

TEN PROBLEMS IN HILBERT SPACE¹

BY P. R. HALMOS

Dedicated to my teacher and friend Joseph Leo Doob with admiration and affection.

TABLE OF CONTENTS

| | |
|---|-----|
| PREFACE | 887 |
| 1. CONVERGENT. Does the set of cyclic operators have a non-empty interior?... | 889 |
| 2. WEIGHTED. Is every part of a weighted shift similar to a weighted shift? . . . | 892 |
| 3. INVARIANT. If an intransitive operator has an inverse, is its inverse also intransitive? | 898 |
| 4. TRIANGULAR. Is every normal operator the sum of a diagonal operator and a compact one? | 901 |
| 5. DILATED. Is every subnormal Toeplitz operator either analytic or normal?... | 906 |
| 6. SIMILAR. Is every polynomially bounded operator similar to a contraction?.. | 910 |
| 7. NILPOTENT. Is every quasinilpotent operator the norm limit of nilpotent ones? | 915 |
| 8. REDUCIBLE. Is every operator the norm limit of reducible ones? | 919 |
| 9. REFLEXIVE. Is every complete Boolean algebra reflexive? | 922 |
| 10. TRANSITIVE. Is every non-trivial strongly closed transitive atomic lattice either medial or self-conjugate? | 926 |
| BIBLIOGRAPHY | 931 |

PREFACE

Machines can write theorems and proofs, and read them. The purpose of mathematical exposition for people is to communicate ideas, not theorems and proofs. Experience shows that almost always the best way to communicate a mathematical idea is to talk about concrete examples and unsolved problems. In what follows I try to communicate some of the basic ideas of Hilbert space theory by discussing a few of its problems.

Nobody, except topologists, is interested in problems about Hilbert space; the people who work in Hilbert space are interested in problems about operators. The problems below are about operators. I do not know the solution of any of them. I guess, however, that

AMS 1969 subject classifications. Primary 4702.

Key words and phrases. Hilbert space, operators, cyclic vectors, weighted shifts, invariant subspaces, compact operators, dilations, contractions, quasinilpotence, reducibility, reflexive lattices, transitive lattices.

¹ Lectures delivered at a Regional Conference in the Mathematical Sciences, sponsored by the National Science Foundation and arranged by the Conference Board of the Mathematical Sciences, at Texas Christian University, Fort Worth, Texas, May 25–29, 1970.

they vary a lot in depth. Some have been around for many years and are known to be both difficult and important; others are untested and may turn out to be trivial.

The problems are formulated as yes-or-no questions. That, I believe, is the only clear way to formulate any problem. What a mathematician usually wants to do is something vague, like study, determine, or characterize a class of objects, but until he can ask a clear-cut test question, the chances are he does not understand the problem, let alone the solution. The purpose of formulating the yes-or-no question, however, is not only to elicit the answer; its main purpose is to point to an interesting area of ignorance.

A discussion of unsolved problems is more ephemeral than an exposition of known facts. Deciding to make a virtue out of necessity, I chose to focus on a narrow segment of the present, rather than on a broad view of history. As often as possible the references are to recent publications, and the proofs presented in detail are more from current folklore than from standard texts. Consequently some results (with appropriate credit lines to their discoverers) are published here for the first time. Most of the results and proofs with no assigned credit are "well known"; only a small number are mine. If a result is attributed to someone but is not accompanied by a reference to the bibliography, that means I learned it through personal communication, and, as far as I could determine, it is not available in print.

The chief prerequisite for understanding the exposition is a knowledge of the standard theory of operators on Hilbert spaces. In some cases that means that not even the spectral theorem is needed, and in other cases every available ounce of multiplicity theory would help to understand things clearly. In any case, the presentation is not cumulative; some readers who find §1 obscure may find §10 obvious.

As an expository experiment I have put some exercises into each section. This is unusual in papers (as opposed to books). The reason I did it is that originally each section was a lecture, and lectures (as opposed to papers) can stand every bit of elasticity that anyone can think of to put in. The audience is trapped and cannot impatiently riffle the pages forward and back, and the lecturer can use something that is easily omitted or inserted without disturbing the continuity.

As for terminology: *Hilbert space* means a complete, complex, inner-product space; *subspace* means a closed linear manifold; *operator* means a bounded linear transformation. The Hilbert spaces of principal interest are the ones that are neither too large nor too small, i.e., they are separable but infinite-dimensional. The problems are stated without any such explicit size restrictions, but in some

cases they become trivial for either the large spaces, or the small ones, or both.

I thank Peter Fillmore, Pratibha Gajendragadkar, and Eric Nordgren for reading the first written version of these lectures and for making many helpful suggestions.

1. CONVERGENT

Problem 1. *Does the set of cyclic operators have a non-empty interior?*

A vector f in a Hilbert space H is cyclic for an operator A on H in case the smallest subspace of H that contains f and is invariant under A is H itself. An operator is cyclic in case it has a cyclic vector.

In the finite-dimensional case, A is cyclic if and only if its minimal polynomial is equal to its characteristic polynomial. (Equivalently: each eigenvalue occurs in only one Jordan block, or the space of eigenvectors corresponding to each eigenvalue has dimension 1 [24, p. 69 et seq]; cf. also [33].) It follows easily that in the finite-dimensional case the set \mathcal{C} of cyclic operators is open. If the dimension is n , then the operators with n distinct eigenvalues constitute a dense set; it follows that \mathcal{C} is dense. (Consequence: \mathcal{C} is not closed; for instance, $\text{diag}\langle 1, 1+1/n \rangle \rightarrow \text{diag}\langle 1, 1 \rangle$.)

There is only one reasonable topology for the operators on a finite-dimensional space; in the infinite-dimensional case there are several. Problem 1 refers to the norm (or uniform) topology, the one according to which $A_n \rightarrow A$ in case $\|A_n - A\| \rightarrow 0$.

If f is a cyclic vector for A , then the countable set $\{f, Af, A^2f, \dots\}$ spans H ; it follows that in the non-separable case \mathcal{C} is empty.

