The Development
of
Certain Aspects
of
Twentieth Century
Analysis

B. Sims

HISTORY OF MATHEMATICS

THE DEVELOPMENT OF CERTAIN ASPECTS OF TWENTIETH CENTURY ANALYSIS

INTRODUCTION:

In writing this essay I have relied heavily on secondary sources, in particular the excellent paper by Bernkopf - "The development of Function Spaces with particular reference to their origins in Integral Equation Theory" and the opus "Mathematical thought from Ancient to Modern Times" by Morris Kline.

Wherever possible I have, however, endeavoured to inspect the original material, for points of notation, terminology and the like. Almost nowhere have I remained faithful to these original works. I have none-the-less tried to use symbolism and terminology which will convey some of the original flavour, consequently the reader will find both the language and notation gradually evolve as the essay does. Allowance must be made for this and I hope it has not unduly obscured the overall story.

The essay is not meant as a piece of historiography not even chronicling, rather it is addressed to final year undergraduate students. The necessarily selective nature of courses offered in modern analysis cannot give the student any clear idea of its unity and certainly not its development. Although applications, to other more applied areas, are often included in such courses, a common reaction of many students meeting abstract methods for the first time, seems to be "but this is not real mathematics". I believe this reaction is in part due to the isolation of such methods from the more "classical" style of mathematics they have been used to.

In any undergraduate mathematics programme the emphasis must, for reasons of time, be on results and methods, however in such an approach the natural evolution of abstract methods out of their "classical" background will inevitably be lost. Indeed, because of the large number of options available to an undergraduate in some programmes, it is quite possible that key topics in this progression will have been missed. For this reason I have devoted the first 30 odd pages to those aspects of pre-twentieth century mathematics which I considered most important for the development of modern functional analysis.

Wherever I felt they would be readily understood I have included mathematical details. Some of the less essential ones have been "boxed" as inserts, the rest need only be skimmed - however I hope the student will do

more. A student reading the essay should not only gain a better under-
standing of how and why modern analysis developed but also learn some
mathematics on the way, at least in a "birds-eye" sense.

B. Sims.

April 1980.

Contents

APOLOGIA:

When writing an historical account of mathematics as a coherently developing body of knowledge the important features of any earlier period are identifiable as results and techniques which have gained a lasting place, or else work which directly contributed to subsequent advances. Inevitably this represents only part of the mathematics done during any era, a significant proportion rapidly fades into oblivion; or is subsumed as simple consequences of later more powerful developments.

Of necessity, our view of current mathematics must lack this "historical perspective".

Undoubtedly the dominant and novel feature of 20th Century Mathematics has been the recognition and study of "abstract" structures and methods. Whether these structures will retain a fundamental place in future mathematics or perhaps become *mere* "techniques", useful in establishing those results for which they prove most suitable, cannot be decided. (Although some will argue that the study of structure *per se* comes nearer to the true essence of mathematics than anything else.) What we can, and will, attempt to show is how these developments followed naturally from 19th century mathematical endeavours. The structures studied were not "created" at random. Only those structures found to underlie already existing areas of mathematics have been extensively developed, particularly when the area involved major unsolved problems. The hope, sometimes realized, being that a change of view-point: from the "concrete" to the "abstract", would provide new lines of attack on recalcitrant problems. (After all, mathematics proper is a living art: the solving of problems. What the student of mathematics studies is merely the "success" story of past mathematics – an apprenticeship so to speak.) The change of view-point, as with any new development, has bred its own generation of problems, problems which have occupied many 20th century mathematicians. Indeed some abstractions have only assumed significance because they were found to underlie already developed structures. For example, the algebraic notion of general vector (or linear) space grew in importance because such structures were present in the abstract notion of "function space" which in turn, as we shall see, provides a setting for the continuation of 19th century work on integral equations and the calculus of variations, while the recently developed theory of Categories (MacLane et al,1948 on) might be described as the "Structure of Structures".

We have already noted how later developments sort out key features from the great volume of mathematics produced in any previous era. Further, this sorting is accomplished by whole generations of subsequent mathematicians. This is a great asset to the mathematical historian. Significant source material is already pin-pointed and systematic accounts of the material are available.

Any would-be-writer on 20th century mathematics lacks this assistance. Systematic accounts are at best fragmentary, much of the material appearing only in the original research papers. Add to this the "exponential" growth in the number of papers published, and the task of surveying any broad area of current mathematics becomes formidable. It is for these reasons that I have elected to confine the subsequent article to one area of analysis (functional analysis and, in particular, material immediately connected to the notion of "function spaces"). The magnitude of this restriction and the truth of the above assertions can possibly be gauged from the following data.

The subdivisions of analysis listed are a slightly compressed version of those used in the international review journal, *Mathematical Reviews*. This journal appears in 12 monthly parts each year. The numbers listed on the right are the number of papers reviewed in the May, 1979 issue and so roughly correspond to the number of research papers on <u>analysis</u> currently published in major international journals (about 55 in number) during <u>one</u> month.

Functions of real variable	9
Measure and Integration	17
Functions of one (or more) complex variables	97
Special functions	24
Ordinary differential equations	60
Partial differential equations	156
Finite differences and functional equations	16
Sequences, Series, Summability	8
Approximations and Expansions	45
Fourier and Harmonic Analysis	50
Integral Transforms and Operational Calculus	14
Integral Equations	28
Functional Analysis and Operator Theory	176
Calculus of Variations and Optimal Control	43
Differential Geometry and Manifold theory	239
General topology	105
Algebraic topology	42
	1,129

The subsequent account, while restricted to functional analysis during the earlier part of this century, is typical of much 20th century analysis and may give the reader some idea of developments. Although it deals with an

area which is more familiar to me than many, still it must inevitably suffer from serious omissions and distortions due to my very incomplete knowledge of the literature. For these inaccuracies I apologise in advance.

19TH CENTURY PRELUDE

We consider some aspects of 19th and early 20th century mathematics which contributed directly to the evolution of "function spaces".
Some of this material may be found in standard accounts of the 19th century, however, in several cases the work did not belong to the main stream of 19th century mathematics and so is omitted from short accounts of that period.

Integral equations

DEFINITION

An <u>integral equation</u> for the unknown function y, is an equation involving an integral whose integrand depends on y. For example;

$$\int_{-1}^{1} \frac{y(t)}{t - x} \, dt = f(x),$$

where $f(x)$ is a given function and $-1 < x < 1$.

The problem of solving such an equation is to determine y. Of course, the simplest example of an integral equation is

$$\int_{0}^{x} y(t) dt = f(x) - f(0)$$

for which the solution $y(x) = \dfrac{d}{dx} (f(x))$ is obvious.

EXAMPLES

Throughout the 19th century many problems in mathematical physics and engineering led naturally to integral equations. Thus, as early as 1782 Laplace considered

$$\int_{0}^{\infty} e^{-xt} y(t) dt = F(x).$$

[You should recognise this as the problem of finding the function y whose "Laplace Transform" is $F(x)$.] Similarly Fourier while studying questions of heat flow was led to consider in 1811 the equation

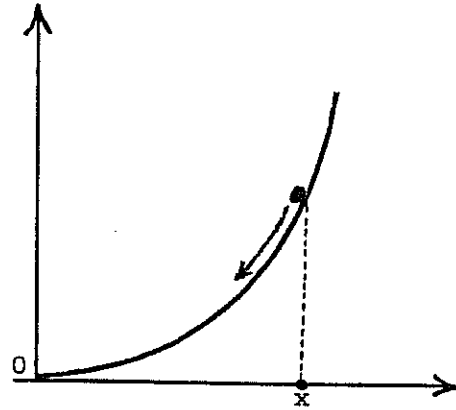$$\int_{0}^{\infty} \cos(xt) \, y(t) dt = f(x)$$

$$\text{(c.f. Fourier Transforms)}.$$

One of Abel's earlier mathematical works (1823 and 1826) concerned the problem of determining the curve (see diagram) through 0 along which a particle,

4.

starting from the point above x, would
take a specified time $T(x)$ to slide to
0. This led to the integral equation

$$T(x) = \int_0^x \frac{s'(t)}{\sqrt{x-t}}\, dt \quad \text{for} \quad s'(t).$$

Liouville was interested in integral equations from 1832 on, and in 1837
observed that certain differential equations could be converted into integral
equations. This was soon to become a powerful technique in the theory of
ordinary differential equations, particularly Boundary Value Problems.

Formulation of the Boundary Value problem

$$y'' + py' + qy = \lambda y, \quad y(0) = y(1) = 0$$

as an integral equation.

If y_1, y_2 are two linearly independent solutions of the
homogeneous problem $y'' + py' + qy = 0$ satisfying
$y_1(0) = y_2(1) = 0$ and $y_1(1) = y_2(0) = 1$, then by the
method of variation of parameters (see for example Boyce
and Di Prima, p.126) the general solution of the non-
homogeneous equation $y'' + py' + qy = g$ is

$$y = c_1 y_1 + c_2 y_2 + \int_0^x \frac{y_1(t)y_2(x) - y_1(x)y_2(t)}{w(y_1,y_2)(t)}\, g(t)dt$$

where $w(y_1,y_2)(t) = y_1(t)y_2'(t) - y_1'(t)y_2(t)$ is the non-
vanishing Wronskian of y_1 and y_2.
Choosing c_1 and c_2 to satisfy

$$y(0) = y(1) = 0$$

we obtain the solution

$$y = \int_0^x \frac{y_1(t)y_2(x)}{w(t)}\, g(t)dt + \int_x^1 \frac{y_1(x)y_2(t)}{w(t)}\, g(t)dt$$

or $\quad y = \int_0^1 G(x,t)\, g(t)dt$

where

JOSEPH LIOUVILLE
1809 - 1882

$$G(x,t) = \begin{cases} \dfrac{y_1(t)y_2(x)}{w(t)} & \text{for } 0 \leqslant t \leqslant x \\[4mm] \dfrac{y_1(x)y_2(t)}{w(t)} & \text{for } x < t \leqslant 1, \end{cases}$$

is the Green's function for the problem.

Replacing g by λy, we now see that the Boundary value problem has been reformulated as:

Determine those values of λ (the eigenvalues) for which the integral equation

$$y(x) - \lambda \int_0^1 G(x,t)\, y(t) dt = 0$$

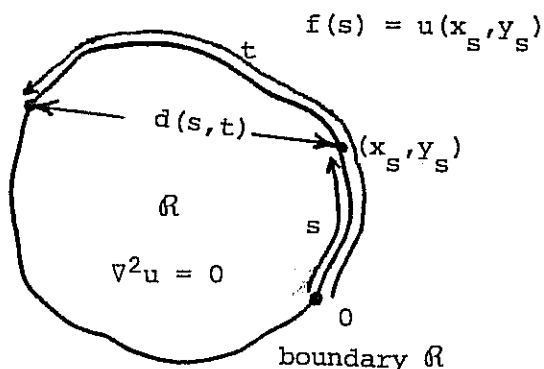*has non-trivial solutions y (the corresponding eigenfunctions).*

Boundary Value problems arise naturally from separating variables in a partial differential equation. By the middle of the 19th century, partial differential equations were a subject of considerable interest (Dirichlet, Riemann, Neumann and many others), and again attention was drawn to integral equations. For example, the function $u(x,y)$ which satisfies the Laplace (or potential) equation

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

in the interior of a given plane region $\mathcal{R}$ and assumes specified values on the boundary of $\mathcal{R}$ may be determined from an auxiliary function ρ which in turn satisfies the integral equation

$$2\pi f(x) + \int_0^\ell P(s,t)\, \rho(t) dt = 0.$$

Here it is assumed that the boundary of $\mathcal{R}$ has been parameterized by arc length along it. ℓ is the perimeter of $\mathcal{R}$, $f(s)$ is the specified value of u at the point on the boundary corresponding to a parameter value of s and $P(s,t) = \log(d(s,t))$ where $d(s,t)$ is the euclidean distance between the two boundary points with parameter values of s and t.

$$f(s) = u(x_s, y_s)$$
$$d(s,t)$$
$$(x_s, y_s)$$
$$\mathcal{R}$$
$$s$$
$$\nabla^2 u = 0$$
$$0$$
boundary $\mathcal{R}$

6.

Late in the century (1894), Poincaré considered the non-homogeneous partial differential equation

$$\nabla^2 u + \lambda u = F(x,y)^*$$

where λ is, in general, a complex number.

Starting with this equation and suitable boundary conditions, an analysis akin to that for the Laplace equation, led Poincaré (1896) to consider the integral equation

$$f(s) = \phi(s) - \lambda \int_a^b P(s,t)\phi(t)dt.$$

DEVELOPMENT OF A THEORY

During the early part of the 19th century individual problems were considered in isolation and solved by particular methods. Thus in 1823 Poisson obtained the solution

$$y(t) = \frac{1}{2\pi i} \int_{R-i\infty}^{R+i\infty} e^{xt} F(x)dx \quad \text{(for a suitable choice of R)}$$

to Laplace's problem

$$\int_0^\infty e^{-xt} y(t)dt = F(x).$$

Such a solution is known as an "inversion formula". Similar inversion formula were obtained by both Fourier and Abel for their respective problems.

---

* Solutions of this equation are important in the study of wave motion in the presence of external forces, described by the equation

$$\nabla^2 a - \frac{1}{c^2}\frac{\partial^2 a}{\partial t^2} = D(x,y,t).$$

To see this, note that, if λ_j, u_j satisfy $\nabla^2 u_j + \lambda_j u_j = f(x,y)$ then $u_j(x,y) e^{c\sqrt{\lambda_j}\, t}$ is a solution of $\nabla^2 a - \frac{1}{c^2}\frac{\partial^2 a}{\partial t^2} = f(x,y) e^{c\sqrt{\lambda_j}\, t}$. By adding together such solutions we are able to solve

$$\nabla^2 a - \frac{1}{c^2}\frac{\partial^2 a}{\partial t^2} = f(x,y) \sum_j \alpha_j e^{c\sqrt{\lambda_j}\, t}.$$

The problem is to be sure of an adequate supply of possible λ_j values for the given $D(x,y,t)$ to be expressible in this form.

The germ of a general method was sown by Liouville who obtained the "solution" of

$$y(x) = \cos kx + \frac{1}{k} \int_0^x \sigma(t) \sin kx \, y(t) dt$$

by a method of successive substitutions.

A similar iterative procedure was employed by A. Beer in 1845 to "solve" for ϕ a special case of

$$f(s) = \frac{1}{2} \phi(s) + \frac{1}{2\pi} \int_0^{\ell} Q(s,t) \, \phi(t) dt$$

which arises in connection with Laplace's equation on a region $\mathcal{R}$.

Neither Beer nor Liouville justified their procedures by establishing convergence of the successive iterates. It was some thirty years later that C. Neumann (1832 - 1925) demonstrated the validity of Beer's solution when $\mathcal{R}$ is a "convex" region.

These methods were brought nearer to maturity by the Italian mathematical physicist, Vito Volterra (1860 - 1940) toward the end of the century.

At this point it is convenient to recognise some basic types of integral equations.

An integral equation of the form

$$y(x) = f(x) + \lambda \int_a^b K(x,t) \, y(t) dt$$

is a <u>Fredholm equation of the *Second Kind*</u>.   (When $f(x) \equiv 0$ the equation is said to be <u>homogeneous</u>.)

$$f(x) = \int_a^b K(x,t) \, y(t) dt$$

is a <u>Fredholm equation of the *First Kind*</u>.
The function $K(x,t)$ is referred to as the <u>Kernel</u>.

Fourier's equation is an example of an equation of the first kind, while those arising from boundary value problems are equations of the second kind. In general, equations of the first kind present the greatest difficulty.

When the fixed upper limit of integration b is replaced by x we have <u>Volterra equations</u> of the first and second kind. Thus Abel's equation is an example of a Volterra equation of the first kind.

Volterra equations are really special cases of Fredholm equations in which the kernel is such that $K(x,t) = 0$ for $t > x$.

A Fredholm equation is said to be <u>symmetric</u> if $K(x,t) = K(t,x)$.   Almost

8.

all the examples given have been of symmetric equations.

For equations of the second kind, Volterra considered the following iterates.

$$\phi_0(x) = f(x)$$

$$\phi_1(x) = \int_a^b K(x,t) \, \phi_0(t)dt$$

$$\phi_2(x) = \int_a^b K(x,t) \, \phi_1(t)dt$$

. . . . .

Provided $K(x,t) = 0$ for $t > x$, $|K(x,t)| \leqslant M$, $|f(x)| \leqslant M$, for some M and all x,t, Volterra was able to establish that the series

$$\sum_{n=0}^{\infty} \phi_n(x)$$

converges to a function $\phi(x)$ which by direct substitution into the equation may be shown to be a solution of

$$y(x) = f(x) + \int_a^b K(x,t) \, y(t)dt.$$

It is worth noting that the substitution leads to

$$\phi(x) = f(x) + \sum_{n=0}^{\infty} \int_a^b K(x,t) \, \phi_n(t)dt$$

$$= f(x) + \int_a^b K(x,t) \, f(t)dt + \int_a^b K(x,t) \int_a^b K(t,s) \, f(x)ds \, dt$$

$$+ \dots$$

$$= f(x) + \int_a^b K(x,t)f(t)dt + \int_a^b \int_a^b K(x,s)K(s,t) \, f(t)ds \, dt$$

$$+ \dots$$

(by interchanging the order of integration and the roles of the dummies s and t)

$$= f(x) + \int_a^b \left[K(x,t) + \int_a^b K(x,s)K(s,t)ds + \dots \right] f(t)dt,^*$$

---

* The "Picard process" of successive approximations

$$y_0(x) = f(x)$$

$$y_1(x) = f(x) + \int_a^b K(x,t)y_0(t)dt$$

$$y_2(x) = f(x) + \int_a^b K(x,t)y_1(t)dt$$

. . . .

also formally leads to this expression. Volterra's work implicitly establishes

VITO VOLTERRA
1860-1940

and so we obtain an "inversion formula" for the equation, namely

$$y(x) = f(x) + \int_a^b k(x,t)\ f(t)\,dt$$

where
$$k(x,t) = K(x,t) + \int_a^b K(x,s)K(s,t)\,ds + \ldots$$

is known as the <u>solving kernel</u>, <u>iterated kernel</u> or <u>resolvent</u> for the equation.

Under certain conditions Volterra was able to reduce a Volterra equation of the first kind to one of the second kind and so apply his methods to it.

Assuming each of the required operations is justified, by differentiating both sides of

$$f(x) = \int_0^x K(x,t)\ y(t)\,dt$$

with respect to x we obtain

$$f'(x) = K(x,x)\ y(x) + \int_0^x \frac{\partial K}{\partial x}(x,t)\ y(t)\,dt$$

dividing throughout by K(x,x) yields a Volterra equation of the second kind for y;

$$y(x) = -\frac{f'(x)}{K(x,x)} + \int_0^x \frac{\frac{\partial K}{\partial x}(x,t)}{K(x,x)}\ y(t)\,dt.$$

In 1896 Volterra also observed that an integral equation may be regarded as a limiting form of a system of n linear algebraic equations in n unknowns. However, while he privately developed this observation into a powerful method for deducing results he did not publish it, instead he simply published verifications for the solutions so obtained.

It remained for the Swedish Mathematician, Ivar Fredholm (1866 - 1927) to develop and publish, in 1900[*], a similar approach to the solution of general integral equations of the second kind.

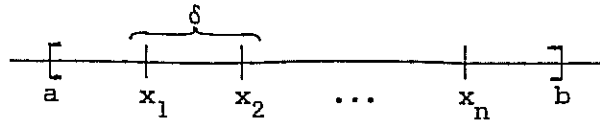Starting with the equation

(0)
$$y(x) = f(x) + \lambda \int_a^b K(x,t)\ y(t)\,dt.$$

Fredholm partitioned the interval [a,b] into n+1 subintervals each of the same length δ (see diagram)

---

[*] translated from Swedish into French in 1903.

