SPECTRAL THEORY

NELSON DUNFORD

Spectral theory has for its origin the classical theory of finite matrices and, in a sense, includes all theories which may be specialized to some particular phase of that classical theory. Such a meaning for the term spectral theory is perhaps too broad, for its acceptance would force one, as is becoming increasingly evident, to admit that spectral theory embraces not a small part of such fields of mathematics as the ergodic theory, the theory of probability, absolutely convergent trigonometric series, Fourier transforms, the theory of continuous functions on a bicompact Hausdorff space, the theory of partially ordered rings and Abelian groups, and perhaps even the well known embedding theorem concerning completely regular topological spaces. In fact, even though by far the greatest part of spectral theory that has been at present developed has for its roots the properties of a very special type of matrix, namely the Hermitian matrix, it is certainly true that modern developments have made significant contacts with all the theories just mentioned.

Although I shall touch briefly upon certain features of these modern developments my chief concern here will not be an attempt to describe the wealth of ideas that have grown out of the properties of an Hermitian matrix, but rather to show how the properties of a general matrix can guide the way to the solution of a special class of problems of vital interest in analysis. The type of problem with which I shall be concerned is illustrated by the following question. How can it be determined whether or not a given sequence of polynomials $f_n(T)$ in a linear operator T on a Banach space converges to a specified type of limit operator? Before making this problem more precise I should like to show how the elementary properties of a finite matrix can, if put in the proper form, give an immediate answer to all such questions of convergence for a finite dimensional space.

Suppose $\mathfrak{X}$ is an m-dimensional linear vector space over the field of complex numbers (or over any algebraically closed field). Let T be a linear operator in $\mathfrak{X}$, that is, T is an $m \times m$ matrix whose elements are complex numbers. For every complex number λ and every

$n = 0, 1, 2, \cdots$ define the linear manifolds

$$\mathfrak{N}_\lambda^n = (\lambda I - T)^n \mathfrak{X}, \qquad \mathfrak{M}_\lambda^n = \underset{x}{\mathcal{E}} \left[x \in \mathfrak{X}, (\lambda I - T)^n x = 0 \right].$$

It is clear that for a fixed λ the manifold $\mathfrak{M}_\lambda^n$ is non-decreasing in n and the manifold $\mathfrak{N}_\lambda^n$ is nonincreasing. It is also clear that each λ defines uniquely a positive integer or zero, $\nu = \nu(\lambda)$, called the *index* of λ, such that $\mathfrak{M}_\lambda^\nu = \mathfrak{M}_\lambda^n$ for $n \geq \nu$ while for $n < \nu$ we have $\mathfrak{M}_\lambda^n$ a proper subset of $\mathfrak{M}_\lambda^{n+1}$. It is also readily shown that $\mathfrak{N}_\lambda^\nu = \mathfrak{N}_\lambda^n$ for $n \geq \nu$ while for $n < \nu$ we have $\mathfrak{N}_\lambda^{n+1}$ a proper subset of $\mathfrak{N}_\lambda^n$. If $n \geq \nu(\lambda)$ we always have

$$(1) \qquad\qquad\qquad \mathfrak{X} = \mathfrak{N}_\lambda^n \oplus \mathfrak{M}_\lambda^n,$$

and conversely if (1) holds then $n \geq \nu(\lambda)$. Let $\lambda_1, \cdots, \lambda_k$ be the distinct characteristic numbers of T and let $\nu_1, \cdots, \nu_k$ be the corresponding indices, then

$$(2) \qquad\qquad\qquad \mathfrak{X} = \mathfrak{M}_{\lambda_1}^{\nu_1} \oplus \cdots \oplus \mathfrak{M}_{\lambda_k}^{\nu_k}.$$

Equation (2) shows that there are uniquely defined projection operators (that is, idempotent matrices) $E_{\lambda_1}, \cdots, E_{\lambda_k}$ such that

$$(3) \qquad I = E_{\lambda_1} + \cdots + E_{\lambda_k}, \quad E_{\lambda_i}\mathfrak{X} = \mathfrak{M}_{\lambda_i}^{\nu_i}, \quad E_{\lambda_i} \cdot E_{\lambda_j} = 0, \qquad i \neq j.$$

Since, in the space $\mathfrak{M}_{\lambda_i}^{\nu_i}$, any polynomial $f(T)$ in T is equal to its residue modulo $(T - \lambda_i I)^{\nu_i}$ we have from (3)

$$(A) \qquad\qquad f(T) = \sum_{i=1}^{k} \sum_{j=0}^{\nu_i - 1} \frac{(T - \lambda_i I)^j}{j!} f^{(j)}(\lambda_i) E_{\lambda_i}.$$

While our chief purpose will be to investigate some of the consequences of this formula (A), or rather the formula corresponding to (A) in the infinite case, we should like to point out very briefly the relation of this formula to the special case which has led to so many fruitful investigations. If the indices ν_i of T are all 1, in particular if T is an Hermitian matrix, then (A) becomes

$$(B) \qquad\qquad\qquad f(T) = \sum_{i=1}^{k} f(\lambda_i) E_{\lambda_i}.$$

The important algebraic difference between cases (A) and (B) is that, as is easily seen, in case (B) the ring of polynomials in T is isomorphic with a direct sum of fields, that is, isomorphic with a linear space of scalar functions, while in case (A) the ring of all polynomials in T is isomorphic with a direct sum of reduced polynomial rings. Another

way of seeing the chief difference between the two cases is to observe that in case (B) we have no nilpotent elements, that is, no radical in the ring, while in case (A) we may have $f^m(T) = 0$, $f(T) \neq 0$. It is formula (B) which symbolizes the classical theory of reduction of quadratic forms. This theory was extended to infinite forms by Hilbert and his school and by E. H. Moore and his students. It is perhaps just to say that it is F. Riesz to whom we owe the advantages of the concept of a linear operator. To F. Riesz we owe the formula