In the separable infinite-dimensional case what little is known is negative. The set \mathcal{C} is surely *not* open; the cyclic operator $\text{diag}\langle 1, \frac{1}{2}, \frac{1}{3}, \dots \rangle$ is the limit of the non-cyclic operators $\text{diag}\langle 1, \frac{1}{2}, \dots, 1/n, 0, 0, 0, \dots \rangle$. It is, of course, possible that the interior of \mathcal{C} is non-empty; that is what Problem 1 is about. A reasonable candidate for an operator in the interior of \mathcal{C} is the unilateral shift U . (According to the most easily accessible definition U is the operator on l^2 that sends $\langle \xi_0, \xi_1, \xi_2, \dots \rangle$ onto $\langle 0, \xi_0, \xi_1, \xi_2, \dots \rangle$.) That candidate happens to fail (J. G. Stampfli). The set \mathcal{C} is also *not* dense; in fact if V is the direct sum of two copies of U and $\|V - A\| < 1$, then A is not cyclic. Indeed:

$$\|1 - A^*V\| = \|V^*V - A^*V\| \leq \|V^* - A^*\| < 1,$$

so that A^*V is invertible; it follows that V^*A is invertible and hence that $\text{ran } A \cap \ker V^* = 0$. Since, thus, $\text{ran } A$ is disjoint from the 2-dimensional space $\ker V^*$, the co-dimension of $\text{ran } A$ must be at least

2. (The proof so far was needed to get this inequality only. An alternative way to get it is to note that it is a special case of a small part of the theory of Fredholm operators [35].) Now it is clear that A cannot be cyclic; for each f , the span of $\{f, Af, A^2f, \dots\}$ has co-dimension at least 1.

The point of the preceding discussion is to call attention to the existence of topological problems of analytic importance in operator theory. There are many; Problem 1 is just a sample. Other examples are Problems 7 and 8, about the closures of certain sets of operators, or, in other words, about approximation theory. Here is a well known example of the same kind: is the set of all invertible operators dense? The answer is yes for finite-dimensional spaces, and furnishes an occasionally useful proof technique; to prove something for all matrices, prove it for the non-singular ones first and then approximate. For infinite-dimensional spaces the answer is no [20, Problem 109]. It is slightly less well known that the answer remains no even for a somewhat larger set, the set of kernel-free operators. (If A is invertible, then A is kernel-free, i.e., $\ker A = 0$, but the converse is not true.) In other words, there exists a non-empty open set every element of which has a non-trivial kernel. Indeed, the open ball $\{A : \|U^* - A\| < 1\}$ is such a set (where U is, as before, the unilateral shift). For the proof, apply the (Fredholm) argument that disproved the density of the cyclic operators to infer that, for each A in that ball, $\text{ran } A^*$ has positive co-dimension, and then conclude that $\ker A \neq 0$.

Here is a classically important question: is the set of invertible operators connected? The answer is yes [20, Problem 110]; the proof is an application of the polar decomposition of operators. A much less important question that takes most people a few seconds longer to answer than it should is this (**Exercise 1**): is the set of non-scalar normal operators connected?

All the questions so far concerned the norm topology; questions about the other operator topologies are likely to be (a) easier and (b) more pathological. The pathology of topology can be fun, however, and, although the ground has been pretty well worked over, new and even useful facts do turn up from time to time.

The strong operator topology is the one according to which $A_n \rightarrow A$ in case $\|A_n f - A f\| \rightarrow 0$ for each vector f . (The indices, here and elsewhere, need not and should not be restricted to integers; any directed set will do.) A famous and sometimes bothersome misbehavior of the strong topology is the discontinuity of the adjoint. Indeed, if $A_k = U^{*k}$ (the unilateral shift again), $k = 1, 2, 3, \dots$, then $A_k \rightarrow 0$ strongly, but $\{A_k^*\}$ is not strongly convergent to anything [20, Problem 90]. This

negative result is old; the following interesting positive observation of Kadison's is quite recent [26]: the restriction of the adjoint to the set of normal operators is strongly continuous. (That is: if $A_n \rightarrow A$ strongly, where A and each A_n is normal, then $A_n^* \rightarrow A^*$ strongly.) The proof is a trick, but a simple trick:

$$\begin{aligned} \|(A_n^* - A^*)f\|^2 &= \|A_n f\|^2 - \|A f\|^2 + (f, (A - A_n)A^*f) + ((A - A_n)A^*f, f) \\ &\leq \|(A - A_n)f\|(\|A f\| + \|A_n f\|) + 2\|(A - A_n)A^*f\| \cdot \|f\|. \end{aligned}$$

Approximation techniques that rely on the strong topology do not seem to have been used much, although there are some interesting questions whose known and surprising answers might turn out to be useful. Example: what is the strong closure of the set of normal operators? Answer: the set of subnormal operators [4]. The proof is not trivial. Another example (**Exercise 2**): what is the strong closure of the set of unitary operators?

The weak operator topology is the one according to which $A_n \rightarrow A$ in case $(A_n f, g) \rightarrow (A f, g)$ for each f and g . A useful recent result is that rank (dimension of the closure of the range) is weakly lower semicontinuous; that is, if $A_n \rightarrow A$ weakly, then $\liminf_n \text{rank } A_n \geq \text{rank } A$ [21]. The proof is not immediately obvious, but it is certainly not deep. Warning: the possible values of rank in this context are the nonnegative integers, together with ∞ ; no distinction must be made among different infinite cardinals. Were such a distinction to be made, the result would become false. Suppose indeed that the underlying Hilbert space has an uncountable orthonormal basis $\{e_j; j \in J\}$. Let D be the set of all countable subsets of J , ordered by inclusion; for each n in D , write A_n for the projection onto the span of $\{e_j; j \in n\}$. Since for each vector f there exists an n_0 in D such that $f \perp e_j$ whenever j is not in n_0 , it follows that $A_n \rightarrow 1$ (not only weakly, but, in fact, strongly). Since $\text{rank } A_{n_0} = \aleph_0$ and $\text{rank } 1 > \aleph_0$, the cardinal version of semicontinuity is false.

What is the weak closure of the set of normal operators? Answer: in the finite-dimensional case it is the set of normal operators [21]; in the infinite-dimensional case it is the set of all operators [18]. What is the weak closure of the set of projections? Answer: in the finite-dimensional case it is the set of projections; in the infinite-dimensional case it is the set of Hermitian operators A with $0 \leq A \leq 1$ [18]. The proofs of the infinite-dimensional statements are not difficult, but they depend on a little dilation theory. The results show why weak approximation theory is not likely to be fruitful; it is too good to be any good. **Exercise 3**: what is the weak closure of the set of scalars?

