; it became known as the *Dirichlet problem*. Their idea is similar to Gauss's: they consider the volume integral

$$\iiint_V \left(\left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2 \right) d\omega \tag{21}$$

and the function v continuous in $\overline{V}$ with continuous and bounded first derivatives in V (v taking the given values on Σ), for which the integral (21) takes its smallest value; applying the standard techniques of the Calculus of variations, they easily show that such a function is indeed harmonic in V.

The great success of this idea is probably due to the imaginative use Riemann almost immediately made of it, in his epoch-making papers on holomorphic functions, Riemann surfaces and abelian integrals. By considering the real and imaginary parts of such functions, he was the first to realize that the existence theorems he needed could be derived from similar existence theorems for these harmonic functions, which he thought he could prove by adapting Dirichlet's argument to similar integrals in 2 variables, called by him "Dirichlet principle" [**182**].

His magnificent results attracted considerable attention, but soon mathematicians realized that they rested on three properties for which W. Thompson, Dirichlet and Riemann did not give any proof at all:

1) For a given continuous function g on Σ, there exist continuous functions v in $\overline{V}$ whose restriction to Σ is g and for which the integral (21) is meaningful.

2) If such functions exist, there is one for which the smallest value of (21) is attained.

3) For that function v, the second order derivatives $\frac{\partial^2 v}{\partial x^2}, \partial^2 v \partial y^2, \partial^2 v \partial z^2$ exist.

However, in 1871, F. Prym presernited an example (for two variables and V the disk $x^2 + y^2 < 1$) where *no function* v satisfying 1) existed [**181**][(*)]. On the other hand, in 1870, Weierstrass observed that in all problems of the Calculus of variations which had been studied since the beginning of the XVIII$^{\text{th}}$ century, properties 2) and 3) had been taken for granted without any proof, and he gave a very simple example in which property 2) does not hold: the problem of minimizing the integral $\int_{-1}^{1} xy'^2 dx$ among all C^1 functions y defined in the interval [-1, 1] and satisfying the boundary conditions $y(-1) = a, y(1) = b$, with $a \neq b$

[(*)] If such a problem is solved, it implies the existence of the Green function: one considers the function $u(M, P)$ harmonic in V (as a function of P) which takes the values $-1/MP$ on Σ; the Green function is then $G(M, P) = u(M, P) + (1/MP)$, provided one shows that $\frac{\partial G}{\partial n}$ exists and is continuous on Σ.

[(*)] The discovery of that fact is usually attributed to Hadamard, who published a similar example in 1906 [**094**]

TODO: add reference: [S, p. 390–391]

. Spurred by these difficulties, Weierstrass and his pupils (P. Du Bois-Reymond, A. Kneser, S. Zaremba) undertook to put the Calculus of variations on sounder foundations and were able to rescue many classical results from the suspicion raised by such counterexamples. But the "Dirichlet principle" eluded their efforts, and it was only in 1899 that Hilbert, using new ideas in what was called his "direct method", was able to give a complete justification of the use Riemann had made of that "principle" [**111**].

§5. The Beer–Neumann method.

We shall see in later chapters how the concepts and tools used by Hilbert and the Weierstrass school contributed to the birth of General Topology and later to the introduction of such notions as "weak" solutions of partial differential equations. Meanwhile, the challenge remained to prove the existence of a solution to the Dirichlet problem and similar boundary values problems for the Laplace equation, at least under conditions such as were used in Riemann's work. Between 1870 and 1890, that challenge was successfully taken up by three mathematicians: H.A. Schwarz around 1870, C. Neumann in 1877 and H. Poincaré in 1887.

We shall not discuss in detail the contributions of Schwarz and Poincaré, which did not influence directly the development of Functional Analysis. Both are based on the idea of *approximation*: starting from known solutions of the Dirichlet problem for special kinds of domains, an approximation process enables one to get solutions for much more general domains. Schwarz limits himself to 2 variables; he first comnsiders domains limited by a convex polygon, for which it is possible to prove directly (by explicit construction) the existence of a conformal mapping on the unit disk, hence the existence of a solution of the Dirichlet problem (by transferring the Poisson formula from the circle to the polygon). Using the maximum principle, it is then possible to prove the existence of the solution for a *convex* domain by approximating it by a sequence of inscribed convex polygons. A little later, he invented an ingenious "alternating process" which enabled him to show that when one can solve the Dirichlet problem for two domains in the plane, it is also possible to solve it for their union, and from that result he finally showed that the Dirichlet problem in the plane is solvable for any domain limited by piecewise analytic curves [**196**].

Poincaré's famous "sweeping-out method" applies to any number of dimensions. To solve the Dirichlet problem for a bounded domain V limited by a surface Σ, he shows (using the maximum principle) that it is enough to consider the case in which the function given on Σ is the restriction to Σ of a func tion Φ defined in a neighborhood W of $\overline{V}$, of class C^2 and such that $\Delta\Phi \geq 0$. By Poisson's equation (12), Φ is the sum of a harmonic function and a potential Φ_0 of masses ≥ 0. The fundamental idea is that if B is a ball contained in W, it is possible to use the Poisson integral (18) extended to the surface of B in order to replace Φ_0 by another potential which coincides with Φ_0 outside B and is smaller than Φ_0 inside B; the masses inside B have been "swept out" on the surface of B. One then takes an infinite sequence of balls B_n whose union is V, and one applies the "sweeping-out" process repeatedly to the B_n in the order $B_1, B_2, B_1, B_2, B_3, B_1, B_2, B_3, B_4, \ldots$ (each B_n, is "swept-out" in finitely many times). The corresponding sequence of potent ials is decreasing, hence has a limit in V; using Harnack's inequalities (consequences of the Poisson formula (18)) and the maximum principle, Poincaré is able to show that this limit is a solution of the Dirichlet problem, provided the boundary Σ satisfies a "regularity" condition, namely, for any point $M \in \Sigma$, there must be a small ball whose intersection with $\overline{V}$ is reduced to M [**177**]; later, Zaremba could replace the small ball by a small cone of vertex M in that condition.

In contrast with Schwarz's and Poincaré's papers, the Beer–Neumann method was a landmark in Functional Analysis by introducing the first example of what was later to be called a "Fredholm integral equation of the second kind". Green's formula (17) naturally introduced still another type of potential:

$$u(M) = \iint_{\Sigma} \rho(P) \frac{\partial}{n} \left(\frac{1}{MP} \right) d\sigma \tag{22}$$

which was harmonic outside the surface Σ. It also occurred in the theory of magnetism, from which it got its name of *double layer potential*: it was there conceived as the limit of a difference of two simple layer potentials, one with density μ on Σ, the other with density μ on a surface Σ' parallel to Σ and at an "infinitely small" distance ε; when ε tends to 0, μ was supposed to increase to $+\infty$ in such a way that the product $\mu\varepsilon$ tended to *rho*.

Such a potential had been shown to have near Σ a behavior quite similar to the normal derivative of a simple layer potential, studied by Gauss: when M tends to a point M_0 of Σ along the normal to Σ at M_0, $u(M)$ tends to a limit on each side of Σ, but these limits are *different* in general; however, $\frac{\partial u}{\partial n}$ is the *same* on both sides.

Formula (22) also had a nice geometric interpretation; one has $\frac{\partial}{\partial n}\left(\frac{1}{MP}\right) = \frac{\cos\varphi}{MP^2}$ where φ is the angle between MP with the normal to Σ at P, and $\frac{\cos\varphi}{MP^2} d\sigma$ is the infinitesimal "solid angle" from which $d\sigma$ is "seen" from the point M.

Around 1860, C. Beer proposed to obtain a solution to the Dirichlet problem by formula (22) for a suitable density ρ on Σ. From the continuity properties of double layer potentials, it follows that if Σ is a smooth surface, and $g(M)$ is the function on Σ to which the solution $u(M)$ must be equal, the unknown density must satisfy the equation

$$2\pi\rho(M) + \iint_{\Sigma} \rho(P) \frac{\partial}{\partial n} \left(\frac{1}{MP} \right) d\sigma = g(M) \quad \text{for } M \in \Sigma. \tag{23}$$

He then concluded that one could compute ρ by the usual device of "successive approximations" (§1) starting with $\rho_0(M) = \frac{1}{2\pi} g(M)$ and defining recursively $\rho_n(M)$ by

$$2\pi\rho_n(M) + \iint_{\Sigma} \rho_{n-1}(P) \frac{\partial}{\partial n} \left(\frac{1}{MP} \right) d\sigma = 0 \quad \text{for } n \geq 1$$

so that the series $\rho(M) = \rho_0(M) + \rho_1(M) + \ldots + \rho_n(M) + \ldots$ would give the solution to (23); but he made no attempt to prove that the series converged.

In 1877, Carl Neumann attempted to give such a proof [**165**]. He restricted himself to the case in which the domain V is bounded and *convex*, but he allowed a non smooth boundary Σ; equation (23) must then be modified to

$$4\pi\rho(M) = \iint_{\Sigma} (\rho(M) - \rho(P)) \frac{\cos\varphi}{MP^2} d\sigma + f(M) \tag{24}$$

with f continuous on Σ, and the successive approximations are given by $4\pi\rho_0(M) = f(M)$ and, for $n \geq 1$,

$$4\pi\rho_n(M) = \iint_{\Sigma} (\rho_{n-1}(M) - \rho_{n-1}(P)) \frac{\cos\varphi}{MP^2} d\sigma \tag{25}$$

Neumann's idea is to consider the maximum value L_n and minimum value l_n of ρ_n, and to show that there is a number q such that $0 < q < 1$ and

$$L_n - l_n \leq (L_0 - l_0)\, q^{n-1} \tag{26}$$

from which he majorizes $|\rho_n(M)|$ by a multiple of q^n using (25), and he can conclude that the series $\sum_{n=0}^{\infty} \rho_n(M)$ converges to a continuous function.

To prove (26), Neumann divides Σ into two parts A_n, B_n respectively defined by the conditions

$$\frac{1}{2}(L_{n-1} + l_{n-1}) \leq \rho_{n-1}(P) \leq L_{n-1} \quad \text{for } A_n$$

$$l_{n-1} \leq \rho_{n-1}(P) < \frac{1}{2}(L_{n-1} + l_{n-1}) \quad \text{for } B_n$$

and he deduces from (25) that for all points M of Σ

$$(L_{n-1} - l_{n-1})\left(A_n(M) + \frac{1}{2}B_n(M)\right) \leq 4\pi\rho_n(M) \leq (L_{n-1} - l_{n-1})\left(\frac{1}{2}A_n(M) + B_n(M)\right)$$

where $A_n(M)$ and $B_n(M)$ are the solid angles from which A_n and B_n are "seen" from M. This implies

$$L_n - l_n \leq (L_{n-1} - l_{n-1})\, q$$

where q is the *least upper bound* of the quantity

$$\Lambda(M, M', A, B) = \frac{1}{4\pi}\left(\frac{1}{2}A(M) + B(M) + A(M') + \frac{1}{2}B(M')\right) \tag{27}$$

when M and M' vary arbitrarily in Σ, A is an arbitrary closed part of Σ and B its complement. One is thus faced with the *purely geometric* problem of showing that $q < 1$. The expression (27) can be written

$$\frac{1}{4\pi}\left(A(M) + B(M) + A(M') + B(M') - \frac{1}{2}(A(M) + B(M'))\right)$$

and also

$$\frac{1}{4\pi}\left(\frac{1}{2}(A(M) + B(M)) + \frac{1}{2}(A(M') + B(M')) + \frac{1}{2}(A(M') + B(M))\right)$$

and as one always has $A(M) + B(M) \leq 2\pi$ (maximum value of the solid angle from which the whole of Σ is "seen" from one of its points), the problem can also be formulated in two equivalent ways:

$$A(M) + B(M') \geq 4\pi r \quad \text{for an } r > 0, \tag{28}$$

$$A(M) + B(M') \leq 4\pi s \quad \text{for an } s > 1, \tag{29}$$

for all points M, M' in Σ in two parts A, B. The form (28) of that condition immediately shows that there is an *exceptional* type of convex set for which it cannot be satisfied, namely the case in which V is the *intersection of two convex cones* ("double cone"): indeed we then have $A(M) = B(M') = 0$ if A is the surface of one of the cones, B the surface of the other, M the vertex of A and M the vertex of B. Furthermore, this particular choice of A, B, M and M' is the *only one* for which $A(M) + B(M')$ may be 0. However, when the exceptional case is excluded, Neumann concludes, from the fact that $A(M) + B(M') > 0$ for all choices of A, B, M and M', that there is an $r > 0$ for which (28) is satisfied for all these choices, and does not give a proof of that assertion valid for all convex sets other than double cones. This gap in Neumann's proof seems to have remained undetected until Lebesgue drew attention to it is 1937 ([**138**], vol. IV, p. 151–166). He shows in addition how one can fill in that gap by a compactness argument:

there are two points M_0, M_0' in Σ, limits of sequences $(M_k), (M_k')$ such that for each k there is a splitting of Σ in two parts A_k, B_k such that $A_k(M_k) + B_k(M_k')$ tends to the l.u.b. $4\pi s$ of $A(M) + B(M')$ for all choices of A, B, M, M'. On the other hand there are a point N of Σ and neighborhoods $V(M_0), V(M_0'), V(N)$ of M_0, M_0', N respectively in Σ such that the planes of support at all points of $V(N)$ do not intersect $V(M_0)$ nor $V(M_0')$ (it is here that the assumption that V is not a double cone is used); an elementary geometrical ar gument then gives an upper bound < 1 for s. Historically, such an argument would have been barely possible in the late 1870's, but I strongly doubt that C. Neumann was familiar enough with the use of the "Bolzano–Weierstrass" theorem (as it was called at that time) to have thought of it. He was apparently satisfied with the fact that for simple convex sets, such as ellipsoids, it was possible to compute explicity an upper bound < 1 for s.

C. Neumann dealt in the same way with the Dirichlet problem in the plane, with a similar gap in his proof.

For a long time, the restrictions on the surface Σ in all the existence proofs of the Dirichlet problem were thought to be imperfections of the methods of proof; but in 1912, Lebesgue gave an example (in 3 dimensions) of a bounded open set V (homeomorphic to a ball) such that there is a continuous function on the boundary Σ of V, for which the Dirichlet problem has no solution ([**138**], vol. IV, p. 131). This was the starting point of modern Potential theory, where, on one hand, the initial formulation of the Dirichlet problem is modified in such a way that it always has a unique "solution" for any bounded domain, the word "solution" being interpreted in some "weak" sense; on the other hand, the behavior of these "weak" solutions on the boundary of the domain is investigated under various conditions [**030**]. The detailed history of that extensive theory is outside the scope of this book.

Chapter III
The equation of vibrating membranes

§1. H.A. Schwarz's 1885 paper.

The same physical arguments which lead to the equation of vibrating strings (Chap. I, §2, equation (7)) apply to the small vibrations of a membrane which at rest is in the plane Oxy, and has a constant density: if $z = u(x, y, t)$ is the equation of its surface at time t, the function u satisfies the equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 u}{\partial t^2} \tag{1}$$

(for suitable units of length and time). The usual method of "separation of variables" consists here in looking for solutions $u(x, y, t) = v(x, y)w(t)$ and one finds for v the equation (also called "Helmholtz's equation")

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \lambda v = 0 \tag{2}$$

for a constant λ. If in addition the membrane at rest is a bounded portion Ω of the plane and is *fixed* at its boundary Σ (which means that $u(x, y, t) = 0$ for all t if $(x, y) \in \Sigma$), λ must be $> 0, w(t) = \sin \sqrt{\lambda} t$, and one has to find a solution v of (2) which vanishes on Σ and is not identically 0. Contrasting with the easy solution of the corresponding problem for the vibrating string, the elucidation of that problem was going to challenge the ingenuity of mathematicians during the whole second half of the XIX$^{\text{th}}$ century.

Experimental evidence, as well as the explicit solution of the problem for very special domains Ω, such as a rectangle or a disk, showed that, just as in the case of the vibrating string, solutions of (2) vanishing on Σ and not identically 0 could only exist when λ was equal to one of an infinite sequence (λ_n) of real numbers > 0 (the "eigenvalues" of the problem), tending to $+\infty$.

The first attempt to prove such a result for general domains Ω was made by H. Weber in 1869 [**247**], by an adaptation of the variational method used by Riemann for the Dirichlet problem. Using Green's formula (Chapter II, formula (15)) he first shows that if μ_1, μ_2 are two distinct eigenvalues, v_1, v_2 corresponding "eigenfunctions", then

$$(\mu_1 - \mu_2) \iint_\Omega v_1(x, y) v_2(x, y) dx dy = 0 \tag{3}$$

from which he deduces, as Poisson had done for ordinary differential equations (Chap. I, §3), that the eigenvalues are necessarily real numbers. To determine the smallest eigenvalue λ_1, he considers the Dirichlet integral

$$F(v) = \iint_\Omega \left(\left(\frac{\partial v}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2 \right) dx dy \tag{4}$$

for C^2 functions v in $\overline{\Omega}$, equal to 0 on Σ and subject to the additional constraint

$$\iint_\Omega v^2 dx dy = 1. \tag{5}$$

He assumes, as Riemann, that in this set $\mathcal{F}_1$ of functions, there is one for which $F(v)$ is equal to its greatest lower bound λ_1 and by the usual methods of the Calculus of variations, he shows that this function v_1 is a solution of (2) for $\lambda = \lambda_1$.

He next considers the subset $\mathcal{F}_2$ of $\mathcal{F}_1$ defined by the additional condition

$$\iint_\Omega v(x,y)v_1(x,y)dxdy = 0, \tag{6}$$

takes the function $v_2 \in \mathcal{F}_2$ for which $F(v_2)$ is equal to its greatest lower bound λ_2 and shows that v_2 is a solution of (2) for $\lambda = \lambda_2$. The induction process is then obvious, and Weber concludes that he has proved the existence of an increasing infinite sequence (λ_n) of positive eigenvalues to each of which there corresponds an eigenfunction v_n normalized by condition (5), and orthogonal to each other. But he does not try to prove that $\lim_{n\to\infty} \lambda_n = +\infty$, nor that functions in $\mathcal{F}_1$ possess a "Fourier expansion" $\sum_n c_n v_n$ defined in the same manner as in the Sturm–Liouville problem (Chap. I, §3, formula (33)) (a result which he states however, without proof).

Weber's proofs were of course subject to the general criticisms of Weierstrass against the Calculus of variations, but no one seems to have tried to find more rigorous ones until 1885, In that year, H.A. Schwarz published a long paper on the theory of minimal surfaces, in which he had to consider a type of equation slightly more general than (2):

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \lambda^2 pv = 0 \tag{7}$$

where p is a continuous function in a domain D, with values $> O$; his arguments apply for any such function, but in fact he is only interested in the particular case $p(x,y) = -8/(1+x^2+y^2)^2$ [**195**].

Schwarz's paper is extremely remarkable by the originality of its methods, which do not seem to have been inspired by any previous work; it may be that the study of the Sturm–Liouville problem led him to arguments which later could be transferred almost verbatim to general integral equations with symmetric kernels (see Chap. V, §2), but there is no hint in his paper of such an influence, and in fact he quotes nobody, not even Weber.

His starting point is not the problem of existence of eigenvalues λ^2 for equation (7), but a "Dirichlet problem" for the equation

$$\Delta w + \xi p w = 0 \tag{8}$$

depending on a parameter ξ; he limits himself to the case where w is subject to the condition of being equal to 1 on the boundary Γ of D. Using the time honored method of representing the solution as a power series in ξ (Chap. II, §1)

$$w = w_0 + \xi w_1 + \ldots + \xi^n w_n + \ldots \tag{9}$$

he takes for w_0 the constant function equal to 1, and imposes on the w_n for $n \geq 1$ to vanish on Γ; they are then determined inductively by the equations

$$\Delta w_n + p w_{n-1} = 0 \quad \text{for } n \geq 1. \tag{10}$$

He assumes that the Green function $G(M,P)$ for the domain D exists (remember that he himself had proved that existence in extensive cases (Chap. II, §4)); the properties of that function implied that for any function f continuous in $\overline{D}$ the equation

$$\Delta w + f = 0 \tag{11}$$

has a unique solution vanishing on Γ, given by the formula

$$w(M) = \frac{1}{2\pi} \iint_D f(P)G(M,P)d\omega \qquad (\text{with } d\omega = dxdy) \tag{12}$$

Therefore his functions are given explicitly by

$$w_n(M) = \frac{1}{2\pi} \iint_D p(P)w_{n-1}(P)G(M,P)d\omega. \tag{13}$$

One must now investigate the convergence of (9) for small enough values of $|\xi|$, and it is here that Schwarz's original contributions begin. His main tool is the inequality named after him(*)

$$\left(\iint_D fgd\omega\right)^2 \leq \left(\iint_D f^2 d\omega\right)\left(\iint_D g^2 d\omega\right)$$

for any two functions f, g continuous in Γ; this gives from (13)

$$\begin{aligned} 4\pi^2 \left(w_n(M)\right)^2 &\leq \left(\iint_D p^2(P)G^2(M,P)d\omega\right)\left(\iint_D w_{n-1}^2(P)d\omega\right) \\ &\leq A\left(\iint_D w_{n-1}^2(P)d\omega\right) \end{aligned} \tag{14}$$

where A is a constant independent of n (due to the properties of the Green function of a bounded domain). Schwarz is thus led to study the numbers

$$W_{n,k} = \iint_D pw_k w_{n-k}d\omega \tag{15}$$

which, using the *symmetry* of the Green function, he shows are independent of k, so that $W_{n,k} = W_{n,0}$, which he writes W_n. He also proves that

$$W_{n,k} = \iint_D \left(\frac{\partial w_{k+1}}{\partial x}\frac{\partial w_{n-k}}{\partial x} + \frac{\partial w_{k+1}}{\partial y}\frac{\partial w_{n-k}}{\partial y}\right) dxdy. \tag{16}$$

Finally, using the Schwarz inequality, he obtains the relation

$$W_n^2 \leq W_{n-1}W_{n+1}, \tag{17}$$

hence the sequence of numbers W_n/W_{n-1} is increasing; on the other hand, integrating (14) gives $W_{2n} \leq BW_{2n-2}$ for a constant B independent of n, and therefore the limit of the sequence (W_n/W_{n-1}) is a finite number $c > 0$. It follows then from (14) that the series (9) is absolutely and uniformly convergent in $\overline{D}$ for $|\xi| < 1/\sqrt{c}$; the properties of the Green function enable one to show that the derivatives of w are also given by convergent series obtained by differentiating (9) termwise, and that w then satisfies (8) and is equal to 1 on the boundary Γ.