(B') $$f(T) = \int_{-\infty}^{+\infty} f(\lambda) dE_\lambda$$

which replaces (B) in the case of a bounded self-adjoint operator in Hilbert space. It was J. von Neuman and M. H. Stone who developed the theory symbolized by (B') in the case where T is an unbounded self-adjoint operator in Hilbert space. In recent years the formula (B') has been rapidly disappearing from the theory of quadratic forms. In fact, the theory of bounded quadratic forms has become, in the hands of Stone,[1] Gelfand,[2] Kakutani,[3] and others, almost identical with the theory of continuous real functions on a bicompact Hausdorff space. Each of these mathematicians has, I believe, taken roughly the point of view just mentioned, that is, that a commutative family of Hermitian matrices may be thought of as a family of scalar functions and each one has been led to a characterization of the family of *all* real continuous functions on a bicompact Hausdorff space. While their starting points are somewhat different there are, in each case, postulates sufficient to rule out a radical in the system, that is, these mathematicians have all been working with systems generalizing formula (B).

With this very incomplete summary of what has been done with formula (B) let us return to (A) and ask the question:—when does a given sequence $f_n(T)$ of polynomials in a general matrix T converge? We shall describe two ways of answering this question, the first being fairly obvious, the second less so but in many ways more practical than the first. The first method is illustrated by the following theorem.

THEOREM 1. *The sequence $f_n(T)$ converges if and only if $f_n^{(j)}(\lambda_i)$ con-*

[1] M. H. Stone, *A general theory of spectra* I, Proc. Nat. Acad. Sci. U. S. A. vol. 26 (1940) pp. 280–283.

[2] I. Gelfand, *Normierte Ringe*, Recueil Math. (Mat. Sbornik) N.S. vol. 51 (1941) pp. 3–24.

[3] S. Kakutani, *Concrete representation of abstract (M) spaces*, Ann. of Math. vol. 42 (1942) pp. 994–1024.

verges for each $i = 1, \cdots, k$ and $j = 0, \cdots, \nu_i - 1$. Also $f_n(T) \to f(T)$ if and only if $\lim_n f_n^{(j)}(\lambda_i) = f^{(j)}(\lambda_i)$, $i = 1, \cdots, k, j = 0, \cdots, \nu_i - 1$.

The sufficiency of the conditions is obvious from (A) while their necessity is an immediate consequence of the definition of ν_i.

We shall reserve any general statement of the second method until later and here merely illustrate it by a typical example.

EXAMPLE. *The sequence $(\sum_{\nu=0}^{n-1} T^\nu)/n$ converges if and only if $T^n/n \to 0$.*

The truth of this statement is readily seen by applying Theorem 1. For if $f_n(\lambda) = \lambda^n/n$ then $f_n(\lambda) \to 0$ if and only if $|\lambda| \leq 1$. Whereas all derivatives $f_n^{(j)}(\lambda) \to 0$ if $|\lambda| < 1$, not even the first derivative $f^{(1)}(\lambda) \to 0$ if $|\lambda| = 1$. Thus by Theorem 1 we can say that $T^n/n \to 0$ if and only if all roots λ_i of T have $|\lambda_i| \leq 1$, and also every root λ_i of T with $|\lambda_i| = 1$ has the corresponding index $\nu_i = 1$. A similar application of Theorem 1 shows that $(\sum_{\nu=0}^{n-1} T^\nu)/n$ converges if and only if the spectral points of T satisfy precisely the same conditions.

It is not pure coincidence that these two statements about convergence have precisely the same spectral interpretation. It is a consequence of a general principle which will be discussed more fully later.

The formal difference that we wish to emphasize between the theorem and the example is obvious; namely, in the example there is no mention of the spectrum of T. It is for this reason that a theorem of the type of the example, that is, a theorem asserting the equivalence of two statements about convergence, is in many ways more practical than one like Theorem 1. For example if a deck of cards is repeatedly shuffled in such a way that in any given shuffle there is a definite probability p_{ij} that the card in the ith place goes into the card in the jth place, and this probability is fixed and independent of preceding shuffles, then one knows immediately that the probability of finding in the jth place, after n shuffles, the card which was originally in the ith place is the element $p_{ij}^{(n)}$ of the nth iterate of the matrix $P = (p_{ij})$. Hence it is seen without any spectral considerations that $P^n/n \to 0$. Similarly there are many examples of linear operators T in analysis where for one reason or another (perhaps because $|T^n| \leq M$) it is known that $T^n/n \to 0$, and hence a theorem of the type of the example could be much more easily applied than one like Theorem 1.

Our main objective will now be the description of a method for arriving at a class of theorems similar to the example, that is, theorems asserting the equivalence of two statements about convergence,

but in the case where T is a bounded linear operator on a complete normed linear vector space not necessarily of finite dimension. In this general situation there are several different but closely related problems, for there are usually three and sometimes four different meanings for convergence. We always have the notions of convergence in the uniform, strong, and weak topologies, and, in case the space in question consists of measurable functions we have the notion of convergence almost everywhere. I wish to outline a unified theory of convergence, in each of the four types, of a sequence $f_n(T)$ of functions of an operator T to a *specified* type of limit operator.

