WHAT DOES THE SPECTRAL THEOREM SAY?

P. R. HALMOS, University of Michigan

Most students of mathematics learn quite early and most mathematicians remember till quite late that every Hermitian matrix (and, in particular, every real symmetric matrix) may be put into diagonal form. A more precise statement of the result is that every Hermitian matrix is unitarily equivalent to a diagonal one. The spectral theorem is widely and correctly regarded as the generalization of this assertion to operators on Hilbert space. It is unfortunate therefore that even the bare statement of the spectral theorem is widely regarded as somewhat mysterious and deep, and probably inaccessible to the nonspecialist. The purpose of this paper is to try to dispel some of the mystery.

Probably the main reason the general operator theorem frightens most people is that it does not obviously include the special matrix theorem. To see the relation between the two, the description of the finite-dimensional situation has to be distorted almost beyond recognition. The result is not intuitive in any language; neither Stieltjes integrals with unorthodox multiplicative properties, nor bounded operator representations of function algebras, are in the daily toolkit of every working mathematician. In contrast, the formulation of the spectral theorem given below uses only the relatively elementary concepts of measure theory. This formulation has been part of the oral tradition of Hilbert space for quite some time (for an explicit treatment see [6]), but it has not been called the spectral theorem; it usually occurs in the much deeper "multiplicity theory." Since the statement uses simple concepts only, this aspect of the present formulation is an advantage, not a drawback; its effect is to make the spirit of one of the harder parts of the subject accessible to the student of the easier parts.

Another reason the spectral theorem is thought to be hard is that its proof is hard. An assessment of difficulty is, of course, a subjective matter, but, in any case, there is no magic new technique in the pages that follow. It is the statement of the spectral theorem that is the main concern of the exposition, not the proof. The proof is essentially the same as it always was; most of the standard methods used to establish the spectral theorem can be adapted to the present formulation.

Let ϕ be a complex-valued bounded measurable function on a measure space X with measure μ. (All measure-theoretic statements, equations, and relations, e.g., "ϕ is bounded," are to be interpreted in the "almost everywhere" sense.) An operator A is defined on the Hilbert space $\mathcal{L}^2(\mu)$ by

$$(Af)(x) = \phi(x)f(x), \qquad x \in X;$$

the operator A is called the *multiplication* induced by ϕ. The study of the relation between A and ϕ is an instructive exercise. It turns out, for instance, that the adjoint A^* of A is the multiplication induced by the complex conjugate $\bar{\phi}$ of ϕ. If ψ also is a bounded measurable function on X, with induced multiplication B, then the multiplication induced by the product function $\phi\psi$ is the product operator AB. It follows that a multiplication is always normal; it is Hermitian if and only if the function that induces it is real. (For the elementary concepts of operator theory, such as Hermitian operators, normal operators, projections, and spectra, see [**3**]. For present purposes a concept is called elementary if it is discussed in [**3**] before the spectral theorem, i.e., before p. 56.)

As a special case let X be a finite set (with n points, say), and let μ be the "counting measure" in X (so that $\mu(\{x\})=1$ for each x in X). In this case $\mathfrak{L}^2(\mu)$ is n-dimensional complex Euclidean space; it is customary and convenient to indicate the values of a function in $\mathfrak{L}^2(\mu)$ by indices instead of parenthetical arguments. With this notation the action on f of the multiplication A induced by ϕ can be described by

$$A\langle f_1, \cdots, f_n\rangle = \langle \phi_1 f_1, \cdots, \phi_n f_n\rangle.$$

To say this with matrices, note that the characteristic functions of the singletons in X form an orthonormal basis in $\mathfrak{L}^2(\mu)$; the assertion is that the matrix of A with respect to that basis is diag $\langle \phi_1, \cdots, \phi_n\rangle$.

The general notation is now established and the special role of the finite-dimensional situation within it is clear; everything is ready for the principal statement.

SPECTRAL THEOREM. *Every Hermitian operator is unitarily equivalent to a multiplication.*

In complete detail the theorem says that if A is a Hermitian operator on a Hilbert space $\mathfrak{H}$, then there exists a (real-valued) bounded measurable function ϕ on some measure space X with measure μ, and there exists an isometry U from $\mathfrak{L}^2(\mu)$ onto $\mathfrak{H}$, such that

$$(U^{-1}AUf)(x) = \phi(x)f(x), \qquad x \in X,$$

for each f in $\mathfrak{L}^2(\mu)$. What follows is an outline of a proof of the spectral theorem, a brief discussion of its relation to the version involving spectral measures, and an illustration of its application.

Three tools are needed for the proof of the spectral theorem.