A simple example that illustrates many of the curious properties of the weak operator topology can be described in terms of matrices. Let A_n be the infinite matrix (or, by a slight abuse of language, the operator) whose entries at the positions $\langle 1, 1 \rangle$, $\langle 1, n \rangle$, $\langle n, 1 \rangle$, and $\langle n, n \rangle$ are $\frac{1}{2}$, and whose remaining entries are all 0 ($n = 2, 3, 4, \dots$). It is easy to verify that A_n is idempotent, and it is plausible (and true, and easy to verify) that if A is the matrix whose $\langle 1, 1 \rangle$ entry is $\frac{1}{2}$ and whose remaining entries are all 0, then $A_n \rightarrow A$ weakly. The example shows how projections can converge to a non-projection, it proves that multiplication (in fact squaring) is not weakly continuous, and, incidentally, it implies that the strong and the weak topologies are distinct. (Here is an unimportant but unsolved teaser: does squaring have any points of weak continuity?)

ANSWERS TO THE EXERCISES. (1) The non-scalar *Hermitian* operators form a (norm) connected set. For a 1-dimensional space this is trivial. In any other case, the set of Hermitian operators is a real vector space of dimension 3 or more. If, therefore, A and B are non-scalar Hermitian operators, then there always exists a non-scalar Hermitian C such that $tA + (1-t)C$ and $tB + (1-t)C$ are non-scalar Hermitian operators whenever $0 \leq t \leq 1$. The *normal* case can be joined to the Hermitian one. Indeed, if $A + iB$ (with A and B Hermitian) is a non-scalar normal operator, then assume, with no loss of generality, that A is not a scalar, and consider $A + itB$, $0 \leq t \leq 1$. (This solution is due to J. J. Schäffer.)

(2) The strong closure of the set of unitary operators is the set of isometries. Since the set of isometries is easily seen to be strongly closed, and since every isometry is a direct sum of a unitary operator and a number of copies of the unilateral shift [20, Problem 118], it is sufficient to prove that U , the unilateral shift, is a strong limit of unitary operators. That is easy: If

$$U_n \langle \xi_0, \xi_1, \xi_2, \dots \rangle = \langle \xi_n, \xi_0, \xi_1, \dots, \xi_{n-1}, \xi_{n+1}, \xi_{n+2}, \dots \rangle$$

for $n = 1, 2, 3, \dots$, then each U_n is unitary and $U_n \rightarrow U$ strongly.

(3) The set of scalars is weakly closed. Indeed, if $\lambda_n(f, g) \rightarrow (Sf, g)$ for each f and g , then $\{\lambda_n\}$ is a Cauchy net of complex numbers, and therefore $\lambda_n \rightarrow \lambda$ for some complex number λ ; it follows that $\lambda_n(f, g) \rightarrow \lambda(f, g)$ and hence that $Sf = \lambda f$ for each f .

2. WEIGHTED

Problem 2. *Is every part of a weighted shift similar to a weighted shift?*

If U is the unilateral shift and if e_n is the sequence $\langle \xi_0, \xi_1, \xi_2, \dots \rangle$

with $\xi_n = 1$ and $\xi_i = 0$ when $i \neq n$, then $\{e_0, e_1, e_2, \dots\}$ is an orthonormal basis for l^2 and $Ue_n = e_{n+1}$, $n = 0, 1, 2, \dots$. If $\langle \alpha_0, \alpha_1, \alpha_2, \dots \rangle$ is a bounded sequence of positive numbers, then the equations $Ae_n = \alpha_n e_{n+1}$, $n = 0, 1, 2, \dots$, unambiguously define an operator A . Such operators are called *weighted shifts*; they are of interest because they can be used to construct examples of many different kinds of operatorial behavior. (To eliminate easy but dull case distinctions, the number 0 is not allowed as a weight.)

For any operator A on any Hilbert space H , a *part* of A is, by definition, the restriction of A to a subspace of H invariant under A . One of the main reasons for the success of the Beurling treatment of the unilateral shift is that every non-trivial part of U is unitarily equivalent to U [20, Problem 123]. Here "non-trivial" excludes the most trivial case only, the restriction to the subspace 0. That should, of course, be excluded in the formulation of Problem 2 also, and in all related contexts; that exclusion is hereby made, retroactively and forward to eternity.

What sense does it make to say that a certain operator "is" a weighted shift? If shifts are defined on the concrete Hilbert space l^2 (as above), then the best that can be hoped about a part of one is that it is "abstractly identical" with or "isomorphic" to another; the correct technical phrase is "unitarily equivalent". If shifts are defined with respect to arbitrary orthonormal bases $\{e_0, e_1, e_2, \dots\}$ in arbitrary separable infinite-dimensional Hilbert spaces (the best way), then "is" means "is". There is a third interpretation, which may seem artificial but has some merit; it is obtained by demanding isomorphism with respect to the linear and the topological structures, but ignoring the metric structure. The correct technical word is "similar"; two operators A and B , on Hilbert spaces H and K , are similar in case there is a bounded linear transformation S from H onto K , with a bounded inverse, such that $A = S^{-1}BS$.

There are two reasons for the insistence on positive weights in the definition of weighted shifts: (a) it seems more natural, and (b) nothing is gained by allowing arbitrary complex weights. The proof of (b) is the following assertion (**Exercise 1**): if A and B are weighted shifts, with complex weight sequences $\langle \alpha_n \rangle$ and $\langle \beta_n \rangle$, then a necessary and sufficient condition that A and B be unitarily equivalent is that $|\alpha_n| = |\beta_n|$ for $n = 0, 1, 2, \dots$. Consequence: if A is a weighted shift and if λ is a complex number of modulus 1, then A is unitarily equivalent to λA . This kind of "circular symmetry" is an unusual but useful property for an operator to have. More along the same lines cannot be asked; modulus 1 is in the nature of things. That is (**Exercise 2**): if a non-nilpotent operator A is similar to λA , then $|\lambda| = 1$.

The unitary equivalence theory of weighted shifts is easy; their similarity theory is slightly more complicated, but still near the surface. In the notation of the preceding paragraph, A is similar to B if and only if the sequence of quotients

$$\left| \frac{\alpha_0 \cdots \alpha_n}{\beta_0 \cdots \beta_n} \right|$$

is bounded away from 0 and from ∞ [20, Problem 76]. It is thus easy to tell when two weighted shifts are similar, but the circle of ideas centered at Problem 2 includes a question of a much more difficult type: when is an operator obtained by certain non-transparent constructions similar to a weighted shift? The formation of a part of a weighted shift is such a construction; there are others. Consider, for instance, the (Cesaro) operator C_0 on l^2 defined by

$$C_0 \langle \xi_0, \xi_1, \xi_2, \dots \rangle = \langle \eta_0, \eta_1, \eta_2, \dots \rangle,$$

where

$$\eta_n = \frac{1}{n+1} \sum_{i=0}^n \xi_i.$$

The known spectral properties of C_0 [7] suggest the question: is $1 - C_0$ similar to a weighted shift? The answer is not known. Another interesting operator (Volterra) is defined on $L^2(0, 1)$ by $Vf(x) = \int_0^x f(y) dy$; i.e., Vf is, for each f , the indefinite integral of f . (Exercise 3): is V similar to a weighted shift?