In brief our method for arriving at this goal consists of two steps. The first and easiest step is the establishment of an operational calculus to replace formula (A). The second and more difficult one is the matter of applying the operational calculus. Here the chief tools are theorems corresponding to (a) the minimal equation theorem for matrices, (b) the theorem of Sylvester concerning the determinant of a polynomial in a matrix and (c) the various decompositions of the whole space into a direct sum of its subspaces determined by the spectrum of a matrix.

Returning our attention for a moment to a finite matrix T and in particular to equation (A) we observe that one formula, and perhaps the only one, which may be used to replace (A) and which does not in any way exhibit the finite character of the space $\mathfrak{X}$ or the operator T is

$$(A') \qquad f(T) = \frac{1}{2\pi i} \int_C \frac{f(\lambda)d\lambda}{\lambda I - T},$$

where C is a set of small circles $C_1, \cdots, C_k$ with C_i surrounding λ_i. That formula (A') is valid for any polynomial $f(T)$ may be verified by direct computation or it may be derived from (A) by using the Hermite interpolation formula. Conversely, it is not difficult to pass from (A') to (A) without going through the preliminary algebraic arguments used to derive (A). The reason for taking C as a set of small circles each one of which surrounds one and only one spectral point of T rather than taking C as one large circle containing the whole spectrum is the following. It is essential for our purposes that we do not have to restrict the operational calculus expressed by (A') to the class of functions $f(\lambda)$ which are regular and single-valued on a *connected domain* containing the spectrum of T. We shall consider the class $\mathcal{J}(T)$ of functions $f(\lambda)$ which are regular and single-valued in the closure of a domain $\mathcal{D}$ (which may vary with the function $f \in \mathcal{J}(T)$) satisfying the conditions

(i) $\mathcal{D}$ is a finite sum of connected open sets $\mathcal{D}_i$ with $\overline{\mathcal{D}}_i\overline{\mathcal{D}}_j=0$, $i\neq j$.

(ii) The boundary C of $\mathcal{D}$ consists of a finite number of disjoint closed rectifiable Jordan curves contained in the resolvent set $\rho(T)$ of T.

(iii) The spectrum $\sigma(T)$ of T is contained in $\mathcal{D}$.

A domain $\mathcal{D}$ satisfying (i) and (ii) but not necessarily (iii) will be called a *T-admissible domain*, in symbols $\mathcal{D}=\mathcal{D}(T)$. These are the requirements whether T be a finite matrix, a bounded linear operator on a complex Banach space, or an element of a normed ring. Thus for example in the case of the finite matrix we may take $f(\lambda)\equiv1$ in and on C_i and $f(\lambda)\equiv0$ in and on C_j, $j\neq i$, and if we do the result is

$$E_{\lambda_i} = \frac{1}{2\pi i}\int_{C_i}\frac{d\lambda}{\lambda I - T}.$$

This result, as well as the general one expressed by (A$'$), was discovered in 1928 by Fantappié,[4] who gave (A$'$) as a basis for an operational calculus on finite matrices. Fantappié proved the standard rules for such a calculus, namely,

(α) $(\alpha f+\beta g)(T)=\alpha f(T)+\beta g(T)$.

(β) $(f\cdot g)(T)=f(T)\cdot g(T)$.

(γ) If $f(\lambda)=\sum_{n=1}^m\alpha_n\lambda^n$ then $f(T)=\sum_{n=1}^m\alpha_n T^n$.

It was pointed out by E. Cartan[5] that formula (A$'$) could be used as a basis for an operational calculus on infinite matrices. In what follows we shall think of T as a bounded linear operator on a complex Banach space and shall write $\rho(T)$ for the resolvent set of T, that is, the set of complex numbers for which $R_\lambda(T)=(\lambda I-T)^{-1}$ exists as an everywhere defined and hence bounded linear operator. The symbol $\sigma(T)$ will be used for the spectrum of T, that is, the complex numbers not in $\rho(T)$. It should be recalled that the index $\nu=\nu(\lambda)$ (as defined for finite matrices) of a point $\lambda\in\sigma(T)$ may be 0 or ∞. Recalling that $R_\lambda(T)$ is analytic on $\rho(T)$ we define for every $f\in\mathcal{F}(T)$ the operator $f(T)$ by the formula (A$'$), where the integral is taken around the boundary C of any T-admissible domain $\mathcal{D}$ which contains $\sigma(T)$ and

[4] L. Fantappié, *La calcul des matrices*, C. R. Acad. Sci. Paris vol. 186 (1928) pp. 619–621. In the case where $f(\lambda)$ is a power series and C is a circle, formula (A$'$) has been given by H. Poincaré, *Sur les groupes continus*, Transactions of the Cambridge Philosophical Society vol. 18 (1900) pp. 220–255.

[5] The suggestion was made in a letter to G. Giorgi. For this reference as well as for others pertaining to an operational calculus see C. C. MacDuffee, *The theory of matrices*, Berlin, 1933.

upon which $f(\lambda)$ is regular and single-valued. We shall add to the basic rules (α), (β), (γ) of Fantappié two more, namely,

(δ) $f(\sigma(T)) = \sigma(f(T)), f \in \mathcal{F}(T)$.