10.

and, by replacing the integral with a sum, obtained an associated equation

$$y_n(x) = f(x) + \lambda \sum_{j=1}^{n} K(x,x_j) y_n(x_j) \delta$$

for the function $y_n(x)$.

In particular then for $i = 1,2,\ldots,n$ we have

$$y_n(x_i) = f(x_i) + \lambda \sum_{j=1}^{n} K(x_i,x_j) y_n(x_j) \delta$$

or, setting

$$\underset{\sim}{y}_n = \begin{bmatrix} y_n(x_1) \\ y_n(x_2) \\ \vdots \\ y_n(x_n) \end{bmatrix} \qquad \underset{\sim}{f}_n = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}$$

and

$$K_n = [k_{ij}] = \begin{bmatrix} K(x_1,x_1) & K(x_1,x_2) & \cdots & K(x_1,x_n) \\ K(x_2,x_1) & K(x_2,x_2) & \cdots & K(x_2,x_n) \\ \cdots\cdots\cdots & \cdots\cdots\cdots & \cdots & \cdots\cdots\cdots \\ K(x_n,x_1) & \cdots\cdots\cdots & \cdots & K(x_n,x_n) \end{bmatrix}$$

we have the matrix equation

$$(1) \qquad\qquad (I - \lambda\delta K_n)\underset{\sim}{y}_n = \underset{\sim}{f}_n .$$

Using the determinant formula for the inverse of a matrix we obtain

$$(2) \qquad\qquad y_n(x_i) = \sum_{j=1}^{n} \frac{\Delta_n(x_i,x_j,\lambda)\, f(x_j)}{\Delta_n(\lambda)}$$

$$\text{(provided } \Delta_n(\lambda) \neq 0),$$

Ivar Fredholm

where $\Delta_n(\lambda)$ denotes the determinant of $(I - \lambda\delta K_n)$ and $\Delta_n(x_i,x_j,\lambda)$ is the 'first minor' of the element in the i, jth place of $(I - \lambda\delta K)$.

Fredholm now chose to expand $\Delta_n(\lambda)$ as a polynomial in λ of degree n;

$$(2) \qquad \Delta_n(\lambda) = 1 - \lambda \sum_{i=1}^{n} k_{ii}\delta + \frac{\lambda^2}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} \begin{vmatrix} k_{ii} & k_{ij} \\ k_{ji} & k_{jj} \end{vmatrix} \delta^2 - \ldots \quad *$$

Letting $n \to \infty$ suggested using

$$(3) \qquad D(\lambda) = 1 - \lambda \int_a^b K(t,t)dt + \frac{\lambda^2}{2!} \int_a^b \int_a^b \begin{vmatrix} K(t,t) & K(t,u) \\ K(u,t) & K(u,u) \end{vmatrix} du\, dt - \ldots$$

in place of $\Delta_n(\lambda)$ in the solution of (0).

Because of the similarity of the expressions for $\Delta_n(\lambda)$ and the minors $\Delta_n(x_i,x_i,\lambda)$, Fredholm also replaced these by $D(\lambda)$. While for $i \neq j$, if $x_i \to x$ and $x_j \to t$, it seemed appropriate to replace $\Delta_n(x_i,x_j,\lambda)/\delta$ by

$$D(x,t,\lambda) = \lambda K(x,t) - \lambda^2 \int_a^b \begin{vmatrix} K(x,t) & K(x,u) \\ K(u,t) & K(u,u) \end{vmatrix} du + \ldots \ .$$

Using a theorem, published by Hadamard in 1893 (though the result seems to have been known as early as 1886):

For any $n \times n$ matrix $[m_{ij}]$, $\left| \det [m_{ij}] \right| < \sqrt{n}^n \left(\max_{i,j} |m_{ij}| \right)^n$;

Fredholm was able to show that the series for $D(\lambda)$ and $D(x,t,\lambda)$ converge and so both of these are well defined functions.

By writing (2) as

$$y_n(x_i) = \frac{\Delta_n(x_i,x_i,\lambda)}{\Delta_n(\lambda)} f(x_i) + \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\Delta_n(x_i,x_j,\lambda)}{\delta\, \Delta_n(\lambda)} f(x_j)\delta$$

* At this point Fredholm may have been influenced by the work of Helge Von Koch on systems of infinitely many equations in infinitely many unknowns. In 1893 Von Koch had used the above expansion for infinite determinants for which (2) becomes a power series in λ.

and using the above observations for $n \to \infty$, Fredholm was naturally led to infer that

$$(4) \qquad y(x) = f(x) + \int_a^b \frac{D(x,t,\lambda)}{D(\lambda)} f(t)dt \qquad (\text{provided } D(\lambda) \neq 0)$$

is a solution of (0).

This final "passage to a limit" is not mathematically rigorous, however by <u>direct substitution</u> into (0) Fredholm was able to verify that the suggested form of solution (4) is indeed correct.

Thus Fredholm established the *existence* of solutions to (0) for any λ which is not a zero of the function $D(\lambda)$ given in (3). [Fredholm also obtained satisfactory conclusions for the case when $D(\lambda) = 0$, however we will not pursue this line.]

In addition, Fredholm established *uniqueness* of the solution when $D(\lambda) \neq 0$. To do this he introduced ideas which appear to have been well ahead of his time and came close to explicitly using "function space methods". Write $D_K(\lambda)$ for $D(\lambda)$ to emphasize the dependence on $K(x,t)$ — see (3). Given a value of λ, Fredholm then considered the set of all possible kernels K for which $D_K(\lambda) \neq 0$. For each such K he regarded (0) as defining a "transformation" of the unknown function y into the given function f. The desired conclusion followed by showing that the set of all such transformations forms a group under composition.

REMARK: In order to easily join onto subsequent work we have explicitly included the parameter λ in our calculations. In the original work Fredholm did not. His calculations correspond to ours with $\lambda = 1$. He did, however, study the effect of such parameters by replacing $K(x,t)$ by $K^*(x,t) = \frac{1}{\lambda} K(x,t)$. Solving $y = f + \int_a^b K(x,t)y(t)dt$ is then, of course, equivalent to solving $y = f + \lambda \int_a^b K^*(x,t)y(t)dt$.

The theory of integral equations and, in particular, Fredholm's methods were brought nearer to a definitive form by the German mathematician, David Hilbert (1862 - 1943). Hilbert, a Professor of Mathematics at the University of Göttingen, had done important work in algebraic number theory, the theory of invariants, and the foundations of geometry and was already recognised as the foremost mathematician of his time, when from 1906 to 1910 he turned his attention to the theory and applications of integral equations.

Hilbert's work on integral equations was both extensive and detailed. Without expanding this section into a fully blown essay on integral equations, we can do little more than hint at Hilbert's methods and state some of the more important conclusions. For a more detailed discussion of Hilbert's work see Kline, pp. 1060-1070, and, in particular, Bernkopf [1], pp. 9-33.

Hilbert began the first of his six papers on the subject by thoroughly "reviewing" the theory of n linear equations in n-unknowns, paying particular attention to the case when the matrix for the system is symmetric. In this work Hilbert makes repeated use of the notion of "inner-product" for two vectors $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ defined by

$$[x,y] = \sum_{p=1}^{n} x_p y_p.^*$$

In particular, he uses the fact that if

$$[x,\emptyset] = [x,f] + \lambda[x,K\emptyset]$$

is an identity in x, then $\emptyset$ is a solution of

(a) $$y = f + \lambda Ky.$$

Like Fredholm he expresses the solution $\emptyset$ using determinants and when K is symmetric, he develops expressions for these in terms of the eigenvalues[**] λ_k and eigenvectors $\emptyset^k$ of the homogeneous problem

(a') $$y = \lambda Ky.$$

For example:

$$\frac{D(\lambda;x,y)}{d(\lambda)} = \sum_{k=1}^{n} \frac{[\emptyset^k,x][\emptyset^k,y]\lambda_k}{[\emptyset^k,\emptyset^k](\lambda-\lambda_k)},$$

here $d(\lambda) = \det(I - \lambda K)$, and

$$D(\lambda;x,y) = \det \begin{bmatrix} 0 & x_1 \ldots x_n \\ y_1 & \\ \vdots & I - \lambda D \\ y_n & \end{bmatrix}.$$

---

* Hilbert's notation
** Note: From the form of (a'), Hilbert's "eigenvalues" are the reciprocals of what we would understand by the term today.

14.

He has already shown that $\emptyset$ is a solution of (a) if

$$[x,\emptyset] = \frac{D(\lambda;x,f)}{d(\lambda)} \quad \text{for all vectors x}$$

Hilbert now applied these results to the "transcendental problem", that is the integral equation.

(b)
$$y(x) = f(x) + \lambda \int_0^1 K(x,t)y(t)dt \quad *$$

which "arises" from (a), with λ replaced by $\frac{\lambda}{n}$, as $n \to \infty$.
Unlike Fredholm, who inferred the form of solution by analogy with the finite case, then verified the correctness of the guess by direct substitution, Hilbert verified various steps in his theory by rigorously taking to the limit expressions occurring in his theory of finite systems of linear equations. For example, he proved that for bounded λ, $d\left(\frac{\lambda}{n}\right)$ converges uniformly to

$$\delta(\lambda) = \sum_{n=0}^{\infty} \frac{(-1)^n \lambda^n}{n!} \int_0^1 \cdots \int_0^1 \det[K(s_i,s_j)] \, ds_1 \cdots ds_n.$$

By these methods Hilbert arrived at the solution

(c)
$$y(x) = f(x) + \lambda \int_0^1 \overline{K(x,t)} \, f(t)dt$$

for any value of λ for which $\delta(\lambda) \neq 0$, where

$$\overline{K}(x,t) = 1 - \frac{\lambda \Delta(\lambda;x,t)}{\delta(\lambda)}$$

and

$$\Delta(\lambda;x,t) = \sum_{n=0}^{\infty} \frac{(-1)^n \lambda^n}{n!} \int_0^1 \cdots \int_0^1 \det \begin{bmatrix} 0 & | & K(x,s_1) \cdots K(x,s_n) \\ \hline K(s_1,t) & | & \\ \cdots\cdots & | & [K(s_i,s_j)] \\ K(s_n,t) & | & \end{bmatrix} ds_1 \cdots ds_n.$$

---

* It is convenient to take the range of integration to be from 0 to 1 and by a suitable change of variable this can always be assumed to be the case.

David Hilbert, 1912 — one of a group of portraits of professors
which were sold as postcards in Göttingen

The eigenvalues of (c) are taken to be the zeros of $\delta(\lambda)$.

Hilbert next proceeded to extend his results for finite systems of linear equations with a symmetric coefficient matrix to the theory of symmetric integral equations (ones for which $K(x,t) = K(t,x)$).

His first result is that the eigenvalues for such a problem are all real and so may be ordered. Next Hilbert used an expression analogous to that for the finite case to define eigenfunctions $\phi_k(x)$ corresponding to the eigenvalue λ_k

$$\left[\phi_k(x) = \left(\frac{\lambda_k}{\Delta^*(\lambda_k;s,s)}\right)^{\frac{1}{2}} \Delta^*(\lambda_k;x,s)\right.$$

where $\Delta^*(\lambda_k,x,t) = -\delta(\lambda_k)\overline{K}(x,t)$ and s is any real number chosen so that $\left.\Delta^*(\lambda_k;s,s) \neq 0\right]$.

The next results show that these eigenfunctions may be selected so as to form an <u>orthonormal</u> family; that is,

$$\int_0^1 \phi_k^2(t)\, dt = 1 \quad \text{and if} \quad \phi_k(x),\ \phi_j(x)$$

correspond to distinct eigenvalues, then

$$\int_0^1 \phi_k(t)\phi_j(t)\, dt = 0,$$

and that $\phi_k(x)$ is a solution of the homogeneous problem

$$\text{(d)} \qquad y(x) = \lambda \int_0^1 K(x,t)y(t)\, dt,$$

with $\lambda = \lambda_k$.

The major result of this study was a generalized "principal axis theorem":

For arbitrary continuous functions $y_1(x),\ y_2(x)$ *on* $[0,1]$ *we have*

$$\int_0^1\int_0^1 K(x,t)y_1(t)y_2(x)\, dx\, dt = \sum_{n=1}^{\infty} \frac{1}{\lambda_n}\left(\int_0^1 \phi_n(t)y_1(t)\, dt\right)\left(\int_0^1 \phi_n(t)y_2(t)\, dt\right).$$

As an application Hilbert then proved an important result, now known as the <u>Hilbert-Schmidt Theorem</u>:

If $f(x)$ *is such that*

$$f(x) = \int_0^1 K(x,t)g(t)\, dt$$

for some continuous function g, *then*

$$f(x) = \sum_{n=1}^{\infty} \left[\int_0^1 f(t)\phi_n(t)\,dt \right] \phi_n(x).$$

That is, such f's *can be expressed as a "Fourier" series in the eigen-functions of (d).*

In the closing few lines we discuss some results of what was probably Hilbert's most extensive and penetrating work.

In his work on finite systems Hilbert was led to consider bilinear forms: expressions of the form $\sum_{i,j=1}^{n} k_{ij} x_i y_j$; and quadratic forms (the result of setting $y_j = x_j$ in the previous expression). He now proceeded to an investigation of infinite quadratic forms (forms corresponding to $n = \infty$). The connection of such forms and integral equations is established as follows.

By using polynomials, Hilbert shows that it is always possible to construct a "complete" set of orthonormal functions; that is, an orthonormal family $\phi_1(x), \phi_2(x), \ldots$ for which the conclusion of the generalized principal axis theorem holds in the particular case when $K(x,t) \equiv 1$ and for each n, $\lambda_n = 1$. Any continuous function $g(x)$ can now be represented by its sequence of " "Fourier" coefficients $g_1, g_2, \ldots, g_n, \ldots$ where $g_n = \int_0^1 \phi_n(t) g(t)\,dt$. Similarly $K(x,t)$ is represented by the sequence (infinite matrix) of double "Fourier" coefficients

$$k_{nm} = \int_0^1 \int_0^1 K(s,t)\phi_n(s)\phi_m(t)\,ds\,dt.$$

It is then shown that $y(x)$ is a solution of (a) if and only if

$$y_n = f_n + \lambda \sum_{m=1}^{\infty} k_{nm} y_m$$

and that this happens if and only if

$$\sum_{n=1}^{\infty} y_n x_n = \sum_{n=1}^{\infty} f_n x_n + \lambda \sum_{n,m=1}^{\infty} K_{nm} y_m x_n$$

holds for every sequence $x_1, x_2, \ldots$ with $\sum_{n=1}^{\infty} x_n^2 < \infty$.

Hilbert's key result on infinite quadratic forms is the following. As with an integral equation, Hilbert associates a set of eigenvalues λ_1, λ_2, ... with the quadratic form

$$\sum_{n,m=1}^{\infty} k_{nm} \, x_m \, x_n,$$

he then shows that there exists an "orthogonal" transformation T such that with respect to the new variables $x_n' = T(x_n)$ the quadratic form becomes

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} x_n'^2 + \int_{\Lambda} \frac{d\sigma(\mu,\xi)}{\mu} \, .$$

We will not describe the integral term, except to note that essentially it corresponds to a "sum" over a "continuous" range of "eigenvalues". Indeed Hilbert was intent on finding conditions under which it would not be present. To this end he introduced the concept of "complete continuity" for the quadratic form

$$F(x_1,x_2,...) = \sum_{n,m=1}^{\infty} k_{nm} \, x_m \, x_n.$$

$F(x_1, x_2, ...)$ is <u>completely continuous</u> if, whenever $\varepsilon_1(t)$, $\varepsilon_2(t)$, ... is a sequence of functions for which $\sum_n \varepsilon_n^2(t) < \infty$ for each t and $\underset{t\to 0}{\text{Limit}} \, \varepsilon_n(t) = 0$, we have

$$\underset{t\to 0}{\text{Limit}} \, F(x_1 + \varepsilon_1(t), \, x_2 + \varepsilon_2(t), \, ...) = F(x_1, \, x_2, \, ...).$$

Hilbert is able to show that when the k_{nm}'s result from a symmetric kernel, the corresponding quadratic form is completely continuous. From these results he re-deduces the "generalized principal axis theorem" and Hilbert-Schmidt Theorem for the eigenfunctions of a symmetric integral equation.

Calculus of Variations

The calculus of variations is concerned with finding those functions $y(x)$ which satisfy certain "constraints" and for which a given expression of the form

$$J = \int_a^b f(x,y,y',...) \, dx$$

is a minimum (or a maximum).

The simplest form of constraint is that y have specified values at a and

b.   In other problems it may be required that a second expression, similar in form to J, be a constant.

Such problems appeared early in the history of the calculus.

In the *Principia*, Newton considered the curve y(x) joining the points A, B (see diagram) for which the surface of revolution obtained by rotation about the x-axis, offers least resistance when moving through a liquid such as water at a constant velocity in the direction of the x-axis.
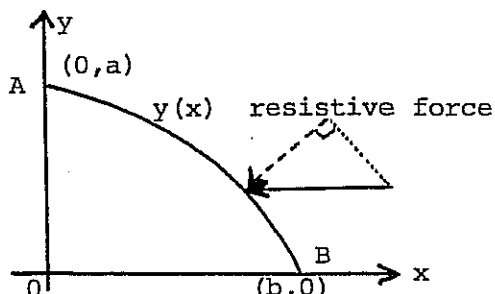
[The resistance at any point on the surface is assumed (quite unrealistically) to be proportional to the component of velocity normal to the surface at that point.]

In 1694 he obtained a solution, which amounts to finding a function y(x) with y(0) = a, y(b) = 0 and for which

$$J = \int_0^b \frac{y(x) \ [y'(x)]^3}{1 + [y'(x)]^2} \, dx$$

is a minimum.

The connection of similar problems with "stream-lined" design is obvious, though seldom used because of mathematical difficulties.

In 1696 John Bernoulli proposed the *brachistochrone* problem as a challenge to other mathematicians of the time.

Here one seeks the curve y(x) joining two given points A, B down which a particle will slide in minimum time. That is, one requires y(x) to be such that y(0) = a, y(b) = 0 and

$$J = \int_0^b \left(\frac{1 + [y'(x)]^2}{a - y(x)} \right)^{\frac{1}{2}} \, dx$$
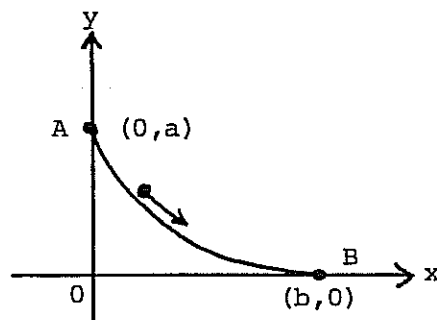
is a minimum.

The unique solution (a segment of a cycloid) was correctly given by Newton, Leibniz, L'Hospital and John and James Bernoulli.

Another problem, first considered in the 18th century, and which may be formulated as a question in the calculus of variations, is the problem of finding the path of minimum length (*geodesic*) joining two points on a given surface.

In the 18th century the concern was with geodesics on the earth's surface. Subsequently, such questions played an important role in the study of non-euclidean geometries and later in the general theory of relativity.