(*) That inequality had been discovered by Buniakowsky in 1859, but does not seem to have been noticed nor used by many mathematicians before 1885. If is of course a direct generalization of the corresponding inequality for finite sums, which goes back at least to Cauchy.

But Schwarz goes one step further. He proves that when $\xi = 1/\sqrt{c}$, the general term of the series (9) tends uniformly to a limit U_1 which is not identically 0 in $\overline{D}$ but vanishes on the boundary and is solution of

$$\Delta w + (1/c)pw = 0. \tag{18}$$

He has thus proved the existence of the *smallest eigenvalue* $\lambda_1^2 = 1/c$ of the equation (8) for functions vanishing on the boundary, and of the corresponding eigenfunction.

It should be observed here that these developments in fact are just another treatment of a "crypto-integral" equation (which Schwarz does not write, however). If one writes $w = w_0 + \xi v$ and "solves" equation (8) by formula (12) (using the same idea as Liouville in 1837 to obtain his "Volterra integral equation" (Chap. II, §1 and Chap. I, §3, equation (35))), one gets for v this time a "Fredholm integral equation"

$$v(M) = g(M) + \frac{\xi}{2\pi} \iint_D p(P)G(M,P)v(P)d\omega \tag{19}$$

with

$$g(M) = \frac{1}{2\pi} \iint_D G(M,P)p(P)d\omega.$$

Schwarz's procedure is therefore essentially the same as C. Neumann's for the Dirichlet problem (Chap. II, §4), at least as a starting point; the main difference is in the emphasis put by Schwarz on the dependence on the parameter ξ.

To appreciate the originalityand power of Schwarz's method, it is perhaps not superfluous to show how it can be translated, almost without change, in the theory of self-adjoint compact operators in a separable Hilbert space E. Suppose U is such an operator in E, which in addition we suppose *positive*, i.e. $(U \cdot f|f) \geq 0$ for all $f \in E$. The spectrum of U then consists in a decreasing sequence (finite or infinite) $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n \geq \ldots 0$, where each μ_n is an eigenvalue counted a number of times equal to its multiplicity; 0 is always in the spectrum but $\mathrm{Ker}(U)$ may be reduced to 0 or have infinite dimension; for each μ_n there is an eigenvector φ_n of norm 1, such that E is the *Hilbert sum* of the one-dimensional spaces $\mathbb{C}\varphi_n$ and of $\mathrm{Ker}(U)$. Let

$$w_0 = \sum_n d_n\varphi_n + w_0'$$

with $w_0' \in \mathrm{Ker}(U)$, be the expression of a vector $w_0 \in E$ for that decomposition. Then, for any $m \geq 1$, we have

$$U^m \cdot w_0 = \sum_n \mu_n^m d_n \varphi_n$$

and therefore the Schwarz series (9) is equal to

$$w = \sum_{m=0}^{\infty} \xi^m U^m \cdot w_0 = \sum_n \left(\sum_{m=0}^{\infty} \xi^m \mu_n^m \right) d_n\varphi_n = \sum_n \frac{d_n}{1-\xi\mu_n}\varphi_n$$

provided $|\xi| < \mu_1^{-1}$, and we have $w = \xi U \cdot w + w_0 - w_0'$. For $\xi = 1/\mu_1$,

$$\xi^m U^m \cdot w_0 = \sum_n \xi^m \mu_n^m d_n\varphi_n = \sum_n (\mu_n/\mu_1)^m d_m\varphi_n$$

tends to $d_1\varphi_1$ if μ_1 is a simple eigenvalue, to the sum of the $d_n\varphi_n$ such that $\mu_n = \mu_1$ in general. Finally, we have

$$W_m = (U^m \cdot w_0|w_0) = (U^{m-k} \cdot w_0|U^k \cdot w_0) = \sum_n \mu_n^m d_n^2$$

from which it follows that if the d_n such that $\mu_n = \mu_1$ are not all 0, the ratio W_m/W_{m-1} tends to μ_1; furthermore, one has

$$W_{2m} = \|U^m \cdot w_0\|^2 = |\,(U^{m+1} \cdot w_0 | U^{m-1} \cdot w_0)\,| \leq \|U^{m+1} \cdot w_0\| \cdot \|U^{m-1} \cdot w_0\| = (W_{2m-2} W_{2m+2})^{\frac{1}{2}}\,.$$

To get the inequality $W_m \leq (W_{2m-2} W_{2m+2})^{\frac{1}{2}}$ for all integers m, it is enough to consider the unique compact positive operator V such that $V^2 = U$, and apply to V the preceding argument. Of course the concept of the "square root" of a positive selfad joint operator was not available to Schwarz, and this is why he had to use the expression (16) for his numbers W_n.

In 1893, E. Picard published a short Comptes-Rendus Note ([**172**], vol. II, p. 545–550) in which he went one step further. For any point $M \in D$, the function $w(M, \xi)$ given by Schwarz's series (9) is holomorphic in the circle $|\xi| < \xi_1 = 1/\sqrt{c}$, and Picard investigates the *analytic continuation* of $\xi \mapsto w(M, \xi)$ beyond that circle: he shows that such a continuation exists in a circle $|\xi| < \xi_2$, of radius independent of $M \in D$, and that it has a *simple pole* with residue $-\xi_1 U_1(M)$ at the point ξ_1. He limits himself to the case in which $p = 1$ and Γ is convex and smooth, and his idea is to adapt the method of C. Neumann (Chap. II, §4) to evaluate the differences $\left|\xi_1^n w_n - \xi_1^{n-1} w_{n-1}\right|$; with apparently the same gap as in Neumann's argument (the details are not given in the Note) he "proves" that there are constants C and $q < 1$ independent of M, such that $\left|\xi_1^n w_n - \xi_1^{n-1} w_{n-1}\right| \leq C q^n$, hence $|\xi_1^n w_n - U_1| \leq C' q^n$ for another constant C', hence his result. writing

$$w = \frac{U}{1 - (\xi/\xi_1)} + v$$

he looks for a power series development

$$v = v_0 + \xi v_1 + \ldots \xi^n v_n + \ldots \tag{20}$$

similar to (9) but which should converge in a circle $|\xi| < \xi_2$ with $\xi_2 > \xi_1$. He determines the v_n by the successive approximations

$$\Delta v_0 - \xi_1 U_1 = 0, \qquad \Delta v_n + v_{n-1} = 0 \quad \text{for } n \geq 1$$

with the boundary conditions: $v_0 = 1$ on Γ and $v_n = 0$ on Γ for $n \geq 1$. Introducing numbers similar to the $W_{n,k}$ of Schwarz, he is able to prove that the radius of convergence ξ_2 of (20) is finite, but he cannot show that there is an eigenfunction corresponding to ξ_2 and vanishing on Γ.

§2. The contributions of Poincaré.

In 1890, H. Poincaré published in the American Journal of Mathematics a long paper developing some of his research done since 1887, which had been announced in three Comptes-Rendus Notes ([**177**], vol. IX, p.15–113). The paper consists of two completely independent parts; in the first, he describes in detail his "sweeping-out" method for the solution of the Dirichlet problem (Chap. II, §4). The second part is devoted to the *cooling off* problem in the theory of heat, which had been treated by Fourier in some particular cases, for instance the cooling off of a sphere when the temperature is a function of the distance to the center (Chap. I, §2). The *general* cooling off problem had been presented by Fourier in the following form: given a solid body V of constant density, isotropic for the propagation of radiations, one has to find the temperature $u(x, y, z, t)$ inside V, as a function of the coordinates x, y, z and the time t, when the outside temperature

is 0. Fourier shows that the function u must satisfy inside V an equation (where a is constant)

$$\frac{\partial u}{\partial t} = a^2 \Delta u \tag{21}$$

and in addition is subject to the boundary condition on the surface Σ of V

$$\frac{\partial u}{\partial n} + hu = 0 \tag{22}$$

where $\frac{\partial u}{\partial n}$ is the normal derivative (towards the exterior) and h is a constant ≥ 0 (see Chap. IV, §4). The usual method of "separation of wvariables" led to solutions of the form $u(x, y, z, t) = e^{-\lambda a^2 t} v(x, y, z)$, where v should be a solution of the Helmholtz equation

$$\Delta v + \lambda v = 0 \tag{23}$$

with a *different* boundary condition from the one deriving from the equation of vibrating membranes, namely

$$\frac{\partial v}{\partial n} + hv = 0 \quad \text{on } \Sigma. \tag{24}$$

In his 1869 paper, H. Weber had also cousidered that problem, but he had only described his variational method to obtain eigenvalues and eigenfunctions for the particular case $h = 0$. Poincaré apparently was unaware of Weber's paper and never mentioned it in his own work; what he does in 1890 is first to repeat Weber's arguments for the general boundary condition (24), replacing the Dirichlet integral by the function

$$F(v) = h \iint_{\Sigma} v^2 d\sigma + \iiint_{V} \left(\left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2 \right) d\omega. \tag{25}$$

Having thus obtained an increasing infinite sequence (λ_n) of eigenvalues and the corresponding sequence (v_n) of eigenfunctions, Poincaré is of course aware of the non rigorous character of his "proof"; however, having for the time being no better arguments at his disposal, he takes for granted the existence of λ_n and v_n and proceeds to study them in more detail, and in the first place to prove that the sequence (λ_n) tends to $+\infty$, a question which Weber had not been able to answer. In his attack on that problem, it is quite remarkable to see Poincaré introducing a whole batch of *completely new ideas*. In the first place, he considers the eigenvalues as functions $\lambda_n(h, V)$ of the constant h and the domain V and begins to study the way in which they *depend* on h and V, a trend of thought which will later blossom in the work of H. Weyl and R. Courant, and even now has not entirely lost its interest. Poincaré first shows that, for V fixed, $\lambda_n(h, V)$ is increasing with h, by an application of Green's formula to the eigenfunctions $v_n(h, V), v_n(h', V)$ corresponding to two values of h; as he wants to prove that $\lambda_n(h, V)$ tends to $+\infty$, he can assume that $h = 0$, which implies that $\lambda_1(0, V) = 0$ and $v_1(0, V)$ is a constant.

The second idea is to *decompose* V into a union of smaller solids $V_1, V_2, \ldots, V_p$; the variational definition of λ_n enables him to prove that if $p \leq n-1$, $\lambda_n(0, V)$ is *at least equal* to the smallest of the numbers $\lambda_2(0, V_1), \ldots, \lambda_2(0, V_p)$. Poincaré is thus led to *minorize* $\lambda_2(0, V)$ by a number depending only on the geometry of V; by definition (since $h = 0$), this means finding a lower bound of the expression

$$\frac{\iiint_V \left(\left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2 \right) d\omega}{\iiint_V v^2 d\omega} \tag{26}$$

where v is a C^2 function in $\overline{V}$, subject to the condition

$$\iiint_V v\,d\omega = 0. \tag{27}$$

He assumes V is *convex*; using polar coordinates and the standard methods of the Calculus of variations, he obtains as lower bound

$$C \cdot \mathrm{vol}(V)/(\mathrm{diam}(V))^5 \tag{28}$$

where C is an absolute constant; one should here stress the fact that this Poincaré inequality is the first example of what we now call *"a priori" inequalities* (cf. Chap. IX, §4). Returning to the minoration of $\lambda_n(0, V)$, he takes $p = n-1$, assumes that V can be decomposed in $n-1$ solids V_j which are convex and have a diameter tending to 0 with $1/n$ and such that the ratio of their volume to the fifth power of their diameter tends to $+\infty$ with n; this gives him his conclusion.

Poincaré's next step is to investigate how the knowledge of the λ_n and v_n gives the solution of the cooling off problem, when the temperature $u(x, y, z, 0)$ is a known function $f(x, y, z)$ in V at time $t = 0$. Fourier's method consists in writing

$$u(x, y, z, t) = \sum_{n=1}^{\infty} c_n \exp\left(-\lambda_n a^2 t\right) v_n(x, y, z) \tag{29}$$

which gives for the unknown coefficients c_n the condition that $f = \sum_{n=1}^{\infty} c_n v_n$, hence, from the orthogonality relations, $c_n = \iiint_V f v_n d\omega$. But Poincaré, no more than Weber, is not at that time able to prove that this Fourier expansion converges to the function f in V. However, taking his cue from Tchebychef's results in approximation theory, he shows (by a clever use of Green's formula) that the integral

$$S_n = \iiint_V \left(u - \sum_{k=1}^{n} c_k \exp\left(-\lambda_k a^2 t\right) v_k \right)^2 d\omega$$

satisfies an inequality $S_n \leq C \cdot \exp\left(-\lambda_{n+1} a^2 t\right)$ where C is independent of n and t; in other words, for $t > O$, he proves the convergence of the series in (29) in what we now call the topology of Hilbert space(*).

(*) The method of least squares of Legendre–Gauss had led Tchebychef to define a "best approximation" to a function F, by a linear combination $\sum_{j=1}^{N} a_j \psi_j$ of given functions $\psi_j (1 \leq j \leq N)$, by the condition that

$$\sum_{k=1}^{n} \rho(x_k) \left(F(x_k) - \sum_{j=1}^{N} a_j \psi_j(x_k) \right)^2$$

be minimum, for given points $x_k (1 \leq k \leq n)$ and given "weight" ρ. Gram, in 1883, generalized the problem by considering instead of a finite sum, an integral

$$\int_a^b \rho(x) \left(F(x) - \sum_{j=1}^{N} a_j \psi_j(x) \right)^2 dx \tag{+}$$

The final section of Poincaré's paper (if we except a kind of postscript which we will discuss later in Chapter IV, §4) is devoted to the general study of the eigenfunctions v_n (their existence being admitted). In general, if v satis fies (23) and (24), use of Green's formula shows that there is a formula similar to Green's expression of the potential (Chap. II, §3, formula (17))

$$-4\pi v(M) = \iint_{\Sigma} v\left(\frac{\partial T}{\partial n} + hT\right) d\sigma \tag{30}$$

where T (replacing the function $1/r$) is now $\exp(i/\sqrt{\lambda}r)/r$. Using that formula, he is able to show, after a rather long discussion (patterned on the study of double layer potentials but more difficult), that v is continuous in $\overline{V}$, and to obtain bounds for its derivatives in V.

The second paper devoted by Poincaré to the equation of vibrating membranes ([**177**], vol. IX, p. 123–196) is even more original. It is likely that in 1890, he was not aware of Schwarz's paper of 1885. The publication of Picard's note in 1893 immediately attracted his attention, and in a few months he had seen that by combining Schwarz's method and his "a priori" inequality of 1890, he could go beyond Picard and prove the analytic continuation of the function $\xi \mapsto w(M,\xi)$ as a *meromorphic function in the whole complex plane*, obtaining at the same time the existence of the long sought eigenvalues and eigenfunctions for the Helmholtz equation (with the same boundary condition as Schwarz).

Poincaré starts with a simplification and an impvrovement of his inequality for the expression (26); using Schwarz's inequality, he is able to replace his lower bound (28) by $C/(\mathrm{diam}(V))^2$ for a convex solid V. He then only assumes that for a general solid V it is possible to decompose it in convex solids having arbitrary small diameters, and uses this idea of decomposition to prove the following crucial lemma: given p arbitrary C^2 functions $F_1, F_2, \ldots, F_p$ in $\overline{V}$, it is possible to choose p numbers $\alpha_1, \ldots, \alpha_p$ in such a way that, for $v = \alpha_1 F_1 + \ldots \alpha_p F_p$, one has $\iiint_V v d\omega = 0$ and the ratio (26) is *at least* L_p, where L_p is a number which only depends on V and p (and not on the F_j) and *tends to* $+\infty$ with p. This is simply done by decomposing V in the union of $p-1$ convex subsets V_j, and choosing the coefficients α_j by the $p-1$ conditions $\iiint_{V_j} v d\omega = 0$ $(1 \leq j \leq p-1)$.

Poincaré, as Picard, limits himself to the case in which the function p in equation (8) is the constant 1, but considers a problem which slightly generalizes Schwarz's, namely he looks for a function v solution of

$$\Delta v + \xi v + f = 0 \tag{31}$$

and vanishing on the boundary Σ, with f an *arbitrary* C^∞ function (if in Schwarz's equation (8) with $p = 1$, one writes $w = w_0 + \xi v$, the equation for v is (31) with $f = w_0$);

---

and he solved the problem in an original way, by applying to the ψ_j the "orthogonalization process" usually attributed to E. Schmidt [**089**]. He was thus reduced to the case in which the ψ_j form an orthonormal system (for the measure ρdx), where he showed that the a_j giving the best approximation a are the "Fourier coefficients" $\int_a^b \rho(x)F(x)\psi_j(x)dx$. He went on to consider an *infinite* orthonormal system (ψ_n) and investigated under which conditions the minimum value μ_n of the integral (+) tends to 0 when n increases to $+\infty$; he was able to see that this was linked to the "completeness" of the system (ψ_n), i.e. the fact that no function other than the constant 0 is orthogonal to all ψ_n. It is unlikely that Poincaré had any knowledge of Gram's paper.

he will make a very clever use of this arbitrariness. He starts by observing that Schwarz's method works just as well for arbitrary f as for $f = 1$, and proves the existence of the solution of (31) vanishing on Σ for *small enough* $|\xi|$; he writes it

$$v = [f, \xi] = v_0 + \xi v_1 + ... + \xi^n v_n + ...$$

with $\Delta v_0 + f = 0, \Delta v_n + v_{n-1} = 0$ for $n \geq 1$, the v_n all vanishing on Σ.

For any given integer p, he introduces p arbitrary coefficients $\alpha_1, ..., \alpha_p$ and forms the function (defined at least for small ξ)

$$w = [\alpha_1 f + \alpha_2 v_0 + ... + \alpha_p v_{p-2}, \xi] = w_0 + w_1 \xi + ... + w_n \xi^n + ... \tag{32}$$

Next, applying his lemma for the evaluation of the Schwarz integrals W_n *corresponding to* w, he is able to show that, for a suitable *choice* of the α_j the series (32) *converges for* $|\xi| \leq L_p$ (uniformly in $\overline{V}$). But if one writes $u_j = \left[v_{j-2}, \xi\right]$, one has

$$\begin{cases} \alpha_1 v + \alpha_2 u_2 + ... \alpha_p u_p = w \\ v - u_2 \xi = v_0 \\ u_2 - u_3 \xi = v_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ u_{p-1} - u_p \xi = v_{p-2} \end{cases} \tag{33}$$

a linear system from which Cramer's formulas give

$$v = P/D \tag{34}$$

with

$$D = \begin{vmatrix} \alpha_1 & \alpha_2 & \alpha_3 & ... & \alpha_{p-1} & \alpha_p \\ 1 & -\xi & 0 & ... & 0 & 0 \\ 0 & 1 & -\xi & ... & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & & 1 & -\xi \end{vmatrix} = \alpha_p - \alpha_{p-1}\xi + \alpha_{p-2}\xi^2 - ... + (-1)^{p-1}\alpha_1 \xi^p$$

and

$$D = \begin{vmatrix} w & \alpha_2 & \alpha_3 & ... & \alpha_{p-1} & \alpha_p \\ v_0 & -\xi & 0 & ... & 0 & 0 \\ v_1 & 1 & -\xi & ... & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ v_{p-2} & 0 & 0 & ... & 1 & -\xi \end{vmatrix}$$

which shows that P, as w, is equal to a series

$$P = P_0 + P_1 \xi + ... + P_n \xi^n + ... \tag{35}$$

where the P_n are C^∞ functions in $\overline{V}$ vanishing on Σ, the series being uniformly convergent in $\overline{V}$ for $|\xi| < L_p$, and all derivatives of P (with respect to ξ or to x, y, z) being obtained by derivating termwise the series. This shows that $\xi \mapsto v(M, \xi)$ extends to a *meromorphic* function in $|\xi| < L_p$ with a finite anumber of poles, which are roots of $D(\xi) = 0$ and *independent* of M; it is easy to show, using Green's formula, that these poles are all *simple*. As this is true for any p, $(\xi \mapsto v(M, \xi)$ extends to a meromorphic function in the

whole complex plane, with simple real and positive poles independent of M; furthermore, for each one of these poles λ_n, the function $P(M, \lambda_n)$ satisfies $\Delta P + \lambda_n P = 0$; in other words, one has found for each λ_n an eigenfunction u_n corresponding to that eigenvalue. In addition, Poincaré's *a priori* inequality enables him to show that $\lambda_n \geq c \cdot n^{2/3}$, where c is a constant,

The remainder of Poincaré's 1894 paper is devoted to two questions:

A) In the last 4 sections of the paper, he takes up again the problem of Fourier expansions (when the boundary condition is $v = 0$). Attaching to the function f its "Fourier coefficients" $c_n = \iiint_V f u_n d\omega$ (where the eigenfunctions u_n have been normalized by $\iiint_V u_n^2 d\omega = 1$), he first deduces from the relations $\Delta u_n + \lambda_n u_n = 0$ and Schwarz's inequality, that $|u_n| \leq A\lambda_n$ in $\overline{V}$ (A constant), and that the c_n are uniformly bounded. From that it follows that for ξ different from the eigenvalues the unique solution of (31) vanishing on Σ is given by the absolutely and uniformly convergent series

$$v = -\sum_n \frac{c_n u_n \xi^2}{\lambda_n^2 (\xi - \lambda_n)} + v_0 + v_1 \xi; \tag{36}$$

in addition, Poincaré shows that if the series $\sum_n c_n u_n$ is absolutely convergent, its sum is equal to f; he cannot prove that for "arbitrary" functions f (probably at least C^4), vanishing on Σ, the series converges, but he proves absolute convergence when in addition Δf and $\Delta^2 f$ also vanish on Σ.