What are weighted shifts good for? The answer is that they can be used for examples and counterexamples to illustrate many properties of operators. Among such properties are the existence of square roots, spectral behavior, irreducibility, the structure of invariant subspaces, and subnormality.

Assertion: no weighted shift has a square root. Suppose, indeed, that A is a weighted shift, and suppose that $B^2 = A^*$. (The adjoint is easier to treat, and the result obviously comes to the same thing.) Since $\ker B \subset \ker A^*$, and since $\ker A^* \neq 0$, it follows that $\ker B = \ker A^*$. Since $\ker A^* \subset \text{ran } A^*$, it follows that $\ker A^* \subset \text{ran } B$. Hence if e is in $\ker A^*$, $e \neq 0$, then there exists a vector f such that $Bf = e$, and it is easy to deduce (from $\ker B = \ker A^*$) that e and f are linearly independent. Conclusion: the dimension of $\ker B^2 (= \ker A^*)$ must be at least 2, which is absurd [29], [20, Problem 114].

A weighted shift provides the easiest example of a kernel-free operator whose spectrum consists of 0 only [20, Problem 80], and the

more delicate spectral behavior of weighted shifts has also received quite a bit of attention [16], [29], [34], [41].

Weighted shifts are irreducible. That is, if A is a weighted shift and if M is a subspace that reduces A , then M is either 0 or the entire space; equivalently, if P is a projection that commutes with A , then $P=0$ or $P=1$. To prove this, all that is needed is that the kernel of A^* is spanned by a single (non-zero) vector e which happens to be cyclic for A . Indeed: $A^*Pe=PA^*e=0$ implies (since e spans $\ker A^*$) that $Pe=\lambda e$, and hence (since P is a projection) that either $Pe=0$ or $Pe=e$. If $Pe=0$, then $A^ne=A^n(1-P)e=(1-P)A^ne$, which belongs to $\ker P$ for all n , and therefore (since e is cyclic for A) $P=0$; if $Pe=e$, then $A^ne=A^nPe=PA^ne$, which belongs to $\text{ran } P$ for all n , and therefore $P=1$. (This proof is due to K. J. Harrison.) A somewhat stronger result is known: every operator similar to a weighted shift is irreducible [29], [51].

To what extent can the structure of the invariant subspaces of an operator be prescribed? Is it, for instance, possible that an operator has exactly one invariant subspace of each dimension between 0 and \aleph_0 inclusive? The answer is yes, and the standard example (due to Donoghue) is given by the adjoint of a suitably weighted shift [20, Problem 151]. (The unilateral shift itself, unweighted, won't do.)

There are many generalizations of the concept of normality; two important ones, in decreasing order of generality and increasing order of importance, are hyponormality and subnormality. An operator A is *hyponormal* in case $A^*A-AA^*\geq 0$; an operator is *subnormal* in case it is part of a normal operator. (The verification that every subnormal operator is hyponormal is the easiest part of this otherwise tricky subject.) Which weighted shifts are normal, which are hyponormal, and which are subnormal? The first question is settled just by looking at kernels; there is no weight sequence that can make a weighted shift normal. The answer to the second question is also easy [20, Problem 160]: the weight sequence must be monotone increasing. The answer to the third question is not easy. One formulation was offered by Stampfli [49]; the hitherto unpublished formulation below (due to C. Berger) is very different. Berger's condition is elegant and easy to state: a weighted shift is subnormal if and only if the squares of the partial products of the weights constitute the moment sequence of a probability measure in the unit interval.

For the proof, suppose first that A is a subnormal weighted shift, and assume, with no loss of generality, that $\|A\|=1$. Let B (on a Hilbert space K) be the minimal normal extension of A ; then $\|B\|=1$. Since e_0 is "bicyclic" for B (i.e., by minimality, there is no proper sub-

space of K that contains e_0 and reduces B), the spectral theorem implies that B is unitarily equivalent to the position operator ($f(z) \rightarrow zf(z)$) on L^2 of the closed unit disc D with respect to some probability measure μ . The standard unitary equivalence, moreover, carries the vector e_0 onto the constant function 1. Identify B with its unitarily equivalent transform on $L^2(\mu)$, and, correspondingly, identify A with the restriction of B to the span of (the images of) the vectors e_0, e_1, e_2, \dots .

Let ν be the "marginal" measure induced by μ in the closed unit interval I . That is: if $S: D \rightarrow I$ is the mapping defined by $Sz = |z|$, then $\nu(E) = \mu(S^{-1}E)$ for each Borel subset E of I . It follows [17, p. 163] that $\int_I f d\nu = \int_D fS \cdot d\mu$ whenever either side of the equation makes sense. If, in particular, $f(\rho) = \rho^{2n}$, $n = 0, 1, 2, \dots$, then

$$\begin{aligned} \int_I \rho^{2n} d\nu(\rho) &= \int_D |z|^{2n} d\mu(z) = \int_D |z^n \cdot 1|^2 d\mu(z) \\ &= \|B^n e_0\|^2 = \|A^n e_0\|^2. \end{aligned}$$

The last term is easy to compute: if $p_0 = 1$ and $p_{n+1} = \alpha_n p_n$, then $A^n e_0 = p_n e_n$, and therefore

$$\int_I \rho^{2n} d\nu(\rho) = p_n^2, \quad n = 0, 1, 2, \dots$$

The proof of the necessity of Berger's condition is complete; p_n^2 is the n th moment of $d\nu(\sqrt{\rho})$.

To prove sufficiency, start with a probability measure ν in I , and write $\int_I \rho^{2n} d\nu(\rho) = p_n$. If λ is normalized Lebesgue measure in the perimeter C of the unit circle, then $\nu \times \lambda$ in $I \times C$ induces (it is tempting to say "is") a measure μ in D . That is: if $T: I \times C \rightarrow D$ is the mapping defined by $(\rho, u) \rightarrow \rho u$, then $\mu(E) = (\nu \times \lambda)(T^{-1}E)$ for each Borel subset E of D . Let B be the position operator on $L^2(\mu)$, and let A be the restriction of B to the subspace $H^2(\mu)$ spanned by $\{f_0, f_1, f_2, \dots\}$, where $f_n(z) = z^n$, $n = 0, 1, 2, \dots$. Since B is normal, it follows that A is subnormal.