Another class of problems which belong to the calculus of variations are the so-called *isoperimetric* problems. A typical problem of this class would be to find the curve y(x) joining A, B (see diagram) which has a given length and for which the area enclosed by it and the x-axis is a maximum. Thus we seek y(x) such that

$$\int_0^b \sqrt{1 + [y'(x)]^2} \, dx = \ell$$

and $\int_a^b y \, dx$ is a maximum.
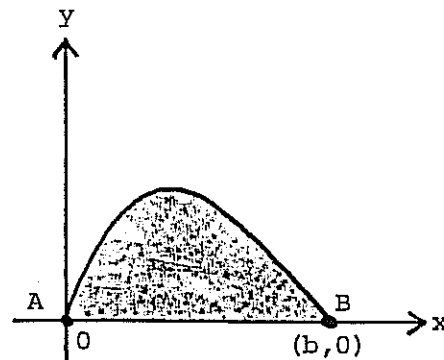
[We are here also assuming that the desired solution y(x) be a <u>function</u> of x.]

During the 19th and 20th centuries the calculus of variations has grown steadily in importance. In part this is due to the discovery of so-called "principles of least action", first in optics and classical mechanics, then in many other areas including quantum mechanics.

Such principles date back to classical greek times. Thus Heron (improving on statements in Euclid's *Catoptrica*) noted that light passing from A to B by reflecting off a plane mirror follows the path of minimum length (time). He then went on to apply this result to spherical mirrors. Fermat adopted this "principle of least time", and in 1661 successfully applied it to problems of refraction.
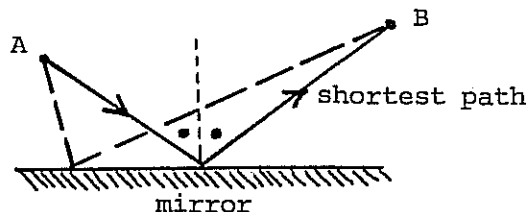
The similarity between Newton's first law for the motion of a particle and the observation that in free space, light travels in straight lines (paths of shortest time) suggested that a similar principle might apply in mechanics. Maupertuis made an attempt in this direction and Euler established that for a single particle (moving in a conservative force field) the observed velocity function v(t) either maximises or minimises the "action"

$$J = \int v^2(t) \, dt.$$

It was Lagrange, however, who first formulated a generally applicable "principle of least action" in about 1755.

Lagrange considered a system of n particles in a force field derivable from a potential function V; that is, the force $F = \nabla V$ is conservative. For such a system Lagrange showed that Newton's Second Law is equivalent to:

(i)   the total energy T + V be a constant; and

(ii)   the <u>action</u> $\int T\, dt$ must be a maximum or a minimum.

Here $T = \tfrac{1}{2} \sum_{j=1}^{n} m_j v_j \cdot v_j$ is the kinetic energy of the system.

Almost immediately Poisson reformulated Lagrange's results in terms of the "Lagrangian" $L = T - V$, however, the most important development was by the Irish scientist, William Hamilton. In a series of papers from 1824 to 1835 he developed the principle of "stationary" action, first for optics and then mechanics. His principle, which asserts that the action

$$S = \int L\, dt$$

is stationary (that is, either a maximum or a minimum) applied even when energy is not conserved. Hamilton expressed necessary conditions for S to be stationary in a very symmetric way in terms of the "Hamiltonian" H, which in effect represents the total energy of the system.

Many generalizations of the calculus of variations have also been studied. For example y may be a function of several variables and the form of J may involve multiple integrals. One class of problems of this type are the so-called "minimal surface" or "Plateau" problems. Here the problem is to determine the surface of minimal area having a given boundary. When a wire frame is fashioned in the shape of the boundary and drawn out of a soap solution the resulting soap film defines a minimal surface, a fact exploited experimentally by the Belgian physicist, Joseph Plateau (1801-1883). Such problems were considered by Lagrange and later Ampère in 1817, however, it is only in the last few decades that major advances have been made.[*]

So far we have only described the type of problem dealt with in the calculus of variations. We now turn to the more important issue of surveying methods developed to solve such problems.

During the 17th century problems in the calculus of variations were treated by individual techniques which exploited special features of each problem. Although some of the approaches used hinted at more general methods it remained for Euler to publish, in 1736, the rudiments of a general method.

[*] See the *Scientific American* article by Almgren and Taylor "The Geometry of soap films and soap bubbles", July 1976.

Euler began by replacing the integral with a finite sum and the derivatives of y appearing in the definition of f by difference quotients. In this way "J" was made to depend on only a finite number of points on the curve y(x). He then derived an expression for the change (variation) in J resulting from variations in the ordinates at these points. By setting this variation to 0, Euler obtained a difference equation for the ordinates of the "solution". A crude limiting process then convinced him that a necessary condition for the curve y(x) to maximize or minimize

(1)
$$J = \int_a^b f(x,y,y')dx$$
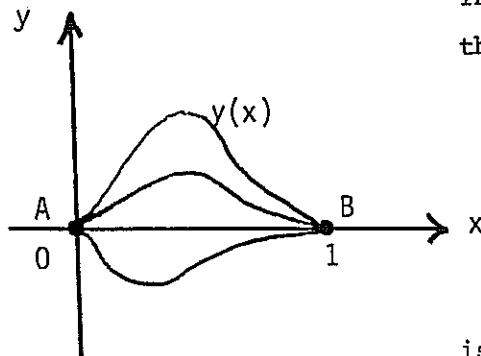
is that y(x) must satisfy the differential equation

(2)
$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0$$

Over the next few years he applied this same approach to a larger and larger variety of problems.

We illustrate the use of Euler's result by applying it to a very simple problem.

Find the curve y(x) of shortest length which joins the two points A,B (see diagram).

Thus we require y(x) such that $y(0) = y(1) = 0$ and

$$J = \int_0^1 ds$$

$$= \int_0^1 \left(1 + [y'(x)]^2\right)^{\frac{1}{2}} dx$$

is a minimum.

Here $f(x,y,y') = (1 + y'^2)^{\frac{1}{2}}$, so $\frac{\partial f}{\partial y} = 0$,

$\frac{\partial f}{\partial y'} = y'(1 + y'^2)^{-\frac{1}{2}}$, and (2) requires that

$$\frac{d}{dx}\left(y'(1 + y'^2)^{-\frac{1}{2}}\right) = 0.$$

Clearly this implies $y'(1 + y'^2)^{-\frac{1}{2}}$ equals a constant c_1, or

$$y' = c_1(1 - c_1^2)^{-\frac{1}{2}}.$$

Integrating yields

$$y = c_1(1 - c_1^2)^{-\frac{1}{2}} x + c_2.$$

Imposing the constraints $y(0) = y(1) = 0$ gives $c_2 = c_1 = 0$ and we conclude that the only possible solution is the straight line $y = 0$!

While Euler's method led to an elegant and applicable necessary condition: that y satisfy a differential equation of the form (2), the derivation, involving as it did a mixture geometric and analytic arguments, was cumbersome and far from rigorous.

It was Lagrange in 1755 who introduced a widely applicable and purely analytic approach.

Lagrange considered the change in J, ΔJ, when $y(x)$ is perturbed by the addition of a "small" function $\delta y(x)$, with $\delta y(a) = \delta y(b) = 0$.
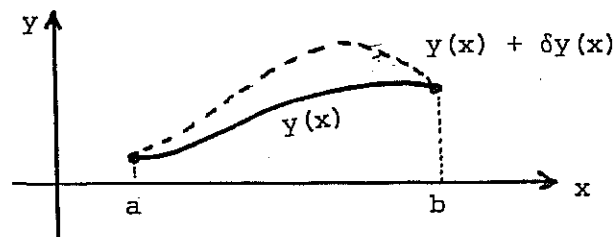
Thus,

$$\Delta J = \int_a^b [f(x, y + \delta y, y' + (\delta y)') - f(x, y, y')]dx.$$

Using Taylor' theorem to expand f about (x,y,y') we obtain

$$\Delta J = \int_a^b \left(\left[\frac{\partial f}{\partial y} \delta y + \frac{\partial f}{\partial y'} (\delta y)' \right] + \frac{1}{2}\left[\frac{\partial^2 f}{\partial y^2} \delta y^2 + 2 \frac{\partial^2 f}{\partial y \partial y'} \delta y (\delta y)' + \frac{\partial^2 f}{\partial y'^2} (\delta y)'^2 \right] + \ldots \right) dx$$

$$= \delta J + \frac{1}{2} \delta^2 J + \ldots$$

where $\delta J = \int_a^b \left[\frac{\partial f}{\partial y} \delta y + \frac{\partial f}{\partial y'} (\delta y)' \right] dx$ etc.

δJ is known as the <u>first variation</u> of J, $\delta^2 J$ the <u>second variation</u>, etc.

If J is a maximum or minimum for y then ΔJ has the same sign regardless of how y is perturbed. By assuming that δJ dominates the expansion for ΔJ and so ΔJ changes signs as δJ does, Lagrange convinced himself that a necessary condition for y to maximize or minimize J is that $\delta J = 0$.[*]

---

[*] This observation, that $\delta J = 0$ for any y which maximizes or minimizes J, was assumed in all subsequent work, however, it was not until 1848 that a correct proof was given by Pierre Frédéric Sarrus. The result is now known as the

From this Lagrange rededuced Euler's necessary condition (2).

$$\delta J = \int_a^b \frac{\partial f}{\partial y} \, \delta y \, dx + \int \frac{\partial f}{\partial y'} (\delta y)' dx$$

Integrating by parts we have

$$\delta J = \int_a^b \frac{\partial f}{\partial y} \, \delta y \, dx + \left[\frac{\partial f}{\partial y'} \, \delta y\right]_a^b - \int_a^b \left(\frac{d}{dx} \frac{\partial f}{\partial y'}\right) \, \delta y \, dx.$$

$$= \int_a^b \left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'}\right) \, \delta y \, dx, \quad \text{as} \quad \delta y(a) = \delta y(b) = 0$$

Thus $\delta J = 0$ implies

$$\int_a^b \left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'}\right) \, \delta y \, dx = 0$$

for every choice of the function δy and so we conclude that

$$\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'} = 0.$$

One cannot help but notice the analogous roles played by the first variation δJ in the calculus of variation and the first derivative of a function in the ordinary calculus. A sufficient condition for a differentiable function $g(x)$ to have a maximum at a point x_0, where the necessary condition $g'(x_0) = 0$ is satisfied, is that $g''(x_0) < 0$. It seems reasonable to expect the second variation $\delta^2 J$ to play a similar role in the calculus of variations. This proved to be a much more difficult question, first considered, unsuccessfully, by Lagrange. In 1786 Legendre was able to prove that for a y maximizing J we have $\delta J = 0$ and $\delta^2 J \leq 0$. Thus the condition $\delta^2 J \leq 0$, which Legendre proved equivalent to $\frac{\partial^2 f}{\partial y^2} \leq 0$, is a necessary condition, however in 1787 he was able to demonstrate that it is not sufficient.

Sufficient conditions were not found until later in the 19th century. The first was found by Jacobi in 1837, though an adequate proof of Jacobi's result was first given by Weierstrass in 1879.

Weierstrass also drew attention to an important limitation in the methods developed by Euler, Lagrange, Legendre and Jacobi.

We have already noted that Lagrange assumed the first variation δJ dominates the variation ΔJ. Since δJ depends linearly on δy and $(\delta y)'$ while the higher order variations depend on terms like δy^2, $\delta y (\delta y)'$ and $(\delta y)'^2$, this assumption requires that both δy and its derivative $(\delta y)'$ be small. Thus the size of J for y was only being compared to the size of J on a limited class of other curves. Perturbations δy for which both δy and $(\delta y)'$ are small were later called *weak variations* by Adolf Kneser (1862-1930).

Functions y which satisfied the necessary (and sufficient) conditions derived from such perturbations are termed *weak solutions*. The theories of Euler, Lagrange, Legendre and Jacobi were concerned with weak solutions. To find solutions which really maximize or minimize J one must allow all possible perturbations, including those for which $(\delta y)'$ need not be small even when δy is, thus one must allow so called *strong variations*.
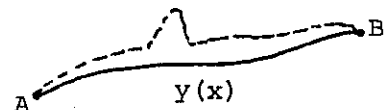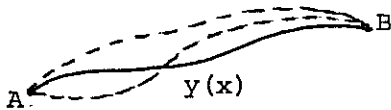
The early theories only allowed weak variations. That is perturbations with both δy and $(\delta y)'$ small

Weierstrass admitted strong variations, for which $(\delta y)'$ need not be small

Weierstrass proved that Jacobi's conditions were sufficient for weak solutions. By introducing a further condition, Weirerstrass also obtained a set of sufficient conditions for y to be a strong solution, his proofs in this connection were greatly simplified by Hilbert in 1900.

In the course of the work Weierstrass was led, naturally, to measure the "nearness" of two functions: f and g are in an <u>ε-neighbourhood of order</u> <u>p</u> (p = 0,1,2,...) if for every x with $a \leqslant x \leqslant b$ we have

$$\left| \frac{d^n f}{dx^n} - \frac{d^n g}{dx^n} \right| < \varepsilon \quad \text{for} \quad n=0,1,\ldots,p.$$

[Roughly speaking, δy is a strong or weak variation of y depending on whether y and $y + \delta y$ belong to an ε-neighbourhood of order 0 or of order 1.]

Although Weierstrass did not publish any of his work on the calculus of variations his lectures aroused fresh interest in the subject. Perhaps as early as 1883, though the first published work did not appear till 1887, Vito Volterra started to develop a theory of "functions of lines". By a <u>line</u> L Volterra understood what would now more usually be termed a curve specified by the equation $y(x)$ – or more generally by a pair of parametric equations $x = \ell_1(t)$, $y = \ell_2(t)$. In modern language, a <u>function of lines</u> U was a mapping from some family of lines into the real numbers. Thus, U is a function which assigns to each line L in the family a real number $U(L)$. Since L is determined by the function $y(x)$ we may think of U as a function of $y(x)$, $U(y(x))$. For convenience of discussion we will use the modern terminology and refer to U as a *functional*, although this term was first introduced by Hadamard after Volterra had established his theory. As an example

$$J \equiv J(y(x)) = \int_a^b f(x,y,y')dx$$

defines a functional on the family of smooth curves joining A and B.

Volterra first defined a functional U to be <u>continuous</u> at the function $y(x)$ if, given $\varepsilon > 0$ there exists a $\delta > 0$ such that for all functions $z(x)$ in a δ-neighbourhood of order 0 of $y(x)$ we have $|U(y(x)) - U(z(x))| < \varepsilon$. He then went on to establish a "differential calculus" for functionals.

In his theory the "<u>derivative</u> of U with respect to the function f, at a point x_0", is defined by

$$U'(f(x),x_0) = \lim_{\substack{h \to 0 \\ \varepsilon \to 0}} \frac{U(f(x) + \eta(x)) - U(f(x))}{\int_{x_0-h}^{x_0+h} (f(t) + \eta(t))\,dt}$$

provided this limit exists uniformly in h and ε, where $|\eta(x)| < \varepsilon$ for all x and $\eta(x) = 0$ for $x < x_0 - h$ or $x > x_0 + h$.

The <u>total variation</u> of U is now taken to be

$$\int_a^b U'(y(x),x_0)\ \delta y(x_0)\ dx_0.$$

For problems in the calculus of variations, Volterra hoped to use the vanishing of his total variation for J in place of the requirement $\delta J = 0$, as a necessary condition for y to be a solution. He also hoped to obtain sufficient conditions in terms of higher derivatives and higher total

variations defined by repeated applications of the above procedures.

Unfortunatley his theory did not prove very satisfactory for this purpose and subsequently the adequacy of his definitions was criticized by the French Mathematician, Jacques Hadamard (1865-1963).

Hadamard, along with several other French mathematicians had taken up the study of functionals toward the end of the century. Again the calculus of variations supplied their motivation, however, their work was in a different direction to that of Volterra's. Their aim was to find simple representations for a functional as a "Fourier" type series of integrals. Hadamard introduced the concept of a linear functional and obtained some results in this case. The functional U is linear if

$$U(\lambda_1 \, f(x) + \lambda_2 \, g(x)) = \lambda_1 \, U(f(x)) + \lambda_2 \, U(g(x))$$

where λ_1 and λ_2 are constants.

Jacques Salomon Hadamard

The development of a theory of point sets

Georg Cantor (1845 - 1918), though born in Russia of Danish-Jewish parents, lived most of his life in Germany where he studied under Weierstrass. His original interest was in Fourier series and, in particular, the question of uniqueness of the representing series. In 1870 he proved that, if for each point x

$$a_0 + \sum_{n=1}^{N} (a_n \cos nx + b_n \sin nx) \rightarrow 0 \quad \text{as} \quad N \rightarrow \infty,$$

then $a_n = b_n = 0$. A year later he extended this result to cover the case when the series converges for all but a finite number of x values. Then, in 1872 Cantor introduced the notion of a set* of points of the "first species" and proved that $a_n = b_n = 0$ provided the set of x values at which the series fails to converge is of the first species. Cantor defined a first species set of real numbers as follows.

A point p is a limit (point) of a set S of real numbers if every (open) interval containing p contains infinitely many points of S. The set of all limit points of S he termed the (first) derived set of S. The second

---

* For Cantor a set is a collection of definite and distinct objects for which we can decide whether or not any given object belongs to it. He also defined the union and intersection of sets.

derived set of S is then the derived set of the derived set of S. In this way Cantor is able to define the n'th derived set of S. A set S is of the <u>first species</u> if for some n the n'th derived set is a finite set of points. (For example, the points of a convergent sequence form a set of the first species, as does a finite union of sets of the first species.) In this work we recognise the beginnings of a theory of point set topology[*] and the reason for Cantor's subsequent study of infinite sets of points.

Using an idea already investigated by Bolzano (in 1851), Cantor said two sets have the same *power* (or, as it subsequently became known, *cardinal number*) if there is a one-to-one correspondence between them. For Cantor a set is 'infinite' if it can be put into a one-to-one correspondence with a (proper) part of itself. For two sets M and N, the cardinal number of M is <u>larger</u> than that of N if there is a one-to-one correspondence between N and a subset of M but no such correspondence between M and N. Cantor proved that the relation "has the same cardinal number" is an equivalence relationship and that larger, as defined above, has the properties of a partial order. For sets with the same cardinal number as the set of natural numbers 1, 2, 3, 4, ... Cantor introduced the term <u>enumerable</u> sets. The cardinal number of such sets he denoted by $\aleph_0$ – the "smallest" transfinite cardinal number.

His first paper on the subject (1874) included the following results:

(1) *The set of rational numbers is enumerable.*

(2) *The set of algebraic numbers is enumerable.*

Recall: A real number is algebraic if it is the root of some polynomial with integer coefficients. The first example of non-algebraic (transcendental) numbers had been given by Liouville in 1844 and the transcedence of e established by Hermite in 1873.

(3) *The "continuum", the set of all real numbers* (or points on a line) *is uncountable* (not enumerable).

From (1) and (3) Cantor readily deduced the existence or irrational numbers and more importantly, from (2) and (3) the existence of transcendental

---

[*] Cantor also introduced other relevant notions. A set is <u>closed</u> if it contains all its limit points. It is <u>open</u> if all its points are <u>interior</u> points, that is belong to open intervals containing only points of the set. If every point of the set is a limit point, then the set is <u>perfect</u> (that is, contains no isolated points).

numbers[*]. Indeed, both the set of irrational numbers and the set of transcendental numbers are "larger" than the set of rationals.

Cantor's proof that the algebraic numbers are enumerable.

For any polynomial equation

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n = 0$$

of degree n with integer coefficients, Cantor defined the "height" to be

$$N = (n-1) + |a_0| + |a_1| + \ldots + |a_n|.$$

Corresponding to each possible value of the height $N = 1,2,3,\ldots$ there are only a finite number of possible polynomials.

For example for $N = 2$ we have the four distinct equations;

$x^2 = 0$, $2x = 0$, $x + 1 = 0$ and $x - 1 = 0$.

Since each polynomial equation has only a finite number of roots, only a finite number of algebraic numbers arise from the polynomials of any given height.