(1) *The equality of norm and spectral radius.* If the spectrum of A is $\Lambda(A)$, then the *spectral radius* $r(A)$ is defined by

$$r(A) = \sup \{ |\lambda| : \lambda \in \Lambda(A)\}.$$

It is always true that $r(A) \leq \|A\|$ ([**3**, Theorem 2, p. 52]); the useful fact here is that if A is Hermitian, then $r(A) = \|A\|$ ([**3**, Theorem 2, p. 55]).

(2) *The Riesz representation theorem for compact sets in the line.* If L is a positive linear functional defined for all real-valued continuous functions on a com-

pact subset X of the real line, then there exists a unique finite measure μ on the Borel sets of X such that

$$L(f) = \int f d\mu$$

for all f in the domain of L. (To say that L is linear means of course that

$$L(\alpha f + \beta g) = \alpha L(f) + \beta L(g),$$

whenever f and g are in the domain of L and α and β are real scalars; to say that L is positive means that $L(f) \geqq 0$ whenever f is in the domain of L and $f \geqq 0$.) For a proof, see [**4**, Theorem D, p. 247].

(3) *The Weierstrass approximation theorem for compact sets in the line.* Each real-valued continuous function on a compact subset of the real line is the uniform limit of polynomials. For a pleasant elementary discussion and proof see [**1**, p. 102].

Consider now a Hermitian operator A on a Hilbert space $\mathcal{H}$. A vector ξ in $\mathcal{H}$ is a *cyclic vector* for A if the set of all vectors of the form $q(A)\xi$, where q runs over polynomials with *complex* coefficients, is dense in $\mathcal{H}$. Cyclic vectors may not exist, but an easy transfinite argument shows that $\mathcal{H}$ is always the direct sum of a family of subspaces, each of which reduces A, such that the restriction of A to each of them does have a cyclic vector. Once the spectral theorem is known for each such restriction, it follows easily for A itself; the measure spaces that serve for the direct summands of $\mathcal{H}$ have a natural direct sum, which serves for $\mathcal{H}$ itself. Conclusion: there is no loss of generality in assuming that A has a cyclic vector, say ξ.

For each real polynomial p write

$$L(p) = (p(A)\xi, \xi).$$

Clearly L is a linear functional; since

$$\begin{aligned} | L(p) | &\leqq \|p(A)\| \cdot \|\xi\|^2 = r(p(A)) \cdot \|\xi\|^2 \\ &= \sup \{ | \lambda | : \lambda \in \Lambda(p(A))\} \cdot \|\xi\|^2 \\ &= \sup \{ | p(\lambda) | : \lambda \in \Lambda(A)\} \cdot \|\xi\|^2, \end{aligned}$$

the functional L is bounded for polynomials. (The last step uses the spectral mapping theorem; cf. [**3**, Theorem 3, p. 55].) It follows (by the Weierstrass theorem) that L has a bounded extension to all real-valued continuous functions on $\Lambda(A)$. To prove that L is positive, observe first that if p is a real polynomial, then

$$((p(A))^2\xi, \xi) = \|p(A)\xi\|^2 \geqq 0.$$

If f is an arbitrary positive continuous function on $\Lambda(A)$, then approximate $\sqrt{f}$ uniformly by real polynomials; the inequality just proved implies that $L(f) \geqq 0$

(since f is then uniformly approximated by squares of real polynomials). The Riesz theorem now yields the existence of a finite measure μ such that

$$(p(A)\xi, \xi) = \int p d\mu$$

for every real polynomial p.

For each (possibly complex) polynomial q write

$$Uq = q(A)\xi.$$

Since A is Hermitian, $(q(A))^*(=\bar{q}(A))$ is a polynomial in A, and so is $(q(A))^*q(A)$ $(=|q|^2(A))$; it follows that

$$\int |q|^2 d\mu = (\bar{q}(A)q(A)\xi, \xi) = ((q(A))^*q(A)\xi, \xi) = \|q(A)\xi\|^2 = \|Uq\|^2.$$

This means that the linear transformation U from a dense subset of $\mathcal{L}^2(\mu)$ into $\mathcal{H}$ is an isometry, and hence that it has a unique isometric extension that maps $\mathcal{L}^2(\mu)$ into $\mathcal{H}$. The assumption that ξ is a cyclic vector implies that the range of U is in fact dense in, and hence equal to, the entire space $\mathcal{H}$.

It remains only to prove that the transform of A by U is a multiplication. Write $\phi(\lambda)=\lambda$ for all λ in $\Lambda(A)$. Given a complex polynomial q, write $\tilde{q}(\lambda) =\lambda q(\lambda)=\phi(\lambda)q(\lambda)$; then

$$U^{-1}AUq = U^{-1}Aq(A)\xi = U^{-1}\tilde{q}(A)\xi = U^{-1}U\tilde{q} = \tilde{q}.$$

In other words $U^{-1}AU$ agrees, on polynomials, with the multiplication induced by ϕ, and that is enough to conclude that $U^{-1}AU$ is equal to that multiplication. This completes the outline of the proof of the spectral theorem for Hermitian operators.