A straightforward computation shows that

$$(f_n, f_m) = \int_I \rho^{n+m} d\nu(\rho) \cdot \int_C u^{n-m} d\lambda(u), \quad n, m = 0, 1, 2, \dots$$

If $n \neq m$, the last factor vanishes, i.e., the f 's are pairwise orthogonal. If $n = m$, the last factor is 1, and therefore

$$\|f_n\|^2 = p_n^2, \quad n = 0, 1, 2, \dots$$

If, therefore, $e_n = f_n/p_n$, $n = 0, 1, 2, \dots$, then the e 's form an ortho-

normal basis for $H^2(\mu)$. Since, finally,

$$Ae_n = A\left(\frac{1}{p_n}f_n\right) = \frac{1}{p_n}f_{n+1} = \frac{p_{n+1}}{p_n}e_{n+1}, \quad n = 0, 1, 2, \dots,$$

it follows that A is (unitarily equivalent to) the weighted shift whose weight sequence has exactly the p 's for its partial products. The proof of Berger's theorem is complete.

The preceding report on weighted shifts is a representative but not exhaustive summary of what is known. There is much that remains to be done, and, in particular, there are at least three directions of generalization that promise to be fruitful.

(a) Two-way shifts ($e_n \rightarrow \alpha_n e_{n+1}$ for all integers n , not the non-negative ones only) have many properties in common with the one-way kind, but not all. (Sample question with a known but not completely trivial answer: is the Volterra operator similar to a two-way weighted shift?)

(b) If the shift $n \rightarrow n+1$ on the integers is replaced by a measure-preserving transformation on a more interesting measure space, the theory of weighted shifts becomes the theory of "weighted translations"; although a few things about them are known [36], their theory is still largely undeveloped.

(c) The space l^2 is the direct sum of copies of a 1-dimensional Hilbert space; a generalization in the direction of higher multiplicities is obtained via the formation of direct sums of more general Hilbert spaces. In that case the role of the numerical weights is played by operator weights. This is essentially virgin territory; the idea has been used to construct counterexamples [20, Problem 164], but a general theory does not yet exist.

ANSWERS TO THE EXERCISES. (1) Given the α 's and the β 's, put $\delta_0 = 1$ and determine δ_n recursively from $\alpha_n \delta_n = \beta_n \delta_{n+1}$, $n = 0, 1, 2, \dots$; if always $|\alpha_n| = |\beta_n|$, then always $|\delta_n| = 1$, and the "diagonal operator" determined by the δ 's transforms A onto B . This proves sufficiency. To prove necessity, suppose that $A = W^*BW$, where W is unitary, so that $A^n = W^*B^nW$, and also $A^* = W^*B^*W$. It follows that W^* carries e_0 (in the kernel of B^*) onto a unit vector in the kernel of A^* ; assume with no loss of generality that $W^*e_0 = e_0$. It follows that $W^*B^n e_0 = A^n e_0$, and hence (form norms and use induction) that $|\beta_n| = |\alpha_n|$ for $n = 0, 1, 2, \dots$ [20, Problem 75; 28].

(2) Since

$$\frac{1}{\|S\|} \|A^n\| \frac{1}{\|S^{-1}\|} \leq \|S^{-1}A^nS\| \leq \|S^{-1}\| \cdot \|A^n\| \cdot \|S\|,$$

it follows that $\|S^{-1}A^nS\|^{1/n}/\|A^n\|^{1/n} \rightarrow 1$. If $S^{-1}AS = \lambda A$, then this implies that $|\lambda| = 1$ [28]. Note that if A is nilpotent, it is quite possible to have A be similar to λA for every non-zero complex number λ ; an example is $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

(3) The Volterra operator is not similar to a weighted shift. Reason: if V is similar to A , then V^* is similar to A^* ; if A is a weighted shift, then $\ker A^* \neq 0$, but an easy calculation shows that $\ker V^* = 0$. (This proof is due to J. G. Stampfli.)

3. INVARIANT

Problem 3. *If an intransitive operator has an inverse, is its inverse also intransitive?*

An operator is called *intransitive* if it leaves invariant some subspace other than 0 or the whole space; in the contrary case it is transitive. The hope that non-trivial invariant subspaces always exist (i.e., that, except on a 1-dimensional space, every operator is intransitive) is perhaps still alive in the hearts of some. That existence theorem would be the first step toward a detailed structure theory for operators (possibly a generalization of the theory of the Jordan form in the finite-dimensional case). Repeated failure has convinced most of those who tried that the truth lies in the other direction, but the proof of that is, so far, just as elusive.

There are many special classes of operators that have been proved intransitive. Problem 3 (due to R. G. Douglas) seems to be the simplest problem of its kind (derive the intransitivity of something from that of something else). For the strong kind of invariance (that is, for reduction) the problem is easy: if a reducible operator has an inverse, then the inverse is also reducible. Indeed, more is true: if a subspace M reduces an invertible operator A , then the same M reduces A^{-1} . (Proof: if a projection commutes with A , then it commutes with A^{-1} .) For plain invariance the more is false; the invariance of a subspace under an invertible operator A does not imply its invariance under A^{-1} . (Consider the bilateral shift.)

A related question concerns square roots instead of inverses. Squaring preserves invariance (and reduction); what about the formation of square roots? If A^2 is reducible, is A reducible? The answer is no (consider weighted shifts). If A^2 is intransitive, is A intransitive? The answer is not known. Equally unknown is the answer to a mixed question (proposed by C. Pearcy): if A^2 is reducible, is A at least intransitive? The only thing along these lines that can be said has an undesirably special hypothesis (**Exercise 1**): if A^2 is normal, then A is intransitive.

If even the relatively elementary algebraic questions are unanswered, there is no immediate hope for the much deeper questions of perturbation (in imitation of the theory of spectral perturbation), but they deserve to be put on record. What is known is that if A and B are Hermitian and B has rank 1, then $A+iB$ is intransitive. The presently hoped-for generalization is obtained by replacing "rank 1" by "compact". Although large classes of compact operators are admissible here [47], not all are known to be. The mildest unproved perturbation statement appears to be this: if A is normal and B has rank 1, then $A+B$ is intransitive. The generalization of this perturbation problem to arbitrary intransitive operators in place of normal ones is not going to be easy to settle. Indeed: if A is an arbitrary operator, and if P is a projection of rank 1, then $A=AP+A(1-P)$; the first summand has rank 1, and the second has a non-trivial kernel.

For normal operators the spectral theorem yields many invariant subspaces. The step from normal to subnormal is, however, large: it is not known whether every subnormal operator is intransitive.