---

*These were the first important examples of "non-constructive existence proofs". As we shall see, such proofs were soon to appear in many other areas of mathematics. Essentially a non-constructive existence proof proceeds as follows. Let E be the set of all objects of the type whose existence we wish to establish, then E is proved to be non-empty by showing that its complement has some property which prevents it from being the whole. In this way the existence of our objects is established without the need to give a single example of such an object. Such a proof should be sharply contrasted with the "constructive" style of proof, in which E is proved to be non-empty by explicitly producing an element of it. For example, Fredholm's proof that the set of solutions $y(x)$ to the integral equation

$$y(x) = f(x) + \lambda \int_a^b K(x,t)\, y(t)\, dt$$

is non-empty, is achieved by "constructing" the explicit solution

$$y(x) = f(x) + \int_a^b \frac{D(x,t,\lambda)}{D(\lambda)}\, f(t)\, dt$$

(see pages 9 to 12).

Non-constructive existence proofs require us to pay careful attention to the formulation of our problem and to the processes of logic. The emergence of such proofs partly explains the growing concern at the end of the 19th century to formalize (axiomatize) mathematics. It also promoted fresh interest in the Foundations of Mathematics and Logic. The profound results of Gödel, Cohen and other 20th century logicians were a direct outgrowth of this work. Indirect existence proofs were not, and are still not, universally accepted. Some mathematicians have worked toward the developments of a purely "constructive mathematics", most notably L. Brouwer (1881 - 1967), E. Bishop (1928 -   ) and currently a Belgium School of Mathematics centred around Garnir (1921 -   )___ the "intuitionist" school of mathematics.*

Let $\phi(N)$ be the number of algebraic numbers which arise from polynomials of height N and which have not arisen from any polynomial of height less than N. Such an algebraic number is also said to have height N. For example; 0, 1 and -1 are the algebraic numbers arising from the four polynomials of height 2. Further 0 is the unique algebraic number arising from the one equation $x = 0$ of height 1, so $\phi(2) = 2$ and the algebraic numbers of height 2 are 1 and -1. Cantor now assigns the numbers 1 to $\phi(1)$ to the algebraic numbers of height 1, the numbers $\phi(1) + 1$ to $\phi(1) + \phi(2)$ to the $\phi(2)$ algebriac numbers of height 2, the numbers $\phi(1) + \phi(2) + 1$ to $\phi(1) + \phi(2) + \phi(3)$ to the $\phi(3)$ algebraic numbers of height 3 etc.

In this way a one-to-one correspondence between the algebraic numbers and the natural numbers is established.

Cantor continued to investigate infinite sets for the next two decades. For example he constructed a subset of real numbers with zero length which may be placed in one-to-one corre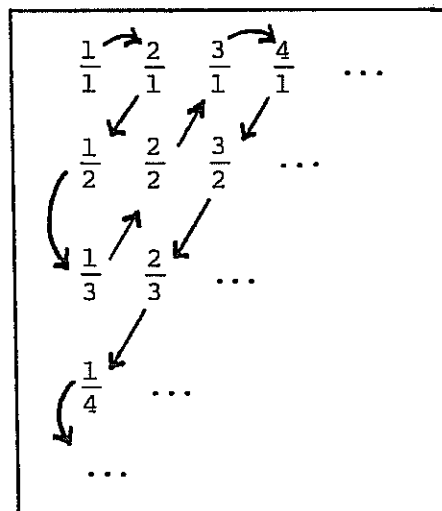spondence with the whole continuum. During this period he developed an "arithmetic" for transfinite cardinal and ordinal numbers. It was in the course of this work that he formulated the celebrated *continuum hypothesis*, however, this apsect of his researches need not concern us here.

He also continued to give "simpler" proofs for many of his 1874 results. Thus, in 1895 he produced the often quoted enumeration of the rationals

| Georg Cantor (1845—1918) |

$1 \leftrightarrow 1/1$

$2 \leftrightarrow 2/1$

$3 \leftrightarrow 1/2$

$4 \leftrightarrow 1/3$

$5 \leftrightarrow 2/2$

$6 \leftrightarrow 3/1$

$7 \leftrightarrow 4/1$

$\cdots \quad \cdots$

$$\frac{1}{1} \rightarrow \frac{2}{1} \quad \frac{3}{1} \rightarrow \frac{4}{1} \quad \cdots$$

$$\frac{1}{2} \quad \frac{2}{2} \quad \frac{3}{2} \quad \cdots$$

$$\frac{1}{3} \quad \frac{2}{3} \quad \cdots$$

$$\frac{1}{4} \quad \cdots$$

$$\cdots$$

Earlier, in 1890, he developed the now famous "diagonal argument" to prove the real numbers (between 0 and 1) are uncountable. This argument has subsequently been adapted for use in many other problems, particularly problems of "computability".

Cantor's diagonal Argument

We begin by writing each real number between 0 and 1 uniquely as a non-terminating decimal. (Thus $\frac{1}{2} = 0.4999 \ldots$, for example.)
We now assume the real numbers can be enumerated

$$1 \leftrightarrow 0.a_{11}\ a_{12}\ a_{13} \cdots$$
$$2 \leftrightarrow 0.a_{21}\ a_{22}\ a_{23} \cdots$$
$$3 \leftrightarrow 0.a_{31}\ a_{32}\ a_{33} \cdots$$
$$\cdots$$

Here, $a_{mn} = 0,1,\ldots,9$ is the n'th digit in the decimal expansion of the real number in correspondence with m.
Now, let $b = 0.b_1\ b_2\ b_3 \ldots$,

$$\text{where} \quad b_k = \begin{cases} 9 & \text{if } a_{kk} = 1 \\ 1 & \text{if } a_{kk} \neq 1, \end{cases}$$

then $0 \leq b \leq 1$ and b differs from any of the real numbers appearing in the above enumeration. Thus b is not in correspondence with any natural number, contradicting our assumption that all the numbers between 0 and 1 have been enumerated and so establishing the result.

Cantor's work was not favourably received by many of his contemporaries. It was strongly opposed by Kronecker and certainly treated unsympathetically by Felix Klein, Poincaré and Hermann Weyl, to mention a few. Indeed, as we have already noted (see previous footnote) some present day mathematicians still prefer to develop mathematics without the use of transfinite methods. It was not, however, without its champions, including Hadamard and later Hilbert and Bertrand Russell. Today his ideas pervade almost every branch of mathematics.

Most of Cantor's work was concerned with subsets of real numbers regarded as points on a line, however, he did attempt to extend some of the ideas to sets of points in n-dimensional euclidean spaces. In this he was partly successful, particularly for some of the more "topological" results. Similar

generalizations were attempted by other mathematicians. Two Italian
mathematicians G. Ascoli (1843-1896) and C. Arzelà (1847-1912) aimed to
extend certain of Cantor's ideas to sets of curves or functions.* In
particular they were concerned with properties of limits of sequences of
functions. Thus, in 1883 Ascoli published a memoir "On the limit curves
of a variety of curves". To illustrate their work we state the following
result of Arzelà.

Let {f} *be a family of functions all defined and continuous on the
same interval* [a,b], *then a necessary and sufficient condition for every
sequence of functions from* {f} *to have a uniformly convergent subsequence
is that* {f} *be bounded and equicontinuous on* [a,b]. ({f} is <u>equicontinuous</u>
if given $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(x')| < \varepsilon$ for all
pairs of points x,x' in [a,b] with $|x-x'| < \delta$ and all f in {f}.)

In 1893 the French mathematician René Baire (1874-1932) travelled to
Italy where he worked toward his doctoral degree (conferred at Milan in
1899). Under the influence of Volterra and Dini his interests turned to
the theory of functions of a real variable, a subject which occupied his
studies for the remainder of his short working life (ill-health forced Baire
to resign the professorship of analysis at Dijon in 1914 and greatly
curtailed his subsequent research activities.)

Baire began with a critical analysis of the classical definition of
continuity due to Cauchy (in the course of which he introduced the notions
of upper and lower semi-continuity). He then proceeded to classify limits
of continuous functions defined on subsets of euclidean n-space R^n, applying
transfinite methods of a type which had previously only been used by Cantor
for sets of real numbers.

Continuous functions, he took to be of class 0. Limits of continuous
functions were of class 1 and so on. He found properties common to all
functions of class n for each finite n, and undertook an extensive study
of the functions in class 1, 2 and 3. Baire used topological notions to
classify subsets of R^n into two categories. Sets of the first category are
'small' (a countable union of sets whose complements are dense). They are
analogous to Cantor's enumerable sets. He proved that the complement of a
first category set is of the second category; that is, not of the first
category, and so is a 'large' set. (A similar classification of subsets
of R had been made two years earlier by Osgood.) Baire then demonstrated
that the set of discontinuities of a function of class 1 is a first category
('small') set.

---

* Volterra's "lines".

THE BIRTH OF ABSTRACT FUNCTION SPACES

As early as 1851, Riemann had "philosophized" about the importance of regarding a collection of functions as a totality endowed with certain intrinsic properties. Thus, in connection with his work on "Dirichlet's principle" we find the statement "...*the totality of these functions forms a connected domain which is closed in itself*..." and later, "*There are however manifolds in which the determination of position requires ... an infinite sequence or even a continuous manifold of determinations ...* ". As an example of this he cites the manifold whose "points" are functions on some common domain.

This idea, of regarding functions as "points" in a "space", was implicit in Volterra's approach to the calculus of variations and was pursued by Ascoli, Arzelà and Baire, particularly as a result of the infusion of Cantor's ideas into the theory of real functions.

Similar ideas were also present, though less obviously, in the work on integral equations; witness Fredholm's uniqueness proof (page 12).

Thus, by the beginning of the 20th century, the notion of "spaces" of functions in which certain operations could be performed was certainly 'in the air'. What was perhaps unclear was how profitable a pursuance of these abstract ideas might be. One of the first to make the "bold" step of finding out was the French Mathematician, Maurice Fréchet (1878 - 1973).

Frechét's contribution

When 12 years old, Fréchet's leaning toward mathematics was recognised by his teacher, at the Lyceum in Buffon, who was non other than Jacques Hadamard. Hadamard proceeded to give Fréchet private instruction and encouraged Fréchet's father, himself a school teacher, to foster these talents. Hadamard continued to instruct Fréchet throughout his student years.

Fréchet's early work was on the representation of functionals, however, in his doctoral thesis of 1906, he sought to encompass the ideas of Volterra, Cantor, Arzelà and others in a general theory of abstract function spaces.

Despite many interruptions due to the war, Fréchet continued to work on abstract analysis until 1930 when, under the influence of Emile Borel, his

interests turned to statistics and probability theory, an area which occupied his attention for the remainder of his long working life.

Fréchet first considered the class L of sets in which an abstract notion of limit is defined.

A set E belongs to the <u>class L</u>, if there is defined in E a *procedure* which assigns to certain infinite sequences $\{A_n\} \equiv A_1, A_2, \ldots, A_n, \ldots$ of elements of E a unique "limit". Further, the procedure must be such that:

(i) a constant sequence, $A_n = A$ for all n, is assigned the limit A; and

(ii) if $\{A_n\}$ is assigned the limit A, then so is every subsequence $A_{n_1}, A_{n_2}, \ldots$ of $\{A_n\}$.

For sets of class L, Fréchet now introduces a number of useful concepts. The <u>derived</u> set S' of a set S is the set of all limits of sequences of elements of S. S is <u>closed</u> if $S' \subseteq S$; <u>perfect</u> if $S' = S$. A is an <u>interior point</u> of S if A is not the limit of any sequence of elements in the complement of S.*

S is <u>compact</u> (in modern terminology, *sequentially precompact*) if either it is finite or it has the "Bolzano-Weierstrass property": every infinite sequence has, at least, one subsequence which is assigned a limit. S is <u>extremal</u> (today's *sequentially compact*) if it is both compact and closed. Compactness is among the more important of the concepts identified by Fréchet. The "Bolzano-Weierstrass property" was important in the work of Ascoli and Arzelà. (Thus, the theorem of Arzelà, given on page 31, provides a necessary and sufficient condition for a subset of continuous functions to have this property, with respect to "uniform convergence".) It was also important for Hilbert's 1900 proof of the existence of a minimum in Dirichlet's principle. Fréchet proved, *if* $\{S_n\}$ *is a nested sequence of closed non-empty subsets of an extremal set than* $\bigcap_n S_n$ *is non-empty.*

Fréchet next considered functionals, that is, real valued functions defined on a set of class L. He defines a functional U to be <u>upper (lower) semi-continuous</u> at A if $\lim_{n \to \infty} U(A_n) \leq (\geq) U(A)$ for all sequences A_n with limit A. U is <u>continuous</u> if it is both upper and lower semi-continuous. For such functionals he proves; *an upper (lower) semi-continuous functional attains a finite maximum (minimum) on an extremal set.* (These results proved to be important in the calculus of variations where the functionals are frequently not continuous but are lower semi-continuous. For example, arc-length is lower semi-continuous but not continuous.)

---

*In these first few definitions Cantor's influence is readily discerned. Indeed, Cantor's notion of set is basic to all of Fréchet's work.

Fréchet then proceeded to develop results for functionals and sequences of functionals analogous to; the intermediate value theorem, Arzelà's theorem of page 31 and those for Baire's class 0,1,2,... functions.

The next step was to introduce a special subclass of L in which the derived set of any set is always closed.

V is the subclass in which the limit process is defined in terms of a real valued function (A,B), termed by Fréchet a *neighbourhood* (voisinage).

(A,B) is required to satisfy

(i) (A,B) = (B,A) $\geq$ 0 for all elements A,B of the set

(ii) (A,B) = 0 if and only if A = B

(iii) there is a real valued function f(ϵ) for which $\lim_{\epsilon \to 0}$ f(ϵ) = 0, such that (A,B) $\leq$ f(ϵ) whenever (A,C) and (C,B) are both less than ϵ.

The sequence $\underline{\{A_n\}\ \text{has limit A}}$ if $(A_n,A) \to 0$ as $n \to \infty$.

[In (iii) we recognise a type of weakened "triangle inequality". Indeed, in an attempt to prove the converse of the previously cited theorem: *If* K *is an extremal set, then every continuous functional* U *attains a finite maximum and minimum on* K, Fréchet considered a true "metric" for which (iii) is replaced by the requirement (A,B) $\leq$ (A,C) + (C,B) for all elements A, B and C of the set. In this case he referred to (A,B) as an *ecart* (literally "écart" translates as "distance apart"). It is worth remarking that while Fréchet's theory was more general, in all the examples he considered an *ecart* was present.]

For sets of class V, Fréchet gave the following definitions and results.

The $\underline{\text{spheroid}}$ of centre A and radius ρ is the set of all B such that (A,B) < ρ. A set is $\underline{\text{bounded}}$ if it is contained in a spheroid of finite radius and *every extremal set is bounded*. A version of the "open-covering" characterization of extremal sets is proved, however to obtain the full result he further restricted attention to what he termed V-normal sets. A set is $\underline{\text{V-normal}}$ if it is perfect, separable (contains a countable dense subset) and every Cauchy sequence has a limit. Here, $\{A_n\}$ is a Cauchy sequence if for every positive integer p, $(A_{n+p},A_n) \to 0$ as $n \to \infty$.

In V-normal sets he also introduced the usual ϵ-δ definition of continuity and uniform continuity for functionals and proved that *every continuous functional on an extremal set is uniformly continuous*.

This work was followed by several examples, some of which are listed below.

(1) The set of all continuous real valued functions on some interval I. Here the *ecart* is (f,g) = $\max_{x \in I} |f(x) - g(x)|$ and "limit" is "uniform limit".

MAURICE FRÉCHET (1954)

(2) The set of all sequences of real numbers E_ω.

Fréchet referred to this as "a space of countably many dimensions", a point of which may correspond to the Taylor coefficients of a function expanded about some fixed origin.

As an *ecart* on E_ω Fréchet used

$$(x, y) = \sum_{p=1}^{\infty} \frac{1}{p!} \frac{|x_p - y_p|}{1 + |x_p - y_p|}$$

where $x = (x_1, x_2, \ldots, x_p, \ldots)$ etc.

(3) The set of all "curves" (directed arcs) in three dimensional space. Such a "curve" γ is represented parametrically by

$$x = \gamma_1(t), \ y = \gamma_2(t), \ z = \gamma_3(t), \quad 0 \leqslant t \leqslant 1,$$

where $\gamma_1, \gamma_2, \gamma_3$ are continuous real valued functions. The *ecart* is defined by

$$(\gamma, \gamma^*) = \text{g.l.b.} \left(\max_t \sqrt{(\gamma_1(t) - \gamma_1^*(t))^2 + (\gamma_2(t) - \gamma_2^*(t))^2 + (\gamma_3(t) - \gamma_3^*(t))^2} \right)$$

where the greatest lower bound is taken over all possible parametric representations of the two curves γ, γ^*.

For each of these examples Fréchet went on to show that the resulting "space" is normal; that is, perfect, separable and, in modern terminology, complete.

The significance of Examples (1) and (3) for the calculus of variations is obvious. Indeed, in 1911 Fréchet considered the "differentiability" of functionals defined on such spaces.

Thus, for the space of example (1), Fréchet said the functional U is differentiable at f if there exists a linear functional L_f (see definition on page 26) such that

$$\lim_{\lambda \to 0} \text{g.l.b.} \left| \frac{U(f+\lambda g) - U(f)}{\lambda} - L_f(g) \right| = 0,$$

here the greatest lower bound is taken over all g with $M(g) = 1$ where $M(g) = \underset{x \in I}{\text{Max}} |g(x)|$.[*]

---

*Fréchet gave the definition in the equivalent form

$$U(f+g) - U(f) = L_f(g) + o(M(g)).$$

In this form you may recognise the definition as that often used in the calculus of several variables to define the total differential of a function.

Differentiability of U at f in Fréchet's sense implies continuity of U at f and so Fréchet's differentiability proved too stringent a condition for use in the calculus of variations (see comment on page 33). The most useful form of differentiation for this purpose appears to have been found by a contemporary of Fréchet, and also a student of Hadamard, R. Gateaux in 1913. Unfortunately Gateaux was killed early in the Great War and a full version of his work did not appear until 1922 when Lévy prepared Gateaux's papers for publication.

Gateaux dropped the uniformity in Frechet's definition. U is Gateaux differentiable at f if for each g the following limit exists

$$\lim_{\lambda \to 0} \frac{U(f + \lambda g) - U(f)}{\lambda}$$

The use of these and similarly defined higher order derivatives in the calculus of variations were considered by Charles Albert Fischer (1884 – 1922) and Elizabeth L. Stourgeon (1881 – 1971), who restricted attention to weak variations by taking $M(g) = \underset{x \in I}{Max} \{ |g(x)|, |g'(x)| \}$.

By 1924, these ideas had been brought near to a definitive form through the efforts of the Italian Mathematician, Leonida Tonelli (1885 – 1946). Tonelli's work brought to a "successful conclusion" the programme started by Volterra in 1887.

Fréchet's general approach gave considerable impetus to the newly emerging theory of "point set topology" (now known as *General Topology*), initiated by Cantor. Thus, in Felix Hausdorff's (1868 – 1942) influential book, *Grundzüge der Mengenlehre* (Essentials of set theory, published in 1914), there is a great deal of dependence on Fréchet's ideas. Hausdorff both refined and further generalized Fréchet's work. We owe much of our modern terminology to Hausdorff. For example, Hausdorff defined a metric space (Fréchet's *ecart*-spaces) and the notion of completeness: every Cauchy sequence is convergent. He also obtained many new results for such spaces including a generalization of Baire's theory of Category I and II sets. He proved that the complement of a first category subset in a complete metric space is of the second category.

By way of generalization, Hausdorff build on an idea used by Hilbert in 1902 in a special axiomatic approach to plane euclidean geometry.