The formulation of the spectral theorem given above yields fairly easily all the information contained in the more common versions. Thus if A is the multiplication on $\mathcal{L}^2(\mu)$ induced by the real function ϕ on X, and if F is a (complex) Borel measurable function that is bounded on $\Lambda(A)$, then $F(A)$ can be defined as the multiplication induced by the composite function $F\circ\phi$. The mapping $F\rightarrow F(A)$ is the homomorphism that is frequently known by the impressive name of "the functional calculus." If, in particular, $F=F_M$ is the characteristic function of a Borel set M in the real line, and if $E(M)$ is the multiplication induced by $F_M\circ\phi$, then E is the spectral measure of A. The verification that E is indeed a spectral measure is easy. To prove that it belongs to A (i.e., that $A=\int\lambda dE(\lambda)$), proceed as follows. Fix f and g in $\mathcal{L}^2(\mu)$ and write

$$\nu(M) = (E(M)f, g)$$

for each Borel set M; it is to be proved that

$$(Af, g) = \int \lambda d\nu(\lambda).$$

Since $(E(M)f, g) = \int (F_M \circ \phi) f\bar{g} d\mu$ and $\nu(M) = \int F_M d\nu$, it follows that

$$\int (F_M \circ \phi) f\bar{g} d\mu = \int F_M d\nu$$

for all Borel sets M. This implies that

$$\int (F \circ \phi) f\bar{g} d\mu = \int F d\nu$$

whenever F is a simple function, and hence, by approximation, whenever F is a bounded Borel measurable function. This conclusion (for $F(\lambda) \equiv \lambda$) is just what was wanted.

The multiplication version of the spectral theorem implies the spectral measure version, but the latter is canonical (E is uniquely determined by A) whereas the former is not. Consider, for instance, the identity operator on a separable infinite-dimensional Hilbert space in the role of A. It is unitarily equivalent to multiplication by the constant function 1 on, say, the unit interval (with Lebesgue measure); it is also unitarily equivalent to multiplication by the constant function 1 on the set of positive integers (with the counting measure).

There is a spectral theorem for normal operators also; its statement can be obtained from the one given above by substituting "normal" for "Hermitian." It is a well-known technical nuisance that the proof of the spectral theorem for normal operators involves some difficulties that do not arise in the Hermitian case. The source of the trouble is that it is not enough just to replace polynomials in a real variable by polynomials in a complex variable; the Weierstrass theorem demands the consideration of polynomials in two real variables. There is a consequent difficulty in extending the spectral mapping theorem to the kind of functions (polynomials in the real and imaginary parts of a complex variable) that arise in the imitation of the proof above. Even the equality of norm and spectral radius, while true for normal operators, requires a proof quite a bit deeper than in the real case. One way around all this is not to imitate the proof but to use the result. In [**3**, p. 72], for instance, the spectral theorem for normal operators (spectral measure version) is derived from the Hermitian theorem (spectral measure version); the only additional tool needed is an essentially classical extension theorem for measures in the plane.

In any case, all this talk about proof is somewhat beside the point in this paper. The reason a proof is outlined above is not so much to induce belief in the result as to clarify it. The emphasis here is not on *how* but on *what*, not on *proof* but on *statement*, not on *How does the spectral theorem come about*? but on *What does the spectral theorem say*?

To see how the multiplication point of view can be used, consider the Fuglede commutativity theorem [**2**]. A possible statement is this: if A is normal and if B is an operator that commutes with A, then B commutes with $F(A)$ for each Borel measurable function bounded on $\Lambda(A)$. (An alternative statement, only apparently weaker, is that if B commutes with A, then B commutes with A^*; for a recent elegant proof see [5].) The spectral theorem shows that there is no loss of generality in assuming that A is the multiplication induced by ϕ, say, on a measure space X with measure μ. If F_M is, for each Borel set M in the complex plane, the characteristic function of M, and if $E(M)$ is the multiplication induced by $F_M \circ \phi$, then it is sufficient to prove that B commutes with each $E(M)$. (Approximate the general F by simple functions, as before.) If $\mathcal{E}(M)$ is the range of the projection $E(M)$, then the desired result is that $\mathcal{E}(M)$ reduces B, but it is, in fact, sufficient to prove that $\mathcal{E}(M)$ is invariant under B. Reason: apply the invariance conclusion, once obtained, to the complement of M, and infer that both $\mathcal{E}(M)$ and $(\mathcal{E}(M))^\perp$ are invariant under B.