(ϵ) If $f \in \mathcal{F}(T)$, $g \in \mathcal{F}(f(T))$, and $F(\xi) = g(f(\xi))$ then $F \in \mathcal{F}(T)$ and $F(T) = g(f(T))$.

A frequently used corollary of (β) is the following. If $f \in \mathcal{F}(T)$ is identically 1 in a neighborhood of a spectral set σ of T (a spectral set σ of T is a subset of $\sigma(T)$ which is open and closed in $\sigma(T)$) and identically zero in a neighborhood of the complement $\sigma' = \sigma(T) - \sigma$ of σ then $f(T)$ is a projection. In other words the operator $E_\sigma[T] = (1/2\pi i)\int_C R_\lambda(T)d\lambda$, where C is the boundary of a T-admissible domain $\mathcal{D}$ with $\sigma = \mathcal{D} \cdot \sigma(T)$, is a projection. This result was proved recently by Lorch[6] who used methods independent of an operational calculus. Lorch also showed that the Boolean algebra of all such projections is isomorphic with the Boolean algebra of all spectral sets of T. The fact that the mapping $\sigma \to E_\sigma[T]$ is a homomorphism is readily seen from (α), (β). That it is actually an isomorphism follows from the following fundamental principle.

THEOREM 2. *When T is considered as an operator in the space $\mathcal{X}_\sigma = E_\sigma[T]\mathcal{X}$ it has σ for its spectrum. Furthermore for $\lambda \in \sigma$ any one of the following statements is true if and only if it is true when T is regarded as an operator in $\mathcal{X}_\sigma$: λ is in the point spectrum of T, λ is in the residual spectrum of T, λ is in the continuous spectrum of T, λ is a pole of order ν for $R_\xi(T)$, λ is an essential singularity for $R_\xi(T)$.*

This is all we shall say about the formal operational calculus. The finite dimensional space has been replaced by a complex Banach space $\mathcal{X}$, the matrix by a linear operator T in $\mathcal{X}$, the class of polynomials by the class $\mathcal{F}(T)$, and the formula (A) by the formula (A'). Now it is perfectly evident that the arguments used to derive Theorem 1 from formula (A) can not be applied immediately to formula (A'). Certain elementary facts can, however, be stated immediately. On the one hand, as a necessary condition for the convergence of a sequence $f_n(T)$ where $f_n \in \mathcal{F}(T)$, we have from (α) and (δ) that $|f_n(\lambda) - f_m(\lambda)| \leq |f_n(T) - f_m(T)|$, $\lambda \in \sigma(T)$, so that if $f_n(T)$ converges in the uniform topology of operators to an operator U we must have $f_n(\lambda)$ converging uniformly for $\lambda \in \sigma(T)$ to a continuous function $f(\lambda)$ defined for $\lambda \in \sigma(T)$. This function $f(\lambda)$ which will be mentioned later is called the *spectral function* of U and depends only on U and not on the par-

[6] E. R. Lorch, *The spectrum of linear transformations*, Trans. Amer. Math. Soc. vol. 52 (1942) pp. 238–248.

ticular sequence $f_n(T)$ with $f_n \in \mathcal{J}(T)$ used to approach U. On the other hand it is clear from (A′) that the uniform convergence of $f_n(\lambda)$ on a *fixed* (that is, independent of n) T-admissible domain containing $\sigma(T)$ is a sufficient condition for the convergence of $f_n(T)$. Unfortunately, however, most of the interesting problems in analysis of this type are concerned with a sequence $f_n(\lambda)$ whose points of divergence have a spectral point of T as a limiting point. Take for example the ergodic theorem and suppose that $\lambda = 1$ is in the spectrum $\sigma(T)$ (this is the only non-trivial case). The sequence of polynomials $(\sum_{\nu=0}^{n-1}\lambda^\nu)/n$ clearly diverges at points in every neighborhood of $\lambda = 1$. Thus it seems necessary to seek for more refined methods. To this end we shall virtually eliminate formula (A′) by deriving from it certain fundamental results of an algebraic nature which can be used as a basis for the convergence theory. We shall describe here three of the principal theorems of this type.

The minimal equation theorem. For a finite matrix T this theorem is well known and asserts that a polynomial $f(T)$ in T vanishes if and only if $f(\lambda)$ contains $\prod_{i=1}^{k}(\lambda - \lambda_i)^{\nu_i}$ as a factor. The statement for a general linear operator is almost identical and reads as follows.

THEOREM 3 (THE MINIMAL EQUATION THEOREM). *If $f \in \mathcal{J}(T)$ then $f(T) = 0$ if and only if*
  (i) *For every pole λ of $R_\xi(T)$ of order ν, $f^{(j)}(\lambda) = 0$, $j = 0, 1, \cdots, \nu - 1$.*
  (ii) *$f(\lambda) \equiv 0$ on a neighborhood containing all the spectrum $\sigma(T)$ except perhaps poles of $R_\xi(T)$.*

A corollary of this is

COROLLARY 1. *If $f, g \in \mathcal{J}(T)$ then $f(T) = g(T)$ if and only if*
  (i) *For every pole λ of $R_\xi(T)$ of order ν, $f^{(j)}(\lambda) = g^{(j)}(\lambda)$, $j = 0, \cdots, \nu - 1$.*
  (ii) *$f(\lambda) = g(\lambda)$ for every λ in a neighborhood containing all the spectrum $\sigma(T)$ except perhaps poles of $R_\xi(T)$.*