Replacement of the algebraic condition of normality by the analytic condition of compactness yields a famous and useful intransitivity theorem: every compact operator (on a space of dimension greater than 1, of course) has a non-trivial invariant subspace [2]. Do two commutative compact operators always have a non-trivial invariant subspace in common? The answer is not known.

Two more invariant subspace theorems deserve mention; they are elementary, but they are sufficiently different from the preceding ones and from each other that they might suggest a new idea to someone.

(i) The strictly algebraic version of the invariant subspace problem has a positive solution (**Exercise 2**): every linear transformation (on a vector space of dimension greater than 1) has a non-trivial invariant linear manifold [46].

(ii) For every operator A there exists a hyperplane M (subspace of co-dimension 1) such that the compression of A to M has an eigenvector. (The compression of A to M is the operator B on M defined as follows: if P is the projection with range M , then $Bf=PAf$ for each f in M .) In fact more is true: for each non-zero vector f , there exists a hyperplane M containing f such that the compression of A to M has f as an eigenvector. Proof: if f is an eigenvector for A , the assertion is trivial. If f is not an eigenvector for A , then f and Af span a 2-dimensional space; let g be a non-zero vector in that space orthogonal to f , and let M be the orthogonal complement of g . Clearly f belongs to the hyperplane M . Since Af is a linear combination of f and g , the projection of Af into M is a scalar multiple of f . This proof is due to L. J.

Wallen. The result was suggested by the following theorem of C. Apostol's: if A is an operator such that 0 is in the spectrum of A but $\ker A = \ker A^* = 0$, then there exists an infinite-dimensional subspace M such that the compression of A to M is compact.

All the results reported on so far concern intransitivity; they assert that under certain conditions certain subspaces are invariant. The last two results to be mentioned in this context go in the direction of transitivity: certain subspaces are not invariant.

(iii) **Exercise 3:** for each countable set of non-trivial subspaces, there exists an operator that leaves none of them invariant.

(iv) There exists a linear transformation on a Hilbert space H that leaves no non-trivial subspace of H invariant. Warning: this does not pretend to be a solution of the invariant subspace problem. "Subspace" here means closed linear manifold, as always, but "linear transformation" does not mean operator, i.e., the linear transformation whose existence is asserted is not necessarily bounded. (There is, however, no fudging about domains: the assertion is about linear transformations that act on the entire space H .) The result is due to A. L. Shields [48]; the proof goes as follows.

Let ψ be the smallest ordinal number with cardinal number c (continuum). If H is a separable infinite-dimensional Hilbert space, then the set of all infinite-dimensional proper subspaces has cardinal number c , and is therefore in one-to-one correspondence $\alpha \rightarrow M_\alpha$ with the predecessors of ψ .

Let f_0 be a non-zero vector in M_0 , and let g_0 be a vector *not* in M_0 . Suppose that f_β and g_β have been defined whenever $\beta < \alpha < \psi$ so that f_β is in M_β , g_β is not in M_β , and the set of all f 's and g 's is linearly independent. The linear (not necessarily closed) manifold V_α spanned by the f 's and g 's cannot include M_α (because the cardinal number of a Hamel basis of M_α is c , and the cardinal number of the predecessors of α is less than c). Let f_α be a vector in M_α that is not in V_α . The linear manifold W_α spanned by V_α and f_α is not H (same argument), and, therefore, it cannot include a non-empty open set (such as the set-theoretic complement of M_α). It follows that there exists a vector g_α that belongs to neither M_α nor W_α . The inductive process so described defines a linearly independent set $E = \{f_\alpha, g_\alpha : \alpha < \psi\}$. Extend E to a Hamel basis for H ; i.e., let F be a set disjoint from E such that $E \cup F$ is a Hamel basis.

Let A be the transformation on E defined by $Af_\alpha = g_\alpha$ and $Ag_\alpha = f_{\alpha+1}$. If F is empty, stop here; if F is not empty but finite, $F = \{h_1, \dots, h_n\}$, write $Ah_i = h_{i+1}$, $i = 1, \dots, n-1$, and $Ah_n = f_0$; and if, finally, F is infinite, let A on F be any permutation of F that has no

finite orbits. In all cases, A is a transformation on the Hamel basis $E \cup F$, and, therefore, A has a unique extension to a linear transformation on H .

The linear transformation A can have no invariant infinite-dimensional proper subspace (because Af_α is never in M_α), and it can have no eigenvector (because it maps each finite linear combination of the vectors in the basis $E \cup F$ onto a linear combination that involves at least one new vector).

ANSWERS TO THE EXERCISES. (1) Since A commutes with A^2 it follows that A commutes with all the spectral projections of A^2 . The only way this comment can fail to yield an invariant (in fact reducing) subspace for A is if A^2 is a scalar. That case has to be examined separately; the examination is straightforward and the desired result follows easily.

(2) Assume that the linear transformation A is such that, for some vector f , the set $\{f, Af, A^2f, \dots\}$ is a linear basis; in all other cases the conclusion is obvious. Since every vector is a (finitely non-zero) linear combination of the form $\sum_{n=0}^{\infty} \alpha_n A^n f$, the mapping $\sum_{n=0}^{\infty} \alpha_n A^n f \rightarrow \sum_{n=0}^{\infty} \alpha_n$ is a linear functional; the kernel of that linear functional is an invariant linear manifold.

(3) If M is a non-trivial subspace, then so is M^\perp , and, therefore, both M and M^\perp are nowhere dense sets. It follows that the set-theoretic union of a countable set of non-trivial subspaces and their orthogonal complements is a set of the first category (meager set), and that, consequently, there exists a vector that does not belong to that union. The projection onto the 1-dimensional space spanned by that vector cannot leave invariant any of the originally given subspaces.

4. TRIANGULAR

Problem 4. *Is every normal operator the sum of a diagonal operator and a compact one?*

An operator A on a Hilbert space H is *diagonal* if H has an orthonormal basis each element of which is an eigenvector of A . On a finite-dimensional space the answer to Problem 4 is trivially yes, because every operator is compact. **Exercise 1:** on a non-separable space the answer is no, even for Hermitian operators.

It is a remarkable theorem of H. Weyl [53] that every *Hermitian* operator on a separable space is the sum of a diagonal operator and a compact one. What follows is first the description of a context within which it is possible to give a proof of Weyl's theorem (different from the classical one), and then a natural generalization that has already

been found useful in other parts of operator theory; it is reasonable to hope that this circle of ideas will have further applications.