B) Before returning to the question of Fourier expansions, Poincaré had tried to extend his results on the existence of eigenvalues and eigenfunctions for the boundary condition (24) of the cooling off problem. He realizes that Schwarz's method would work, and therefore also his own existence theorem, *provided* one could prove the existence of a "Green function" for the Laplace equation with that new boundary condition, i.e. a function $G(M, P)$ having the same properties as the usual Green function, with the exception that, for $M \in V, P \mapsto G(M, P)$ satisfies (24) on the boundary.

In the special case $h = 0$, C. Neumann, in his work on the Dirichlet problem, had shown how to obtain such a "Green function" (also named "Neumann function") when V is convex and not a double cone. He had observed that by changing the sign before the integral in the Beer–Neumann equation (Chap. II, §4, formula (23)), the solution of that new equation gave a density ρ such that the corresponding double layer potential, in the *exterior* of V (complement of $\overline{V}$) is harmonic, tends to 0 at infinity and to $-g$ on the boundary Σ (a solution to what is called the "exterior Dirichlet problem"). From this result, he had shown how to obtain a solution of what is now called the *Neumann problem* for the Laplace equation: find in V a harmonic function u such that u is continuous in $\overline{V}$ and has on Σ a normal derivative $\frac{\partial n}{\partial u}$ equal to a *given* continuous function g; a necessary condition for the existence of the solution (deduced from Green's formula applied to u and the constant 1) is that $\iint_\Sigma g d\sigma = 0$. Neumann proves that this condition is sufficient (the solutions being determined up to an additive constant): he considers the *simple layer* potential w defined by the density $\frac{1}{4\pi} g$; it is continuous on Σ and its normal derivative jumps by $-g$ when crossing Σ from the interior to the exterior. Neumann next takes the *double layer* potential v, solution of the *exterior* Dirichlet problem which tends to $-w$ on Σ. Then the function $u = v + w$ is harmonic outside Σ, and 0 in the exterior of V; as the normal derivative of v is the same on both sides of Σ, it follows at once that $\frac{\partial u}{\partial n}$ tends to g from Σ from the interior of V, and therefore solves the Neumann problem.

C. Neumann had not been able to solve the corresponding problem when the boundary condition is $\frac{\partial u}{\partial n} + hu = g$ for a an constant $h > 0$. Poincaré tried to solve the problem by representing u as a power series in h, $u = u_0 + hu_1 + \ldots + h^n u_n + \ldots$, and was indeed

able to obtain in that way (using Neumann's results) a series convergent for all $h \geq O$, uniformly in $\overline{V}$; however, for the first derivatives, his method could omnly prove uniform convergence in compact subsets of V, so that it was impossible to give meaning to $\frac{\partial u}{\partial n}$ on the boundary Σ, and to show that u was indeed a solution of the problem. The most interesting result in this attempt is that Poincaré, probably for the first time in history, arrives at the idea of "weak" solution of a boundary problem; he shows that his function u is such that, for *any* function v which is C^2 in $\overline{V}$, one has

$$\iiint_V u\Delta v d\omega + \iint_\Sigma gv d\sigma = \iint_\Sigma \left(\frac{\partial v}{\partial n} + hv\right) u d\sigma \tag{37}$$

and adds that "physically" this is equivalent to a genuine solution.

The last of the three long papers of Poincaré on partial differential equations was witten in 1895 ([**177**], vol. IX, p. 202–272). Although it is the one which contains the smallest number of new results, it probably had a greater influence than the others. From his work both on the Dirichlet problem and on the equation of vibrating membranes, Poincaré had become convinced that there were also "eigenvalues" and "eigenfunctions" linked to the Dirichlet problem. For us this is completely obvious, for if we look for a solution of $\Delta u = 0$ taking given values on the boundary Σ of V, we extend the function g given on Σ to a C^2 function h in $\overline{V}$ (when this is possible); replacing u by $v = u - h$, we have to find a solution of $\Delta v + f = O$, with $f = \Delta h$, which vanishes on Σ, and this is just the special case of Schwarz's problem for the equation (31) with $\xi = 0$.

At that time, however, nobody had yet thought of this simple argument[(*)], and Poincaré's reasoning is quite different and much more circuitous. He observes that one can formulate both the interior and exterior Dirichlet problems as special cases of the problem which consists in finding a double layer poten tial W (for a density on Σ) such that, for $s \in \Sigma$,

$$W(s^-) - W(s^+) - \lambda\,(W(s^-) + W(s^+)) = 2\Phi(s) \tag{38}$$

where $W(s^-)$ is the limit of W at s along the interior normal, $W(s^+)$ its limit along the exterior normal, λ is a complex parameter and Φ a given function on Σ; the values $\lambda = 1$ and $\lambda = -1$ correspond respectively to the interior and the exterior Dirichlet problem. To this general problem Poincaré associates a new variational problem: for any simple layer potential Ψ defined by a density on Σ, he considers the ratio J/J', where J is the Dirichlet integral $\iiint (\operatorname{grad}\Psi)^2 d\omega$ extended over V, and J' the integral of the same function, extended to the exterior of V. The usual non rigorous arguments lead him to conjecture: 1° the existence of an increasing sequence $0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_n \leq \ldots$ of eigenvalues, and: 2° for each λ_i, the existence of a simple layer potential Φ_i, such that, on Σ,

$$\frac{\partial \Phi_i}{\partial n}(s^-) + \lambda_i \frac{\partial \Phi_i}{\partial n}(s^+) = 0; \tag{39}$$

in addition, for $i \neq j$, $\iiint_V \operatorname{grad}(\Phi_i) \cdot \operatorname{grad}(\Phi_j) d\omega = 0$. Normalizing the Φ_i by $\iiint_V (\operatorname{grad}\Phi_i)^2 d\omega = 1$, he assumes that there is a Fourier expansion $\Phi = \sum_i c_i \Phi_i$ of

---

(*) It is explicitly mentioned in 1909 by E.E. Levi [**145**]; the first statement and proof of the existence of a continuous function in the whole space $\mathbb{R}^3$ extending a given function defined and continuous in a closed subset (*i.e.* what we now call the Tietze–Urysohn theorem) is due to Lebesgue in 1907 [**138**].

the given function Φ on Σ, and "solves" the equation (39) by

$$W(s^-) = \sum_i A_i \Phi_i(s), \qquad W(s^+) = -\sum_i \lambda_i A_i \Phi_i(s)$$

with $A_i = 2c_i/(1 + \lambda_i - \lambda(1 - \lambda_i))$.

All this is of course presented by Poincaré as purely conjectural, and as a motivation for his detailed study (by methods inspired by those of Schwarz) of the ratio J/J', which forms the central part of his 1896 paper; but the only positive result he is able to deduce from his study is that the Beer–Neumann series (Chap. II, §4) converges, not only for convex domains V, but also for domains $V \subset \mathbb{R}^3$ having the following property: when $\mathbb{R}^3$ is imbedded in the 3-dimensional sphere $\mathbb{S}_3$ by adjoining a point at infinity, V can be transformed into a ball by a homeomorphism of $\mathbb{S}_3$ onto itself, leaving fixed the point at infinity, and which is C^2 in $\mathbb{S}_3$ as well as the inverse homeomorphism(*).

Almost immediately after the publication of Poincaré's papers, several mathematicians were able to complete and extend his results. In 1898, E. Le Roy [**144**] proved the existence of the simple layer potentials Φ_i conjectured by Poincaré in his 1896 paper; he replaced the ratio J/J' by $(J+J')/I$, where I is the surface integral $\iint_\Sigma \rho^2 d\sigma$, ρ being the density on Σ corresponding to the simple layer potential Ψ, and adapted the methods of Schwarz and Poincaré to the correspond ing variational problem. In 1899, S. Zaremba [**232**], by a modification of the method of solution of Neumann's problem used by Poincaré, could complete the latter's solution of the "cooling off" problem, proving that the "weak" solution of Poincaré was a genuine one. In 1901, Zaremba and W. Stekloff, independently, finally showed that one could drop the global topological property of the domain V which Poincaré and Le Roy had used, and even weaken the "smoothness" conditions on Σ; they made essential use of a paper of Liapounoff published 3 years earlier [**147**], in which he was able to prove the existence of the normal derivative on Σ of the solution of Dirichlet's problem under these less stringent conditions.

(*) Without the slightest justification, Poincaré claims as "clear" the fact that this property holds for *any* bounded domain V such that the boundary Σ is a smooth simply connected surface ([**177**], vol. IX, p. 223–224). With the tools of modern Differential Topology, it is now possible to prove that theorem. But in 1895, Poincaré was just beginning to formulate the first notions of that theory, and one wonders if he realized the difficulties which lay in the way of a rigorous proof if he had tried to write it down (when smoothness conditions on Σ and on the homeomorphism are dropped, the result is known to be false, a counterexample being the famous "Alexander horned sphere").

Chapter IV
The idea of infinite dimension

§1. Linear algebra in the XIX$^{\text{th}}$ century.

I think that in order to understand the trend of ideas which led to Functional Analysis, it is useful to summarize the evolution of linear algebra during the XIX$^{\text{th}}$ century. Until around 1830, it had consisted in the study of systems of linear equations in any number of variables, with real or complex coefficients, most of the times limited to the case in which the number of equations was equal to the number of variables; the Cramer formulas gave the unique solution when the deter minant of the system was not 0, but not much effort was spent on the elucidation of the other cases; the only result which was used occasionally was the fact that a system of m homogeneous equations in $n > m$ variables always had a non trivial solution (obvious by induction on m).

Linear changes of variables

$$y_j = \sum_{k=1}^{n} a_{jk} x_k \quad (1 \leq j \leq m) \tag{1}$$

had been familiar since the XVITI$^{\text{th}}$ century (mostly for $m = n \leq 3$). It naturally led to computations done, not on numbers, but on *rectangular arrays* (a_{jk}) of numbers, which "represented" these changes of variables. Beginning with Gauss, this trend was systematized in the 1850's by Sylvester and Cayley in the theory of *matrices*.

Ever since the invention of cartesian coordinates ("analytic geometry", as it came to be called in the XVIII$^{\text{th}}$ century), mathematicians had known how to interpret geometrically computations on systems of 2 or 3 variables, and many had envisioned the possibility of similarly interpreting computations on systems of any number n of vaviables in a "geometry in n dimensions", which however would be devoid of "reality". After 1840, mainly under the influence of Hamilton and Cayley, this geometrical language was gradually adopted by more and more mathematicians, and had become commonplace at the end of the century. But in the XIX$^{\text{th}}$ century, after 1822 "geometry" essentially meant *projective* geometry, and most "geometric" interpretations of computations were done, not in the vector space $\mathbb{R}^n$ or $\mathbb{C}^n$, but in the complex projective spaces $\mathbb{P}_n(\mathbb{C})$; for instance, the relations (1) for $m = n$ were interpreted as defining also a *projective transformation* in $\mathbb{P}_{n-1}(\mathbb{C})$ sending the point of homogeneous coordinates (x_j) to the point of homogeneous coordinates (y_j), and the efforts of Grassmann and Peano to introduce vector spaces in an axiomatic way were persistently ignored until 1900.

Between 1850 and 1880 are proved the main theorems of linear algebra, concerning what are called the "reductions" of square matrices. One of these is the problem of finding, for agiven square matrix U, an invertible matrix P such that PUP^{-1} has a "reduced" unique canonical form, which here (for complex matrices U) means a diagonal array of Jordan matrices; this is the way Jordan himself treats the problem, improving on a previous result of Grassmann, who had proved the existence of a "reduced" triangular matrix PUP^{-1} for any U (using already the intrinsic notion of endomorphism instead of the notion of square matrix).

Unfortunately, another type of "reduction" interfered with the preceding one. To a quadratic form $\sum_{j\leq k} a_{jk}x_jx_k$ corresponds the *symmetric* matrix $(a_{jk}) = U$, and it was well known since Cauchy that if U is *real* it is possible to find an in vertible real matrix P (which may even be supposed to be *orthogonal*) such that PUP^{-1} would be a (real) *diagonal* matrix; this is equivalent to finding an orthogonal change of variables for which the quadratic form became equal to a linear combination $\sum_{j=1}^{n} \lambda_j y_j^2$ of squares, the λ_j being the elements of the diiginal matrix PUP^{-1}, or equivalently the roots (with their multiplicity) of the "characteristic equation" $\det(U - \lambda I) = 0$. Weierstrass, who was the first to find the "Jordan normal form" of a square complex matrix (which Jordan only discovered independently 2 years later(*)), presented it as a generalization of the "reduction" of a quadratic form, by considering a *bilinear form* $\sum_{j,k} a_{jk}x_jy_k$ (with $U = (a_{jk})$ an *arbitrary* square matrix) and applying to the x_j and y_j two "contragredient" changes of variables, i.e. such that the bilinear form $\sum_{j,k} x_jy_k$ remains invariant; this amounts to replacing U by a matrix PUP^{-1}. When, in 1878, Frobenius gave a systematic account of these results ([**078**], vol. I, p. 343–405), he deliberately abandoned the language of matrices in favor of the language of bilinear forms, defining the "product" (*Faltung*) of two bilinear forms $A(x, y), B(x, y)$, as $\sum_{k=1}^{n} \frac{\partial A}{\partial y_k}\frac{\partial B}{\partial x_k}$![(*)]

Finally, the concept of duality in *vector spaces* was completely foreign to mathematicians until 1900. Duality was well understood in the realm of *projective geometry* (it had been one of the big discoveries of the early XIX$^{\text{th}}$ century), as a bijection of points on planes (in projective space of 3 dimensions) and later as a bijection of points on hyperplanes in any number of dimensions. But linear forms were identified with the systems of their coefficients, "vectors" and "forms" being thus both "n-tuples" of numbers, which one had to distinguish, according to the way they behaved under changes of variables, by the awkward concepts of "contragredient" and "cogredient" systems. This identification of a vector space and its dual was reverberated in the identification of endomorphisms with bilinear forms, mentioned above (*).

To sum up, at the end of the XIX$^{\text{th}}$ century, the main results of linear and multilinear Algebra had been found but were expressed through insufficiently clarified notions. They could therefore be of no help to the generalizations of linear Algebra to infinite dimensional spaces which were called forth by the development of Functional Analysis; these had to go through the same painful stages, first linear equations, then determinants, later bilinear forms, matrices, and only at the very end vector spaces and linear maps; in other words, the *historical evolution*, just as for finite dimensional linear algebra, was exactly in the reverse order of what we not consider to be the *logical* order!

(*) Jordan was not dealing with matrices having elements in $\mathbb{R}$ or $\mathbb{C}$, but with matrices having elements in a *finite field* ([**123**], p. 114–126).

(*) In 1896, Pincherle reinterpreted Weierstrass's results in terms of endomorphisms ([**173**], vol. I, p. 358–367).

(*) In modevrn linear algebra, the space of endomorphisms of a finite dimensional vector space E is identified with the tensor product $E^* \otimes E$, whereas the space of bilinear forms on $E \times E$ is identified with $E^* \otimes E^*$.

§2. Infinite determinants.

The first appearance of infinite systems of linear equations in infinitely many unknowns seems to occur in Fourier's work on the theory of heat. He has to determine an infinite sequence $(a_m)_{m\geq 1}$ of coefficients such that the relation

$$1 = \sum_{m=1}^{\infty} a_m \cos(2m-1)y \tag{2}$$

holds for all y ([**067**], vol. I, p. 149). Fourier's idea is to take derivatives of all orders of both sides of (2) and identify them for $y = 0$, which gives him the infinite system of linear equations for the a_m

$$\begin{cases} 1 = \sum_{m=1}^{\infty} a_m \\ 0 = \sum_{m=1}^{\infty} (2m-1)^2 a_m \\ 0 = \sum_{m=1}^{\infty} (am-1)^4 a_m \\ \dots\dots\dots\dots\dots \end{cases} \tag{3}$$

To solve it, he considers the first k equations where he replaces the a_m for $m > k$ by 0; he then solves that system by Cramer's formulas, which give him a system of k numbers $a_1^{(k)}, a_2^{(k)}, \dots, a_k^{(k)}$, and lets k tend to infinity in each expression of $a_m^{(k)}$ for fixed m. Using the formulas giving Vandermonde determinants, he obtains

$$a_1^{(k)} = \frac{3^2 \cdot 5^2 \dots (2k-1)^2}{8 \cdot 24 \dots (4k^2-4k)}$$

tending to $a_1 = 4/\pi$, and

$$\frac{a_{m+1}^{(k)}}{a_m^{(k)}} = \frac{2m-1}{2m+1} \cdot \frac{m+k}{m-k}$$

which gives him $a_m = (-1)^{m-1} 4/\pi(2m-1)$; when later in his book he proves the general formula giving the Fourier coefficients, he can of course check that these values of the a_m are correct. But he never bothered to give any justification of his procedure, where all questions of convergence are completely disregarded; that procedure could of course be repeated for any infinite system

$$\sum_{k=1}^{\infty} a_{jk} x_k = b_j \qquad (j = 1, 2, \dots) \tag{4}$$

but nobody undertook to justify it before 1885(*). In that year, P. Appell met such a system with $a_{jk} = a_k^j$ for a given sequence (a_k), in a question relative to elliptic functions, and used the same method as Fourier; his paper attracted Poincaré's attention, and he showed that for such a "generalized Vandermonde system", the procedure was

(*) During that period, Fourier's method was used in two little-known papers, one by Fürstenau in 1860 on the computation of roots of an algebraic equation, and another by Kötteritzsch in 1870, for a system (4) in which the a_{jk} are 0 for $j > k$ (see [**184**], p. 8–12).

justified provided the infinite product $F(z) = \prod_{k=1}^{\infty} \left(1 - \frac{z}{a_k}\right)$ was convergent for all complex numbers z.

The next year, he returned to the subject, in relation with a paper published in 1877 by the American astronomer and mathematician G.W. Hill on the lunar theory [**114**]. Hill proposed a new approach which rested on the integration of a second order differential equation

$$w'' + \left(\sum_{n=-\infty}^{+\infty} \theta_n e^{nit} \right) w = 0 \tag{5}$$

where the θ_n are constants, and one looks for a solution of period 2π. Hill writes such a solution as a trigonometric series

$$w = \sum_{n=-\infty}^{+\infty} b_n e^{i(n+c)t} \tag{6}$$

and substituting in (5), obtains for the coefficients b_n the infinite system of equations

$$\sum_{k=-\infty}^{+\infty} \theta_{n-k} b_k - (n+c)^2 b_n = 0, \qquad -\infty < n < +\infty. \tag{7}$$

He probably was unaware of Fourier's procedure, but used a similar one, keeping this time the equations (7) for $-p \leq n \leq p$, replacing in these equations the b_m by 0 for $m < -p$ or $m > p$, and letting p tend to infinity in the solutions of the system thus obtained.

Poincaré considers a general system (4), where he supposes that $a_{jj} = 1$ for all j (one can always reduce (4) to such a system by dividing the j-th equation by a_{jj}, when $a_{jj} \neq 0$). His idea is to compare the determinant $D_n = \det(a_{jk})_{1 \leq j,k \leq n}$ to the product $P_n = \prod_{j=1}^{n} \left(\sum_{k=1}^{n} |a_{jk}| \right)$. It is clear from the definition of a determinant that $|D_n| \leq P_n$, and from the assumption on the diagonal terms, one has also $|D_m - D_n| \leq P_m - P_n$; this inequality immediately gives Poincaré's *sufficient* condition for the existence of $D = \lim_{n \to \infty} D_n$, namely that the double sum $\sum_{j \neq k} |a_{jk}|$ be *finite*. Furthermore, Poincaré shows that, when the k-th column of D is replaced by a sequence (b_j) which is *bounded*, there is still convergence for the new "infinite determinant", and that there is a unique bounded solution x_k of (4) given by the usual Cramer formulas (with "infinite determinants" of course).(*) Finally he extends his results to doubly infinite systems

$$\sum_{k=-\infty}^{+\infty} a_{jk} x_k = b_j \qquad (-\infty < j < +\infty) \tag{8}$$

with the same restriction $a_{jj} = 1$ for all j; in particular he shows that Hill's method is justified for the system (7) ([**177**], vol. V, p. 95–107).

(*) One must beware of the fact that the Fourier method (when no condition is imposed on the a_{jk}) may very well give convergent "infinite determinants", but the values given by the Cramer formulas may be such that the left hand sides of (4) are *divergent* series. An example is given by taking $a_{jk} = 0$ if $j > k$, $a_{jk} = 1$ if $j \leq k$, $b_j = (-1)^j$; one finds as a "solution" $x_k = 2(-1)^k$.

Ten years later, H. von Koch [**220**] refined and generalized Poincaré's results. Insteadof making assumptions on the diagonal terms, he writes the coefficients $\delta_{jk} + c_{jk}$ instead of a_{jk} (with the Kronecker delta), and uses the expression of a determinant $\Delta_n = \det(\delta_{jk} + c_{jk})_{1\leq j,k\leq n}$ as a sum of principal minors

$$\Delta_n = 1 + \sum_{s=1}^{n} c_{ss} + \frac{1}{2!}\sum_{s_1,s_2}\begin{vmatrix} c_{s_1s_1} & c_{s_1s_2} \\ c_{s_2s_1} & c_{s_2s_2} \end{vmatrix} + \tag{9}$$
$$+\frac{1}{3!}\sum_{s_1,s_2,s_3}\begin{vmatrix} c_{s_1s_1} & c_{s_1s_2} & c_{s_1s_3} \\ c_{s_2s_1} & c_{s_2s_2} & c_{s_2s_3} \\ c_{s_3s_1} & c_{s_3s_2} & c_{s_3s_3} \end{vmatrix} + \dots$$

(an expression which will be the starting point of Fredholm's theorems on integral equations 4 years later (Chap. V, §1)). He is thus able to replace Poincaré's criterion for convergence by a weaker one: it is enough that the sums $\sum_j |c_{jj}|$ and $\sum |c_{i_1i_2}c_{i_2i_3}\dots c_{i_pi_1}|$ (extended to all sequences $(i_1, i_2, \dots, i_p)$ of distinct indices) be finite. Another convergence criterion is that the sum $\sum_j |c_{jj}|$ and $\sum_{j,k} |c_{jk}|^2$ be finite.

§3. Groping towards function spaces.