COROLLARY 2. *Let $\lambda_1, \cdots, \lambda_k$ be poles of $R_\xi(T)$ of orders $\nu_1, \cdots, \nu_k$ respectively. Let σ' be the complement of the spectral set $\sigma = (\lambda_1, \cdots, \lambda_k)$. Then for every $f \in \mathcal{J}(T)$ we have*

$$f(T) = f(T)E_{\sigma'}[T] + \sum_{i=1}^{k} \sum_{j=0}^{\nu_i - 1} \frac{(T - \lambda_i I)^j}{j!} f^{(j)}(\lambda_i) E_{\lambda_i}.$$

Corollary 2 follows from Corollary 1 as follows. If $\psi(\lambda)$ is the characteristic function of a small neighborhood of λ_i then by Corollary 1 and (β) we have $\psi(T)f(T) = \psi(T)P(T) = E_{\lambda_i}[T]P(T)$ where $P(\lambda) = \sum_{j=0}^{\nu_i - 1}(\lambda - \lambda_i)^j f^{(j)}(\lambda_i)/j!$

As is suggested by these theorems we have also the fact that if λ is a pole of order ν for $R_\xi(T)$ then λ has index ν and also the manifold $\mathfrak{N}_\lambda^\nu$ is closed and $\mathfrak{N}_\lambda^\nu = \mathfrak{N}_\lambda^n$ for $n \geq \nu$, while $\mathfrak{N}_\lambda^{n+1}$ is a proper subset of $\mathfrak{N}_\lambda^n$ for $n < \nu$. Just as in the case of finite matrices (see equation (1)) there is the decomposition[7]

(1') $$\mathfrak{X} = \mathfrak{N}_\lambda^n \oplus \mathfrak{M}_\lambda^n, \qquad\qquad n \geq \nu,$$

and *conversely if for some n we have $\mathfrak{N}_\lambda^n$ a closed complement of $\mathfrak{M}_\lambda^n$ then λ is a pole of order $\nu \leq n$.*

When λ is a pole of order ν we always have $\mathfrak{M}_\lambda^\nu = \mathfrak{X}_\lambda \equiv E_\lambda[T]\mathfrak{X}$. When λ is not a pole but has index ν we frequently have the decomposition

(1'') $$\mathfrak{X} = \overline{\mathfrak{N}_\lambda^\nu} \oplus \mathfrak{M}_\lambda^\nu.$$

The Sylvester theorem. This theorem generalizes the formula $\sigma(f(T)) = f(\sigma(T))$. For finite matrices this formula merely means that if $\lambda_1, \cdots, \lambda_k$ are the distinct roots of an $n \times n$ matrix T then $f(\lambda_1), \cdots, f(\lambda_k)$ are the roots of $f(T)$. A theorem due to Sylvester asserts more, namely, if $\lambda_1, \cdots, \lambda_k$ are the distinct roots of T and if m_i is the multiplicity of λ_i then

(i) $$\mathrm{Det}\,(f(T) - \lambda I) = \prod_{i=1}^{k} (f(\lambda_i) - \lambda)^{m_i}.$$

This means that if $\lambda_1, \cdots, \lambda_n$ are the roots of T, each repeated according to its multiplicity, then $f(\lambda_1), \cdots, f(\lambda_n)$ are the roots of $f(T)$, where the number of repetitions of a given number in this array is its multiplicity as a root of $f(T)$. Recalling that the multiplicity $m(\lambda, T)$ of a root λ of T of index ν is the number of linearly independent solutions of $(\lambda I - T)^\nu x = 0$, that is, $m(\lambda, T)$ is the dimension of $\mathfrak{M}_\lambda^\nu = \mathfrak{X}_\lambda[T] = E_\lambda[T]\mathfrak{X}$, we see that the Sylvester Theorem, that is, the formula

(ii) $$m(\mu, f(T)) = \sum_{\lambda \in \sigma} m(\lambda, T), \qquad \sigma = f^{-1}(\mu),$$

is an immediate consequence of

(iii) $$\mathfrak{X}_\mu[f(T)] = \sum_{\lambda \in \sigma} \mathfrak{X}_\lambda[T], \qquad \sigma = f^{-1}(\mu),$$

(where the sum on the right is a vector sum). Whereas formula (iii)

---

[7] At this point reference should be made to F. Riesz, *Über lineare Funktional-gleichungen*, Acta Math. vol. 41 (1918) pp. 71–98, who proved similar theorems for a compact operator.

states the equality of two manifolds, formula (ii) merely states the equality of their dimensions. A more desirable form of this theorem than (iii) is

(iv) $$E_\mu[f(T)] = \sum_{\lambda \in \sigma} E_\lambda[T] = E_\sigma[T],$$

which states the equality of two projections rather than, as (iii) does, the equality of the manifolds upon which they project. It is, roughly speaking, the form (iv) that the Sylvester theorem takes in the infinite case. To be more precise we have

THEOREM 4 (THE SYLVESTER THEOREM). *Let* $f \in \mathcal{F}(T)$ *and let* τ *be a spectral set of* $f(T)$. *Then* $\sigma = \sigma(T) \cdot f^{-1}(\tau)$ *is a spectral set of* T *and* $E_\tau[f(T)] = E_\sigma[T]$.