On a separable space (the only kind to be considered in this section from now on) if an operator A is diagonal, then there exists an increasing sequence $\{E_n\}$ of projections of finite rank such that $E_n \rightarrow 1$ strongly and $AE_n = E_nA$ for each $n = 1, 2, 3, \dots$. (The converse is not true. What the condition implies is that A is "block-diagonal". That is: the space is the direct sum of finite-dimensional reducing subspaces, which, however, need not be 1-dimensional.) A surprisingly rich class of operators is obtained if commutativity is replaced by asymptotic commutativity. Call an operator A *quasi-diagonal* in case there exists an increasing sequence $\{E_n\}$ of projections of finite rank such that $E_n \rightarrow 1$ strongly and $\|AE_n - E_nA\| \rightarrow 0$.

There are three kinds of basic results about quasidiagonal operators: characterization, closure, and inclusion. Characterization theorems are equivalences; they replace the definition by something else that is sometimes easier to apply. Closure theorems assert that certain operations on quasidiagonal operators lead to other operators of the same kind. Inclusion theorems assert that certain other, more familiar, classes of operators are included among the quasidiagonal ones.

The most elegant characterization theorem replaces the unnatural existence requirement in the definition of quasidiagonality by a simple equation. To describe that equation, note first that the set of all projections of finite rank is partially ordered by range inclusion, and, endowed with that order, it becomes a directed set. If ϕ is a real net (i.e., a real-valued function) on that directed set, then the convergence of ϕ to a limit α , denoted by

$$\phi(E) \rightarrow \alpha \quad \text{or} \quad \lim_{E \rightarrow 1} \phi(E) = \alpha,$$

means that for every positive number ϵ there exists a projection E_0 (of finite rank, of course) such that $|\phi(E) - \alpha| < \epsilon$ whenever $E_0 \leq E$. Similarly, the assertion that

$$\liminf_{E \rightarrow 1} \phi(E) = \alpha,$$

means that α is the smallest number such that for every positive number ϵ and for every E_0 there exists a projection E with $E_0 \leq E$ and $|\phi(E) - \alpha| < \epsilon$. Theorem: an operator A is quasidiagonal if and only if $\liminf_{E \rightarrow 1} \|AE - EA\| = 0$.

As for closure theorems, there are several. The most important one takes the word "closure" seriously: the set of quasidiagonal operators

is (norm) closed. There are also algebraic closure theorems; e.g., every polynomial in a quasideagonal operator, the adjoint of a quasideagonal operator, and countable direct sums of quasideagonal operators are again quasideagonal.

The most trivial inclusion theorem is that diagonal operators are quasideagonal; together with norm closure (and the spectral theorem) this yields the valuable inclusion theorem that normal operators are quasideagonal. Another trivial inclusion theorem is that operators of finite rank are quasideagonal; together with the characterization of compact operators as the limits of operators of finite rank [20, Problem 137] this yields the result that compact operators are quasideagonal. The latter result can be strengthened; in fact, if A is compact, then

$$\lim_{E \rightarrow 1} \|AE - EA\| = 0.$$

(That is: \liminf can be replaced by \lim .) This, in turn, implies that a compact operator will never interfere with the \liminf that establishes the quasideagonality of some other operator; precisely, if A is quasideagonal and C is compact, then $A + C$ is quasideagonal. Noteworthy special case: the sum of a normal operator and a compact one is always quasideagonal. Is the converse true? That is (**Exercise 2**): is every quasideagonal operator the sum of a normal operator and a compact one?

One of the facts that make quasideagonal operators interesting is that every quasideagonal operator is the sum of a block diagonal operator and a compact one. A slightly stronger statement is true: every quasideagonal operator has a matrix such that, for a suitable way of dividing it into finite blocks, the result of replacing the blocks on the diagonal by 0's is compact. To prove this, suppose that $\{E_n\}$ is an increasing sequence of projections of finite rank such that $E_n \rightarrow 1$ strongly and $\|AE_n - E_nA\| \rightarrow 0$. Choose a subsequence, if necessary, to justify the assumption that $\sum_n \|AE_n - E_nA\| < \infty$. Write D for the matrix formed of the diagonal blocks $(E_n - E_{n-1})A(E_n - E_{n-1})$ (where $E_0 = 0$) and write $C = A - D$. It is not obvious by inspection that D and C are operators, and it is even less obvious that C is compact, but that is what the proof is about to show. Indeed: put

$$C_n = E_{n+1}(AE_n - E_nA)E_n - E_n(AE_n - E_nA)E_{n+1}.$$

Clearly $\|C_n\| \leq 2\|AE_n - E_nA\|$, and therefore $\sum_n C_n$ converges in the norm to a compact operator; an obvious computation shows that that operator is equal to C .

Weyl's theorem (the affirmative answer to Problem 4 for Hermitian operators) is an immediate consequence. Reason: if A is Hermitian, then so is each diagonal block in any matrix of A ; it follows that the block diagonal summand that the preceding result yields is, in fact, diagonal. The proof shows also what blocks the proof in the normal case; a diagonal block in a normal matrix may fail to be normal. For unitary operators J. A. Dyer (simplifying and generalizing an argument of I. D. Berg) proved that Weyl's theorem remains true; the trick is to represent a unitary operator as e^{iA} , with A Hermitian, and apply the Hermitian result to A .

Quasidiagonal operators are a special case of a concept that has received some attention in the literature [10], [12], [22]; the more general operators are called quasitriangular. By definition an operator A is *quasitriangular* in case there exists an increasing sequence $\{E_n\}$ of projections of finite rank such that $E_n \rightarrow 1$ strongly and such that $\|AE_n - E_nAE_n\| \rightarrow 0$. (Motivation: replace asymptotic reduction by asymptotic invariance. Note that the appropriate definition of a *triangular* operator requires the existence of a sequence $\{E_n\}$ of projections of finite rank such that $AE_n = E_nAE_n$.)