It should not be believed that set-theoretic concepts in mathematics were unknown before Boole (1847) or Cantor; they can be traced at least as far back as Aristotle. The use of the word "class" (or, in German, "Gebiet", "Inbegriff", "Mannigfaltigkeit", "System") to designate a set of objects having a common property, becomes frequent among mathematicians since the beginning of the XIX$^{\text{th}}$ century. But it is only after Boole, in the second half of the century, that using letters to denote more or less arbitrary sets, and *computing* with these letters, will become a widespread practice.

In particular "classes" of functions where very often considered in Analysis, even if their description lacks precision most of the time. Even more widespread was the use, since the XVIII$^{\text{th}}$ century, of sequences of functions, or of functions depending on one or several real parameters (for instance in the Calculus of variations). It was of course dimly realized that such families of functions were "much smaller" than the "class" of *all* functions under consideration; the first attempt to give a clearer expression to that feeling is probably due to Riemann. In his famous inaugural lecture on the foundations of geometry, after having tried to give an idea of what he means by a "finite dimensional multiplicity (i.e. manifold)" where the position of a point is determined by a finite set of numbers, he adds that there are "multiplicities" (*Mannigfaltigkeiten*) for which such a determination is not possible, but needs "an infinite sequence or a continuous multiplicity of numbers", and gives as an example "all the possible determinations of a function in a given domain" ([**182**], p. 276).

The extension of the concepts of limit and of continuity to mathematical objects other than numbers or points, such as curves, surfaces or functions, is also very old. However, the applications of that idea dealt with sequences of such ob jects, or families depending on a finite number of real parameters; again, Riemann seems to have been the first to conceive that a whole "class" of functions might be given some kind of "geometrical" structure (what we now would call a topology), for when he speaks of the functions for which the Dirichlet integral (Chap. II, formula (21)) has a meaning, he says that "this set of functions constitutes a connected domain, closed in itself" ([**182**], p. 30), and although it is not quite clear what he means by that, we may see in that statement a first glimpse of the notion of compactness, which will emerge in the last part of the century (see below).

The rigorous study of limits of sequences of functions, which began around 1820, brought to light a phenomenon which had no counterpart for sequences of numbers or of points in $\mathbb{R}^n$: there are *several distinct ways* for a sequence (f_n) of functions to tend to a limit f. The first problem occurred with the distinction between simple and uniform convergence, which was only quite cleared up around 1850. This was followed in the last third of the XIX$^{\text{th}}$ century by a deeper study of these notions, chiefly due to the Italian school (Dini, Ascoli, Arzelà); the most important step taken by that school was the introduction by Ascoli in 1883 of the notion of *equicontinuity*. He discovered that the unpleasant phenomenon of a sequence of continuous functions (in a bounded closed interval I), converging simply to a discontinuous function, would disappear if one assumed on the sequence the following additional property: for each $\varepsilon > 0$, there exists a $\delta > 0$ such that, if $|x' - x''| \leq \delta$, then $|f_n(x') - f_n(x'')| \leq \varepsilon$ for *all indices* n (in other words, the continuity is "uniform", not only with respect to x, but also with respect to n) [8].

One of the fundamental properties of equicontinuous sequences is that, when in addition the fn are uniformly bounded, it is possible to find a *subsequence* (f_n) which converges uniformly, a generalization of the "Bolzano–Weierstrass" theorem for sequences of numbers, which was well-known after 1880. This "compactness" property (which holds for functions defined in a closed bounded set of $\mathbb{R}^n$) was thrust in the limelight by Hilbert, who apparently rediscovered it independently in a special case (he does not quote the Italians) and used it as an essential tool in his famous 1900 paper where he invented the "direct method" in the Calculus of variations ([**111**], vol. III, p. 10–14) and thus was able to justify Riemann's use of the "Dirichlet principle" (chap. II, §3) (*loc.cit.*, p. 15–37).

It is also from the Calculus of variations that another notion of "neighborhood" for a function emeirged during the last years of the XIX$^{\text{th}}$ century. Already at the end of the XVIII$^{\text{th}}$ century mathematicians investigated the problem of deciding if a solution y of the Euler equation for an integral $\int_a^b F(x, y, y')dx$ actually gave a "relative extremum" for that integral. Legendre tried to give a solution to that problem by replacing y in the integral by $y + \varepsilon u$, where $u = \delta y$ is an arbitrary "variation" of class C^1; he thus obtains a function $\Phi(\varepsilon)$ of the real parameter ε and if $\Phi''(0) > 0$ (resp. $\Phi''(0) < 0$) that function reaches a relative minimum (resp. maximum) for $\varepsilon = 0$. This yields the condition $\frac{\partial^2 F}{\partial y'^2} > 0$ (resp. < 0); but it was soon realized that this condition was not sufficient to guarantee that the integral would actually be smaller (resp. larger) than *all* numbers obtained by replacing y by $y + \delta y$ for a "small" variation δy. Clearly this hinges on the question of what exactly is meant by the word "small". Ever since Lagrange, it had been taken for granted that the derivative $(\delta y)' = \delta y'$ is "small" whenever δy itself is "small"; but Weierstrass and his school realized that this was an additional assumption, and this led them to distinguish between "strong extremum" and "weak extremum": the second corresponds to a notion of "neighborhood" of a C^1 function y, where z is "close" to y when the maximum of $|z - y|$ is small, whereas for the first z is only considered as "close" to y if *both* the maximum of $|z - y|$ and the maximum of $|z' - y'|$ are small.

Finally, we have noticed earlier that Gram and Poincaré were naturally confronted with the notion of "convergence in the mean square" in their study of "Fourier expansions" (chap. III, §2). We may therefore say that in the last years of the XIX$^{\text{th}}$ century, the idea of "function spaces" with various "topologies" was so to speak "in the air", and ready to blossom forth as soon as it could be expressed in sufficiently general and simple terms.(*)

The concept of *mapping* of a set of functions into $\mathbb{R}$, or into another set of functions,

(*) It is, however, typical of the unpredictability of mathematical developments that nobody

is also much older than the general definition of a mapping of an arbitrary set into an arbitrary set, which does not seem to have been formulated be fore Dedekind's famous "*Was sind und was sollen die Zahlen*", written in 1872 (although only published in 1888) ([**048**], vol. III, p. 335–391). Ever since the beginning of the Calculus of variations, mathematicians were familiar with the idea of attaching for instance to each C^1 function y in an interval $[a, b]$ a number $\int_a^b F(x, y, y')dx$ depending on y; such mappings would receive the name of "functional" at the end of the XIX$^{\text{th}}$ century. Similarly, as soon as the concept of function emerged at the end of the XVII$^{\text{th}}$ century together with its use in Calculus, the concept of *operator*, yielding a new function when applied to a given function, was in evidence with the examples of the derivatives $f \mapsto Df$

TODO: did he write $f \mapsto D^\nu f$?

or the translation operator $f \mapsto \gamma(a)f$ (function $x \mapsto f(x - a)$); and from Leibniz to Pincherle (end of the XIX$^{\text{th}}$ century) many analysts were led to ponder on the *algebraic* properties of these operators, and their similarity with results of ordinary algebra (which was originally conceived as applying to numbers only). For instance, the similarity of Leibniz's formula for the iterated differential $d^n(uv)$ of a product, with the binomial theorem, probably gave him the idea of attempting to introduce differentials d^α with negative or irrational exponents, a problem to which many mathematicians (such as Liouville, Riemann, Pincherle) later returned, and which has only finally been put to rest with the modern theory of distributions. Other examples are the expression of Taylor's formula given by Lagrange as a relation $\gamma(-a) = e^{aD}$ between operators, or the factoring of a differential polynomial $D^n + a_1 D^{n-1} + ... + a_n$ on the model of the factoring of an ordinary polynomial $z^n + a_1 z^{n-1} + ... + a_n$.

Such ideas, abundantly developed in the period 1790–1830, had much to do with the new conception of Algebra as dealing with symbols rather than with numbers, and later with the axiomatic and formalist conception of the whole of mathematics (see [**054**], chap. XIII, §III); but they had no perceptible influence on Analysis, probably because they did not pay much attention to questions of continuity. It is only in the last yvears of the XIX$^{\text{th}}$ century that such questions appear, in a very episodic way, in papers by Pincherle, Bourlet aund Volterra.

The first two of these authors only consider one "space" E, the set of all holomorphic functions in a domain Δ of the complex plane, and they are exclusively concerned with *linear* operators in that space. In 1886, Pincherle studies operators which, to a holomorphic function[(*)] φ, associate the function $x \mapsto \int_\Gamma A(x, y)\varphi(y)dy$, where Γ is a curve in Δ and A is holomorphic, and he writes that function $\text{¤}\varphi$, but he limits himself to special cases, of the type of the Laplace transform ([**173**], vol. I, p. 92–141). He

---

seems to have been able to foresee, even conjecturally, the direction which was taken by Functional Analysis in the fateful years 1900–1910, This is clear in the communication made by Hadamard in the first In ternational Congress of mathematicians in 1897 ([**094**], vol. I, p. 311–312); he was keenly interested in these "set-theoretical" ideas, and had great expectations of what was to come; but he could think of no serious applications beyond the rehabilitation of the "Dirichlet principle" and some vague ideas on what we now call "precompactness".

(*) After Grassmann (1862), Pincherle seems to have been one of the first mathematicians to write a function with a single letter φ, when all his contemporaries wrote $\varphi(x)$. In his later papers, he repeatedly insists on the fact that a function should be considered as a "point" in some set.

several times returned later to such questions, but failed to obtain any substantial results(*). In 1897, Bourlet [**029**], limiting himself to the case in which A is a disk $|z| < r$, explicitly determines the linear operators in E which are "continuous" (by which he means continuity for what we now call the topology of compact convergence), showing that they are integral operators of the form considered by Pincherle.

We must finally mention the first attempts at "Functional Analysis" of the young Volterra in 1887 ([**219**], vol. I, p. 294–314), to which, under the influence of Hadamard, has been attributed an exaggerated historical importance. Volterra had in mind a generalization of analytic functions, which may be considered as a prefiguration of Hodge's theory(*); for this he needs what he calls "functions of lines". Although, from our point of view, his definitions are not very precise(*), he apparently considers the set E of C^1 mappings of an interval $I \subset \mathbb{R}$ into R^3 (the "lines"), and the mappings $y : E \to \mathbb{R}$, continuous for the topology of uniform convergence. For these "functions of lines" he immediately wants to generalize the classical notion of derivative; in a manner reminiscent of the Calculus of variations, he considers a "varia tion" $\delta y = y(\varphi+\theta)-y(\varphi)$, where the increment θ is supposed to vanish outside of an interval $[a, b]$, and then the quotient $\delta y/\sigma$, where $\sigma = \int_a^b |\theta(t)|dt$; this should tend to a limit when $b-a$ and the maiimum of $|\theta|$ tend to 0. With our experience of 50 years of Functional Analysis, we cannot help feeling that, without even the barest notions of general topology, these *ad hoc* definitions were decidedly premature. Nevertheless, they caught the fancy of Hadamard, who tried to apply similar ideas to Green's functions and encouraged his students to work in that direction (see [**094**] and [**146**]). But these ideas have not, up to now, produced anything comparable to the applications of spectral theory and distribution theory, which we will describe in chap. VII and IX; it might be worthwhile to reexamine them in the light of recent progress in the theory of infinite dimensional manifolds, which could be their natural setting.

§4. The passage "from finiteness to infinity".

The urge to deal with "infinity" has been present from the very beginnings of Greek mathematics, in spite of all philosophical preconceptions and objections, and has taken various forms. The simplest and most "matural" passage "from finiteness to infinity" is the "indefinite repetition" of the arithmetical operation of addition, on smaller and smaller summands, giving birth to the concept of convergent series, of which one can already find examples in Archimedes. Replace addition by multiplication, and you have the infinite

---

(*) He should however be credited with what is probably the first conception of a closed hyperplane in E as the kernel of a continuous linear form, and of closed subspaces of finite codimension as intersections of hyperplanes ([**173**], vol. I, p. 39). In 1897–98, he also has the idea of generalizing Lagrange's "adjoint" of an operator (chap. I, §1, formula (5)) by considering two vector subspaces S, S' of E, and a nondegenerate bilinear form φ, ψ on $S \times S'$; to a linear mapping of S into S', he then associates the "adjoint" A, a linear mapping of S' into S such that $(A \cdot \varphi, \psi) = (\varphi, A \cdot \psi)$, and he observes the relation between the kernel of A and the image of A ([**173**], vol. II, p. 77–84).

(*) See A. Weil, *Oeuvres Scientifiques*, vol. II, Commentaires sur [1952 e], p. 532 of the correct edition (or vol. III, p. 450 of the first printing), Springer, Berlin-Heidelberg-New York, 1979.

(*) This can be said of practically all mathematicians before 1906.

product, born with Calculus in the XVII$^{\text{th}}$ century; and still more sophisticated algebraic manipulations would lead to continued fractions and to the infinite determinants which we have discussed in §2.

Another line of thought goes back at least to Eudoxus's "method of exhaustion", and was to lead in the first place to the concept of integral. But in the hands of the mathematicians of the XVII$^{\text{th}}$ and XVIII$^{\text{th}}$ century, this idea of decomposing an object into "infinitesimal" parts in which the phenomenon they studied became much easier to describe "in a first approximation", was developed into a more and more sophisticated method to discover the differential or partial differential equations which governed the phenomenon "in the large". It is in that way that the equation of vibrating strings (chap. I, §2, formula (7)) was established, either by considering, as D. Bernoulli, a massive string as a limit (for n tending to infinity) of a system of n massive points distributed on a massless string, or by analyzing, as d'Alembert, the forces which are exerted on an "infinitesimal" portion of the string by its neighbors.

It is this second method that Fourier applied to obtain the heat equation; he takes for granted that in a system of small "molecules", a given molecule M receives in an "infinite simal" time dt a quantity of heat from another molecule M' equal to the difference of temperatures of M and M', multiplied by dt and by a coefficient depending only on the distance MM'; the molecule M, if situated at the surface separating the system of molecules from the external world, also radiates a quantity of heat equal to the difference of its temperature and of the external temperature, multiplied by dt and another coefficient depending on M. He then derives the equation of the "cooling off" process (chap. III, §2, equation (21)) by decomposing the solid body V in "infinitesimal" cubes and evaluating the amount of heat received by one of them from its 6 neighbors in time dt, which he takes as proportional (with a constant coefficient) to the variation du of the temperature of that cube the boundary condition (chap. III, §2, equation (22)) is similarly obtained by evaluating the amount of heat lost (by radiation) by an in finitesimal cube at the surface of V.

At the end of his 1890 paper on the cooling off problem (chap. III, §2), Poincaré suggests another method reminiscent of D. Bernoulli's procedure. He first considers a large number N of molecules M_i; following Fourier's physical considerations, and denoting by $v_i(t)$ the temperature of M_i at time t, these functions satisfy the system of linear differential equations

$$\frac{dv_i}{dt} + \sum_{k\neq i} C_{ik}(v_i - v_k) + C_i v_i = 0 \qquad (1 \leq i \leq N), \tag{10}$$

$C_{ik}(v_i - v_k)$ being the quantity of heat received from M_k, and $C_i v_i$ the quantity of heat radiated by M_i outside the system. But instead of letting the number of molecules increase to infinity, Poincaré *first* integrates the system (10) by the classical Euler–Lagrange method: he writes $v_i(t) = u_i e^{-\alpha t}$, and, using the fact that the matrix (C_{ik}) is *symmetric*, he recognizes in the equation he obtains for α the equation giving the eigenvalues of the symmetric matrix corresponding to the non degenerate positive *quadratic form*

$$\Phi(u_1, u_2, \dots, u_N) = \sum_{i\neq k} C_{ik}(u_i - u_k)^2 + \sum_i C_i u_i^2. \tag{11}$$

Let $\xi_1 \leq \xi_2 \leq \dots \leq \xi_N$ be these eigenvalues; the classical theory of quadratic forms shows that one may write

$$\Phi = \xi_1 \varphi_1^2 + \dots + \xi_N \varphi_N^2 \tag{12}$$

where the φ_i are linear forms in the variables $u_1, \dots, u_N$ such that $\varphi_1^2 + \dots + \varphi_N^2 = u_1^2 + \dots + u_N^2$; if for two such forms $f = \alpha_1 u_1 + \dots + \alpha_N u_N, g = \beta_1 u_1 + \dots + \beta_N u_N$ one writes $(f|g) = \sum_k \alpha_k \beta_k$, the N forms φ_i are mutually orthogonal for that scalar prnduct. It is then clear that ξ_1 is the smallest value of the function of $u_1, \dots, u_N$

$$\frac{\Phi(u_1, \dots, u_N)}{u_1^2 + u_2^2 + \dots + u_N^2} = \frac{\Phi}{\varphi_1^2 + \dots + \varphi_N^2} \tag{13}$$

where the u_i are arbitrary; similarly ξ_2 is the minimum of (13) for $\varphi_1 = 0, \xi_3$ the minimum for $\varphi_1 = \varphi_2 = 0$ as relations between the u_i; and so on. This is of course the analgous procedure in N dimensions to the classical determination of the "axes" of an ellipsoid in 3-dimensional space. Poincaré's idea is that the expression (13) corresponds exactly to the quotient

$$\frac{\iiint_V \left(\left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2 \right) d\omega + h \iint_\Sigma v^2 d\sigma}{\iiint_V v^2 d\omega}$$

in his (or rather Weber's) procedure for the definition of the eigenvalues in the cooling off problem; these eigenvalues (the poles λ_m of his function $[f, \xi]$ (chap. III, §2)) correspond to the ξ_j in (12) and the eigenfunctions $U_m(M) = P(M, \lambda_m)$ to the φ_j, the orthogonality of the φ_j corresponding to the relations

$$\iiint_V U_p U_q d\omega = 0 \quad \text{for } p \neq q$$

between the U_m. Finally, he realizes that the same ideas apply as well to other problems and gives as an example the theory of elasticity, and he suggests that a rigorous proof of the existence of the λ_m and the U_m, which he had not been able to give, might be obtained by simply letting N tend to $+\infty$ in the formula (12). He never came back to the question; but we cannot fail to see that this is exactly the program which Hilbert in 1904 followed to its successful conclusion for integral equations with symmetric kernel (chap. V, §2).

A similar "passage from finiteness to infinity" emerged in the first *general* theory of integral equations, beginning with the papers of Le Roux in 1894 and Volterra in 1896. In addition to the *particular* integral equations which had been met by Liouville in the Sturm–Liouville problem (chap. I, §3, equation (34) and chap. II, §1, equation 6) and by Beer and Neumann in the Dirichlet problem (chap. II, §4, equation (23)) (not to speak of what we have called "crypto-integral" equations, where the equation is not written down explicitly but the method exactly amounts to solving it), other *particular* equations involving integrals had come up in connection with problems not directly related to differential or partial differential equations. The first one (chronologically) was the "inversion" problem for the "transform" introduced by Fourier in 1822 (and to which we shall return in chap. VII, §6); it associates to a function f in $[0, +\infty[$ the function

$$\varphi(t) = \int_0^{+\infty} f(x) \cos txdx \tag{14}$$

and the problem consisted in finding f when the transform φ is a given function. It was solved by Fourier's inversion formula ([**067**], vol. I, p. 392)

$$f(x) = \frac{2}{\pi} \int_0^{+\infty} \varphi(x) \cos txdt \tag{15}$$

where, as usual with Fourier, both formulas are obtained by a purely formal calculation. A little later, one of the first published papers of Abel ([1], vol. I, p. 11–27 and 97–101) was devoted to a problem of mechanics, which amounted to finding a function φ such that

$$\int_0^x \frac{\varphi(y)dy}{\sqrt{x-y}} = \psi(x) \tag{16}$$

is a given function; he obtains the solution by the formula

$$\varphi(x) = \frac{1}{\pi}\int_0^x \frac{\psi'(y)dy}{\sqrt{x-y}} \tag{17}$$

and extends his result to the case in which $\sqrt{x-y}$ is replaced by $(x-y)^\alpha$ for $0<\alpha<1$. In a letter to Holmboe, he even hinted at more general results, but nothing was found on the subject in his papers. After Abel, a few papers, giving partial generalizations of his results, were published until 1890(*); but it was only in 1894 that Le Roux attacked the *general* problem of "inversion of a definite integral" (as it was called), *i.e.* finding a C^1 function φ in an interval $[a,b]$ satisfying an equation

$$\int_a^b \varphi(x)H(x,y)dx = f(y) \tag{18}$$

where f and H are C^1 (in $[a,b]$ and $[a,b]\times[a,b]$ respectively) and $f(a)=0$ (*). In contrast with his predecessors, Le Roux is not trying to find a "closed formula" similar to (15) and (17) for the unknown function. He assumes that $h(y)=H(y,y)$ does not vanish in $[a,b]$, takes the derivative of both sides of (18), obtaining

$$h(y)\varphi(y) + \int_a^y \frac{\partial H}{\partial y}(x,y)\varphi(y)dx = f'(y) \tag{19}$$

and then applies the method of successive approximations which Picard had popularized a few years earlier:

$$u_0(y) = \frac{f'(y)}{h(y)}, \qquad u_n(y) = \frac{f'(y)}{h(y)} - \frac{1}{h(y)}\int_a^y \frac{\partial H}{\partial y}(x,y)u_{n-1}(x)dx \quad \text{for } n\geq 1,$$

proving easily the convergence of the sequence (u_n) to a solution of (18) ([**143**], p. 244–246).