Essentially, he introduced the notion of a "neighbourhood" of a point which was to replace the spheroids of a metric space.

Hausdorff defined a <u>topological space</u> to be a set of points x together with a specified family of subsets, termed <u>neighbourhoods</u> and satisfying:

(i)   each point x is contained in at least one neighbourhood U_x;

Operationen mit topologischen Räumen.

Topologischer Raum = Menge E, wo ein System offener Mengen G mit den Bedingungen definiert ist:

(1) 0 und E sind offen

(2) Die Summe beliebig vieler offener Mengen ist offen.

(3) Der Durchschnitt endlich vieler offener Mengen ist offen.

$A \subset E$ ist topologischer Teilraum von E, wenn die in A offenen Mengen mit den Mengen AG (G in E offen) identisch sind.

A_t ($t \in T$) sei ein System topologischer Räume.

I. Topologisierung des Durchschnitts $A = \prod_t A_t$ $(\neq 0)$. [x)]

Die Topologisierung von A ist bereits durch die jeder einzelnen A_t eindeutig bestimmt, indem die in A offenen G mit den $A G_t$ (G_t in A_t offen) identisch sein müssen: die notwendige und hinreichende Bedingung ist also, dass $A G_t$ ein von t unabhängiges Mengensystem durchläuft, oder dass jedes $A G_s$ auch ein $A G_t$ ist (für beliebige $s, t \in T$).

x) natürlich so, dass A topologischer Teilraum von jedem A_t ist.

A PAGE FROM HAUSDORFF'S NOTE-BOOK IN WHICH THE MORE MODERN DEFINITION OF A TOPOLOGICAL SPACE IS USED

(ii)   The intersection of two neighbourhoods of a point x is itself
a neighbourhood of x.   (Thus the family of neighbourhoods is
closed under non-empty finite intersections.)

(iii) If U_x is a neighbourhood of the point x and y is any
point in U_x, then there exists a neighbourhood U_y of y
with $U_y \subseteq U_x$.

(iv) [Hausdorff separation axiom]   If x and y are distinct points
then there exist disjoing neighbourhoods U_x, U_y of x and y
respectively.
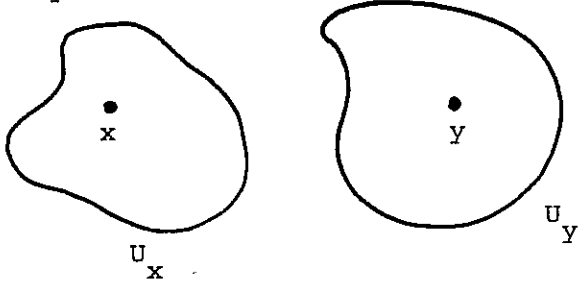
Using this more general structure, Hausdorff is able to redefine many
of Fréchet's terms. Thus: x is a <u>limit point</u> of a set of points if every
neighbourhood of x contains other points of the set; a set is <u>open</u> if
every point of it is an <u>interior point</u>; that is has a neighbourhood con-
taining only points of the set, etc.

The analogy of the usual ε-δ definition of continuity is as follows.
A mapping f   from one topological space into another is <u>continuous</u> at
x if given any neighbourhood $E_{f(x)}$ of f(x) there is a neighbourhood
Δ_x of x whose points all map into $E_{f(x)}$, that is $f(\Delta_x) \subseteq E_{f(x)}$.
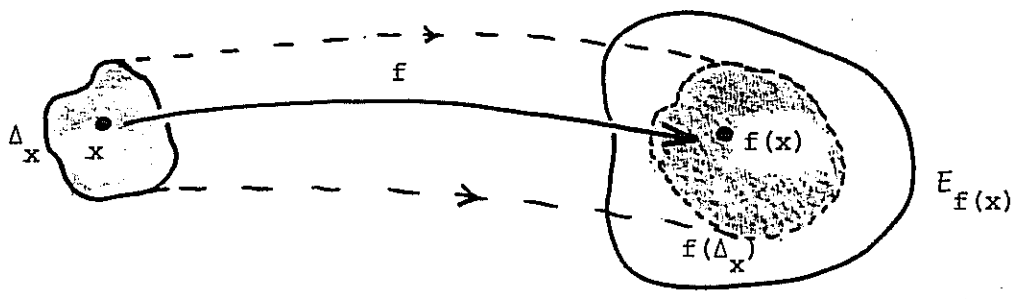
Hausdorff's work was the point of departure for considerable later
research. In particular we had the work of a Russian school of
topologists – Alexandroff (1896 –   ), Urysohn (1898 – 1924), Tychonoff
(1906 –   ) and an American school which developed around Robert L.
Moore (1882 – 1974).

The Russian school was largely concerned with the question of when
a given topological space could be realised as a metric space. For separable
spaces this question was answered by Urysohn (published in 1925)* Urysohn,
together with Karl Menger (1902 –   ) also developed the first generally
acceptable definition of "dimension" in topological spaces.

---

* However, complete answers were not given till the late forties, by a
number of mathematicians including one of R.L. Moore's students, R.H. Bing.

We will not pursue general topology further except to remark that it has remained an area of active research to the present day.

The theory of integral equations, as developed by Volterra, Fredholm and, in particular, Hilbert, inspired an alternative and, in many ways, independent development of function spaces. In this work algebraic as well as topological properties were important.

Hilbert's study of a finite number of equations in the same number of unknowns made frequent use of the fact that these equations are embedded in a more general vector structure in which "addition", "inner product" etc. are defined and in which "geometrical" reasoning is possible. He did not, however, explicitly attempt to develop a similar structural setting for the transcendental problem. There, the appropriate equations and manipulations were arrived at by analogy (or rigorous passages to the limit) from individual results in the finite theory.

The possibility of developing a structural setting for the transcendental problem was not overlooked for long. In 1909 a faltering and isolated attempt to provide a general theory was made by the American Mathematician, Eliakim H. Moore (1862 - 1932). Moore aimed at developing a broad axiomatic theory which would include not only Hilbert's results but also other classical results on infinite systems of linear equations. On the other hand, starting in 1907, the German Mathematician, Erhard Schmidt (1876 - 1959) undertook to simplify and illuminate what Hilbert had done. Schmidt's approach proved both successful and influential.

The contribution of E. Schmidt

In relating his theory of infinite quadratic forms to integral equations, Hilbert had used the fact that the continuous functions on an interval, with which he worked, are uniquely determined by their Fourier coefficients. Using this, Schmidt identified these functions with "points" in the space of square summable infinite sequences of complex numbers.

Thus, a point of Schmidt's space is

$$z = \{z_p\} \equiv z_1, z_2, \ldots, z_p, \ldots,$$

where

$$\sum_{p=1}^{\infty} |z_p|^2 < \infty.$$

Following Hilbert, Schmidt defines

$$[z, w] = \sum_{p=1}^{\infty} z_p w_p',$$

he also introduces the notation

$$\|z\| = \sqrt{\sum_{p=1}^{\infty} z_p \bar{z}_p} = \sqrt{[z, \bar{z}]}^{\,*}$$

[Note: By the square summability assumption $\|z\| < \infty$.]

Schmidt now introduces "geometric" language (and hence thinking) into his space.

Two elements z, w are called *orthogonal* if $[z, \bar{w}] = 0$. From this follows the generalized Pythagorean theorem: *if z and w are orthogonal then* $\|z+w\|^2 = \|z\|^2 + \|w\|^2$. Schwarz's inequality:
$|[z,\bar{w}]| \leqslant \|z\| \|w\|$ and the triangle inequality: $\|z + w\| \leqslant \|z\| + \|w\|$, are also proved.

ERHARD SCHMIDT

Schmidt then considers sequences of mutually orthogonal elements and sets of linearly independent elements. Here, Bessel's inequality, Parseval's identity for finite sums, and the recursion formula now known as the Gram-Schmidt orthogonalization procedure are established.

Finally, Schmidt proceeds to what are undoubtedly the most significant and novel aspects of his work.

The notions of strong convergence and strong Cauchy sequence are introduced. The sequence $\{z^{(n)}\}$ strongly converges to z if $\|z^{(n)}-z\| \to 0$. Schmidt then proves that his sequence space is complete (that is, every strong Cauchy sequence is strongly convergent).

By a closed subspace, Schmidt understands a subset A which is both "topologically closed" (contains all its strong limit points) and "algebraically closed" (that is, if z, w are elements of A and λ_1, λ_2 are any pair of complex numbers, then $\lambda_1 z + \lambda_2 w$ is an element of A).

For any closed subspace A and any point z, Schmidt establishes the existence of unique points $w^{(1)}$, $w^{(2)}$ such that $z = w^{(1)} + w^{(2)}$ and $w^{(1)}$ belongs to A, $w^{(2)}$ is orthogonal to A (that is, $[w^{(2)}, \bar{a}] = 0$

---

*It is $[z,\bar{w}]$ which accords with the modern notion of inner-product. Thus through Schmidt's work we find $[z,\bar{w}]$ being used where today we would use the complex inner-product $(z,w) = \sum_{p=1}^{\infty} z_p \bar{w}_p$.

for all points a of A).

He then proves that, for every point a of A $\|w^{(2)}\| \leqslant \|a - z\|$, with equality only when $a = w^{(1)}$. On the basis of this result he refers to $\|w^{(2)}\|$ as the *distance* between z and A.

The work of Friedrich Riesz

In 1902 the French Mathematician, Henri Lebesgue (1875 – 1941) introduced a very general and powerful theory of integration.

The Lebesgue integral is basic to modern real analysis and the development of his ideas is a major area of active research.

Under Lebesgue's definition many more functions proved to be integrable than had previously been the case.

For example, the *Dirichlet function*

$$f(x) = \begin{cases} 0 & \text{for } x \text{ irrational} \\ 1 & \text{for } x \text{ rational} \end{cases}$$

Henri LEBESGUE

is Lebesgue integrable, $\displaystyle\int_0^1 f(x) \, d\mu(x) = 0$, but not integrable in the sense of Riemann.

One consequence of this is that, under quite mild conditions, Lebesgue was able to show that the "limit" of a sequence of Lebesgue integrable functions is itself Lebesgue integrable. In earlier theories the integrability of the limit usually had to be assumed. This had been an impediment in the work of Ascoli and Arzelà for example.

The Lebesgue integral proved especially useful in the theory of Fourier series.

LEBESGUE INTEGRATION IN ONE DIMENSION

Lebesgue based his theory on the notion of the "measure" of a set of real numbers. Let E be a set of real numbers contained in the interval $[a,b]$. The outer measure of E is

$$\mu(E) \;=\; \text{g.l.b.} \; \sum_{i=1}^{\infty} \ell_i$$

where the ℓ_i are the lengths of a countable collection of intervals whose union contains E, the greatest lower bound is taken over all such collections of intervals.

E is said to be measurable, with (Lebesgue) measure $\mu(E)$ if
$\mu(E) \;=\; (b-a) - \mu([a,b]\backslash E)$.

After showing μ is well defined and "countably additive" Lebesgue defines a real valued function f on $[a,b]$ to be "measurable" if $\{x : f(x) > a\}$ is measurable for all real numbers a.

f is now said to be (Lebesgue) integrable (or summable) if

$$\text{g.l.b.} \quad \sum_i a_i \, \mu(\{x : a_{i-1} < f(x) < a_i\}) \quad < \quad \infty$$

where the greatest lower bound is taken over all possible partitions $a_1, a_2, \ldots, a_n$ of the *range* of f. When it exists, the value of this greatest lower bound is the Lebesgue integral of f and is denoted by

$$\int_a^b f(x)\,d\mu(x).$$

After establishing that the Lebesgue integral is well defined and agrees with the Riemann integral of f when the latter is defined, a number of important results may be established, for example:
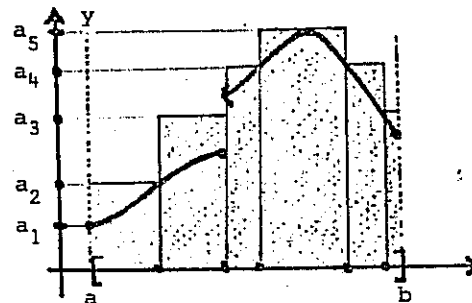
Dominated Convergence Theorem:

Let $\{f_n\}$, f and g *be measurable functions such that*

i) *For each* $n=1,2,\ldots.$
$\mu\{x : |f_n(x)| > |g(x)|\} = 0$ $\left\{\begin{array}{l}\text{that is, } |f_n| \le |g| \text{ except on a set of}\\\text{measure zero.}\end{array}\right.$

ii) *For each* $\varepsilon > 0$, as $n \to \infty$
$\mu\{x : |f_n(x) - f(x)| > \varepsilon\} \to 0.$

and iii) g *and each* f_n *is integrable.*

Then, f *is integrable and, as* $n \to \infty$

$$\int_a^b f_n(x)\,d\mu(x) \quad \to \quad \int_a^b f(x)\,d\mu(x).$$

For example, in 1906 Pierre Fatou (1878 – 1929), generalizing a result of Lebesgue, proved that

$$\frac{1}{\pi} \int_{-\pi}^{\pi} f(x)g(x)d\mu(x) = 2a_0\alpha_0 + \sum_{n=1}^{\infty} (a_n\alpha_n + b_n\beta_n)$$

where a_n, b_n and α_n, β_n are the Fourier coefficients for $f(x)$ and $g(x)$ respectively:

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, d\mu(x) \qquad \alpha_n = \frac{1}{\pi} \int_{-\pi}^{\pi} g(x) \cos nx \, d\mu(x)$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, d\mu(x) \qquad \beta_n = \frac{1}{\pi} \int_{-\pi}^{\pi} g(x) \sin nx \, d\mu(x),$$

and $f(x)$, $g(x)$ are assumed to be Lebesgue square integrable; that is

$$\int_{-\pi}^{\pi} (f(x))^2 d\mu(x) < \infty \quad \text{and} \quad \int_{-\pi}^{\pi} (g(x))^2 d\mu(x) < \infty.$$

We might recognise this as the condition necessary for the orthonormal family, $\dfrac{\cos x}{\sqrt{\pi}}, \dfrac{\sin x}{\sqrt{\pi}}, \dfrac{\cos 2x}{\sqrt{\pi}}, \dfrac{\sin 2x}{\sqrt{\pi}}, \ldots,$ to be *complete* in the sense used by Hilbert (see page 16) to develop his theory of symmetric integral equations.

These results led the Hungarian mathematician, Friedrich Riesz (1880 – 1956) to extend Hilbert's results for integral equations of the form

$$y(x) = f(x) + \lambda \int_a^b K(x,t) \, y(t) \, dt$$

where $f(x)$ and $K(x,t)$ are <u>continuous</u> functions to the case when f and K are only assumed to be Lebesgue square integrable.

In the course of this work, Riesz was led to consider the "moment problem" of determining (if it exists) a function $f(x)$ which has as its "Fourier" coefficeints a given sequence of numbers.[*]

In 1907, Riesz proved the following basic result.

(R) *Let $\{\phi_p\}$ be an orthonormal sequence of Lebesgue square integrable functions on the interval $[a,b]$. Then, given a sequence of real numbers $\{a_n\}$ a necessary and sufficient conditon for there to exist a function f such that*

---

[*]The moment problem for arbitrary families of functions has proved basic to certain questions in statistics and was one of the problems considered by Fréchet in his later work.

$$\int_a^b f(x) \, \phi_p(x) \, d\mu(x) = a_p \quad \text{for} \quad p = 1,2,3,\dots$$

is that

$$\sum_{p=1}^{\infty} a_p^2 < \infty.$$

Further if $\{\phi_p\}$ *is "complete" then* f *is unique up to the addition of a function* $\eta(x)$ *with* $\int_a^b |\eta(x)| \, d\mu(x) = 0$ *such a function is known as a* null *function.*

Less than a month later and in the same journal (*Comptes Rendus*) the German Mathematician, Ernst Fischer (1875 - 1959) published the following result.

Fischer said that a sequence of Lebesgue square integrable functions $\{f_n\}$ is <u>convergent in the mean</u>[*] on [a,b] if

$$\lim_{n,m \to \infty} \int_a^b (f_n(x) - f_m(x))^2 \, d\mu(x) = 0$$

and $\{f_n\}$ <u>converges in mean</u> to f if

$$\lim_{n \to \infty} \int_a^b (f_n(x) - f(x))^2 \, d\mu(x) = 0.$$

Fisher then proves

(F) *If* $\{f_n\}$ *is a sequence of Lebesgue square integrable functions which converges in mean, then there exist a function* f *(unique up to the addition of a null function) such that* $\{f_n\}$ *converges in mean to* f.

Riesz's theorem (R) is then deduced as a consequence. Indeed (F) and (R) are equivalent and are now known as the Riesz-Fischer Theorem.

Via a given complete orthonormal family (the existence of which had been established by Hilbert, page 16), Riesz's theorem establishes a one-to-one correspondence between Lebesgue square integrable functions and square summable sequences.

Following the publication of Riesz's result, both Schmidt and Fréchet remarked that the space of Lebesgue square integrable functions on the interval [a,b], now denoted by $L^2[a,b]$, has a "geometry" completely analogous to

---

[*]Today we would more likely use "Cauchy in mean" instead of "convergent in mean".

Friedrich Riesz

Schmidt's space of square summable sequences. Indeed Fréchet introduced the *ecart*

$$(f, g) = \sqrt{\int_a^b (f(x) - g(x))^2 \, dx}$$

and used this and Riesz's result to characterize the extremal subsets of $L^2[a,b]$.

[Note: In order to satisfy property (ii) of an *ecart* (page 34) it is necessary to identify two functions which differ by a null function.]

Fréchet also proved that for every continuous linear functional U on $L^2[a,b]$ there exists $u(x) \in L^2[a,b]$ such that

$$U(f) = \int_a^b f(x) \, u(x) \, d\mu(x).$$

It is worth remarking that, in terms of Fréchet's *ecart* the Riesz-Fisher Theorem may be interpreted as: $L^2[a,b]$ *is a complete metric space.*

These ideas were further elucidated by Riesz, who, in 1910, was able to realize $L^2[a,b]$ as a special case of a more general class of spaces; the so called L^p spaces.*

For $1 < p < \infty$, Riesz considered the set $L^p[a,b]$ of functions f for which

$$\int_a^b |f(x)|^p d\mu(x) < \infty.$$

(Two functions are identified if they only differ by a null function.)

Using inequalities developed by Hölder (1859 - 1937) and Minkowski (1864 - 1909), Riesz establishes that $L^p[a,b]$ is closed under finite linear combinations and that, *for any* $f \in L^p[a,b]$ *the product* $f(x)g(x)$ *is Lebesgue integrable if and only if* $g \in L^q[a,b]$ *where* q *is such that* $\frac{1}{p} + \frac{1}{q} = 1$. (Henceforth, it will be assumed that p and q are related in this way.) Two types of convergence are introduced into $L^p[a,b]$. The sequence $\{f_n\}$ <u>converges strongly</u> to f if

$$\int_a^b |f_n(x) - f(x)|^p d\mu(x) \rightarrow 0.$$

$\{f_n\}$ <u>converges weakly</u> to f if

(i) $\int_a^b |f_n(x)|^p d\mu(x) < M$ for all n and some $M > 0$; and

---

* "Riesz' idea may have been motivated by the earlier work of H. Minkowski who in connection with his work on the 'geometric theory of numbers', considered, for $1 \leq p < \infty$, $\left(\sum_{i=1}^n |x_i - y_i|^p\right)^{1/p}$ as a distance function between points $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in n-dimensional space."

(ii) $\displaystyle\int_{a}^{t} (f_n(x) - f(x))d\mu(x) \to 0$ for all t with $a \leqslant t \leqslant b$.