Observe now that $\mathcal{E}(M)$ is the set of all those functions in $\mathcal{L}^2(\mu)$ that vanish outside $\phi^{-1}(M)$, and consider first the case of the closed unit disc,

$$M = \{\lambda: |\lambda| \leqq 1\};$$

then

$$\phi^{-1}(M) = \{x: |\phi(x)| \leqq 1\}.$$

Assertion: $\mathcal{E}(M)$ consists of all f in $\mathcal{L}^2(\mu)$ for which the sequence $\{\|Af\|, \|A^2f\|, \|A^3f\|, \cdots\}$ is bounded. Indeed, if f vanishes outside $\phi^{-1}(M)$, then

$$\|A^nf\|^2 = \int |\phi^nf|^2 d\mu = \int_{\phi^{-1}(M)} |\phi^n|^2 \cdot |f|^2 d\mu \leqq \int |f|^2 d\mu.$$

If, on the other hand, there is a set S of positive measure on which $f \neq 0$ and $|\phi| > 1$, then

$$\|A^nf\|^2 = \int |\phi^nf|^2 d\mu \geqq \int_S |\phi^2|^n |f|^2 d\mu \to \infty.$$

The assertion is proved, and the invariance of $\mathcal{E}(M)$ under B follows: if $\|A^nf\| \leqq c$ for all n, then $\|A^nBf\| = \|BA^nf\| \leqq \|B\| \cdot \|A^nf\| \leqq \|B\| \cdot c$ for all n.

If M is any closed disc, $M = \{\lambda: |\lambda - \lambda_0| \leqq r\}$, then

$$\phi^{-1}(M) = \{x: |\phi(x) - \lambda_0| \leqq r\} = \left\{x: \left|\left(\frac{\phi - \lambda_0}{r}\right)(x)\right| \leqq 1\right\}.$$

Since B commutes with multiplication by ϕ, it commutes with multiplication by $(\phi - \lambda_0)/r$ also, and it follows from the preceding paragraph that $\mathcal{E}(M)$ is invariant under B.

The rest of the proof is easy measure theory; from this point of view spectral measures behave even better than numerical measures. Since $\mathcal{E}(M)$ is invariant under B whenever M is a disc, the same is true whenever M is the union of countably many discs. This implies that $\mathcal{E}(M)$ is invariant under B whenever M is open, and hence (regularity) for arbitrary Borel sets M.

References

1. R. P. Boas, A primer of real functions, Math. Assoc. of America, Carus Monograph no. 13, 1960.
2. B. Fuglede, A commutativity theorem for normal operators, Proc. Nat. Acad. Sci. U.S.A., 36 (1950) 35–40.
3. P. R. Halmos, Introduction to Hilbert space, Chelsea, New York, 1951.
4. ———, Measure theory, Van Nostrand, New York, 1950.
5. M. Rosenblum, On a theorem of Fuglede and Putnam, J. London Math. Soc., 33 (1958) 376–377.
6. I. E. Segal, Decomposition of operator algebras, II, Memoirs, Amer. Math. Soc., 1951.

AN EXTENSION OF THE FERMAT THEOREM

L. CARLITZ, Duke University

1. If p is a prime, $e \geqq 1$ and $(a, p)=1$ then

$$a^w \equiv 1 \pmod{p^e} \tag{1}$$

provided $p^{e-1}(p-1) \mid w$. It follows from (1) that

$$\sum_{s=0}^{r} (-1)^{r-s} \binom{r}{s} a^{n+sw} = a^n(a^w - 1)^r \equiv 0 \pmod{p^{re}} \tag{2}$$

for all $r \geqq 0$. This congruence is useful for example in deriving Kummer's congruence for the Euler and Bernoulli numbers [1, Ch. 14].

It may be of interest to examine the sum

$$\Delta^r a^{n^k} = \sum_{s=0}^{r} (-1)^{r-s} \binom{r}{s} a^{(n+sw)^k}, \tag{3}$$

where k is a fixed integer $\geqq 1$. We remark that (3) is suggested by some recent work [2] on colored graphs. We shall prove

THEOREM 1. *Let $p^{e-1}(p-1) \mid w$ and let $\lambda \geqq e$ be the largest integer such that*

$$a^w \equiv 1 \pmod{p^\lambda}. \tag{4}$$

Then

$$\sum_{s=0}^{r} (-1)^{r-s} \binom{r}{s} a^{(n+sw)^k} \equiv 0 \pmod{p^{\lambda r_k}}, \tag{5}$$

where

$$r_k = [(r + k - 1)/k], \tag{6}$$

the greatest integer $\leqq (r+k-1)/k$.