In 1896, Volterra (who apparently was unaware of Le Roux's paper) tackles exactly the same problem by the same method, in a series of 4 notes ([**219**], vol. II, p. 216–262). He goes a little beyond Le Roux, by giving an explicit expression of the solution

$$\varphi(y) = \frac{f'(y)}{h(y)} - \frac{1}{h(y)}\int_a^y \left(\sum_{i=0}^{\infty} S_i(x,y)\right) f'(x)dx \tag{20}$$

(*) See the long historical introduction given by Volterra in his 1897 paper on integral equations ([**219**], vol. II, p. 279–287)

(*) As these conditions are *not* satisfied for Abel's equation, Le Roux's results (which for him are auxiliary properties which he needs in a study of partial differential equations) do not directly generalize those of Abel.

where the S_i are defined by induction:

$$S_0(x,y) = \frac{1}{h(x)}\frac{\partial H}{\partial y}(x,y), \qquad S_i(x,y) = \int_x^y S_0(\xi,y)S_{i-1}(x,\xi)d\xi \quad \text{for } i \geq 1. \tag{21}$$

In the later notes, he discussed the cases in which $h(y)$ may vanish at a finite number of points, and the case in which $H(x,y) = G(x,y)/(x-y)^\alpha$ with $0 < \alpha < 1$ and G is continuous (the generalization of Abel's equation). But the most influential part of his notes was the following remark he made immediately after obtaining formula (20): "If one considers the system

$$\begin{cases} b_1 = a_{11}x_1 \\ b_2 = a_{12}x_1 + a_{22}x_2 \\ \dots\dots\dots\dots\dots\dots\dots \\ b_n = a_{1n}x_1 + \dots + a_{nn}x_n \end{cases} \tag{22}$$

the concept of integral easily leads to look at the question of functional Analysis represented by equation (18) as a limiting case of the solution of a system similar to (22), in which the a_{ij} and a_{ii} are the analogous of $H(x,y)$ and $H(y,y)$." Although he limited himself to that (somewhat vague) statement, it seems obvious that what he had in mind was replacing in (18) the variable y by its values

$$y_k = a + \frac{k}{n}(b-a) \quad \text{for } 1 \leq k \leq n$$

and replacing the integral by the corresponding "Riemann sum" for the subdivision of $[a,b]$ by the points y_k obtaining the system of type (22)

$$f(y_j) = \frac{b-a}{n}\sum_{k=1}^{n} \varphi(y_k)H(y_k,y_j) \quad \text{for } 1 \leq j \leq n. \tag{23}$$

Finally, although he does not mention the product of matrices, Volterra develops in these notes the formalism which to two "kernels" $H_1(x,y), H_2(x,y)$ associates the kernel

$$H(x,y) = \int_x^y H_1(x,\xi)H_2(\xi,y)d\xi$$

(which much later he will write $H = H_1 * H_2$). If, for simplicity, we adopt this notation, he shows that, for an *arbitrary* continuous function $S_0(x,y)$ if we define for $i \geq 1$ the "kernel" S_i by $S_i = S_{i-1} * S_0$, one has also $S_i = S_{i-j} * S_{j-1}$ for $1 \leq j \leq i$, and a majoration

$$|S_i(x,y)| \leq \frac{M^{i+1}}{i!}|x-y|^i.$$

This implies uniform convergence for the series

$$F_0 = \sum_{i=0}^{\infty} S_i \tag{24}$$

and the relation

$$F_0 - S_0 = S_0 * F_0. \tag{25}$$

He observes that one may "invert" that relation: if the F_i are defined for $i \geq 1$ by $F_i = F_0 * F_{i-1}$, one has

$$S_0 = \sum_{i=0}^{\infty} (-1)^i F_i. \tag{26}$$

And finally, at the end of his notes, he arrives at the general concept of what Hilbert will call an "integral equation of the second kind"

$$\varphi(y) - \int_a^y S_0(x, y)\varphi(x)dx = f(y) \tag{27}$$

for which the solution is given by

$$\varphi(y) = f(y) + \int_a^y F_0(x, y)f(x)dx \tag{28}$$

as it follows immediately from (25), the "kernel" and the "resolvent kernel" playing completely symmetric parts in these formulas.

Chapter V
The crucial years and the definition of Hilbert space

Between 1900 and 1910, there was a sudden crystallization of all the ideas and methods which had been slowly accumulating during the XIXth century and which we have described in the previous chapters. This was essentially due to the publication of *four fundamental papers*:

Fredholm's 1900 paper on integral equations;
Lebesgue's thesis of 1902 on integration;
Hilbert's paper of 1906 on spectral theory;
Fréchet's thesis of 1906 on metric spaces.

§1. Fredholm's discovery.

The name "integral equation" (*Integralgleichung*) was used for the first time by P. du Bois-Reyfiond in 1888, in a paper on the Dirichlet problem [**061**]; he has in mind equations of the Beer–Neumann type (chap. II, §4) and considers that a general theory of such equations presents "insuperable difficulties"; he is convinced that much progress would come out of such a theory but acknowledges that "almost nothing is known on this question". The later work of Poincaré, which we have discussed above (chap. III, §2), and of his immediate followers, did nothing to dispel that impression; their results seemed linked to delicate estimates from potential theory. It therefore came as a complete surprise when, in a short Note published in 1900, Fredholm showed that the general theory of all integral equations (or "crypto-integral" equations) considered before him was in fact extremely simple (much simpler than anything known at the time in the theory of partial differential equations).

Ivar Fredholm (1866–1927) was a student of Mittag-Leffler in Stockholm in 1888–1890; he only published a few papers during his lifetime, mostly concerned with partial differential equations (we shall return to his thesis of 1898 in chapter IX, §5). After a visit to Paris, where he had been in contact with all the French analysts and had become familiar with the recent papers of Poincaré, he communicated in August 1899 his first results on integral equations to his former teacher; they were published in 1900 [**074**] and completed 2 years later in a paper published in *Acta Mathematica* ([**074**] and [**075**]).

Fredholm's 1900 note is entitled "On a new method for the solution of Dirichlet's problem", but it is characteristic that from the start, he brushes aside all the particular features of the Beer–Neumann equation, and (as Le Roux and Volterra had done with Abel's equation (chap. IV, §4)) begins with a *general* "integral equation of the second kind" (that name will only be given by Hilbert)

$$\varphi(s) = f(s) + \lambda \int_a^b K(s,t) f(t) dt \tag{1}$$

where K is supposed to be bounded and piecewise continuous in $(a,b] \times [a,b]$, and φ continuous in $[a,b]$, λ being a complex parameter. He briefly mentions the analogy with systems of linear equations and starts right away with the formulas describing his "determinants" (see below). But in a lecture given in 1909 [**074**], he acknowledges: 1°

the inspiration derived from Volterra's idea of a "passage to the limit" from a system of linear equations to an integral equation; 2° the help he found in von Koch's work on infinite determinants (chap. IV, §2). From these indications, sparse as they are, it seems one can reconstruct his procedure, with great probability, as consisting in *putting together three simple ideas*:

I) Replacing the integral in (1) by Riemann sums, one obtains, with the notations of chap. IV, §4, the system of n linear equations for the $f(y_j)$

$$f(y_j) + \frac{\lambda(b-a)}{n}\sum_{k=1}^{n} K(y_k, y_j) f(y_k) = \varphi(y_j) \quad (1 \leq j \leq n). \tag{2}$$

II) Writing the determinant of that system according to von Koch's formula (chap. IV, §2, formula (9))

$$1 + \frac{\lambda(b-a)}{n}\sum_{k=1}^{n} K(y_k, y_j) + \frac{\lambda^2(b-a)^2}{2!n^2}\sum_{k_1,k_2}\begin{vmatrix} K(y_{k_1}, y_{k_1}) & K(y_{k_1}, y_{k_2}) \\ K(y_{k_2}, y_{k_1}) & K(y_{k_2}, y_{k_2}) \end{vmatrix} + \ldots$$

and then letting n tend to $+\infty$, which gives the formula for what Fredholm calls the "determinant" of the integral equation (1)

$$\begin{aligned} \Delta(\lambda) = 1 + \lambda\int_a^b K(s,s)ds + \frac{\lambda^2}{2!}\int_a^b\int_a^b K\begin{pmatrix} s_1 & s_2 \\ s_1 & s_2 \end{pmatrix} ds_1 ds_2 + \ldots \\ + \frac{\lambda^m}{m!}\int_a^b \ldots \int_a^b K\begin{pmatrix} s_1 & s_2 & \ldots & s_m \\ s_1 & s_2 & \ldots & s_m \end{pmatrix} ds_1 ds_2 \ldots ds_m + \ldots \end{aligned} \tag{3}$$

where he has written

$$K\begin{pmatrix} x_1 & x_2 & \ldots & x_m \\ y_1 & y_2 & \ldots & y_m \end{pmatrix} = \begin{vmatrix} K(x_1,y_1) & K(x_1,y_2) & \ldots & K(x_1,y_m) \\ K(x_2,y_1) & K(x_2,y_2) & \ldots & K(x_2,y_m) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ K(x_m,y_1) & K(x_m,y_2) & \ldots & K(x_m,y_m) \end{vmatrix}. \tag{4}$$

III) Proving the uniform convergence of the series (3) in any compact set of the complex plane, for which it is enough to majorize the determinants (4) in a suitable way; in his 1899 letter, Fredholm had given the majoration $n^{n/2}M^n$, where M is the upper bound of $|K|$; he had apparently arrived independently to this result, but was made aware that it was a special case of an inequality published by Hadamard in 1893 ([**094**], vol. I, p. 239–243) for an arbitrary square matrix $A = (a_{ij})$ of order n:

$$|\det(A)|^2 \leq \prod_{i=1}^{n}\left(\sum_{j=1}^{n}|a_{ij}|^2\right). \tag{5}$$

The next "natural" steps are of course to apply Cramer's formulas to the system (2) and let again n tend to infinity in the numerators; the result is described by Fredholm in the following elegant way: a development of the determinant (4) according to the first row yields the formula

$$\begin{aligned} K\begin{pmatrix} s & x_1 & \ldots & x_m \\ t & x_1 & \ldots & x_m \end{pmatrix} = K(s,t)K\begin{pmatrix} x_1 & \ldots & x_m \\ x_1 & \ldots & x_m \end{pmatrix} - \\ - K(s,x_1)K\begin{pmatrix} x_1 & x_2 & \ldots & x_m \\ t & x_2 & \ldots & x_m \end{pmatrix} + K(s,x_2)K\begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_m \\ t & x_1 & x_3 & \ldots & x_m \end{pmatrix} - \\ - \ldots + (-1)^m K(s,x_m)K\begin{pmatrix} x_1 & x_2 & \ldots & x_m \\ t & x_1 & \ldots & x_{m-1} \end{pmatrix}. \end{aligned} \tag{6}$$

On the other hand, Fredholm defines the "minor"

$$\Delta(s,t;\lambda) = K(s,t) + \lambda \int_a^b K\begin{pmatrix} s & x_1 \\ t & x_1 \end{pmatrix} dx_1 + \ldots + \tag{7}$$
$$+\frac{\lambda^m}{m!} \int_a^b \ldots \int_a^b K\begin{pmatrix} s & x_1 & \ldots & x_m \\ t & x_1 & \ldots & x_m \end{pmatrix} dx_1 dx_2 \ldots dx_n + \ldots$$

and replaces each integrand by its expression (6), which gives him the simple relation

$$\Delta(s,t;\lambda) = K(s,t)\Delta(\lambda) - \lambda \int_a^b K(s,\xi)\Delta(\xi,t;\lambda)d\xi. \tag{8}$$

He then introduces the function

$$\Phi(s) = \varphi(s)\Delta(\lambda) - \lambda \int_a^b \Delta(s,\xi;\lambda)\varphi(\xi)d\xi \tag{9}$$

and derives from (8) the equation

$$\Phi(s) + \lambda \int_a^b K(s,t)\Phi(t)dt = \varphi(s)\Delta(\lambda). \tag{10}$$

The conclusion is then immediate: if $\Delta(\lambda) \neq 0$, the function $f(s) = \Phi(s)/\Delta(\lambda)$ is a solution of (1). Furthermore, he shows that one has

$$\frac{d\Delta(\lambda)}{d\lambda} = \int_a^b \Delta(s,s;\lambda)ds \tag{11}$$

and from this he deduces that if λ_0 is a zero of order ν of the entire function $\Delta(\lambda), \Phi(s)$, for a suitable choice of φ, cannot be divisible by a power of $\lambda - \lambda_0$ greater than $(\lambda - \lambda_0)^{\nu-1}$; if $\Phi(s) = (\lambda - \lambda_0)^k \Phi_1(s)$, one then deduces, from (10), that

$$\Phi_1(s) + \lambda_0 \int_a^b K(s,t)\Phi_1(t)dt = 0; \tag{12}$$

in other words, if there is no nontrivial solution of the homogeneous equation (12), necessarily $\Delta(\lambda_0) \neq 0$, hence the solution of (1) for $\lambda = \lambda_0$ exists and is unique. However, at that time, he does not yet prove that the existence of a non trivial solution of (12) implies that $\Delta(\lambda_0) = 0$. But the end of the Note is startling: he considers the Beer–Neumann equation for a bounded plane domain with a C^3 boundary; the kernel of that integral equation is then bounded and continuous, and for $\lambda_0 = 1$ it is very easy to deduce from the properties of double layer potentials that the homogeneous equation (12) has no nontrivial solution. Therefore the existence and uniqueness of the soluation of Dirichlet's problem is proved, doing away, with a single strooke of the pen, so to speak, with all the complications of the Neumann–Poincaré solution!

In his 1903 paper, Fredholm completed his results on some important points. He first defines more general "minors"

$$\Delta\begin{pmatrix} s_1 & s_2 & \ldots s_m \\ t_1 & t_2 & \ldots t_m \end{pmatrix}; \lambda = K\begin{pmatrix} s_1 & s_2 & \ldots s_m \\ t_1 & t_2 & \ldots t_m \end{pmatrix} + \tag{13}$$
$$+\sum_{n=1}^{\infty} \frac{\lambda^n}{n!} \int_a^b \ldots \int_a^b K\begin{pmatrix} s_1 & s_2 & \ldots s_m & x_1 & \ldots & x_n \\ t_1 & t_2 & \ldots t_m & x_1 & \ldots & x_n \end{pmatrix} dx_1 \ldots dx_n.$$

Developing this time the determinants *both* according to the first row and the first column, he obtains the identities

$$\begin{aligned}(14)\qquad &\Delta\begin{pmatrix}s_1 & s_2 & \dots & s_m\\ t_1 & t_2 & \dots & t_m\end{pmatrix};\lambda\Big) + \lambda\int_a^b K(s_1,\xi)\Delta\begin{pmatrix}\xi & s_2 & \dots & s_m\\ t_1 & t_2 & \dots & t_m\end{pmatrix};\lambda\Big)\,d\xi = \\ &= K(s_1,t_1)\Delta\begin{pmatrix}s_2 & \dots & s_m\\ t_2 & \dots & t_m\end{pmatrix};\lambda\Big) - K(s_1,t_2)\Delta\begin{pmatrix}s_2 & s_3 & \dots & s_m\\ t_1 & t_3 & \dots & t_m\end{pmatrix};\lambda\Big) + \dots\end{aligned}$$

and

$$\begin{aligned}(15)\qquad &\Delta\begin{pmatrix}s_1 & s_2 & \dots & s_m\\ t_1 & t_2 & \dots & t_m\end{pmatrix};\lambda\Big) + \lambda\int_a^b K(\xi,t_1)\Delta\begin{pmatrix}s_1 & s_2 & \dots & s_m\\ \xi & t_2 & \dots & t_m\end{pmatrix};\lambda\Big)\,d\xi = \\ &= K(s_1,t_1)\Delta\begin{pmatrix}s_2 & \dots & s_m\\ t_2 & \dots & t_m\end{pmatrix};\lambda\Big) - K(s_2,t_1)\Delta\begin{pmatrix}s_1 & s_3 & \dots & s_m\\ t_2 & t_3 & \dots & t_m\end{pmatrix};\lambda\Big) + \dots\end{aligned}$$

which in particular, for $m = 1$, reduce to (8) and to

$$(16)\qquad \Delta(s,t;\lambda) = K(s,t)\Delta(\lambda) - \lambda\int_a^b K(\xi,t)\Delta(s,\xi;\lambda)\,d\xi.$$

The use he makes of these formulas is a little more sophisticated than in his first Note. He introduces the operator corresponding to the kernel K, $f \mapsto S_K f$ such that $S_K f(s) = f(s) + \int_a^b K(s,t)f(t)\,dt$, and, for two kernels K, K', writes the composite $S_K S_{K'}$ as $S_{K''}$ with

$$(17)\qquad K''(x,t) = K(x,t) + K'(x,t) + \int_a^b K(s,\xi)K'(\xi,t)\,d\xi.$$

Suppose now that $\Delta(\lambda) \neq 0$, and write

$$(18)\qquad R(s,t;\lambda) = -\Delta(s,t;\lambda)/\Delta(\lambda)$$

(the *resolvent kernel* in the later terminology of Hilbert). It then follows from (17), (8) and (16) that we have

$$(19)\qquad S_{\lambda K}S_R = S_R S_{\lambda K} = \mathrm{Id}$$

and Fredholm has thus shown that the necessary and sufficient condition for the existence and uniqueness of a solution of (1) is $\Delta(\lambda) \neq 0$, the kernel λK and the resolvent kernel R playing completely symmetric parts as in the formulas of Volterra (chap. IV, §4, formulas (27) and (28)).

Next he examines what happens when $\Delta(\lambda) = 0$. First he generalizes (11) to

$$(20)\qquad \frac{d^m\Delta(\lambda)}{d\lambda^m} = \int_a^b \dots \int_a^b \Delta\begin{pmatrix}s_1 & \dots & s_m\\ s_1 & \dots & s_m\end{pmatrix};\lambda\Big)\,ds_1 \dots ds_m$$

and from this he deduces that if $\Delta(\lambda) = 0$, there is always an integer m such that $\Delta\begin{pmatrix}s_1 & \dots & s_m\\ s_1 & \dots & s_m\end{pmatrix};\lambda\Big)$ is not identically 0. If m is the smallest integer having that property

(which is exactly the order of λ as a zero of Δ) he exhibits, using (14), m solutions of the homogeneous equation

$$\Phi_1(s) = \frac{\Delta\begin{pmatrix} s & s_1 & \dots & s_m \\ t & t_1 & \dots & t_m \end{pmatrix}}{\Delta\begin{pmatrix} s_1 & \dots & s_m \\ t_1 & \dots & t_m \end{pmatrix}}, \quad \Phi_2(s) = \frac{\Delta\begin{pmatrix} s_1 & s & s_3 & \dots & s_m \\ t_1 & t_2 & t_3 & \dots & t_m \end{pmatrix}}{\Delta\begin{pmatrix} s_1 & \dots & s_m \\ t_1 & \dots & t_m \end{pmatrix}}, \dots \tag{21}$$

for which he shows that they are linearly independent and that every other solution of the homogeneous equation is a linear combination of the $\Phi_j (1 \leq j \leq m)$. He concludes the theory (which one often calls the "Fredholm alternative") by giving necessary and sufficient conditions on φ for the existence of a solution of (1) when λ is a zero of order m of Δ. He observes that the "transposed equation" obtained from (1) by replacing $K(s,t)$ by $K(t,s)$ has the same "determinant", and therefore the corresponding homogeneous equation has exactly m linearly independent solutions $\psi_1, \dots, \psi_m$; the condition φ must satisfy are then

$$\int_a^b \varphi(x)\psi_j(x)dx = 0 \quad \text{for } 1 \leq j \leq m. \tag{22}$$

Finally, Fredholm shows that for any two kernels K, K', if Δ_K and $\Delta_{K'}$ are the corresponding "determinants", then for the "composed" kernel K'' defined by (17) one has

$$\Delta_{K''} = \Delta_K \Delta_{K'} \tag{23}$$

which justifies the name "determinant". He also points out that his results can be generalized when the kernel K is not bounded any more, but such that $(x-y)^\alpha K(x,y)$ remains bounded, with $0 < \alpha < 1$; and he mentions that the extension of his theorems to any number of variables is immediate.

This beautiful paper may be considered as the source from which all further developments of spectral theory are derived. It made a deep and lasting impression on the mathematical world, and almost overnight the theory of integral equations became a favorite topic among analysts ([23], [**175**], [**107**]).

§2. The contributions of Hilbert.

One of the most active proponents of the new theory was David Hilbert. As soon as he heard of Fredholm's results, he started doing himself research work on these questions, made them one of the main subjects discussed in his Seminar at Göttingen(*) and supervised many dissertations on the various aspects of the theory. Between 1904 and 1906, he published six papers on integral equations in the *Göttingen Nachrichten*, later brought together in a single volume entitled "Grundzüge einer allgemeinen Theorie der Integralgleichungen" [**112**].

(*) It is reported (by Hellinger) that Hilbert inaugurated a session of his Seminar by announcing the development of a method which would lead to the proof of the Riemann hypothesis: the problem is to prove that a particular entire function has all its zeroes on the real line, and Hilbert hoped that this function would be expressed as the "determinant" of an integral equation wita symmetric kernel. However, nobody has yet been able to find such an equation.

In his first paper [**112**], Hilbert starts by doing explicitly what had only been hinted at by Volterra and Fredholm, the "passage to the limit" in the system (2), restricting himself (as he will do in almost all his results) to the case in which the kernel K is *symmetric*, i.e. a real continuous function such that $K(t,s) = K(s,t)$. He soon realized that in that particular case he might obtain much more precise results than Fredholm. In the first place, the symmetric matrix $(K(y_k, y_j))$ is then the matrix of the quadratic form $\sum_{j,k} K(y_k, y_j)\xi_k\xi_j$, and Hilbert undertook to apply also his "passage to the limit" to that form. He thus obtained the results which Poincaré had foreseen in the particular case he had considered (chap.IV, §4): the roots of the Fredholm determinant are then real; if they are written as a sequence (λ_n), each being counted with its multiplicity, then, for each n there is an eigenfunction φ_n such that $\int_a^b \varphi_m(t)\varphi_n(t)dt = 0$ for $m \neq n$. Finally, if one normalizes the φ_n by the condition $\int_a^b \varphi_n(t)^2 dt = 1$, and if for each a continuous function x din $[a,b]$, one defines the "Fourier coefficients" $(x|\varphi_n) = \int_a^b x(t)\varphi_n(t)dt$, Hilbert proves that

$$\int_a^b \int_a^b K(s,t)x(s)y(t)dsdt = \sum_n \frac{1}{\lambda_n}(x|\varphi_n)(y|\varphi_n) \tag{24}$$

for any two continuous functions x, y, a relation which he rightly considers as the natural generalization of the classical reduction of a quadratic form to its "axes". What is particularly interesting in the way Hilbert considers this formula is that he shows that the righthand side of (24) is *uniformly convergent* when the functions x and y are allowed to vary arbitrarily, subject only to the conditions $\int_a^b x(t)^2 dt \leq 1$ and $\int_a^b y(t)^2 dt \leq 1$, the first prefiguration of what will become "the unit ball in Hilbert space" a few vears later. Of course Hilbert also justifies for his integral equation the variational deiinition of the eigenvalues λ_n, first proposed by Weber (chap. III, §1). He shows that the set of the λ_n is infinite, except when $K(x,y)$ is a linear combination of a finite number of functions of type $u(x)v(y)$. He also proves that the resolvent kernel $R(s,t;\mu)$ (in the sense of Fredholm) has the eigenvalues $\lambda_n - \mu$, the corresponding eigenfunctions being $\varphi_n/(\lambda_n - \mu)$ (μ distinct from the λ_n) and writes the identity

$$R(s,t;\mu) - R(s,t;\nu) = (\mu - \nu)\int_a^b R(s,\xi;\mu)R(\xi,t;\nu)d\xi \tag{25}$$

for μ and ν distinct from the λ_n. Finally, he shows that if a function f can be written in the form

$$f(s) = \int_a^b K(s,t)g(t)dt \tag{26}$$

for a continuous function g, then the corresponding "Fourier expansion"

$$f(s) = \sum_n (f|\varphi_n)\varphi_n(s) \tag{27}$$

is absolutely and uniformly convergent, and one has the "Parseval identity"

$$\int_a^b f(s)^2 ds = \sum_n (f|\varphi_n)^2. \tag{28}$$

However, he could only give that proof undeir the restrictive assumption that any continuous function could be approximated (in the sense of mean square value, or, as we would now say, for the topology of Hilbert space!) by functions of the form (26).