(Later, he effectively shows that this last definition is equivalent to the modern one:

$$f_n \xrightarrow{\text{weakly}} f \quad \text{if} \quad \int_{a}^{b} (f_n(x) - f(x))\ g(x)d\mu(x) \to 0$$

for every $g \in L^q[a,b]$.)

Riesz notes that strong convergence always implies weak convergence and gives an example $(f_n(x) = \cos nx$ in $L^p[0,1])$ to show that the converse does not hold.

Fischer's theorem (F) of page 42 is now generalized to prove: $L^p[a,b]$ is complete with respect to strong convergence; that is, if $\{f_n\}$ is a strong Cauchy sequence then it is strongly convergent to some f in $L^p[a,b]$.

For weak convergence, Riesz proves that every "bounded sphere" in $L^p[a,b]$ is weakly compact; that is, if $\{f_n\}$ satisfies (i) above then $\{f_n\}$ has a weakly convergent subsequence.

He also answers the moment problem:
Can one find f *in* $L^p[a,b]$ *such that*

$$\int_{a}^{b} f(x)g_n(x)d\mu(x) = a_n$$

where the g_n *are a given sequence of functions in* $L^q[a,b]$ *and the* a_n *are a given sequence of real numbers.* In the course of this he proves what is now known as the <u>Riesz Representation theorem</u>: A *is a bounded linear functional on* $L^p[a,b]$ (that is,

$$A(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 A(f_1) + \lambda_2 A(f_2)$$

for all real numbers λ_1, λ_2 and all f_1, f_2 in $L^p[a,b]$ and

$$|A(f)| \leqslant M\left[\int_{a}^{b} |f(x)|^p d\mu(x)\right]^{\frac{1}{p}}$$

for all f in $L^p[a,b]$ and some $M > 0$) *if and only if*

$$A(f) = \int_a^b a(x)f(x)d\mu(x)$$

for some function a *in* $L^q[a,b]$. *Further for any given* A, a *is unique up to the addition of a null function.*

We should recognise this as a generalization of Fréchet's result on page 43.

Riesz's attention now turns to the study of integral equations in L^p spaces. While the problem under consideration is rather specialized, it is in this work that we find the beginnings of "abstract" operator theory.

He introduces the notion of an <u>operator</u>[*] T form $L^p[a,b]$ into itself and defines T to be <u>linear</u> if

$$T(\lambda_1 f_1(x) + \lambda_2 f_2(x)) = \lambda_1 T(f_1(x)) + \lambda_2 T(f_2(x))$$

and <u>bounded</u> if

$$\int_a^b |T(f(x))|^p d\mu(x) \leqslant M^p$$

for all $f \in L^p[a,b]$ with $\int_a^b |f(x)|^p d\mu(x) \leqslant 1$, and some $M > 0$.

An example of such an operator might be

$$T(f(x)) = \int_a^b K(x,t)f(t)d\mu(t)$$

for suitable K, and so in this notation the study of integral equations is included in the study of "operator equations" of the form

$$\phi(x) = f(x) + \lambda T(\phi(x)).$$

Riesz now introduces the important concept of <u>adjoint operator</u>. For any fixed $g \in L^q[a,b]$

$$A(f) = \int_a^b T(f(x))g(x)d\mu(x)$$

defines a bounded linear functional on $L^p[a,b]$ and so, by his representation theorem there exists a unique $a \in L^q[a,b]$ such that

---

[*] Riesz uses the term "functional transformation".

$$A(f) = \int_a^b f(x)a(x)d\mu(x).$$

He now defines the <u>adjoint operator</u> T^* by

$$T^*(g(x)) = a(x).$$

T^* is then a bounded linear operator on $L^q[a,b]$ such that

$$\int_a^b T(f(x))g(x)d\mu(x) = \int_a^b f(x)T^*(g(x))d\mu(x)$$

for all $f \in L^p[a,b]$ and $g \in L^q[a,b]$.

The question of invertibility of T is considered next; that is, the solvability of the homogeneous problem

$$T(\phi(x)) = f(x) \quad \text{for all} \quad f \quad \text{in} \quad L^p[a,b].$$

Noting that T is invertible if and only if T^* is $((T^*)^{-1} = (T^{-1})^*)$ he derives a necessary and sufficient condition for the existence of these inverses; namely, that there exists $m > 0$ such that

$$\int_a^b |f(x)|^p d\mu(x) \leqslant m^p \int_a^b |T(f(x))|^p d\mu(x)$$

and

$$\int_a^b |g(x)|^q d\mu(x) \leqslant m^q \int_a^b |T^*(g(x))|^q d\mu(x).$$

Restricting himself to the case $p = q = 2$, Riesz now tackles the eigenvalue problem for

$$\phi(x) = f(x) + \lambda T(\phi(x)).$$

Noting that the symmetric integral equations studied by Hilbert correspond to the case when $T = T^*$, he further specializes to this class of problems for which he is able to parallel many of Hilbert's results. To match Hilbert's result for complete continuity (see page 17), Riesz defines an operator T to be <u>completely continuous</u> if T maps every weakly convergent sequence of functions into a strongly convergent one, that is, if $f_n \xrightarrow{\text{weakly}} f$ then $T(f_n(x)) \xrightarrow{\text{strongly}} T(f(x))$. He then establishes that if $T = T^*$ is completely continuous, then

Umkehrung der linearen Funktionaltransformation.

Es bedeute $T[f(x)]$ irgend eine lineare Transformation der Klasse $[L^p]$, $\mathfrak{X}[g(x)]$ die zu ihr transponierte Transformation der Klasse $\left[L^{\frac{p}{p-1}}\right]$. Wir fragen nach der Lösbarkeit der Funktionalgleichung

(34) $$T[\xi(x)] = f(x).$$

Darin bedeutet $f(x)$ die gegebene, $\xi(x)$ die gesuchte Funktion aus der Klasse $[L^p]$. Das Gleichheitszeichen ist bis auf eine additive Nullfunktion zu deuten.

Laut der Entwicklungen des vorhergehenden Paragraphen ist die Gleichung (34) *gleichwertig dem mit sämtlichen Funktionen $g(x)$ der Klasse* $\left[L^{\frac{p}{p-1}}\right]$ *gebildeten Gleichungssystem*

(35) $$\int_a^b \xi(x)\, \mathfrak{X}[g(x)]\, dx = \int_a^b f(x)\, g(x)\, dx.$$

Das Gleichungssystem kann durch endlich oder abzählbar unendlich viele Gleichungen ersetzt werden; man kann hierfür z. B. wieder die Gesamtheit der in § 12 benutzten speziellen streckenweise konstanten Funktionen heranziehen; denn zufolge der Linearität der Transformation $\mathfrak{X}$ kann man jede Funktion $\mathfrak{X}[g(x)]$ durch jene, die nach Anwendung derselben Transformation aus diesen speziellen Funktionen hervorgehen, $\left(\text{in bezug auf}\frac{p}{p-1}\right)$ stark approximieren. Man darf daher das in § 11 formulierte Kriterium auch auf das System (35) anwenden. Beachtet man noch, daß das System der $\mathfrak{X}[g(x)]$ alle linearen Verbindungen dieser Funktionen mit enthält, so ergibt sich:

Eine notwendige und hinreichende Bedingung für die Lösbarkeit der Funktionalgleichung (34) durch eine Funktion $\xi(x)$, für welche

$$\int_a^b |\xi(x)|^p\, dx \leqq M^p$$

ausfällt, besteht darin, daß die Ungleichung

$$\left| \int_a^b f(x)\, g(x)\, dx \right| \leqq M \left[\int_a^b |\mathfrak{X}[g(x)]|^{\frac{p}{p-1}}\, dx \right]^{\frac{p-1}{p}}$$

für alle Funktionen $g(x)$ der Klasse $\left[L^{\frac{p}{p-1}}\right]$ erfüllt ist.

$$T(f(x)) = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} P_n(f(x))$$

where the λ_n's are the eigenvalues of T (real and countable in number) and each P_n is an operator analogous to a projection to the "eigenspace" corresponding to λ_n.

Before leaving the work of Riesz we mention a further paper, published in 1918. In this paper he proceeds to a study, similar to that undertaken for L^p spaces, of the space of all continuous functions on a closed interval $[a,b]$ (Example (1) of Fréchet's on page 34). In this paper he introduces much of the terminology later used in general "Banach spaces". For example, for f in this space he calls $\max_{a \leqslant x \leqslant b} |f(x)|$ the *norm* of f and, following Schmidt, denotes it by $\|f\|$. Up to this point, he had not, however, used similar notations in his study of the L^p spaces.*

THE EMERGENCE OF AN AXIOMATIC APPROACH

Fréchet developed axiom systems which were sufficiently broad to encompass "topological structures" such as those considered by Volterra, Ascoli, Arzelà, Osgood and Baire. Hausdorff's system of axioms allowed for even more general structures.

The spaces discussed by Schmidt and Riesz certainly fitted into these axiom schemes, however the axioms were too general to fully capture the rich structure of these spaces, in which "algebraic" operations were combined with topological ones.

Axiom systems more attuned to these richer structures were presented by: Stefan Banach (1892 - 1945), in his doctoral thesis of 1920; Norbert Wiener (1894 - 1964) in a paper presented at the 1920 International Congress of Mathematicians held in Strasbourg; Eduard Helly (1884 - 1943) in 1921; and Hans Hahn (1879 - 1934) in 1922. The first three men almost certainly worked independently of one another, while Hahn based much of his work on that of Helly. It was however, Banach who more fully developed the work and certainly it was his contributions which proved most influential. At the time, Wiener's interest was in logic and the foundations of mathematics. His set of axioms was presented as an example of an axiom system which might describe various sets of functions occurring in analysis. Wiener did not investigate any consequences of his rather complicated system of axioms.

In contrast, both Helly and Hahn presented their axioms in the context of particular problems; in Helly's case, the solution of systems of infinitely many equations in infinitely many unknowns. Hahn's axioms are

---

* "In all of this work Riesz made no explicit reference to Fréchet's work or the notion of an *ecart*, though in 1908 Riesz had contributed a paper generalizing Fréchet's axioms along similar lines to those of Hausdorff (see page 36) and so certainly was aware of these ideas."

similar to those of Helly and are almost identical to those given by Banach.
Hahn regarded his axioms as a tool for studying the representation of certain
integrals as limits of other integrals and derived results in a number of
specific spaces. For example, in the space of continuous functions on [a,b]
with norm $\mathcal{D}(f) = \underset{a\leqslant x \leqslant b}{\text{Max}} |f(x)|$, he establishes:

A necessary and sufficient condition for

$$\int_a^b f(x)\phi_n(x)\,dx \to \int_a^b f(x)\phi(x)\,dx$$

for all f in the space, is that the result holds for $f(x) = 1$ *and for*
$f(x) = x$.

Neither Helly or Hahn regarded their axioms as the basis for an abstract
theory, however, the important notion of "dual space" is to some extent
latent in both their papers.

Stefan Banach and his work

The departure of the Russian army from Warsaw in 1915 marked the end
of more than a century of occupation for Poland. With independence came the
establishment of a Polish university and the formation of a number of learned
societies. One in particular, the Mianowski Fund, actively encouraged the
development of mathematics in Poland. This Society created a series of books
for self-learning by the young, as well as publishing journals for the
dissemination of more advanced work. It was in one of their publications,
Nauka Polska, that an article by the young Polish mathematician, Zydumunt
Janiszewski "On the needs of Polish mathematics" appeared in 1918.
Janiszewski called for the creation of a "Polish school of mathematics"
concentrating on one field and supported by the publication of a journal
which would gain international recognition by accepting articles written in
the more widespread foreign languages.

Janiszewski's programme was implemented by the mathematician-teachers
of the time: Janiszewski, Stanislaw Mazurkiewicz, Waclaw Sierpiński, Hugo
Steinhauss, S. Zaremba and K. Żorawski.

In 1920 (the same year as Janiszewski's death) the first volume of
Fundamenta Mathematicae appeared. This journal is still published today
and soon gained international status, attracting papers on Set Theory, the
Foundations of Mathematics, Topology and related material. It was followed

shortly afterwards by another journal, *Studia Mathematica**, specializing in functional analysis.

On a walk through the Cracow Green Belt during the summer of 1916, Steinhaus overheard part of a conversation between two youths, Otto Nikodym and Stefan Banach - in which unexpectedly the words "Lebesgue integral" occurred. Banach proved to be an engineering student whose mathematics was largely self-taught. (Banach barely knew his parents; he had been given to a washerwoman for bringing-up, and by the age of fifteen had to earn his own living, giving private lessons, preferably on mathematics.) In 1920 Banach was appointed as an "Assistant in Mathematics" at the Lwów Technological University and so began his career as a "professional" mathematician.

In his thesis of 1920 (a "summary" of which appeared in the 1922 volume of *Fundamenta Mathematicae*), Banach introduces spaces E satisfying a set of axioms identical with those now used to define a complete normed linear (or Banach) space:

E is a (real) vector (or linear) space on which a norm function $\|x\|$ is defined and satisfies

$$\|x\| \geqslant 0,$$

$$\|x\| = 0 \text{ if and only if } x = 0,$$

$$\|ax\| = |a| \|x\| \text{ for all real numbers } a,$$

$$\|x + y\| \leqslant \|x\| + \|y\|$$

and the <u>completeness axiom</u>; if $\{x_n\}$ is a *Cauchy sequence*, that is, $\|x_n - x_p\| \to 0$ as $n, p \to \infty$, then there is an x in E such that $\|x_n - x\| \to 0$ as $n \to \infty$.

The spaces of Schmidt, Riesz and many of Fréchet's examples are Banach spaces.

For example,

$$\|f\| = \left(\int_a^b |f(x)|^p d\mu(x) \right)^{\frac{1}{p}}$$

defines a norm for $L^p[a,b]$ while $C[a,b]$ is a Banach space with respect to the norm

$$\|f\| = \underset{a \leqslant x \leqslant b}{\text{Max}} |f(x)|.$$

Banach develops a number of theorems concerning the underlying space E. Typical of these is the following.

---

*Founded by Banach in 1929

If $\{x_n\}$ *is such that* $\sum_{n=1}^{\infty} \|x_n\| < \infty$, *then* $\sum_{n=1}^{N} x_n$ *converges to a*

unique element of E *as* N $\to \infty$.

Following Fréchet, Banach also introduced "geometry" into the spaces. For example, the sphere with centre x_1 and radius r, $K(x_1, r)$ is, by definition, the set of all elements x satisfying $\|x - x_1\| \leqslant r$. He proves that if $K(x_n, r_n)$ is a nested sequence of spheres then the sequence of centres converges to an element which belongs to every sphere.

The title of Banach's paper includes a reference to the application of his theory to integral equations. While he does not explicitly mention such equations in the paper itself, he does develop results which are "obviously" relevant to the study of integral equations. Banach introduces *operations* (today termed, operators) F which map one Banach space E into another E_1. F is said to be <u>continuous</u> at x_0 if essentially $\|F(x_n) - F(x_0)\| \to 0$ whenever $\|x_n - x\| \to 0$. Banach then rephrases this in terms of spheres. After proving that sums and scalar multiples of continuous operators are continuous, Banach specializes to <u>additive</u> operators where $F(x_1 + x_2) = F(x_1) + F(x_2)$.

As Banach observes, if F is also continuous, then $F(ax) = aF(x)$ for all real numbers a and so, since he mainly considers continuous additive operators, we see that Banach is really working with what would today be termed "linear operators".

Banach next proves that F is continuous on E if it is bounded on some sphere. From this we have the corollaries:
an additive operator is continuous on E *if it is continuous at any one point;*
and
an additive operator is continuous if it is <u>*bounded*</u>*; that is, if for some* M > 0 *we have* $\|F(x)\| \leqslant M\|x\|$ *for all* x *in* E.
He also proves:
If $\{F_n\}$ *is a sequence of continuous additive operators such that* $\|F_n(x) - F(x)\| \to 0$ *for each* x *in* E, *then* F *is also a continuous additive operator and* $\{F_n\}$ *is* <u>*uniformly bounded*</u>*; that is,* $\|F_n(x)\| \leqslant M\|x\|$ *where* M *is independent of* n.

Banach concludes his 1922 paper with two results on operator equations.

(1) *Let* U *be an operator (not necessarily additive) with range and*
   domain both E. *Furthermore, let there exist a real number* M *with*

§ 1. Axiomes et définitions fondamentales.

Soit E une classe composée tout au moins de deux éléments, d'ailleurs arbitraires, que nous désignerons p. e. par $X, Y, Z, \ldots$

a, b, c désignant les nombres réels quelconques. nous définissons pour E deux opérations suivantes:

1) l'addition des éléments de E

$$X + Y, \quad X + Z, \ldots$$

2) la multiplication des éléments de E par un nombre réel

$$aX, \quad bY, \ldots$$

Admettons que les propriétés suivantes sont réalisées:

I_1] $X + Y$ est un élément bien déterminé de la classe E,

I_2] $X + Y = Y + X$,

I_3] $X + (Y+Z) = (X+Y) + Z$,

I_4] $X + Y = X + Z$ entraîne $Y = Z$,

I_5] Il existe un élément de la classe E déterminé θ et tel qu'on ait toujours $X + \theta = X$,

I_6] $a \cdot X$ est un élément bien déterminé de la classe E,

I_7] $a \cdot X = \theta$ équivaut à $X = \theta$ ou $a = 0$,

I_8] $a \neq 0$ et $a \cdot X = a \cdot Y$ entraînent $X = Y$,

I_9] $X \neq \theta$ et $a \cdot X = b \cdot X$ entraînent $a = b$,

I_{10}] $a \cdot (X+Y) = a \cdot X + a \cdot Y$,

I_{11}] $(a+b) \cdot X = a \cdot X + b \cdot X$,

I_{12}] $1 \cdot X = X$,

I_{13}] $a \cdot (b \cdot X) = (a \cdot b) \cdot X$.

II_1] $|X| \geqslant 0$,

II_2] $|X| = 0$ *équivaut à* $X = \theta$.

II_3] $|a \cdot X| = |a| \cdot |X|$,

II_4] $|X+Y| \leqslant |X| + |Y|$,

III] *Si* 1° $\{X_n\}$ *est une suite d'éléments de* E,

$\qquad 2^o$ $\lim\limits_{\substack{r=\infty \\ p=\infty}} |X_r - X_p| = 0$,

il existe un élément X *tel que*

$$\lim\limits_{n=\infty} |X - X_n| = 0.$$

$0 < M < 1$ *such that for all pairs of elements* x, x' *in* E,
$\|U(x) - U(x')\| \leqslant M\|x - x'\|$, *then there exists a (unique) element* x *of*
E *such that* U(x) = x.

This is Banach's contraction mapping theorem which is important in many applications, particularly to differential and integral equations.

(2) *Let* F *be a continuous additive operator with range and domain in* E. *Let* M *be the greatest lower bound of all those numbers* M' *satisfying* $\|F(x)\| \leqslant M'\|x\|$ *for all* x *in* E (today we call M the norm of F and denote it by $\|F\|$). *Then for fixed* y *in* E *the equation*

$$x = y + \lambda F(x)$$

has a (unique) solution x *for each value of* λ *such that* $|\lambda M| < 1$. *Further this solution is given by*

$$x = y + \sum_{n=1}^{\infty} (-1)^{n+1} \lambda^n F^n(y).$$

With suitable definitions of F, E and the norm, the iterative solutions for integral equations given by Neumann and Volterra (see page 8) are special cases of this.

THE THEORY OF ADJOINT SPACES. Banach retained an interest in normed spaces and the theory of linear operators for the remainder of his life.[*] The publication in 1932 of his influential book "Théorie des Opérations Linéaires" did much to disseminate the theory and methods throughout Europe and America.