This theorem has a further generalization which will be needed later. Let $\mathcal{R}(T)$ be the ring of all operators $f(T)$ where $f \in \mathcal{F}(T)$ and let $\overline{\mathcal{R}(T)}$ be the ring of all operators which are limits in the uniform topology of operators of elements in $\mathcal{R}(T)$. Then, as we have seen earlier, with each $U \in \overline{\mathcal{R}(T)}$ is associated a unique continuous function $f(\lambda)$ defined on $\sigma(T)$. This spectral function also has the property that $f(\sigma(T)) \subset \sigma(U)$. Of course if $U = g(T) \in \mathcal{R}(T)$ then $f(\lambda) = g(\lambda)$ on $\sigma(T)$. In terms of these notions we have

THEOREM 4' (THE GENERAL SYLVESTER THEOREM). *Let* f *be the spectral function of an operator* $U \in \overline{\mathcal{R}(T)}$ *and let* τ *be a spectral set of* U. *Then* $\sigma = f^{-1}(\tau)$ *is a spectral set of* T *and* $E_\sigma[T] = E_\tau[U]$.

A typical decomposition theorem. There are many types of decomposition theorems that seem to be needed for a general theory of convergence. We shall content ourselves here with one illustration. This theorem to follow generalizes the theorem mentioned for finite matrices in connection with equation (1), and its comprehension needs the following concepts. Let $f \in \mathcal{F}(T)$ and define the manifolds

$$\mathfrak{N}[f] = f(T)\mathfrak{X}, \qquad \mathfrak{M}[f] = \underset{x}{\mathcal{E}}\,[x \in \mathfrak{X}, f(T)x = 0].$$

THEOREM 5. *Let* $f \in \mathcal{F}(T)$ *and suppose that* $f(\lambda)$ *is not identically zero on any of the domains in which it is regular. Let* $\lambda_1, \cdots, \lambda_k$ *be the roots of* $f(\lambda)$ *and* $m_1, \cdots, m_k$ *their multiplicities. Then the following statements are equivalent.*

   (1) $\mathfrak{X} = \mathfrak{M}[f] \oplus \mathfrak{N}[f]$, $\mathfrak{N}[f]$ *is closed.*

   (2) *For* $i = 1, \cdots, k$, λ_i *is either in the resolvent set* $\rho(T)$ *or else a pole of order* $\nu_i \leqq m_i$ *of* $R_\xi(T)$.

(3) *The finite set* $\sigma = (\lambda_1, \cdots, \lambda_k)\sigma(T)$ *is a spectral set of* T *and* $\mathfrak{X}_\sigma = \mathfrak{M}[f]$, $\mathfrak{X}_{\sigma'} = \mathfrak{N}[f]$.

When the roots λ_i of $f(\lambda)$ are not poles of $R_\xi(T)$, perhaps not even isolated in the spectrum $\sigma(T)$, but have indices $\nu_i \leq m_i$ we often have a decomposition of the form $\mathfrak{X} = \mathfrak{M}[f] \oplus \overline{\mathfrak{N}[f]}$.

Some convergence theorems. While it is clear that a certain type of convergence theorem may be obtained from Corollary 2 of the minimal equation theorem just as Theorem 1 was obtained from formula (A), we wish to restrict our attention here to a special case which presents unexpected features not found in the general situation. We shall be interested in when a given sequence $f_n(T), f \in F(T)$, that is, a given sequence $f_n(T) \in R(T)$, converges to a special type of projection E. We shall first discuss the case of convergence in the uniform topology of operators. As simple examples show we do not in general have $\mathcal{R}(T) = \overline{\mathcal{R}(T)}$, that is, in general $U = \lim_n f_n(T)$ is not in $\mathcal{R}(T)$. However in case U is a projection E it is necessarily in $\mathcal{R}(T)$. This is the first of the two fundamental necessity theorems, and it is the necessity of the conditions in the case of convergence in the uniform topology of operators that presents the most difficulty.

THEOREM 6 (THE FIRST NECESSITY THEOREM). *Every projection* $E \in \overline{\mathcal{R}(T)}$ *is in* $\mathcal{R}(T)$ *and* $E = E_\sigma[T]$ *where* σ *is a spectral set of* T *and consists of all* $\lambda \in \sigma(T)$ *where the spectral function* $f(\lambda)$ *of* E *has the value* 1. *Thus* $E = 0$ *if and only if there are no such* λ.

This theorem is an immediate consequence of the general Sylvester theorem (Theorem 4'). To see this we observe the readily verified fact that the resolvent $R_\xi(E)$ of a projection E is given by the equation $R_\xi(E) = E/(\xi-1) + (I-E)/\xi$, $\xi \neq 0, 1$. Thus $E_1[E] = E$, and if we take τ in Theorem 4' to be the point $\lambda = 1$ then $E = E_1[E] = E_\sigma[T]$ where σ consists of all $\lambda \in \sigma(T)$ with $f(\lambda) = 1$.

THEOREM 7 (THE SECOND NECESSITY THEOREM). *Let* P *be a polynomial not identically zero and let* $f_n \in \mathcal{J}(T)$. *If* $f_n(T)$ *converges to a projection* E *with* $E\mathfrak{X} \subset \mathfrak{M}[P]$ *then either there are no* $\lambda \in \sigma(T)$ *where* $f_n(\lambda) \to 1$ *in which case* $E = 0$ *or else*
   (1) *The set* σ *of points* $\lambda \in \sigma(T)$ *where* $f_n(\lambda) \to 1$ *consists of a finite number of poles* $\lambda_1, \cdots, \lambda_k$ *of* $R_\xi(T)$.
   (2) *If* $\nu_1, \cdots, \nu_k$ *are the orders of the poles* $\lambda_1, \cdots, \lambda_k$ *then* $f_n(\lambda_i) \to 1$, $f_n^{(j)}(\lambda_i) \to 0$, $j = 1, \cdots, \nu_i - 1$, $i = 1, \cdots, k$.