The proof that this last condition is superfluous was given in 1905 in the dissertation of Erhard Schmidt, one of the best students of Hilbert [**191**]; it contained otherwise no startling new results, but it deserves some comments, since it is the first attempt to do away with the Fredholm "determinants", and substitute to them a more conceptual approach(*).

E. Schmidt begins by proving the Bessel identity

$$\int_a^b \left(f(s) - \sum_{n=1}^{N} (f|\varphi_n)\varphi_n(s) \right)^2 ds = \int_a^b f(s)^2 ds - \sum_{n=1}^{N} (f|\varphi_n)^2 \tag{29}$$

for an *arbitrary* orthonormal system (φ_n), from which he deduces that for any continuous function f, g, the series $\sum_n (f|\varphi_n)(g|\varphi_n)$ is absolutely convergent, and the convergence is uniform when f is allowed to vary subject to the condition $\int_a^b f(s)^2 ds \leq A$ for a fixed constant A.

Next he assumes the existence of the eigenvalues λ_n and of the corresponding normalized eigenfunctions φ_n, and using the Bessel inequality, he proves that

$$\sum_n \frac{1}{\lambda_n^2} \leq \int_a^b \int_a^b K(s,t)^2 ds dt \tag{30}$$

from which it follows that each λ_n has finite multiplicity and that $|\lambda_n|$ tends to $+\infty$ with n if there is an infinity of eigenvalues.

To prepare for the proof of the existence of the eigenvalues, he introduces, as Fredholm and Volterra had done, the iterated kernels

$$K_m(s,t) = \int_a^b K_{m-1}(s,\xi) K(\xi,t) d\xi \quad \text{for } m > 1, \text{with } K_1 = K \tag{31}$$

and shows that, if φ is an eigenfunction for K_m, it is also an eigenfunction for K if m is odd, and is sum of two eigenfunctions for K if m is even. This allows him to apply Schwarz's method to prove the existence of at least an eigenvalue when K is not identically 0, as we have shown in chapter III, §1, because what he gets in this way is an eigenvalue of K_2.

Finally, for functions f given by (26), he obtains the convergence of the Fourier expansion (27) by applying his initial lemma to the functions $t \mapsto K(s,t)$ and g; and from that he derives Hilbert's formula (24) by multiplying the formula $x(s) = \sum_n (x|\varphi_n)\varphi_n(s)$ by $K(s,t)y(t)$ and integrating.

Hilbert's interest in integral equations with symmetric kernels of course stemmed from the possibility of applying them to questions of Analysis such as the Dirichlet problem; it is to such applications that he devoted the second and third of his papers on integral equations, We shall bypass them for the time being, as well as most results in his two last papers on the subject (see chapters VII and IX), to concentrate on his fourth paper, published in 1906, a masterpiece and one of the best papers he ever wrote. By the depth and novelty of its ideas, it is a turning point in the history of Functional Analysis, and indeed deserves to be considered as the very *first* paper published in that discipline.

(*) Some of the results of E. Schmidt were also obtained independently by W. Stekloff [**204**].

Hilbert's new departure in that paper is clear from the beginning: he deliberately *abandons* the point of view of integral equations, to return to the older conception of the infinite systems of linear equations (chap. IV, §2), but with a new twist, This is because he realizes that the theory of integral equations can be *subsumed* as a special case of that older theory: indeed, let (ω_n) be a complete orthonormal system of continuous functions in $[a,b]$, and suppose the continuous function f is a solution of (1) for $\lambda = 1$; then, if we consider the "Fourier coefficients"

$$k_{pq} = \int_a^b \int_a^b K(s,t)\omega_p(s)\omega_q(t)dsdt,$$
$$b_p = \int_a^b \varphi(s)\omega_p(s)ds, \qquad x_p = \int_a^b f(s)\omega_p(s)ds \tag{32}$$

the $x_p (p = 1, 2, ...)$ satisfy the infinite system of linear equations

$$x_p + \sum_{q=1}^{\infty} k_{pq}x_q = b_p \qquad (p = 1, 2, ...). \tag{33}$$

The new twist is that, due to the Bessel identity, one has

$$\sum_{p,q} k_{pq}^2 < +\infty, \quad \sum_p b_p^2 < +\infty, \quad \sum_p x_p^2 < +\infty. \tag{34}$$

Conversely, suppose we have a solution (x_p) of (33) (with conditions (34)), and observe that if $k_q(s) = \int_a^b K(s,t)\omega_q(t)dt$, the functions k_q are continuous, and

$$\sum_p k_p(s)^2 \leq \int_a^b K(s,t)^2 dt;$$

the series $u(s) = \sum_p x_p k_p(s)$ is then absolutely and uniformly convergent; hence u is continuous and one has $(u|\omega_p) = b_p - x_p$; therefore, if $f = \varphi - u, (f|\omega_p) = x_p$ and from the completeness of the system (ω_p) it follows that f is a solution of (1) for $\lambda = 1$.

Hilbert then embarks into completely uncharted territory:

1° He exclusively considers sequences $x = (x_p)$ (for $p = 1, 2, ...$) of real numbers such that $\sum_p x_p^2 < +\infty$.

2° On the contrary, with regard to the double sequence (k_{pq}) of real numbers, he *abandons* at first *any* restrictive condition such as the first condition (34), and only retains the *symmetry* conditions $k_{qp} = k_{pq}$.

3° The center of interest is not any more the solution of the system (33), but the "symmetric bilinear form"

$$K(x,y) = \sum_{p,q} k_{pq}x_p y_q \tag{35}$$

which he wants to "reduce" by a formula which would generalize (24).

Of course, even under the restrictions $\sum_p x_p^2 < +\infty, \sum_p y_p^2 < +\infty$, the right hand side of (35) is usually meaningless(*); proceeding as Fourier, Poincaré and von Koch (chap. IV,

(*) To this rather awkward formulation, Hellinger and Toeplitz [**106**] substituted the consideration of "infinite matrices" $(k_{pq})_{1\leq p,q<+\infty}$ and of their "calculus" inspired from Frobenius, but without associating an endomorphism to a matrix.

§2), Hilbert considers, for each integer n, the symmetric bilinear form in $2n$ variables ("sections" (Abschnitte) of K)

$$K_n(x,y) = \sum_{p=1}^{n}\sum_{q=1}^{n} k_{pq}x_p y_q, \tag{36}$$

but instead of investigating the determinants of these forms, he "reduces" each one to its "axes" and is confronted with the problem of "passing to the limit" for these "reduced" forms when n tends to $+\infty$. We postpone the detailed examination of the original method by which he was able to solve that problem, to chapter VII, which is devoted to the history of modern spectral theory, of which this paper of Hilbert is the starting point; we shall only discuss here the various new notions he is led to introduce in that paper.

A) Hilbert is not yet using the geometrical language which will become prevalent among his immediate successors (cf. §3), but it is obvious that everything he does in inspired by the analogy with n-dimensional FEuclidean space. In particular one of his main tools is the generalization of *orthogonal transformations*: by that he means that, to every sequence (x_p) with $\sum_p x_p^2 < +\infty$, he associates the sequence (x'_p), where

$$x'_p = \sum_q a_{pq}x_q \qquad (p = 1, 2, ...) \tag{37}$$

and where he imposes on the double sequence (a_{pq}) the conditions

$$\begin{aligned} &\sum_q a_{pq}^2 = 1, \qquad \sum_p a_{pq}a_{pr} = 0 \quad \text{for } q \neq r \\ &\sum_p a_{pq}^2 = 1, \qquad \sum_q a_{pq}a_{np} = 0 \quad \text{for } n \neq p \end{aligned} \tag{38}$$

from which he immediately deduces that conversely (x_p) is deduced from (x'_p) by appling the "inverse" orthogonal transformation defined by (a'_{pq}) with $a'_{pq} = a_{qp}$.

B) Hilbert restricts himself to forms (35) which he calls *bounded*: they are the (not necessarily symmetric) forms such that there exists an $M > 0$ for which one has $|K_n(x,y)| \leq M$ for $\sum_p x_p^2 \leq 1, \sum_p y_p^2 \leq 1$ and *for all* n; he also introduces *bounded linear forms* $L(x) = \sum_p a_p x_p$ with $\sum_p a_p^2 < +\infty$, so that for any x (resp. y), and any bounded bilinear form K, the linear forms $K(x,\cdot) : x \mapsto K(x,y)$ and $K(\cdot,y) : x \mapsto K(x,y)$ are bounded. One of the things he wants to do (inspired of course by the "reduction" of bilinear forms in a finite number of variables) is to operate an orthogonal transformation on x and y, substituting the expressions (37) for the x_p and doing the same for the y_p. Unfortunately, he follows Frobenius in his conception of the ""*Faltung*" of bilinear forms (instead of the natural idea of "composing" transformations). So, for two bounded bilinear forms A, B, he has to show that the forms $A_n(x,\cdot)B_n(\cdot,y)$ (*Faltung* of $A_n(x,y)$ and $B_n(x,y)$, these forms being defined as in (36)) are the forms $C_n(x,y)$ corresponding to a bounded form $C(x,y)$ which he calls again the "*Faltung*" of $A(x,y)$ and $B(x,y)$ and writes $A(x,\cdot)B(\cdot,y)$. He can then express the action of an orthogonal transformation on a bounded bilinear form $K(x,y)$ as a "*Faltung*"

$$K'(x',y') = K(\cdot,\cdot)O(\cdot,x')O(\cdot,y') \tag{39}$$

where $O(x,y) = \sum_{p,q} a_{pq}x_p y_q$ is the bounded bilinear form which he associates to the orthogonal transformation (37).

C) For the development of Functional Analysis, the most important concepts introduced by Hilbert were what he calls "continuity" and "complete coutinuity", which correspond to what will later be called the "strong" and "weak" topologies on Hilbert space. If $F(x)$ is a complex-valued function defined for all sequences $x = (x_p)$ such that $\sum_p x_p^2 < +\infty$, Hilbert says that F is *continuous* if $F\left(x^{(n)}\right)$ tends to $F(x)$ when $\sum_p \left(x_p - x_p^{(n)}\right)^2$ tends to 0, and that F is *completely continuous* if $F\left(x^{(n)}\right)$ tends to $F(x)$ when $\sum_p x_p^2 \leq 1, \sum_p \left(x_p^{(n)}\right)^2 \leq 1$ and each coordinate $x_p^{(n)}$ tends to x_p. He shows that a bounded bilinear form $K(x,y)$ is continuous, and that $K_n(x,y)$ tends to $K(x,y)$ when n tends to ∞. But he pays special attention to the completely continuous symmetric bilinear forms, and gives a separate proof that an orthogonal transformation can reduce any such form to the type

$$K(x,y) = \frac{1}{\lambda_1}x_1y_1 + \frac{1}{\lambda_2}x_2y_2 + \ldots + \frac{1}{\lambda_n}x_ny_n + \ldots \tag{40}$$

where the sequence $(|\lambda_n|)$ is either finite or tends to $+\infty$. He realizes that this is a genuine generalization of formula (24), which is the special case in which $\sum_{p,q} k_{pq}^2 < +\infty$ (corresponding to what will later be called the *Hilbert–Schmidt operators*); he also mentions another special case, the one in which $K(x,x) > 0$ and $\sum_p k_{pp} < +\infty$ (corresponding to the positive *nuclear operators* of a later date). This formula (40) enables him to go beyond Fredholm by solving a system (33) which is not derived any more from an integral equation, but in which the k_{pq} are only supposed to be such that the symmetric bilinear form (35) is completely continuous. A final remark is that he repeatedly uses with great power what he calls a "principle of choice", which is equivalent to what will later be called the compactness of the unit ball for the weak topology, and that he extends his results to *hermitian* sesquilinear forms

$$K(x,y) = \sum_{p,q} k_{pq}x_p\overline{y}_q \tag{41}$$

where this time the sequences $(x_p), (y_p)$ and (k_{pq}) consist of complex numbers, with $k_{qp} = \overline{k}_{pq}$.

§3. The confluence of Geometry, Topology and Analysis.

It may seem obvious to us that the results of Hilbert are but one step removed from what we now call the theory of Hilbert space; but if, in fact, the birth of that theory almost immediately followed the publication of Hilbert's papers, it seems to me that it is due to the fact that this publication precisely occurred during the emergence of a new concept in mathematics, the concept of *structure*.

Until the middle of the XIX$^{\text{th}}$ century, mathematicians had been dealing with *well determined* mathematical "objects": numbers, points, curves, surfaces, volumes, functions, operators. But the fact that algebraic manipulations on different kinds of "objects" had a strikingly similar appearance soon attracted attention (cf. chap. IV, §3), and after 1840 it gradually became clear that the essence of these maunipulations did not lie in the *nature* of the objects, but in the *rules* to be followed in handling them, which might be the same for very different types of objects. However, a precise formulation of this idea had to wait for the adoption of the set-theoretic concepts and language; and it is only in 1895 that our definition of a *group*, on an *arbitrary* underlying set, was formulated by Weber [**225**]. The trend towards the definition of algebraic structures then gained

momentum, and around 1920 all fundamental notions of present-day Algebra had been defined.

In Analysis, no similar development had yet occurred in 1900. The extensions of the ideas of limit and continuity which had been formulated always were relative to *special* objects such as curves, surfaces or functions. The possibility of defining such notions in an *arbitrary* set is an idea which undoubtedly was first put forward by Fréchet in 1904 [**069**], and developed by him in his famous thesis of 1906 [**071**]. The simplest and most fruitful method which he proposed for such definitions was the introduction of the notion of *distance* (which he called "écart") on a set E, a function $d(x, y)$ defined for any pair (x, y) of elements of E, with values ≥ 0 and such that: 1) the relation $d(x, y) = 0$ is equivalent to $x = y$; 2) $d(y, x) = d(x, y)$; 3) $d(x, z) \leq d(x, y) + d(y, z)$ for any three elements of E. It is extremely remarkable that with such simple axioms it is possible to extend most notions and arguments relative to neighborhoods, limits and continuity in the space $\mathbb{R}^n$, which usually are introduced in relation to euclidean distance. But the greatest merit of Fréchet lies in the emphasis he put on three notions which were to play a fundamental part in all later develop ments of Functional Analysis: compactness, completeness and separability. Moreover, he did not limit himself to deriving general theorems in an abstract setting, but more than half of his thesis is devoted to very "concrete" metric spaces (as they came to be called later) closely linked to Analysis: the space of continuous real functions on a compact interval of $\mathbb{R}$ with the topology of uniform convergence, the space $\mathbb{R}^{\mathbb{N}}$ of all sequences $n \mapsto x_n$, with the topology of simple convergence, the space of holomorphic functions in the disc $|z| < 1$, with the topology of uniform convergence in compact subsets, and finally the space of all continuous "curves", images of $[0, 1]$ in $\mathbb{R}^3$ by continuous maps, with a "distance" which is a special case of what was later called the Hausdorff distance between two compact sets.

Clearly Hilbert's work immediately lent itself to application of these ideas, and even invited a bodily transfer of euclidean geometry in "infinite dimension". This is exactly what was done by Fréchet himself [**072**] and by E. Schmidt [**192**] in 1908. In E. Schmidt's paper, we find the definition of what we now call the (complex) space l^2 (or $l^2_{\mathbb{C}}$), with the notions of scalar product and of norm (already written $\|A\|$), the definition of orthogonality, of closed sets, and of vector subspaces (called "*lineares Funktionengebilde*"). The most interesting feature of that paper is the proof of the existence of the orthogonal projection of a point on a closed vector subspace, and the purely geometric way in which Schmidt uses this result to discuss the *most general* system of linear equations in Hilbert space

$$(x|a_n) = c_n \qquad (n = 1, 2, ...) \tag{42}$$

where the a_n are arbitrary vectors of l^2 and the c_n arbitrary complex numbers,

For each n, Schmidt considers the closed linear affine varieties F_n of l^2 defined by the equatioas $(x|a_j) = c_j$ for $1 \leq j \leq n$, and the orthogonal projection $x^{(n)}$ of the origin on F_n; the necessary and sufficient condition of existence of a solution of the system (42) is that the increasing sequence $\left(\|x^{(n)}\|\right)$ be bounded; the sequence $\left(x^{(n)}\right)$ then has a weak limit x in l^2, which is the solution of (42) of smallest norm. Of course, each F_n must be different from the empty set, which means that any linear relation $\sum_{k=1}^{n} \lambda_k a_k = 0$ between the vectors a_n must imply $\sum_{k=1}^{n} \lambda_k c_k = 0$; one can then assume (by dropping some of the equations (42)) that the a_n are linearly independent, and in that case, Schmidt easily obtains the explicit expression of $\|x^{(n)}\|$:

$$\|x^{(n)}\|^2 = \Delta_n / D_n \tag{43}$$

with

$$D_n = \begin{vmatrix} (a_1|a_1) & (a_1|a_2) & \dots & (a_1|a_n) \\ (a_2|a_1) & (a_2|a_2) & \dots & (a_2|a_n) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ (a_n|a_1) & (a_n|a_2) & \dots & (a_n|a_n) \end{vmatrix},$$

(44)

$$\Delta_n = \begin{vmatrix} 0 & c_1 & c_2 & \cdots & c_n \\ \overline{c}_1 & & & & \\ \overline{c}_2 & & & & \\ \vdots & & D_n & & \\ \overline{c}_n & & & & \end{vmatrix}$$

This geometric outlook was already shared in 1906–1907 by two other young mathematicians, E. Fischer and F. Riesz, in the remarkable work which led them (independently) to what is now called the Fischer–Riesz theorem, introducing a hitherto unsuspected link between Hilbert space and the theory of in tegration ([**066**], [**183**]). The latter, from Cauchy to Jordan and Peano, had evolved in a manner completely independent from spectral theory [**103**]. When Fredholm and E. Schmidt had tried to enlarge the scope of their results on integral equations by weakening the assumptions on the kernel $K(x, y)$, they had nothing else at their disposal beyond the horrible and useless so-called "Riemann integral"(*), and it is likely that progess in Functional Analysis might have been appreciably slowed down if the invention of the Lebesgue integral had not appeared, by a happy coincidence, exactly at the beginning of Hilbert's work on integral equations. With the help of this marvellous new tool, Fischer and F. Riesz could define the space $L^2(I)$ over a compact interval $I \subset \mathbb{R}$, consisting of square integrable functions, when two functions are identified if they only differ in a set of measure 0. Their fundameutal result is that, if to each function $f \in L^2(I)$ one associates the sequence (x_p) of its Fourier coefficients with respect to a complete orthonormal system (equations (32)), this defines an *isomorphism* of $L^2(I)$ onto l^2; from that it follows that $L^2(I)$ is complete and separable. A byproduct was of course that the results of Fredholm and E. Schmidt could be applied without change to any integral equation where the kernel is only supposed to belong to $L^2(I \times I)$, since it is then equivalent to a system of linear equations corresponding to a "completely continuous" bilinear form in the sense of Hilbert.

But the most important consequence of the Fischer–Riesz theorem is that it opened the way to the definition of the L^p spaces and to the general theory of normed spaces, which will be the subject of the next chapter.

(*) As a function $K(x, y)$ of two variables may be "Riemann integrable" even if the partial functions $x \mapsto K(x, y)$ are not, Fredholm is compelled to assume that integrability both for the kernel K and all its partial functions! Although E.Schmidt wrote his dissertation in 1905, he probably had no knowledge of Lebesgue's thesis at that time.

Chapter VI
Duality and the definition of normed spaces

§1. The search for continuous linear functionals.

In chap. IV, §3, we saw that in 1897 C. Bourlet solved for the first time the problem of the determination of a linear map $U : E \to F$ between "function spaces" by conditions of *continuity*. In a short Note published in 1903 ([**094**], vol. I, p. 405–408) Hadamard attacked the same problem with $E = \mathcal{C}([a,b])$, space of real continuous functions in an in terval $[a,b]$, $F = \mathbb{R}$, and "continuity" means for him that $U(f_n)$ tends to $U(f)$ when f_n tends to f *uniformly*. He chooses a fixed function F such that for any continuous function f, one has

$$f(x) = \lim_{n\to\infty} n \int_a^b f(t)F(n(t-x))dt \tag{1}$$

uniformly in x; one has then

$$U(f) = \lim_{n\to\infty} \int_a^b f(t)\Phi_n(t)dt \tag{2}$$

where $\Phi_n(t)$ is the value of U at the function $x \mapsto nF(n(t-x))$ one may take $F(x) = e^{-x^2}$, so that Φ_n is continuous, but the choice of F is largely arbitrary (the argument is a typical case of what later will be called a "regularization" process).