A fundamental advance came in 1929 when Banach introduced and studied the adjoint (or dual) space R^* of a normed linear space R. R^* is the set of all continuous linear functionals from R into the real numbers. With addition and scalar multiplication defined in the obvious "point-wise" way, R^* is a vector space. Further, R^* is a complete normed linear space if we define $\|f\|$ by

$$\|f\| = \text{g.l.b. } \{m: |f(x)| \leqslant m\|x\| \text{ for all } x \text{ in } R\}.$$

[This is a special case of a more general result of Banach's:
The set of continuous additive operators from one normed linear space E *into another* E_1 *is itself a normed linear space with*
$\|F\| = \text{g.l.b. } \{m: \|F(x)\| \leqslant m\|x\| \text{ for all } x \text{ in } E\}$. *Further, this space of operators is complete if* E_1 *is complete.*]

---

[*]In later years, much of Banach's time and energies were consumed in the writing of school and university text books (a task frequently undertaken to disentangle himself from accumulating debts). Many of these texts have enjoyed a wide popularity.

52.

The relationship of a space and its adjoint may be seen as a generalization of that between L^p and L^q. Indeed if $R = L^p[a,b]$ then Riesz's representation theorem (see page 44) gives $R^* = L^q[a,b]$. (Here, as before, $\frac{1}{p} + \frac{1}{q} = 1$.)

After establishing basic facts about adjoint spaces (in particular, that R^* is sufficiently "large" to separate the points of R; that is, if $f(x) = 0$ for all f in R^* then $x = 0$), Banach proceeds to generalize Riesz's notion of an adjoint operator (see page 45). If U is a continuous additive operator from R to S, the adjoint operator U^* from S^* to R^* is defined by

$$U^*(f)(x) = f(U(x)).$$

Adjoint operators are used by Banach to study operator equations in much the same way as Riesz had done for operators on L^p spaces.

The vital result that R^* separates the points of R follows from the celebrated "Hahn-Banach" extension theorem (stated by Banach as follows): *Let* p *be a real valued functional defined on a complete normed linear space and satisfying*

      (i) $p(x + y) \leqslant p(x) + p(y)$

      (ii) $p(\lambda x) = \lambda p(x)$ *for* $\lambda \geqslant 0$.

Then there is a continuous additive functional f *on* R *such that*

$$-p(-x) \leqslant f(x) \leqslant p(x).^*$$

Banach then introduces the notion of a "Banach-Mazur limit functional" and uses this to define a *weakly closed* subset. He then uses the above theorem to derive a number of "geometric" results. For example: *If* L *is a weakly closed subspace of* R^* *and* ϕ *is an element of* R^* *not in* L *then there exists a sequence of norm* 1 *elements* $\{x_n\}$ *in* R *such that*

---

* Banach's <u>proof</u> in fact establishes the modern statement of the theorem:

    Let E *be a complete normed linear space and* p *a mapping from* X *to the real numbers such that*

$$p(x + y) \leqslant p(x) + p(y)$$

and $p(\lambda x) = \lambda p(x)$ *for all* $\lambda > 0$.

    Then, if G *is a closed subspace of* E *and* f_0 *an element of* G^* *satisfying* $f_0(x) \leqslant p(x)$ *for all* x *in* G, *there exists* f *in* E^* *such that* $f(x) = f_0(x)$ *for all* x *in* G *and*

$$f(x) \leqslant p(x) \text{ for all } x \text{ in } E.$$

Hahn had proved this theorem two years earlier in the case when $p(x) = \|x\|$.

Helen Barach

$$f(x_n) = 0 \quad \textit{for all} \quad f \quad \textit{in} \quad L \quad \textit{and all} \quad n$$

and

$$\phi(x_n) \to \text{distance from } \phi \text{ to } L$$
$$= \text{g.l.b. } \{ \|\phi - f\| : f \text{ is in } L \}.$$

The study of "weak topologies" in both R and R^* is also initiated in Banach's 1929 paper.

Many of the basic facts for adjoint spaces, including a special case of the "Hahn-Banach" Theorem (see previous footnote), had previously been developed by Hahn in 1927, however, it was Banach's work which again proved most influential (although Banach did not reference Hahn, it is likely that he was nonetheless influenced by his work). Unlike Banach, Hahn only developed enough of the theory to answer the specific problem which interested him:

> If the system of equations $f_\lambda(x) = c$, $\lambda \in \Lambda$ has a solution x, where $\{f_\lambda\}$ is a given family of functionals and $\{c_\lambda\}$ is a given family of real numbers, what conditions must the family of functionals $\{v_\lambda\}$ satisfy to ensure that the perturbed system $f_\lambda(x) + v_\lambda(x) = c_\lambda$, $\lambda \in \Lambda$ also has a solution?

Banach also obtained many other results, for example: in 1931 he demonstrated that the set of continuous functions differentiable at at least one point is a first category subset of the complete metric space of all continuous functions on the interval $[a,b]$. Thus Banach gave a non-constructive proof for the existence of continuous nowhere differentiable functions - they form a "large" set (see p.35 and p.31)

Banach did more to create a "Polish school" of mathematics than any other individual. His 'peculiar' working habits, possibly a product of his turgid youth, led naturally to his sharing with fellow mathematicians and students not only his discoveries but also their invention. He daily spent long hours in a café alternately sipping coffee and vodka while he worked, along with others, on a mathematical problem.[*] Noise did not seem to distract Banach, indeed when an orchestra was present he preferred a table near to them. Consequently much of his later work was joint work with colleagues: Stanisław Mazur, Władysław Orlicz and Hugo Steinhaus to mention a few. The mathematical group centred on Banach continued to expand functional analysis until their work was interrupted by the war. Many of Banach's contemporaries died during the war[**] - others were dispersed throughout Europe

---

[*] For an interesting personal account of mathematical life in Lwów during the early twenties the reader should refer to Chapter 2 of Stanisław Ulam's book, "Adventures of a Mathematician".

[**]One, V.L. Šmul'yan, continued to write mathematical papers while in the trenches. Several were published and several more manuscripts were found on his person when he was shot in 1944.

and America – Banach himself died very shortly after the war, during which he had at one stage been used as a subject for Nazi medical experiments (although his death appears to have been a natural one).

Banach and his fellow workers bequeathed many "hard" problems to future generations of analysts. Work on Banach spaces and operators has progressed steadily since that time and within the last two decades many of the "classical" problems have been answered, often by 'unexpected' counter-examples. This has meant a resurgence of interest in Banach space theory over the last few years.

Other axiom systems have been introduced into analysis. Some, like those of von Neumann's, discussed below, are specializations of those given by Banach. For example, to capture the richer structure of spaces of operators (or real valued functions) in which a further algebraic operation, composition (or point-wise multiplication), is possible, I.M. Gelfand introduced the study of "Banach Algebras" in 1941.

Other systems have generalised the notion of normed linear spaces, for example; the theory of *distributions*, essentially started by Sobolev in 1936 and largely completed by Laurent Schwartz (1915–) in 1945, fits naturally into the structure of "locally convex spaces" founded by Von Neumann, Moore, Köthe and Toeplitz in the mid thirties and extensively developed by G. Mackey and others from the mid forties onward.

None-the-less the axiomative method had reached a degree of maturity by the early 1930's and it is with this that we will halt our discussion.

The axiomatization of Hilbert spaces

Examples of Hilbert space were among the first spaces to be considered, ironically however, their structure was among the last to be axiomatized.

Both the sequence space considered by Schmidt (see pages 38-40) and the spaces $L^2[a,b]$ studied by Riesz are Banach spaces, however they are distinguished by the presence of an "inner-product" $[z,w]$ from which the norm is derived according to the formula $\|z\| = \sqrt{[z,\bar{z}]}$.* The presence of an inner-product permits a richer theory than is possible for a general Banach space; for example, an adequate definition of "orthogonality" is possible.

By the early 1920's the work of Hilbert, Wiener and Hermann Weyl (1885 – 1955) had shown that Hilbert's "eigen-theory" (or, as it was more usually called, spectral theory) for symmetric "operators" (see pages 15 to 17) provided a possible model for the newly emerging *quantum mechanics*. Symmetric linear "operators" correspond to the "observables" of a physical system. Each

---

* In the case of $L^2[a,b]$ the inner-product is defined by

$$[f,g] = \int_a^b f(x)g(x)d\mu(x).$$

eigenvalue of such an operator represents a permissible value of the observable, while the corresponding eigen-vector (function) determines the "state" (interpreted by Schrödinger as the likelihood of observing the value at different spacial locations).

In 1925 Werner Heisenberg published his "matrix mechanics" in which the "operators", represented by infinite matrices, act on Schmidt's sequence space. Then, in 1926, Erwin Schrödinger presented a theory of quantum mechanics based on differential equations. Here the "operators" were essentially "differential operators" - for example, energy corresponds to $i\hbar\frac{\partial}{\partial t}$. Schrodinger demonstrated the equivalence of his theory with that of Heisenberg (a not unexpected result, since differential equations (operators) may be converted in integral equations (operators) and Riesz' theorem (R) on page 41 allows us to identify the space of functions $L^2[a,b]$, on which such integral operators may act, with Schmidt's sequence space - see comment on page 42.) What was lacking was an abstract frame-work into which these varying approaches would fit naturally and from which a general theory of quantum mechanics could be developed.

In 1929 John von Neumann (1903-1957) presented an axiomatic approach to inner-product spaces. He then went on to develop the theory of linear operators on such spaces. In particular, he considered Hermitian (self-adjoint or symmetric) operators for which he developed rich spectral and representation theories. His subsequent book "Mathematische Grundlagen der Quantenmechanik", published in 1932, is a classic, not only on quantum

John von Neumann

mechanics, but also on Hilbert space theory.

Von Neumann's axioms are the following.

A) H is a vector space, elements denoted by f,g etc., over the complex scalars denoted by a,b etc.

B) There exists on H an <u>inner-product</u> (f,g) satisfying

    1) $(af,g) = a(f,g)$

    2) $(f_1+f_2,g) = (f_1,g) + (f_2,g)$

    3) $(f,g) = \overline{(g,f)}$

    4) $(f,f) \geq 0$ and $(f,f) = 0$ if and only if $f = 0$.

Before stating the remaining axioms von Neumann notes that

$$|f| = \sqrt{(f,f)} \quad \text{defines a norm on } H \text{ and so } |f-g| \text{ is}$$

a metric for H.

C) In the metric defined above, H is separable; that is, H contains a countable dense subset.[*]

D) For every $n=1,2,\ldots$ H contains a set of n linearly independent vectors. [It would appear that the sole purpose of this axiom is to ensure H is "infinite-dimensional".]

E) H is complete with respect to the above metric. That is, if $\{f_n\}$ is such that $|f_n - f_m| \to 0$ as $n,m \to \infty$, then there exists an f in H such that $|f_n - f| \to 0$.[**]

For clarity we will henceforth write $\|f\|$ in place of von Neumann's $|f|$.

Von Neumann derives a number of simple consequences: The Cauchy-Schwarz inequality, $|(f,g)| \leq \|f\|\|g\|$ with equality holding if and only if $f = g$; every complete orthonormal set (see pages 16 and 41) is countable – this is a consequence of (C) and the Gram-Schmidt orthogonalization process, and the following are equivalent for an orthonormal set $\{\phi_n\}$

i) $\{\phi_n\}$ is complete [that is, $(f,g) = \sum\limits_{p=1}^{\infty} (f,\phi_p)(g,\phi_p)$]

ii) $f = \sum\limits_{p=1}^{\infty} (f,\phi_p)\phi_p$ for all f in H

iii) The closed span of $\{\Phi_n\}$ equals H.

The analogue of Schmidt's theory of closed subspaces (see pages 39 and 40) is now developed. Von Neumann shows that if N is a closed subspace of H, then the set of all elements orthogonal to every element of N is itself a closed subspace which von Neumann denotes by $H-N$ (today we would more likely use $N^{\perp}$ and refer to it as the *orthogonal complement* of N). Further H is seen to be a direct sum of N and $H-N$; that is, every closed subspace of H is "complemented". Using this von Neumann is able to define the natural projection operator from H to N. This initiates a general discussion of projection operators on H.

---

[*] The modern definition of Hilbert spaces does not include (C). It is the presence of (C) which distinguishes between "Hilbert space" and the more general notion of a Hilbert space. As was later shown, any two separable Hilbert spaces are naturally isomorphic and so in a sense there is only one such space. $L^2[a,b]$ and Schmidt's sequence space are simply different realizations of the one space – Riesz' theorem on page 41.

[**] Today, it is the presence of (E) which distinguishes Hilbert spaces from among the wider class of inner-product spaces.

After defining a bounded linear operator R in the same way as Banach had done, von Neumann defines the adjoint operator R* by

$$(Rf,g) \quad = \quad (f,R*g).$$

(c.f. Riesz definition given on page 46.)

An operator is said to be <u>Hermitian</u> if R* = R and it is on these operators that von Neumann concentrates. We list a few of his results. $U = e^{iR}$ is a unitary operator; that is U*U = UU* = I where I is the identity operator on H and such unitary operators are precisely the isometries of H. I - λR is invertible for all values of λ (real and complex) exterior to the closed interval [m,M] where m = g.l.b.{(Rf,f): $\|f\|$ = 1} and M = l.u.b.{(Rf,f): $\|f\|$ = 1}.

R is bounded if, whenever $f_n \to f$ and $Rf_n \to g$, we have $g = Rf$[*].

Von Neumann also developed a number of deeper structural results for Hermitian operators.

In later papers he studied the weak topology in Hilbert space, for which he gave a neighbourhood definition (see page 36). This topology has proved particularly valuable for the study of certain classes of operators.

Hilbert spaces and Hilbert space operators have been intensively studied since the time of von Neumann and today an extensive theory exists and continues to grow. The quantum mechanical motivation for Hilbert space theory has been completely overshadowed by its importance in a great variety of applications including the study of ordinary and partial differential equations, probability theory and approximation theory.

Bourbaki

Starting in the late 1930's a "collective" of predominantly French mathematicians began to write a systematic and encyclopedic account of modern mathematics. More than sixty chapters of this treatise, entitled *Eléments de Mathématique*, have already been published, the first in 1939. The work is still incomplete, and by its very nature must presumably remain so.

The authors have preferred to remain anonymous, writing, as a team, under the pseudonym "Nicolas Bourbaki". Reasons for the choice of name are obscure[**] Since several of the authors were at one time or another associated with the University of Nancy, one suggestion is that the name derives from that of a somewhat colourful French general, Charles-Denis Sauter Bourbaki (1816 - 1897), a statue of whom is said to reside in Nancy.

---

[*] We should here recognise a special case of the closed graph theorem, **proved by Banach in his 1932 book.**

[**] For some novel ideas on this see the article by S.K. Berberian in "The Mathematical Intelligencer" Vol. 2. No. 2. 1980.

Regardless of the mystery surround-
ing Bourbaki, his works have proved
extremely influential. They are written
for working mathematicians by a col-
lective of working mathematicians which
has included Jean Dieudonné (1906 –   ),
André Weil (1906 –   ), Henri Cartan
(1904 –   ), Jacques Dixmier
(1924 –   ), Alexandre Grothendieck
(1928 –   ), Laurent Schwartz (1915 –
) and Réne Thom (1923 –   ) to
mention only a few.

Bourbaki chose a formal axiomatic
approach. Starting with set theory he
went on to consider Algebra, General
Topology, Topological Vector Spaces,
etc. His precise language and
inventive systematization of material
has done much to consolidate the role
of abstract methods both in mathematics,
in general, and in analysis, in
particular.

Bourbaki's own terminology has not always proved popular, however,
he has influenced the development of a more or less standard terminology.
For example, the modern meaning of the term "compact" derives from
Bourbaki and replaces Fréchet's "extremal" or the Russian School's
"bicompact".

Some mathematicians have objected to Bourbaki's mercilessly abstract
view of mathematics. They believe that the real problems and applications
of mathematics have been lost sight of. This is very likely due to the
incompleteness of the enterprise. Bourbaki has not yet completed
the central portion of his writings on analysis.

Only the fundamental Structure of Analysis has been dealt with, the
links with particular problems and applications are still to come (?).
Bourbaki has, however, been the first to emphasise that axiomatization
must be "purposeful". Axiom systems should be founded on extensive
bodies of "concrete" mathematics. Such systems should reveal structures
which simplify and organise existing knowledge and should serve to sharpen
the working mathematicians "intuition", thereby aiding the creative process.
As we have seen, this is how axiom systems evolved in the past and this is
how it should continue to be in the future.
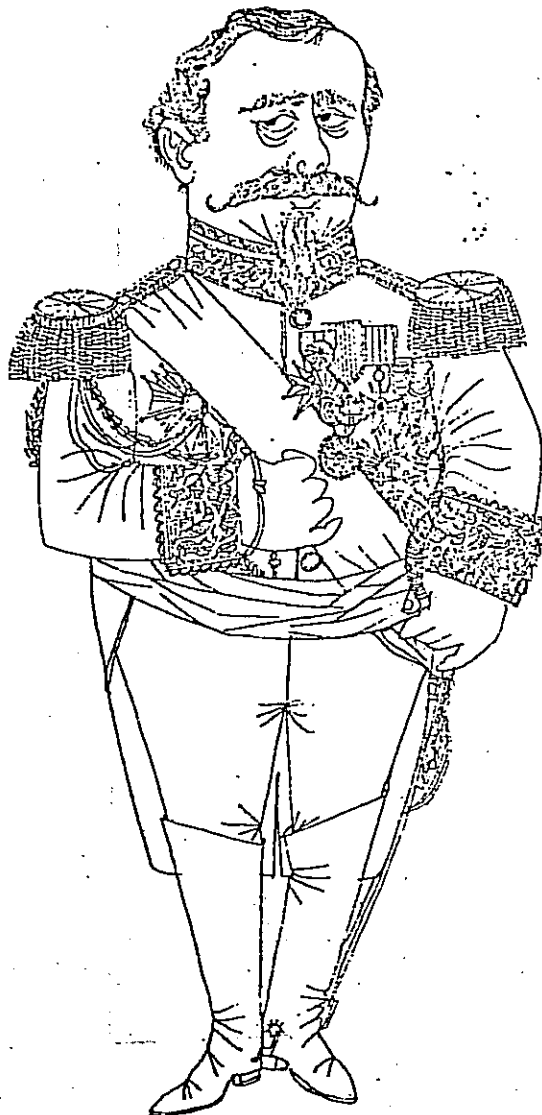
BIBLIOGRAPHY

Suggestions for further reading

BERNKOPF, Michael, "The development of Function Spaces with particular
    reference to their origins in Integral Equation Theory",
    Archive for History of Exact Sciences, Vol. 3, 1966/67,
    pp. 1-96.

HALMOS, Paul R., "Nicolas Bourbaki", Scientific American, May, 1957.
    [Reprinted in Readings from Scientific American "Mathematics in
    the Modern World"]

KLINE, Morris, "Mathematical Thought from Ancient to Modern Times",
    Oxford Univ. Press, 1972.

MANHEIM, "The Genesis of Point set Topology", Pergamon, 1964.

REID, Constance, "Hilbert", Springer-Verlag, 1970.

STEINHAUS, Hugo, "Stefan Banach", Scripta Mathematica, Vol. XXVI, 1961,
    pp. 93-100.

Other sources consulted during the preparation of this essay

AVERBUKH, V.I. and SMOLYANOV, O.G., "The various definitions of the
    derivatives in linear topological spaces, Russian M. Survey, 23:4
    (1968), 67-113.

BANACH, Stefan, "Théorie des Operations Linéarires", Chelsea.

BANACH, Stefan, "Oeuvres, Volume I", PWN, Warszawa, 1967.

BERNKOPF, Michael, "A History of Infinite Matrices", Archive for History
    of Exact Sciences", Vol. 4, 1967/68, pp.308-358.

BOURBAKI, "Historical Notes".

CATELL PRESS, "American Men and Women of Science", Bowker, 1976.