(3) *For $i=1, \cdots, k$, λ_i is a root of $P(\lambda)$ and its multiplicity is at least ν_i.*

(4) $\prod_{i=1}^{k}(\lambda_i I - T)^{\nu_i} E = 0$.

This theorem is a consequence of the minimal equation theorem together with the preceding theorem. The proof goes as follows. Let σ be the set of $\lambda \in \sigma(T)$ where $f_n(\lambda) \to 1$. Then, by Theorem 7, σ is a spectral set of T and $E = E_\sigma[T]$. Since $E\mathfrak{X} \subset \mathfrak{M}[P]$ we have $P(T)E_\sigma[T] = 0$ and so (1), (3) and (4) follow immediately from the minimal equation theorem whereas (2) follows from its Corollary 2.

We have now outlined all of the difficult steps encountered in the proof of the following basic theorem. We shall omit the rest of the details.[8]

THEOREM 8. *Let $P(\lambda) = \prod_{i=1}^{k}(\lambda_i - \lambda)^{\nu_i}$ be a polynomial whose distinct roots are $\lambda_1, \cdots, \lambda_k$. Let $f_n \in \mathcal{F}(T)$ satisfy*

(1) $f_n(\lambda_i) \to 1$, $f_n^{(j)}(\lambda_i) \to 0$, $i = 1, \cdots, k$, $j = 1, \cdots, \nu_i - 1$.

(2) $P(T)f_n(T) \to 0$.

Then the following assertions are equivalent.

(3) $f_n(T) \to E$, $E^2 = E$, $E\mathfrak{X} = \mathfrak{M}[P]$.

(4) *For each $i = 1, \cdots, k$, λ_i is either in $\rho(T)$ or else a pole of $R_\xi(T)$.*

(5) *For $i = 1, \cdots, k$, λ_i is either in $\rho(T)$ or else a pole of $R_\xi(T)$ of order at most ν_i.*

(6) $\mathfrak{X} = \mathfrak{N}[P] \oplus \mathfrak{M}[P]$, $\mathfrak{N}[P]$ *is closed.*

(7) $\mathfrak{N}_{\lambda_i}^{\nu_i+1}$, $i = 1, \cdots, k$, *is closed.*

It is the two assertions "$P(T)f_n(T) \to 0$" and "$f_n(T)$ converges" (for if $f_n(T)$ converges it must, in view of (1), converge to a projection on $\mathfrak{M}(P)$) which in this situation (that is, in view of (1) and (4) for example) have precisely the same spectral significance. This theorem serves to illustrate the general principle mentioned in connection with the example following Theorem 1, for if $P(T) = I - T$ and $f_n(T) = (\sum_{\nu=0}^{n-1} T^\nu)/n$ then the condition $P(T)f_n(T) \to 0$ is precisely $T^n/n \to 0$. Thus for a transformation T which satisfies either

(4') *The point $\lambda = 1$ is either in the resolvent set $\rho(T)$ or a pole of $R_\xi(T)$,*

or

(7') $(I - T)^2 \mathfrak{X}$ is closed,

we can say that the uniform ergodic theorem holds if and only if $T^n/n \to 0$, and conversely if the uniform ergodic theorem holds then

---

[8] The details of many of the results discussed here appear in the Trans. Amer. Math. Soc. vol. 54 (1943) pp. 185–217.

so do (4′) and (7′). In fact if the uniform theorem holds we have

(6′) $(I-T)\mathfrak{X}$ is closed,

and, as we shall see, this condition is also necessary and sufficient for the uniform theorem providing the strong (that is, mean) theorem holds and $T^n/n \rightarrow 0$.

We shall now pass to the case of strong convergence and state

THEOREM 9. *Let* $P(\lambda) = \prod_{i=1}^{k} (\lambda_i - \lambda)^{\nu_i}$ *be a polynomial whose distinct roots are* $\lambda_1, \cdots, \lambda_k$. *Let* $f_n \in \mathcal{F}(T)$ *satisfy*

(1) $f_n(\lambda_i) \rightarrow 1, f_n^{(j)}(\lambda_i) \rightarrow 0, i=1, \cdots, k, j=1, \cdots, \nu_i-1.$

(2) $P(T)f_n(T) \rightarrow 0$ *strongly.*

Then the following assertions are equivalent.

(3) $f_n(T) \rightarrow E$ *strongly,* $E^2 = E, E\mathfrak{X} = \mathfrak{M}[P].$

(4) $f_n(T)x$ *is weakly compact,* $x \in \mathfrak{X}.$

(5) $\mathfrak{X} = \overline{\mathfrak{N}[P]} \oplus \mathfrak{M}[P], |f_n(T)| \leqq M.$

Thus in a reflexive space if (1) and $|f_n(T)| \leqq M$ are satisfied we can say that (2) and (3) are equivalent. This is another example where two statements about convergence have the same spectral significance. Taking $P(T) = I - T, f_n(T) = (\sum_{\nu=0}^{n-1} T^\nu)/n$, as before, we can say that in a reflexive space if $|(\sum_{\nu=0}^{n-1} T^\nu)/n| \leqq M$ then $(\sum_{\nu=0}^{n-1} T^\nu)/n$ converges strongly if and only if $T^n/n \rightarrow 0$ strongly. The usual condition $|T^n| \leqq M$ assumed in ergodic theorems is not a necessary one. Hille and Szegö[9] have recently constructed operators satisfying the strong (that is, mean) as well as the almost everywhere ergodic theorem, and for which $|T^n|$ is of the order of $n^{1/4}$. The essential algebraic difference between the strong and the uniform theorem is seen by a comparison of (6) of Theorem 8 and (5) of Theorem 9.