In two papers published in 1904 and 1905, Fréchet gave another proof of Hadamard's theorem, and, what is more interesting, began to investigate the similar problems when $\mathcal{C}([a,b])$ is replaced by another "function space"; for instance [**070**], he remarked that if one takes for E the space $\mathcal{B}([a,b])$ of all bounded integrable functions in $[a,b]$ (continuous or not) with the topology of uniform convergence, there were other continuous linear functionals than those given by Hadamard's formula, for instance the mappings $f \mapsto c_1 f(x_1) + ... + c_m f(x_m)$, where the x_j are arbitrary points of $[a,b]$ and the c_j constants. Similarly, if one takes for E the space of all C^r functions in $[a,b]$, where convergence means uniform convergence for the function and its derivatives up to order r, Fréchet showed that the continuous linear functionals could then be written

$$f \mapsto c_0 f(a) + c_1 f'(a) + ... + c_{r-1} f^{(r-1)}(a) + \lim_{n\to\infty} \int_a^b f^{(r)}(t)\Phi_n(t)dt.$$

As soon as the study of Hilbert space began (chap. V, §3), Fréchet [**072**] and F. Riesz ([**183**], vol. I, p. 386–388) independently showed that continuous "linear functionals" on Hilbert space $l^2_{\mathbb{R}}$ (for the strong topology) could be written uniquely as $x \mapsto (x|a)$ for a vector $a \in l^2$.

Finally, in 1909, F. Riesz ([**183**], vol. I, p. 400–402) was able to give a better form to Hadamard's theorem by removing the arbitrariness of the sequence (Φ_n); his idea was to use the Stieltjes integral, as Hilbert had done in his work on spectral theory (see chap.

VIII, §2): he showed that any continuous linear functional $U : \mathcal{C}([a,b]) \to \mathbb{R}$ could be written *uniquely*

$$U : f \mapsto \int_a^b f(x)d\alpha(x) \tag{3}$$

where α is a function of bounded variation in $[a,b]$, provided one imposed on α the additional conditions of being continuous on the left and such that $\alpha(a) = 0$. His method consists in considering, for any $t \in [a,b]$, the function $f_t \in \mathcal{C}([a,b])$ equal to $x - a$ for $a \leq x \leq t$, and to $t - a$ for $t \leq x \leq b$, and the function $A : t \mapsto U(f_t)$; he shows that this function is Lipschitzian, and takes for $-\alpha(t)$ one of the "derived numbers" of A at the point t; it is then easy to show that α is a function of bounded variation, and it is a standard procedure to modify it in such a way that it satisfies the additional conditions meuntioned above without changing U.

Although the contemporaries did not realize the novelty of F. Riesz's approach, we are justified in seeing in his results (as he himself did) a radical departure from the conceptions of linear algebra prevalent in his time:

1° Whereas, even for the space L^2, it was possible, due to the Fischer–Riesz theorem, to identify the elements of the space with sequences of numbers, generalizing the dominant Cayley concept of linear algebra as a theory of "n-tuples", no such identification was possible for $\mathcal{C}([a,b])$, where one had to work directly on vectors, and not on their "coordinates"

2° Functions of bounded variation may be discontinuous at a denumerable set of points, and therefore it was impossible to *identify* any more the continuous linear functionals on $\mathcal{C}([a,b])$ to the elements of that space (again in contrast to what happened in l^2 according to the Riesz–Fréchet theorem).

These features would be still more conspicuous in the theory of L^p and l^p spaces, which F. Riesz began to investigate in 1910 ([**183**], vol. I, p. 403).

§2. The L^p and l^p spaces.

Once the L^2 spaces had been defined, it was a natural generalization to define similarly the function spaces $L^p(I)$ for any interval $I \in R$, as the set of all complex valued measurable functions f defined in I and such that $|f|^p$ is integrable, for any $p > 0$ (two functions being identified if they are almost everywhere equal). The study of these spaces was begun by F. Riesz in a fundamental paper ([**183**], vol. I, p. 441–497), second only in importance for the development of Functional Analysis to Hilbert's 1906 paper (chap. V, §2).

Riesz limited himself from the start to the casze $p > 1$, in order to be able to use the Hölder and Minkowski inequalities

$$\left|\sum_{k=1}^n a_k b_k\right| \leq \left(\sum_{k=1}^n |a_k|^p\right)^{1/p} \left(\sum_{k=1}^n |b_k|^q\right)^{1/q} \quad \text{for } \frac{1}{p} + \frac{1}{q} = 1 \tag{4}$$

$$\left(\sum_{k=1}^n |a_k + b_k|^p\right)^{1/p} \leq \left(\sum_{k=1}^n |a_k|^p\right)^{1/p} + \left(\sum_{k=1}^n |b_k|^p\right)^{1/p} \tag{5}$$

which he first extended to measurable functions, showing that if $f \in L^p(I), g \in L^q(I)$ then fg is integrable and

$$\left|\int_I f(x)g(x)dx\right| \leq \left(\int_I |f(x)|^p dx\right)^{1/p} \left(\int_I |g(x)|^q dx\right)^{1/q}, \tag{6}$$

and that if $f \in L^p(I), g \in L^p(I)$, then $f+g \in L^p(I)$ and

$$\left(\int_I |f(x)+g(x)|^p dx\right)^{1/p} \leq \left(\int_I |f(x)|^p dx\right)^{1/p} + \left(\int_I |g(x)|^p dx\right)^{1/p}. \tag{7}$$

His central theme is the study of infinite systems of linear equations

$$\int_I f(x) g_\alpha(x) dx = c_\alpha \tag{8}$$

where the g_α belong to $L^q(I)$ and one looks for a solution $f \in L^p(I)$; this may be considered as the generalization of the problem E. Schmidt had treated in l^2, due to the Fischer–Riesz theorem (chap. V, **??**, equations (42)). In order to adapt Schmidt's method to this problem, F. Riesz begins by extending a number of definitions and results from the theory of Hilbert space: strong convergence of a sequence (f_n) of functions of $L^p(I)$ to $f \in L^P(I)$ is defined as meaning that $\int_I |f(x)-f_n(x)|^p dx$ tends to 0. For weak convergence, he first takes as definition that $\int_a^x f_n(t)dt$ tends to $\int_a^x f(t)dt$ for all numbers $x \in I$; and although he proves a little later that this definition is equivalent to the fact that the integrals $\int_I (f(x)-f_n(x))g(x)dx$ tend to 0 for *all* $g \in L^q(I)$, he essentially uses the first definition to prove the genera lization of Hilbert's "principle of choice" (*i.e.* the weak compactness of the unit ball in $L^p(I)$), which will be one of his main ingredients in the solution of (8). The other ingredient is derived from a result obtained by E. Landau in 1907 [**136**]; in 1906, Hellinger and Toeplitz had shown that if a sequence (a_n) is such that the series $\sum_n a_n x_n$ is convergent for *all* sequences (x_n) in l^2, then (a_n) itself be longed to l^2 [**106**]; Landau proved that, more generally, if $\sum_n a_n x_n$ is convergent for all sequences x_n such that $\sum_n |x_n|^p < +\infty$, then $\sum_n |a_n|^q < +\infty$. Approximating functions of L^p by functions having only denumerably many values, F. Riesz deduced from Landau's result that if, for a measurable function g, the product fg is integrable for all functions $f \in L^p$, then necessarily $g \in L^q$.

His solution of (8) then proceeds along the same lines as E. Schmidt; he starts with a finite system (8), for which, using the standard method of analysis (Lagrange multipliers) he proves the existence and uniqueness of a solution $f \in L^p$ for which $\int_I |f(x)|^p dx$ is minimum. The problem is then to find a necessary and sufficient condition on the c_α such that, when one picks from (8) a finite system corresponding to the indices α in an arbitrary finite subset H, the corresponding minima M_H of the integral $\int_I |f(x)|^p dx$ taken for the "minimal" solutions, are *uniformly bounded* (independently of H); the use of the two ingredients mentioned above then leads to the existence of a solution of (8) by an argument similar to E. Schmidt's. Of course, an explicity expression of M_H (similar to formula (43) of chap. V, §3) is not available here, and the originality of F. Riesz lies in having found a completely different type of condition, namely the existence of a number $M > 0$ such that, for *any* finite subset H of indices, and *any* family $(\lambda_\alpha)_{\alpha in H}$ of scalars, one has the inequality

$$\left|\sum_{\alpha \in H} \lambda_\alpha c_\alpha\right| \leq M \cdot \left(\int_I \sum_{\alpha \in H} |\lambda_\alpha g_\alpha(x)|^q dx\right)^{1/q}. \tag{9}$$

F. Riesz in particular applies this conditions to the special case in which the g_α are *all* the functions of $L^q(I)$; (9) is then equivalent to the *continuity* in L^q of the linear functional L defined by $L(g_\alpha) = c_\alpha$, and he has thusgene ralized his previous results

on l^2 and $\mathcal{C}(I)$, proving what we would now express by the statement that *the dual of* $L^q(I)$ *can be identified with* $L^p(I)$.

Of course the name "dual" is not yet used by F. Riesz, but he explicitly considers, for a "bounded" linear mapping T of L^p into itself (defined by the condition that $\int_I |T(f)(x)|^p dx$ remains bounded for all f such that $\int_I |f(x)|^p dx < 1$), the *transposed* mapping T' defined by the equation

$$\int_I T(f)(x)g(x)dx = \int_I f(x)T'(g)(x)dx \quad \text{for } \textit{all } f \in L^p(I). \tag{10}$$

Indeed, for a function $g \in L^q(I)$, this defines (up to a null set) a unique function $T'(g)$, which (by F. Riesz's previous results) also belongs to $L^q(I)$; furthermore, it is easy to show that the mapping T' of L^q into itself is also linear and "bounded". F. Riesz then used this concept to obtain a necessary and sufficient condition for the mapping T to be *bijective*: he showed that such a condition is the existence of a number $m > 0$ such that *both* inequalities

$$\begin{aligned} \int_I |T(f)(x)|^p dx &\geq m \cdot \int_I |f(x)|^p dx \\ \int_I |T'(g)(x)|^q dx &\geq m \cdot \int_I |g(x)|^q dx \end{aligned} \tag{11}$$

are satisfied for all $f \in L^p(I)$ and all $g \in L^q(I)$. F. Riesz had thus given, for the first time, examples of what we now call *reflexive* Banach spaces not isomorphic to their dual(*). In his 1913 book on infinite systems of linear equations ([**183**], vol. II, p. 835–1016 and [**184**]) he treated in a similar way the l^p spaces for $p > 1$ (defined as the set of sequences (x_n) of complex numbers such that $\sum_n |x_n|^p < +\infty$); in addition he stated without proof that for $p \neq 2$, no isomorphism of l^p and L^p existed any more, in contradistinction to the Fischer–Riesz theorem (*ibid.* Vol. I, p. 444-445).

§3. The birth of normed spaces and the Hahn–Banach theorem.

In 1911, F. Riesz combined his methods for the treatment of the system (8) in L^p with the Hadamard–Riesz theorem on linear functionals in $\mathcal{C}([a,b])$ in order to study the systems of linear equations

$$\int_a^b g_\alpha(x)d\xi(x) = c_\alpha \tag{12}$$

where $[a,b]$ is a compact interval in $\mathbb{R}$, the g_α, are continuous in $[a,b]$, the c_α are given scalars, and one has to determine a function ξ of bounded variation in $[a,b]$ satisfying the equations (12) for all α. This may be considered as a generalization of a problem which had first been proposed and solved by T. Stieltjes in 1894, the "moment problem": it consists in determining an increasing function ξ in $[0,+\infty[$ such that

$$\int_0^\infty x^n d\xi(x) = c_n \geq 0 \quad \text{for } n = 0,1,2,... \tag{13}$$

(*) The dual of $L^1(I)$ for a compact interval $I \in \mathbb{R}$ was shown to be isomorphic to $L^\infty(I)$ by H. Steinhaus [**202**]; he uses the fact that in that case $L^2(I) \subset L^1(I)$, and therefore a continuous linear functional on $L^1(I)$ is also continuous on $L^2(I)$.

(the left hand sides are called the "moments" of the function ξ, a terminology stemming from probability theory) [**205**]; the same problem was later considered when the interval $[0, +\infty[$ is replaced by $]-\infty, +\infty[$ (the "Hamburger moment problem") or by a compact interval $[a, b]$ (the "Hausdorff moment problem") [3].

The solutions to these "moment problems" consist in giving explicit conditions on the c_n involving existence (or existence and uniqueness) of the function ξ (or rather of the measure $d\xi$). The condition given by F. Riesz for the existence of a solution ξ of the general system (12) is similar to condition (9), namely the existence of a number $M > 0$ such that, for any finite family $(\lambda_\alpha)_{\alpha\in H}$ of scalars, one has

$$\left|\sum_{\alpha\in H}\lambda_\alpha c_\alpha\right| \leq M\cdot\sup_x\left|\sum_{\alpha\in H}\lambda_\alpha g_\alpha(x)\right|; \tag{14}$$

(he explicitly observed that the right hand side of this inequality is the limit of the right hand side of (9) when q tends to $+\infty$). His proof is similar to the proof for (8); he first restricts himself to the case of finite systems (12), obtains the existence of a "minimal" solution of such a system, and then, using a "principle of choice" (in our language, the weak compactness of the unit ball in the space of Stieltjes measures), he shows that the condition (14) is sufficient for an arbitrary system (12); his procedure is more complicated than for (8), because even in the case of a finite system (12), there is no more uniqueness for the "minimal" solutions ([**183**], vol. II, p. 798–827).

We now interpret condition (9) in the following way: first, if $\sum_\alpha \lambda_\alpha g_\alpha = 0$ in $L^q(I)$, then $\sum_\alpha \lambda_\alpha c_\alpha = 0$; this implies that, if F is the vector subspace of $L^q(I)$ generated by the g_α, there is a well determined linear form L defined in F such that $L(g_\alpha) = c_\alpha$ for every α. Condition (9) then means that this linear form L is *continuous* in F; the existence of an $f \in L^p(I)$ such that $L(g_\alpha) = \int_I f(x)g_\alpha(x)dx$ for all α then means that L can be *extended* to a continuous linear form defined in the *whole* space $L^q(I)$; in other words, it is a special case of what we now call the Hahn–Banach theorem. There is a similar intecrpretation of condition (14), replacing $L^q(I)$ by $\mathcal{C}([a, b])$.

Such an interpretation of his results was *not* given by F. Riesz; the first mention of that point of view appears in a paper written in 1912 by the Austrian mathematician E. Helly (1884–1943), in which he gives a different proof of F. Riesz's results on the systems (12) [**107**].

TODO: he wrote [107 bis].

After an interval of 9 years (due to the first World War, in which he was a prisoner of war in Russia), Helly returned to his method in a paper of 1921 [**108**] which again should be considered as a landmark in the history of Functional Analysis, since instead of considering *special* spaces such as the l^p, L^p or $\mathcal{C}([a, b])$, he for the first time deals with general "normed sequence spaces" by methods which do not depend on special features of the space, contrasting with the ones used by E. Schmidt and F. Riesz.(*)

Helly considers vector subspaces of the vector space $\mathbb{C}^N$ of all sequences of complex numbers, and assumes that on such a subspace E there has been defined a *norm* $\|x\|$ (he does not use that name nor the notation) such that: 1) $\|x\| \geq 0$ and the relation $\|x\| = 0$ is equivalent to $x = 0$; 2) $\|\lambda x\| = |\lambda|\cdot\|x\|$ for any scalar λ; 3) $\|x + y\| \leq \|x\| + \|y\|$; this defines on E a distance $d(x, y) = \|x - y\|$ in the sense of Fréchet. Of course norms

(*) It seems that during the period 1910–1920, F. Riesz always had in mind possible axiomatic generalizations of his results, although he did not publish anything in that direction ([**183**], vol. I, p. 452).

had been defined in the spaces l^p, L^p anda $\mathcal{C}([a,b])$; but Helly seems to be the first to have noticed the relations of that notion with the concepts of *convexity* introduced earlier by Minkowski in his "Geometry of numbers ([**161**] and [**162**]). He had shown that the concept of norm on a *finite* dimensional space $\mathbb{R}^n$ (with the scalars limited to real values) was equivalent to the notion of "symmetric convex body", i.e. a closed, symmetric, bounded convex set in which the origin O is an interior point: such a set B can be defined by an inequality $p(x) \leq 1$, where p is a uniquely determined norm. The boundary of such a set is defined by the equation $p(x) = 1$, and Minkowski had proved that for each point x_0 of that boundary, there existed at least one *hyperplane of support* H, containing x_0, and such that B lies entirely *on one side* of H. If, for an n-tuple of real numbers $u = (u_1, \ldots, u_n)$ and a point $x = (x_1, \ldots, x_n)$, one writes $\langle u, x\rangle = u_1x_1 + \ldots + u_nx_n$, the equation of H has the form $\langle u, x\rangle = 1$ for a suitable u, and one has the inequality $\langle u, x\rangle \leq p(x)$ for all $x \in \mathbb{R}^n$, with $\langle u, x_0\rangle = p(x_0)$; the n-tuples u being identified with the corresponding linear forms $x \mapsto \langle u, x\rangle$ on $\mathbb{R}^n$, Minkowski had also defined the "support function" $q(u) = \sup_{x\neq 0}\langle u, x\rangle/p(x)$, and shown that it was also a norm on $\mathbb{R}^n$, "dual" to p and such that the hyperplanes of support of B are the hyperplanes $\langle u, x\rangle = 1$ with $q(u) = 1$; furthermore, p is the norm "dual" to q, in other words $p(x) = \sup_{u\neq 0}\langle u, x\rangle/q(u)$.

To transfer to spaces of sequences these concepts and definitions, Helly associates to E the subspace E' of $\mathbb{C}^{\mathbb{N}}$ con sisting of all the sequences $u = (u_n)$ such that the series $\sum_n u_nx_n$ converges for *all* $x = (x_n)$ in E [(*)], and he then considers $\langle u, x\rangle = \sum_n u_nx_n$. For any $u \in E'$, the number $\|u\| = \sup_{x\neq 0}\langle u, x\rangle/\|x\|$ defines a norm on E' provided it is not 0 for some eiements $u \neq 0$. Excluding that case, Helly first obtains a weak generalization of Minkowski's result on the hyperplanes of support; if B is the subset of E defined by $\|x\| \leq 1$, he shows that the hyperplane H defined by $\langle u, x\rangle = 1$ meets B if $\|u\| < 1$, does not meet B for $\|u\| > 1$, but if $\|u\| = 1$, the intersection $H \cap B$ may very well be empty: an example is given by taking $E = l^1, E' = l^\infty$ and for H the hyperplane $\sum_{n=1}^{\infty}\left(1 - \frac{1}{n}\right)x_n = 1$.

The central problem in Helly's paper is the solution of a system

$$\langle u^{(\nu)}, x\rangle = c_\nu \qquad (\nu = 1, 2, \ldots) \tag{15}$$

where the $u^{(\nu)}$ belong to E' and one looks for a solution $x \in E$. The inequality $|\langle u, x\rangle| \leq \|u\| \cdot \|x\|$ immediately yields the necessary condition similar to Riesz's conditions (9) and (14), namely the existence of a number $M > 0$ such that

$$\left|\sum_{\nu=1}^{n}\lambda_\nu c_\nu\right| \leq M \cdot \left\|\sum_{\nu=1}^{n}\lambda_\nu u^{(\nu)}\right\| \tag{16}$$

for any n and *all* choices of scalars λ_ν; but the example given above (for a single equation) shows that there may well be no solution such that $\|x\| = M$, even when condition (16) is satisfied.

Helly, as Schmidt and F. Riesz had done, first considers the case of a *finite* system (15) of N equations, where as usual the $u^{(\nu)}$ may be supposed to be linearly independent. The mapping $f : x \mapsto (\langle u^{(\nu)}, x\rangle)_{1\leq\nu\leq N}$ of E into $\mathbb{C}^{\mathbb{N}}$ is then surjective; Helly shows that on $\mathbb{C}^{\mathbb{N}}$, $\|y\| = \inf_{f(x)=y}\|x\|$ is a norm (for us it is the natural norm on $E/f^{-1}(0)$ deduced

---

(*) This is not always the dual of E as we now understand that word.

fram the norm on E); condition (16) then guarantees the existence of a solution x of (15) such that $\|x\| < M_1$ for *any* $M_1 > M$ (if not necessarily for $M_1 = M$).

The passage from finite systems (15) to the general case is the most original idea of Helly; he splits the problem in two:

A) Given $M_1 > M$, find a linear form $L : E' \to \mathbb{C}$, such that $|L(u)| \leq M_1 \cdot \|u\|$ for all $u \in E'$ and such that $L(u^{(\nu)}) = c_\nu$ for all ν.

B) When such a linear form L has been found, find if possible an element $p \in E$ such that $\langle p, u\rangle = L(u)$ for all $u \in E'$.

To treat problem A), Helly assumes the additional condition that E' is *separable* as a metric space; he then proves the existence of a solution (a special case of the Hahn–Banach theorem) in the following way. Let $(p^{(\nu)})$ be a sequence of elements of E' which is dense in that space. Helly chooses an increasing sequence $M < M^{(1)} < M^{(2)} < \ldots < M_1$ of numbers, and the main point of his proof consists in showing that there exists a family (γ_ν) of complex numbers such that, for any pair of integers $m \geq 1, n \geq 1$, and any pair of families $(\lambda_\nu), (\mu_\nu)$ of scalars, one has

$$\left|\sum_{\nu=1}^{n} \lambda_\nu c_\nu + \sum_{\nu=1}^{m} \mu_\nu \gamma_\nu\right| \leq M^{(m)} \cdot \left\|\sum_{\nu=1}^{n} \lambda_\nu u^{(\nu)} + \sum_{\nu=1}^{m} \mu_\nu p^{(\nu)}\right\|. \tag{17}$$

It is then easy to show that there exists a linear form L on E' such that $L(p^{(\nu)}) = \gamma_\nu$ for all indices ν, and that it is a solution of problem A).