DAVIS, Harold T., "The Theory of Linear Operators", Principia Press, 1936.

DIEUDONNÉ, Jean, "Recent developments in Mathematics", Amer. Math. Monthly,
    Vol. 71, 1964, pp.239-248.

DIEUDONNÉ, Jean (Editor), "Abrégé d'Histoire des Mathématiques", Hermann,
    Paris, 1978.

DUGAS, René, "Histoire Abrégée des Sciences Mathématiques", Librairie Vuibert,
    Paris, 1963.

DUGUÉ, Daniel, "Notices Nécrologuques – Maurice Fréchet", Int. Stat. Rev.,
    Vol. 42, No. 1, 1974, pp.113-114.

FANG, J., "Bourbaki", Paideia, 1970.

FANG, J., "Hilbert", Paideia, 1970.

FANG, J. and TAKAYAMA, K.P., "Sociology of Mathematics and Mathematicians",
    Paideia, 1975.

FREDHOLM, Ivar, "Oeuvres Complètes", Litos Reprotryck, Malmö, 1955.

GILLISPIE, C.C. (Editor), "Dictionary of Scientific Biography" (14 volumes),
    Charles Scribners, N.Y., 1970.

GRACE, Sister Mary, "Poland's Contribution to Mathematics", The Mathematics
    Teacher, April, 1967, pp.383-386.

HAUSDORFF, Felix, "Set Theory", Chelsea, 1962.

HAUSDORFF, Felix, "Nachgelassene Schriften", Teubner-Stuttgart, 1969.

HILBERT, David, "Grundzüge Einer Allgemeinen Theorie der Linearen Integralgleichungen", Chelsea, 1953.

KURATOWSKI, K., "Mathematics in Poland Today".

LOVITT, Vernon, "Linear Integral Equations", Dover, 1950.

MANDELBROJT, Szolem M., "Notice Nécrologiques sur Maurice Fréchet", C.R. Acad. Sc., Paris, t 277 (19 Novembre 1973) Vie Academique.

MONNA, A.F., "Functional Analysis in Historical Perspective", John Wiley, 1973.

POGGENDORFF, "Biographisch-Literarisches Handwörterbuch Der Exalsten Naturwisseschaften, (4 volumes), Akademie Verlag, 1967.

REID, Constance, "Courant", Springer-Verlag, 1976.

RIESZ, Friedrich, "Gesammelte Arbeiten", Gauthier-Villars, Paris, 1960.

"Royal Society's Index of Scientific Papers" (discontinued from 1914).

TURKEVICH, "Soviet Men of Science", Van Nostrand, 1963.

WHITTAKER, Sir Edmund, "Biography of Vito Volterra" in the Obituary Notices of Fellows of the Royal Society", in 1941.
[Reprinted in the 1959 Dover Edition of Vito Volterra's "Theory of Functionals and of Integral and Integro-differential equations".]

"Who's Who in Science in Europe" (4 volumes), Hodgson, 1978.

ADDITIONAL SOURCES

DIEUDONNÉ, J. "History of Functional Analysis", North-Holland, Mathematical Studies 49 (1981).

FLETT, "Differential Analysis", Cambridge Univ. Press, 1980

HOCHSTADT, HARRY, "Eduard Helly, Father of the Hahn Banach Theorem", The Mathematical Intelligencer, Vol. 2, No. 3., 1980.

KURATOWSKI, "Some Remarks on the Origins of the Theory of Functions of a Real Variable and of the Descriptive set Theory" Rocky Mountains Journal of Math. 10 No. 1, 1980 (pp.25-33).

KURATOWSKI, "A Half Century of Polish Mathematics: Rememberances and Reflections" Pergamon, 1979.

The Scottish Book (LASL Monograph).

APPENDIX

Throughout this essay I have made reference to the published papers of many mathematicians. Many students, particularly those isolated from a University library may never have had cause to examine a "research" paper in mathematics. For this reason one is reproduced here.

It is impossible to choose one paper which is typical of the thousands published each year. I have selected a paper by

Shizo Kakutani

which appeared in

The Tôhoku Math. Journal
Vol. 45 (1938) pp.188-193.

The reasons for my choice include the following:

It is shorter than many;

It is on one problem, which grew directly out of material considered in the essay;

The necessary background is less than that required for many papers and so it might be accessible to an undergraduate student with a minimum of preparation.

Weak Convergence in Uniformly Convex Spaces,

by

Shizuo Kakutani, Ôsaka.

S. Banach and S. Saks proved(¹) that the space $L^p(p>1)$ has the following property:

(P) *every weakly convergent sequence of points contains a subsequence whose arithmetic mean converges strongly to the same limit.*

As is shown by J. Schreier(²), this theorem does not hold in the space C. What then is the reason why the same theorem holds in $L^p(p>1)$, while it fails in C? As an answer to this question, we shall prove in the present paper the

Theorem. *If a Banach space is uniformly convex, then it has the property (P).*

We must first explain the meaning of uniform convexity. This notion was introduced by J. A. Clarkson(³) in his paper on the theory of integration of a function whose range lies in a Banach space.

Definition. *A Banach space is called to be uniformly convex, if it has the following property:*

To each $\epsilon > 0$ there corresponds a $\delta(\epsilon) > 0$ *such that the conditions*

(U) $\|x\| = \|y\| = 1, \ \|x - y\| \geq \epsilon$

imply

$$\left\|\frac{x+y}{2}\right\| \leq 1 - \delta(\epsilon).$$

Since, as is shown by J. A. Clarkson(³), the space $L^p(p>1)$ is uniformly convex, the result of S. Banach and S. Saks is a direct consequence of our theorem; and the counter-example of J. Schreier(²) has its reason in the fact that the space C is not uniformly convex. That the converse of the theorem is not true (that is, that there is a Banach space which has the property (P) without being uniformly convex), may be easily seen from the fact that the same Banach space may become uniformly convex or not depending on its norm (without changing the topology)(⁴).

Before entering into the proof of the theorem, let us replace the condition (U) by the equivalent one:

To each $\epsilon > 0$ there corresponds a $\delta'(\epsilon) > 0$ *such that the condition*

(U') $\|x - y\| \geq \epsilon \cdot \max(\|x\|, \|y\|)$

implies

$$\left\|\frac{x+y}{2}\right\| \leq (1 - \delta'(\epsilon)) \cdot \max(\|x\|, \|y\|).$$

Putting $\|x\| = \|y\| = 1$ and $\delta'(\epsilon) = \delta(\epsilon)$ in (U') we have (U), so that (U') implies (U). To prove the converse we proceed as follows:

(i) Since x and y are symmetric in (U'), we may assume $\|x\| \geq \|y\|$.

(ii) Since $\|ax\| = |a| \cdot \|x\|$ for any real number a and since the relations in (U') are homogeneous in x and y, we may assume without the loss of generality that $\|y\| \leq \|x\| = 1$(⁵).

By (i) and (ii) we have only to prove that (U) implies (U''):

To each $\epsilon > 0$ there corresponds a $\delta''(\epsilon) > 0$ *such that the conditions*

(U'') $\|x\| = 1, \ \|y\| \leq 1, \ \text{and} \ |x - y| \geq \epsilon$

imply

(¹) S. Banach and S. Saks: Sur la convergence forte dans les champs L^p, Studia Math. 2 (1930).

(²) J. Schreier: Ein Gegenbeispiel zur Theorie der schwachen Konvergenz, Studia Math. 2 (1930).

It is to be mentioned that the following theorem holds true:

Theorem (Z. Zalcwasser). If $\{f_n(x)\}$ is a sequence of continuous functions which converges weakly to a continuous function $f(x)$ (that is, if $\{f_n(x)\}$ is a sequence of uniformly bounded continuous functions which converges pointwise to a continuous function $f(x)$) in $a \leq x \leq b$, then there is a schema of constants $\lambda_i^n(i=1,2,...,n; n=1,2,...)$ such that

$$\lambda_i^n \geq 0, \ \lambda_1^n + \lambda_2^n + ... + \lambda_n^n = 1$$

and for which

$$\varphi_n(x) \equiv \lambda_1^n f_1(x) + \lambda_2^n f_2(x) + ... + \lambda_n^n f_n(x)$$

converges strongly (uniformly) to $f(x)$.

This is also a consequence of a general theorem of S. Mazur.

Z. Zalcwasser: Sur une propriété du champ des fonctions continues, Studia Math. 2 (1930).

S. Mazur: Uber konvexe Mengen in linearen normierten Räumen, Studia Math. 4 (1933).

(³) J. A. Clarkson: Uniformly convex spaces, Trans. Amer. Math. Soc., 40 (1930).

(⁴) Indeed, the (x,y)-plane is not uniformly convex if we define $\|(x,y)\| = \max(|x|, |y|)$, although it is uniformly convex in the Euclidean norm: $\|(x,y)\| = \sqrt{x^2+y^2}$.

(⁵) If $\|x\| > 0$, replace x and y by $x/\|x\|$ and $y/\|x\|$ respectively; if $\|x\| = 0$

$$\left\|\frac{x+y}{2}\right\| \leq 1-\delta''(\varepsilon).$$

Proof of $(U)\to(U'')$: Let $\eta(<\varepsilon)$ be a sufficiently small positive number, which we shall determine later, and let x and y be such that $\|x\|=1$, $1\geq\|y\|\geq 1-\eta$ and $\|x-y\|\geq\varepsilon$. Putting $z=y/\|y\|$ the conditions

$$\|x-z\|\geq\|x-y\|-\|y-z\|\geq\varepsilon-\eta>0 \text{ and } \|x\|=\|z\|=1$$

imply (by (U))

$$\left\|\frac{x+z}{2}\right\| \leq 1-\delta(\varepsilon-\eta),$$

and consequently

$$\left\|\frac{x+y}{2}\right\| \leq \left\|\frac{x+z}{2}\right\| + \left\|\frac{y-z}{2}\right\| \leq 1-\delta(\varepsilon-\eta)+\frac{\eta}{2}.$$

Since the right hand side is <1 for a sufficiently small $\eta>0$, say $\eta=\eta_0$ (ε being fixed !), we have

$$\left\|\frac{x+y}{2}\right\| \leq 1-\delta(\varepsilon-\eta_0)+\frac{\eta_0}{2}$$

for any x and y with $\|x\|=1$, $1\geq\|y\|\geq 1-\eta_0$ and $\|x-y\|\geq\varepsilon$. Since in case $\|y\|\leq 1-\eta_0$ the inequality

$$\left\|\frac{x+y}{2}\right\| \leq 1-\frac{\eta_0}{2}$$

is evident, we have for any x and y with $\|x\|=1$, $\|y\|\leq 1$ and $\|x-y\|\geq\varepsilon$ the relation :

$$\left\|\frac{x+y}{2}\right\| \leq \max\left(1-\delta(\varepsilon-\eta_0)+\frac{\eta_0}{2}, 1-\frac{\eta_0}{2}\right)=1-\delta''(\varepsilon)<1,$$

$\delta''(\varepsilon)$ being a positive number, independent of x and y. Thus the proof that (U) implies (U'') (and also (U')) is completed.

Now we shall proceed to the

Proof of the theorem: Let $\{x_n\}(n=1,2,\ldots)$ be a sequence of points of a uniformly convex space, which converges weakly to x. We shall prove that there is a subsequence $\{x_{n_k}\}(k=1,2,\ldots)$ of $\{x_n\}$ such that

$$\lim_{k\to\infty}\left\|\frac{x_{n_1}+x_{n_2}+\ldots+x_{n_k}}{k}-x\right\|=0.$$

Without the loss of generality we may assume that $x=0$. Since the sequence $\{x_n\}$ is weakly convergent, there is a constant M such that

$\|x_n\|\leq M$ for $n=1,2,\ldots$. We shall show first that there is a subsequence $\{x_{m_n}\}$ $(n=1,2,\ldots; 1<m_1<m_2<\ldots<m_{2n-1}<m_{2n}<\ldots)$ of $\{x_n\}$ such that

$$\left\|\frac{x_{m_{2n-1}}+x_{m_{2n}}}{2}\right\| \leq \theta\cdot M$$

for $n=1,2,\ldots$, where

$$\theta=\max\left(\frac{3}{4}, 1-\delta\left(\frac{1}{2}\right)\right)<1$$

is a constant indepedent of the given sequence (that is, depending only on the property of the space in question).

Put $m_1=2$ and consider the following two cases :

(i) If $\|x_2\|\leq\frac{M}{2}$, put $m_2=3$; then we have

$$\left\|\frac{x_{m_1}+x_{m_2}}{2}\right\|=\left\|\frac{x_2+x_3}{2}\right\|\leq\frac{1}{2}\left(\frac{M}{2}+M\right)=\frac{3}{4}M\leq\theta\cdot M.$$

(ii) If $\|x_2\|>\frac{M}{2}$, take as m_2 the least integer $n(>2)$ such that $\|x_2-x_n\|>\frac{M}{2}$. The existence of such an n is a consequence of the fact, that if $\|x_2-x_n\|\leq\frac{M}{2}$ for all $n>2$, then $\|x_2\|=\|x_2-0\|=\lim\|x_2-x_n\|\leq\frac{M}{2}$, which is contradictory to our assumption that $\|x_2\|>\frac{M}{2}$. For these m_1 and m_2 we have

$$\|x_{m_1}-x_{m_2}\|>\frac{M}{2}\geq\frac{1}{2}\max(\|x_{m_1}\|, \|x_{m_2}\|),$$

and therefore, by the property (U'),

$$\left\|\frac{x_{m_1}+x_{m_2}}{2}\right\|\leq\left(1-\delta'\left(\frac{1}{2}\right)\right)\cdot\max(\|x_{m_1}\|, \|x_{m_2}\|)\leq\theta\cdot M.$$

After m_2 is determined, put $m_3=m_2+1$ and determine m_4 precisely in the same manner as m_2 is determined from m_1; and then put $m_5=m_4+1$, and so on.

It will be almost obvious, that, proceeding in this way, we shall have the desired subsequence $\{x_{m_n}\}(n=1,2,\ldots)$.

For the sake of simplicity we put

$$x_n^{(1)}=\frac{x_{m_{2n-1}}+x_{m_{2n}}}{2}, \quad n=1,2,\ldots.$$

(i) Clearly the sequence $\{x_n^{(1)}\}$ $(n=1,2,\ldots)$ has the following properties:

SHIZUO KAKUTANI:

(ii)₁ $\|x_n^{(1)}\| \leq \theta \cdot M$, $n=1,2,\ldots$,

(iii)₁ $\{x_n^{(1)}\}$ $(n=1,2,\ldots)$ converges weakly to 0.

In general, after the sequence $\{x_n^{(p)}\}$, $(n=1,2,\ldots)$ is already defined so as to satisfy the conditions:

(i) $$x_n^{(p)} = \frac{x_{2n-1}^{(p-1)} + x_{2n}^{(p-1)}}{2}, \quad n=1,2,\ldots,$$

$$1 < m_1^{(p-1)} < m_2^{(p-1)} < \ldots < m_{2n-1}^{(p-1)} < m_{2n}^{(p-1)} < \ldots,$$

(ii) $\|x_n^{(p)}\| \leq \theta^p \cdot M$, $n=1,2,\ldots$,

(iii) $\{x_n^{(p)}\}$, $(n=1,2,\ldots)$ converges weakly to 0,

we define a sequence $\{x_n^{(p+1)}\}$, $(n=1,2,\ldots)$ so as to satisfy the conditions:

(i)$_{p+1}$ $$x_n^{(p+1)} = \frac{x_{2n-1}^{(p)} + x_{2n}^{(p)}}{2}, \quad n=1,2,\ldots,$$

$$1 < m_1^{(p)} < m_2^{(p)} < \ldots < m_{2n-1}^{(p)} < m_{2n}^{(p)} < \ldots,$$

(ii)$_{p+1}$ $\|x_n^{(p+1)}\| \leq \theta^{p+1} \cdot M$, $n=1,2,\ldots,$

(iii)$_{p+1}$ $\{x_n^{(p+1)}\}$ $(n=1,2,\ldots)$ converges weakly to 0.

This may be done precisely in the same manner as $\{x_n^{(1)}\}$ is defined from $\{x_n\}$. Clearly is satisfied the condition:

Repeating the same method, we shall have a sequence of sequences $\{\{x_n^{(p)}\}$, $(n=1,2,\ldots)\}$, $(p=1,2,\ldots)$, which satisfies the conditions (i)$_p$, (ii)$_p$ and (iii)$_p$ for $p=1,2,\ldots$

Now consider the sequence $\{x_1^{(p)}\}$, $(p=1,2,\ldots)$. As is easily seen from the construction of $\{x_n^{(p)}\}$, each $x_1^{(p)}$ is of the form:

$$x_1^{(p)} = \frac{x_1^{(p)} + x_2^{(p)} + \ldots + x_{2^p}^{(p)}}{2^p}, \quad p=1,2,\ldots,$$

with

$$1 < l_1^{(q)} < l_2^{(q)} < l_3^{(q)} < \ldots < l_{2p-1}^{(q)} < l_{2p}^{(q)} < \ldots < l_p^{(q)} < \ldots;$$

moreover, for any $q<p$ and $1 \leq i \leq 2^{p-q}$,

$$x_i^{(q)} = \frac{x_{(i-1)2^q+1}^{(p)} + x_{(i-1)2^q+2}^{(p)} + \ldots + x_{i2^q}^{(p)}}{2^q},$$

is an element of $\{x_n^{(0)}\}$ and hence of norm not greater than $\theta^q \cdot M$.

If we put $n_1=1$ and $n_{2^p+i-1}=l_i^{(p)}$ $(i=1,2,\ldots,2^p;\ p=1,2,\ldots,$

(that is, $n_1=1$, $n_2=l_1^{(1)}$, $n_3=l_2^{(1)}$, $n_4=l_1^{(2)}$, $\ldots$, $n_7=l_4^{(2)}$, $n_8=l_1^{(3)}$, $\ldots$, $n_{2^{p-1}}=l_{2^{p-1}}^{(p-1)}$, $n_{2^p}=l_1^{(p)}$, $\ldots$, $n_{2^p+i-1}=l_i^{(p)}$, $\ldots$), then n_k is uniquely

determined for $k=1,2,\ldots$, and the subsequence $\{x_{n_k}\}$ $(k=1,2,\ldots)$ of $\{x_n\}$ $(n=1,2,\ldots)$ thus obtained is the required one. In order to prove that

$$\lim_{k\to\infty} \frac{x_{n_1} + x_{n_2} + \ldots + x_{n_k}}{k} = 0,$$

let us consider for $r\cdot 2^q \leq k < (r-1)\cdot 2^q$ the inequality:

$$\|x_{n_1} + x_{n_2} + \ldots + x_{n_k}\| \leq \|x_{n_1} + x_{n_2} + \ldots + x_{n_{r\cdot 2^q}}\|$$

$$+ \sum_{j=r+1} x_{n_{(j-1)2^q+1}} + x_{n_{(j-1)2^q+1}} + \ldots + x_{n_{j\cdot 2^q-1}}$$

Since $x_{n_{(j-1)2^q+1}} + x_{n_{(j-1)2^q+2}} + \ldots + x_{n_{j\cdot 2^q-1}}$ is an element of $\{x_n\}$

$i=2,3,\ldots,r$; for some p and i; for some p and i, we have

and therefore

$$\varlimsup_{k\to\infty} \frac{x_{n_1} + x_{n_2} + \ldots + x_{n_k}}{k} \leq \varlimsup_{k\to\infty} \frac{(r-1)\cdot 2^q}{k} \cdot \theta^q \cdot M = \theta^q \cdot M,$$

Since q is arbitrary and since $\theta < 1$ is fixed, the proof is hereby completed.

Mathematical Institute,
Osaka Imperial University,
Osaka, Japan.