We shall omit any discussion of weak convergence and pass to the case of almost everywhere convergence. Here we have as a basis a space S of points t. The space $\mathfrak{X}$ consists of scalar functions which are defined on S and which are measurable with respect to a completely additive non-negative measure function $|e|$ defined for e in a Borel field including S. It is assumed that $x = 0$ if and only if $x(t) = 0$ almost everywhere, and that addition and scalar multiplication of the functions correspond to the similar operations in $\mathfrak{X}$. It is assumed that if $x_n \rightarrow x$ in $\mathfrak{X}$ and $x_n(t) \rightarrow y(t)$ almost everywhere then $x(t) = y(t)$ almost everywhere. It is also assumed in case $|S| = \infty$ that S is the denumerable union of sets of finite measure. For a linear operator U in $\mathfrak{X}$ we shall write $U(x, t)$ for the value of the function Ux at the point t.

[9] E. Hille and G. Szegö, *Remarks on ergodic theorems*, to appear in the Trans. Amer. Math. Soc.

A fundamental theorem of Banach[10] enables us to state the following theorem as an immediate corollary of Theorem 9.

THEOREM 10. *Let P, f_n, E be as in the preceding theorem and satisfy* (1), (2), (4) *of that theorem. Then*

(1) *For every $x \in \mathfrak{X}$ we have $\lim_n f_n(T)(x, t) = E(x, t)$ almost everywhere if and only if*

(2) *For every $x \in \mathfrak{X}$ $\lim \sup_n |f_n(T)(x, t)| < \infty$ almost everywhere and*

(3) *For every x in a fundamental set in $\mathfrak{X}$ the $\lim_n f_n(T)P(T)(x, t)$ exists almost everywhere.*

Again let us interpret this theorem by taking $P(T) = I - T$, $f_n(T) = (\sum_{\nu=0}^{n-1} T^\nu)/n$ where T is now a linear operator in the Lebesgue space $L(S)$, $|S| < \infty$, defined by a 1-1 point map ϕ of S into itself according to the formula $Tf = g$, $g(t) = f(\phi t)$. Clearly (3) is trivially satisfied with no restrictions on ϕ by merely taking the fundamental set to be the set of bounded measurable functions. Thus (3) is redundant and (2) becomes the necessary and sufficient condition for almost everywhere convergence. But it has been proved[11] that T satisfies the strong (that is, mean) ergodic theorem if and only if

$$(4) \qquad \frac{1}{n} \sum_{\nu=0}^{n-1} |\phi^{-\nu}e| \leqq M |e|,$$

and that this condition (4) implies (2). *Thus* (2) *is also redundant and* (4) *alone is the single condition which is equivalent to the strong theorem and which implies the almost everywhere theorem.* In the case where ϕ is measure preserving (2) has been established by G. D. Birkhoff,[12] N. Wiener,[13] K. Yosida and S. Kakutani,[14] and H. R. Pitt[15] by proving an inequality of the type

---

[10] S. Banach, *Sur la convergence presque partout des fonctionelles linéaires*, Bull. Sci. Math. (2) vol. 50 (1926) pp. 36–43. This theorem of Banach has also been used by Yosida in generalizing the ergodic theorem of G. D. Birkhoff, see K. Yosida, *Ergodic theorems of the Birkhoff-Khintchine's type*, Jap. J. Math. vol. 17 (1940) pp. 31–36.

[11] This was discovered by D. S. Miller and the author who will publish the proof elsewhere.

[12] G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. U. S. A. vol. 17 (1931) pp. 656–660.

[13] N. Wiener, *The ergodic theorem*, Duke Math. J. vol. 5 (1929) pp. 1–18.

[14] K. Yosida and S. Kakutani, *Birkhoff's ergodic theorem and the maximal ergodic theorem*, Proc. Imp. Acad. Tokyo vol. 15 (1939) pp. 165–168.

[15] H. R. Pitt, *Some generalizations of the ergodic theorem*, Proceedings of the Cambridge Philosophical Society vol. 38 (1942) pp. 325–343.

$$(5) \qquad\qquad\qquad | e_\alpha | \leqq \frac{1}{\alpha} \int_S f(t) dt,$$

where e_α is the set where $\sum_{\nu=0}^{n-1} f(\phi^\nu t)/n > \alpha$ and f is positive. While the works of these authors make it clear that an inequality similar to (5) is valid providing $|\phi^{-\nu} e| \leq M |e|$ it is not at all clear whether or not (4) will give anything like (5). Nevertheless (4) implies (2) which is all that is needed for the almost everywhere ergodic theorem. I believe that the ergodic theorems outlined in these pages contain most of the known[16] ergodic theorems in the cyclic case.

YALE UNIVERSITY

---

[16] K. Yosida and S. Kakutani, *Operator theoretical treatment of Markoff's process and mean ergodic theorem*, Ann. of Math. (2) vol. 42 (1941) pp. 188–228.