The proof of (17) is done by induction on m, the case $m = 0$ being the assumption (16). One has then to prove the existence of a point $\gamma_{m+1} \in \mathbb{C}$ which, for any integer $n \geq 1$ and any pair of families of scalars $(\lambda_\nu)_{1\leq\nu\leq n}, (\mu_\nu)_{1\leq\nu\leq m}$, belongs to *all* disks defined in $\mathbb{C}$ by

$$\begin{aligned}&\left|\sum_{\nu=1}^{n} \lambda_\nu c_\nu + \sum_{\nu=1}^{m} \mu_\nu \gamma_\nu + \gamma_{m+1}\right| \leq \\ &\leq M^{(m+1)} \left\|\sum_{\nu=1}^{n} \lambda_\nu u^{(\nu)} + \sum_{\nu=1}^{m} \mu_\nu p^{(\nu)} + p^{(m+1)}\right\|.\end{aligned} \tag{18}$$

However, a general result on convex sets in a finite dimensional space, proved by Helly himself, reduces that question to proving that *any three* of the disks (18) have a common point; and this is shown by Helly to be a consequence of the result proved before for *finite* systems (15).

Turning to problemn B), Helly discovers that it is quite possible that it has *no* solution; in our language, he gives the first example of *non reflexive* Banach spaces(*). That example is the space E of all sequences (x_n) such that the series $\sum_{k=1}^{\infty} x_k$ converges, with the norm $\|x\| = \sup_n \left|\sum_{k=n}^{\infty} x_k\right|$; Helly proves that E' consists of all sequences (u_k) such that $\|u\| = |u_1| + \sum_{k=1}^{\infty} |u_{k+1} - u_k|$ is finite, $\|u\|$ being the natural norm on E'; then if one takes $L(u) = \lim_{k\to\infty} u_k$, L is continuous on E' but there is no $p \in E$ such that $L(u) = \langle p, u\rangle$.

(*) F. Riesz had already observed that one could define on the space of functions of bounded variation continuous linear funcionals which were not of the form $\xi \mapsto \int_a^b f(x)d\xi(s)$ for a continuous function f (for instauce one can take for f an increasing discontinuous function) ([**183**], vol. II, p. 827).

Starting from the work of F. Riesz and Helly, it was a natural generalization to define norms on *arbitrary* vector spaces over $\mathbb{R}$ or $\mathbb{C}$, and not only on spaces of functions or on subspaces of $\mathbb{C}^{\mathbb{N}}$. This was done independently by H. Hahn [**097**] and S. Banach [12], who restrict themselves to *complete* spaces.

Banach's paper is his thesis, written in 1920: although he does not mention convexity, he is careful to develop and extensively use a geometric language. He is mainly interested in continuous linear operators $u : E \to F$, where E and F are arbitrary normed complete spaces, and in limits of sequences of such operators. Hahn's point of view is similar, although he is only concerned with linear forms; neither he nor Banach are at that moment interested in the problem of extension of linear forms, and we postpone a more detailed discussion of their papeirs of 1922–23 to §4. We should however mention that in his thesis Banach gives the "abstract" formulation of the method of successive approximations (chap. II, §1) as a "contraction principle": if F is a mapping of a complete normed space E into itself such that $\|F(x) - F(y)\| \leq k\|x - y\|$ with $0 < k < 1$, then the sequence (x_n) defined by induction as $x_{n+1} = F(x_n)$ (x_0 arbitrary) converges to the unique "fixed point" x, such that $F(x) = x$.

It was only in 1927 that Hahn returned to Helly's paper, in the general context of complete normed spaces, and completely solved the extension problem for such spaces [**098**]. He proceeds by induction as Helly had done, but at the same time he greatly simplifies and generalizes the method by introducing, for the first time in general problems of Functional Analysis [(*)], *transfinite induction* instead of the ordinary kind. In a complete normed space E, one has a vector subspace V and there is defined on V a (real valued) linear form f such that $|f(x)| \leq M\|x\|$ for $x \in V$; the problem is to extend f to a linear form F on E such that $|F(y)| < M\|y\|$ for $y \in E$. Hahn begins by showing the existence of an ordinal γ, and of a mapping $\xi \mapsto V_\xi$, which, to very ordinal $\xi < \gamma$ associates a vector subspace V_ξ of E such that $V_0 = V, V_\xi \subset V$ for $\xi < \eta, V_\xi$ has codimension 1 in $V_{\xi+1}$, and E is the union of the V_ξ for $\xi < \gamma$. The problem is then easily reduced to the case in which V has codimension 1 in E, and then E is generated by V and an element $a \notin V$; Hahn coasiders the l.u.b B of the numbers $f(x) - M\|x - a\|$ for $x \in V$, and the g.l.b. A of the numbers $f(x) + M\|x - a\|$ for $x \in V$, and, using the assumption $|f(x)| \leq M\|x\|$ for $x \in V$, he easily shows that $A \leq B$; the extension F is then defined by $F(x + \lambda a) = f(x) + \lambda c$ for all $\lambda \in \mathbb{R}$, where c is any number such that $A \leq c \leq B$.

As a particular case of his theorem, Hahn shows that for any vector $a \neq 0$ in E there exists a continuous linear form L on E such that $\|L\| = 1$ and $L(a) = \|a\|$; he then formally introduces the *dual* space E' of E ("polare Raum" in his terminology) which is not reduced to 0 due to the preceding result; he writes $B(u, x)$ instead of $u(x)$ for $x \in E, u \in E'$, and considers for any $x \in E$, the linear form $c(x) : u \mapsto B(u, x)$ on E', for which he shows that $\|c(x)\| = \|x\|$. In other words, he has defined a linear isometry c of E into its second dual E'', and he says a space E is "regulär" if c is bijective (our reflexive spaces). It may therefore rightly be said that with this paper of Hahn, duality theory at last has come into its own.

Two years later, Banach, who apparently was not aware of Hahn's paper, published the same theorem with the same proof (he later acknowledged Hahn's priority); in addition, he re cognized that the argument could be generalized: if p is a real valued function

---

[(*)] Transfinite induction had been used by analysts ever since Cantor, but the application of transfinite induction closest to Hahnt's is probably the method by which Banach, in 1923, had proved the existence on $\mathbb{R}$ of a "measure" defined on all subsets of $\mathbb{R}$ and simply additive [13].

defined in a vector space E and such that $p(x+y) \leq p(x) + p(y)$ and $p(\lambda x) = \lambda p(x)$ for $\lambda \geq 0$, and if f is a linear form defined in a vector subspace V of E and such that $f(x) \leq p(x)$ in V, then it is possible to extend f to a linear form F defined in E and such that $F(x) \leq p(x)$ in E. This extension was to play later an important role in the development of the theory of locally convex spaces (cf. chapter VIII).

§4. The method of the gliding hump and Baire category.

In his 1922 paper [**097**], Hahn proved the following theorem: let E be a complete normed space, (u_n) a sequence of continuous linear forms on E, and suppose that for each $x \in E$, the sequence of numbers $|u_n(x)|$ is bounded by a number *depending on* ; then the sequence of the norms $\|u_n\|$ is *bounded.* The proof is by contradiction; assuming that the sequence $(\|u_n\|)$ is unbounded, one determines by induction a sequence (x_k) in E and a sequence (n_k) of integers such that:

1° the series $\sum_{k=1}^{\infty} x_k$ converges to an element $x \in E$;

2° $\sum_{j=k+1}^{\infty} \left|u_{n_k}(x_j)\right| \leq 1$;

3° $\left|u_{n_k}(x_k)\right| \geq k + \sum_{j=1}^{k-1} \left|u_{n_k}(x_j)\right|$.

Then one has for each k,

$$\left|u_{n_k}(x)\right| \geq \left|u_{n_k}(x_k)\right| - \sum_{j=1}^{k-1} \left|u_{n_k}(x_j)\right| - \sum_{j=k+1}^{\infty} \left|u_{n_k}(x_j)\right| \geq k-1$$

which contradicts the assumption. To do this, one assumes the u_{n_j} have been determined for $j < k$, and one considers a ball

$$B_k : \|x\| \leq 2^{-k} \cdot \inf_{j<k} \left(\|u_{n_j} + 1\|\right)^{-1}$$

in E; the assumption that $(\|u_n\|)$ is unbounded guarantees the existence of an index n_k and a point $x_k \in B_k$ for which condition 3° holds; conditions 1° and 2° are then deduced from the choice of the radius of the ball B_k.

This is often called the "method of the gliding hump": in the sequence of values $|u_{n_k}(x_j)|$ when j varies from 1 to $+\infty$, the index $j = k$ corresponds to a "hump" much bigger than the sum of the contributions of the other indices.

The result can be put in a different form: if the sequence $|u_{n_k}(x_j)|$ is unbounded, there exists at least one $x \in E$ such that the sequence $|u_{n_k}(x)|$ is unbounded.

In this form, the first example of the method of the gliding hump is probably the way in which Lebesgue, in 1905 ([**138**] and [**139**]) constructed a continuous periodic function $F(x)$ in $[0, 2\pi]$ whose Fourier series diverges at the point 0. He had proved that, if one writes $S_n(g)$ for the sum of the first n terms of the Fourier series of a continuous function g, it is possible to find a sequence (g_n) of continuous periodc functions of bounded variation such that $|g_n(t)| \leq 1$ in $[0, 2\pi]$ and that the sequence of values $S_n(g_n)(0)$ tends to $+\infty$. He then defines

$$F(x) = \varepsilon_1 f_1(n_1 x) + \varepsilon_2 f_2(n_2 x) + \ldots + \varepsilon_k f_k(n_k x) + \ldots$$

where the ε_k are > 0 and such that $\sum_{k=1}^{\infty} \varepsilon_k = 1$, the f_k are continuous periodic functions of bounded variation such that $f_k(t)| \leq 1$ in $[0, 2\pi]$ and that $\left|S_{p_k}(f_k)(0)\right| \geq k/\varepsilon_k$, for an increasing sequence (p_k) of integers. Finally the increasing seguence (n_k) of integers

is chosen in such a way that $n_k > n_{k-1}p_{k-1}$ and that, for the continuous function of bounded variation $F_k(x) = \varepsilon_1 f_1(n_1 x) + \ldots + \varepsilon_{k-1} f_{k-1}(n_{k-1}x)$, all the sums $S_n(F_k)(0)$ are ≤ 2 in absolute value for $n \geq n_k$, (they converge to $F_k(0)$). This choice implies that, for $j > k$, the sum of the first $n_k p_k$ terms of the Fourier series of $f_j(n_j x)$ is reduced to the *first* term of the series, hence is ≤ 1 in absolute value; using these definitions it is easy to check that $\left|S_{n_k p_k}(F)(0)\right| \geq k - 3$ for all k.

One year later, Hellinger and Toeplitz, two students of Hilbert, found a rather surprising complement to the definition he had given of a *bounded* bilinear form (chap. V, §2); instead of assuming that $|K_n(x,y)| \leq M$ for all n and all $x = (x_p)$ and $y = (y_p)$ such that $\sum_p x_p^2 \leq 1$ and $\sum_p y_p^2 \leq 1$, they showed that it was enough to assume that for each such pair (x,y), one had $|K_n(x,y)| < M_{x,y}$ for all n, where the number $M_{x,y}$ *might depend* on x, y *in an arbitrary way*. Independently of Lebesgue, they proved that result by a "gliding hump" method, constructing a pair (x,y) for which the sequence $(|K_n(x,y)|)$ is unbounded if K is not a *bounded* form in Hilbert's sense [**106**].

During the next 20 years, many more examples of the "gliding hump" method appeared in the literature: Lebesgue used it repeatedly in a 1909 paper on "singular integrals" ([**138**], vol. III, p. 259–351), where one looks for conditions on "kernels" K_n insuring that the integrals $\int_a^b f(t)K_n(t,x)dt$ tend to $f(x)$ when n tends to $+\infty$, for various kinds of function f. The method was also prominent in the study of "summation processes", where one "transforms" a sequence (x_n) into a sequence (y_n) by the formulas $y_n = \sum_{m=1}^{\infty} a_{nm}x_m$, and has to look for conditions on the a_{nm} insuring that when (x_n) has a limit, (y_n) tends to the same limit ([**193**], vol. II, p. 389–321). Hahn's paper of 1922 [**097**] was written to give a general background to all these results, showing that they all were consequences of his general theorem. Independently, Banach, in his thesis, proved a theorem more general than Hahn's, the u_n being now continuous linear operators from a complete normed space E into a complete normed space F; he showed that the assumption that the norms $\|u_n(x)\|$ are bounded for each x by a number depending on x, implies that the sequence of the norms $\|u_n\|$ is bounded.

Finally, in 1927, Banach and Steinhaus (using an idea of Saks) discovered that this theorem could be proved without using the "gliding hump" method, by an application of a theorem Baire had proved in 1899 [11]: he had shown that in $\mathbb{R}^n$, the intersection of a denumerable family of dense open subsets is itself dense(*); this implies that if u is a real function defined and *lower semi-continuous* in $\mathbb{R}^n$, and if $u(x) < +\infty$ for each $x \in \mathbb{R}^n$, then any non empty open subset U of $\mathbb{R}^n$ contains a non empty open subset V such that $\sup_{x\in V} u(x) < +\infty$. These results and their proofs immediately generalize when $\mathbb{R}^n$ is replaced by an arbitrary *complete metric space*. If now H is a set of linear mappings from a complete normed space E into a complete normed space F, and if for each $x \in E$, $\sup_{u\in H}\|u(x)\| < +\infty$, the function $p(x) = \sup_{u\in H}\|u(x)\|$ is lower semi-continuous, and from the Baire theorem it follows that p is bounded in a neighborhood of 0, which implies that $\sup_{u\in H}\|u\|$ is finite [**16**].

§5. Banach's book and beyond.

In 1932 S. Banach published a book [15] containing a comprehensive account of all results known at that time in the theory of normed spaces, and in particular the theorems he had published in his papers of 1923 and 1929. A large part was devoted to

(*) For $n = 1$, the same result had been proved two years earlier by W. Osgood [**170**].

the concept of weak convergence and its generalizations, which he had begun to study in 1929; we shall postpone to chap. VIII, §1 the discussion of these questions. The most remarkable result contained in that book is another consequence of Baire's theorem, discovered by Banach, and much deeper than the Banach–Steinhaus theorem: if u is a continuous linear mapping from a complete normed space E into a complete normed space F, then either $u(E)$ is *meager* in F (a set "of first category" in the terminology of Baire), or $u(E) = F$. An immediate consequence is the famous *closed graph theorem*: if u is a linear mapping from E to F having a closed graph in $E \times F$, then u is continuous. These surprising results have become two of the most powerful tools in all applications of Functional Analysis.

These features, as well as many applicatiouns to classical Analysis, gave the book a great appeal, and it had on Functional Analysis the same impact that van der Waerden's book had on Algebra two years earlier. Analysts all over the world began to realize the power of the new methods and to apply them to a great variety of problems; Banach's terminology and notations were universally adopted, complete normed spaces became known as *Banach spaces*, and soon their theory was considered as a compulsory part in most curricula of graduate students. After 1935, the theory of normed spaces became part of the more general theory of *locally convex spaces*, which we shall discuss in chapter VIII; more recently however, there has been a renewed surge of interest in the special properties of normed spaces and their "geometry"; it is too soon, as yet, to have a clear idea of the scope of these results and of their relation to other parts of mathematics, and we refer the interested reader to [4], [17], [**047**], [**050**], [**116**], [**134**], [**149**], [**150**] and [**185**].

Chapter VII
Spectral theory after 1900

Chapter VIII

Locally convex spaces and the theory of distributions

§1. Weak convergence and weak topology.

In his thesis, Fréchet had already noticed that convergence in a metric space could not always correspond to some classical types of "convergence" for functions. For instance, if $\mathcal{B}(\mathbb{R})$ is the vector space of *all* bounded real functions on $\mathbb{R}$, it is not possible to define a distance on that space such that *simple* convergence in $\mathcal{B}(\mathbb{R})$ would be identical with convergence for that distance. This results from the fact that if A is a subset of a metric space E, the closure $\overline{A}$ of A in E is identical to the set of limits of all convergent sequences of elements of A. However, if one takes in $\mathcal{B}(\mathbb{R})$ the set $A = \mathcal{C}(\mathbb{R})$ of bounded continuous functions, the limits of sequences of elements of A for simple convergence are the Baire functions of class 1, and it is known that there are Baire functions of class 2 which are not of class 1, so that $\overline{A}$ (for the hypothetical distance) could not consist only of functions of class 1 [**071**].

There was thus an obvious need for a generalization of the concept of metric space, but none proved adequate for Functional Analysis until Hausdorff, in 1914, created "General topology" as we understand it now, based on the concept of neighborhood [**100**]; but surprizingly enough, it took some time to become aware of that adequacy. Ever since Hilbert, "weak convergence" of *sequences* had become a ceniral theme, first in Hilbert spaces, then with F. Riesz and Helly in some types of normed spaces (chap. VI), and one would have thought that Hausdorff's concept of topology would have been tested on that notion; but ultil 1934 the only mathematician who seems to have had that idea was von Neumann: he defined weak neighborhoods of a point x_0 in a Hilbert space E by a finite number of conditions $|\,(x - x_0|a_j)\,| \leq \varepsilon$ for points $a_j \in E$, and then went on to define similarly, in the algebra (E) of endomorphisms of E, "strong neighborhoods" of an operator U_0 by a finite number of conditions $\|(U - U_0) \cdot x_j\| \leq \varepsilon$ for $x_j \in E$, and "weak neighborhoods" by a finite number of con ditions $\|\big((U - U_0) \cdot x_j|y_j\big)\| \leq \varepsilon$ [**221**]. But he did not try to extend these ideas to other Banach spaces.

§2. Locally convex vector spaces.

§3. The theory of distributions.

Chapter IX

Applications of functional analysis to differential and partial differential equations

§1. Fixed point theorems.

§2. Carleman operators and generalized eigenvectors.

§3. Boundary problems for ordinary differential equations.

§4. Sobolev spaces and *a priori* inequalities.

§5. Elementary solutions, parametrices and pseudo-differential operators.

References

[1] N.H. Abel, *Oeuvres*, ed. by Sylow and Lie, 2 vols., Christiania, 1881.

[2] N. Adasch *et al.*, *Topological vector spaces*, Lecture Notes in Math. 639, Berlin-Heidelberg-New York: Springer, 1978, DOI: 10.1007/BFb0068514.

[3] N. Akhiezer, *The classical moment problem*, Edinburgh-London: Oliver and Boyd, 1965.

[4] E. Akin, *The metric theory of Banach manifolds*, Lecture Notes in Math. 662, Berlin-Heidelberg-New York: Springer, 1978, DOI: 10.1007/BFb0082264.

[5] L. Alaoglu, "Weak topologies of normed linear spaces", in: *Ann. of Math.* 41 (1940), p. 252–267, DOI: 10.2307/1968829.

[6] P. Alexandroff and P. Urysohn, "Zur Theorie der topologischen Räume", in: *Math. Ann.* 92 (1924), p. 258–266, DOI: 10.1007/BF01448008.

[7] N. Aronszajn and K.T. Smith, "Invariant subspaces of completely continuous operators", in: *Ann. of Math.* 60 (1954), p. 345–350, DOI: 10.2307/1969637.

[8] Giulio Ascoli, "Le curve limiti di una varietà data di curve", in: *Mem. Acc. dei Lincei*, (3) 18 (1883), p. 521–586.

[9] Guido Ascoli, "Sugli spazi lineari metrici e le loro varietà lineari", in: *Ann. di Mat.*, (4) 10 (1932), p. 33–81 and 203–232.

[10] M. Atiyah, "Elliptic operators, discrete groups and von Neumann algebras", in: *Astérisque* 32–33 (1976), p. 43–72, URL: https://www.numdam.org/item/AST_1976__32-33__43_0/.

[11] R. Baire, "Sur les fonctions de variables réelles", in: *Ann. di Mat.*, (3) 3 (1899), p. 1–273, DOI: 10.1007/BF02419243.

[12] S. Banach, "Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales", in: *Fund. Math.* 3 (1923), p. 133–181, DOI: 10.4064/fm-3-1-133-181.

[13] S. Banach, "Sur le problème de la mesure", in: *Fund. Math.* 4 (1923), p. 7–33, DOI: 10.4064/fm-4-1-7-33.

[14] S. Banach, "Sur les fonctionnelles linéaires", in: *Studia Math.* 1 (1929), p. 211–216 and 223–229, DOI: 10.4064/sm-1-1-211-216.

[15] S. Banach, *Théorie des opérations linéaires*, Warszawa, 1932.

[16] S. Banach and H. Steinhaus, "Sur le principe de condensation des singularités", in: *Fund. Math.* 9 (1927), p. 50–61, DOI: 10.4064/fm-9-1-50-61.

[17] *Banach spaces of analytic functions,* Kent, Ohio, 1976 Proceedings, Lecture Notes in Math. 604, Berlin-Heidelberg-New York: Springer, 1977, DOI: 10.1007/BFb0069197.

[18] K. Barbey and H. König, *Abstract analytic function theory and Hardy algebras*, Lecture Notes in Math. 593, Berlin-Heidelberg-New York: Springer, 1977, DOI: 10.1007/BFb0086694.

[19] R. Beals, "A general calculus of pseudodifferential operators", in: *Duke Math. J.* 42 (1975), p. 1–42, DOI: 10.1215/S0012-7094-75-04201-5.

[20] R. Beals and C. Fefferman, "Spatially inhomogeneous pseudodifferential operators I", in: *Comm. Pure Appl. Math.* 27 (1974), p. 1–24, DOI: 10.1002/cpa.3160270102.

[21] J. Benedetto, *Spectral synthesis*, New York: Academic Press, 1975, DOI: 10.1007/978-3-322-96661-2.

[22] G.D. Birkhoff and O. Kellogg, "Invariant points in function space", in: *Trans. Amer. Math. Soc.* 23 (1922), p. 96–115, DOI: 10.1090/S0002-9947-1922-1501192-9.

[23] M. Bôcher, *An introduction to the theory of integral equations*, Cambridge Tracts 10, Cambridge Univ. Press, 1909.

[24] S. Bochner, "Darstellung reellvariabler und analytischer Funktionen durch verallgemeinerte Fourier- und Laplace-Integrale", in: *Math. Ann.* 97 (1927), p. 635–662, DOI: 10.1007/BF01447886.

[25] N. Bourbaki, "Sur les espaces de Banach", in: *C.R. Acad. Sci.* 206 (1938), p. 1701–1704.

[26] N. Bourbaki, *Eléments de Mathématique, Espaces vectoriels topologiques*, vol. Livre V, Chap. I–II, n° 1189; Chap. III–V, n° 1229, Hermann, Paris: Actual. Scient. Ind., 1953–55.