Most of the characterization, closure, and inclusion theorems stated above for quasidiagonal operators are true for quasitriangular operators also [22]; the proofs for quasidiagonal operators are easy adaptations of the proofs in [22]. A case in point is the characterization theorem: A is quasitriangular if and only if $\liminf_{E \rightarrow 1} \|AE - EAE\| = 0$. A notable exception is the closure theorem: the class of quasidiagonal operators is closed under the formation of adjoints. For quasitriangular operators that is not true. In order to support the latter (negative) statement, it is, of course, necessary to have a technique for proving that something is *not* quasitriangular. The tool in [22] is the assertion that if $A^*A = 1$ and $\ker A^* \neq 0$, then A is not quasitriangular. A slightly sharper version of this is true [12]: if $A^*A \geq 1$ and $\ker A^* \neq 0$, then A is not quasitriangular. Suppose, indeed, that $e \neq 0$ and $A^*e = 0$; let E_0 be the projection onto the (1-dimensional) span of e . If E is a projection of finite rank with $E_0 \leq E$, then the restriction of EA^*E to $\text{ran } E$ has e in its kernel; the finite-dimensionality of $\text{ran } E$ implies the existence of a unit vector f in $\text{ran } E$ such that $EAEf = 0$. It follows that

$$\|(AE - EAE)f\| = \|AEf\| = \|Af\| \geq \|f\|,$$

whence $\|AE - EAE\| \geq 1$; this proves that A cannot be quasitriangular. Very small consequence: since the adjoint of the unilateral

shift is quasitriangular (in fact, triangular), the adjoint of a quasitriangular operator can fail to be one.

For all that has been said so far it could be true that, for every operator A , either A or A^* is quasitriangular. A natural candidate for a counterexample is $U \oplus U^*$ (where U , of course, is the unilateral shift), but the candidate fails; $U \oplus U^*$ is, in fact, quasidiagonal. (Proof: it differs from the bilateral shift, which is normal, by an operator of rank 1.) **Exercise 3:** is there an operator A such that neither A nor A^* is quasitriangular?

The sum theorem for quasidiagonal operators (block diagonal plus compact) is not the specialization of a theorem in [22], but it might as well be; a slight complication of the proof serves to show that every quasitriangular operator is the sum of a triangular operator and a compact one.

Two further facts deserve at least brief mention. (i) Douglas and Percy [12] have proved that every operator with a finite spectrum is quasitriangular, and they conjectured that the same is true for operators with a countable spectrum. (ii) What happens if the \liminf in the definition of quasitriangularity is replaced by \lim ? Answer: a necessary and sufficient condition that $\lim_{E \rightarrow 1} \|AE - EAE\| = 0$ is that A be the sum of a compact operator and a scalar [11].

ANSWERS TO THE EXERCISES. (1) For every compact operator C there exists a separable reducing subspace whose orthogonal complement is included in $\ker C$. This fact is usually stated for compact *normal* operators only [20, Problem 133]; the general case follows by applying the usual statement to C^*C and forming the smallest subspace that reduces C and includes the separable part. Another application of the same technique shows that if $A = D + C$, where C is compact, then there exists a separable subspace that reduces both D and C and whose orthogonal complement is included in $\ker C$. If the space is non-separable and D is diagonal, then it follows that A has many eigenvectors. If, therefore, A has no point spectrum, then A admits no sum representation of the required kind.

(2) The operators of the form "normal plus compact" do not exhaust all quasidiagonal operators. For an example, consider the direct sum A of infinitely many copies of the operator on C^2 with matrix $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$; since A is block diagonal, it is quasidiagonal. Then commutator $A^*A - AA^*$ is the direct sum of infinitely many copies of $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, which is about as far as possible from being compact; if, on the other hand, N is normal and C is compact, then $(N+C)^*(N+C) - (N+C)(N+C)^*$ is compact.

(3) If U is the unilateral shift and $A = (U - i) \oplus (U^* + i)$, the

neither A nor A^* is quasitriangular. For the proof, consider $A+i = U \oplus (U^* + 2i)$ and note that it is bounded from below by 1 but its adjoint has a non-zero kernel. This answer to a question in [22] appears in [12].

5. DILATED

Problem 5. *Is every subnormal Toeplitz operator either analytic or normal?*

Toeplitz operators are the best known and analytically most important examples of compressions; they are compressions of the simplest “continuous” generalizations of diagonal matrices. If C is the unit circle ($\{z: |z|=1\}$) in the complex plane, endowed with normalized Lebesgue measure, then the functions e_n , defined by $e_n(z) = z^n$, $n = 0, \pm 1, \pm 2, \dots$, form an orthonormal basis for $L^2 (= L^2(C))$; let H^2 be the subspace of L^2 spanned by $\{e_0, e_1, e_2, \dots\}$. If ϕ is a bounded measurable function on C , then the *Toeplitz operator* T_ϕ induced by ϕ is the compression to H^2 of the multiplication operator induced by ϕ on L^2 . Explicitly, if P is the projection with domain L^2 and range H^2 , then $T_\phi f = P(\phi \cdot f)$ for each f in H^2 . The Toeplitz operator T_ϕ is *analytic* in case ϕ itself belongs to H^2 . For the basic facts about Toeplitz operators see [6].

One of the first things noticed about Toeplitz operators was that their matrices (with respect to the basis $\{e_0, e_1, e_2, \dots\}$) have an exceptionally simple form. Indeed, if $i, j = 0, 1, 2, \dots$, then

$$(T_\phi e_j, e_i) = (P(\phi \cdot e_j), e_i) = (\phi \cdot e_j, e_i);$$

since the multiplication operator induced by e_i is the unilateral shift U , and hence isometric, it follows that

$$\begin{aligned} (T_\phi e_j, e_i) &= (U(\phi \cdot e_j), Ue_i) = (\phi \cdot e_{j+1}, e_{i+1}) \\ &= (P(\phi \cdot e_{j+1}), e_{i+1}) = (T_\phi e_{j+1}, e_{i+1}). \end{aligned}$$

The conclusion is that the diagonals of the matrix of T_ϕ are constants; the entries in row number 0 and column number 0 determine all others. Those entries are, in fact, the Fourier coefficients of ϕ (the ones with non-positive and non-negative indices respectively). From this observation it is easy to prove that not only does every Toeplitz operator have a Toeplitz matrix (i.e., one with constant diagonals), but the converse is true also; for each (bounded) Toeplitz matrix, the ϕ that induces it is given by the Fourier series whose coefficients appear in row 0 and column 0. (An algebraic way of expressing the same conclusion is this: an operator A on H^2 is a Toeplitz operator

if and only if $U^*AU = A$.)

If a Toeplitz operator T_ϕ is analytic (which, by the way, happens if and only if its matrix is lower triangular), then $\phi \cdot f$ is in H^2 whenever f is in H^2 ; the projection P is not needed. Consequence: if T_ϕ is analytic, then it is a part of the multiplication operator induced by ϕ on L^2 , and hence it is subnormal.

An easy way to get a normal Toeplitz operator is to take a Hermitian one (i.e., to take ϕ real). It is somewhat surprising but true [6] that every normal Toeplitz operator has the form $\alpha + \beta T_\phi$, where ϕ is real.

These two trivial ways of getting subnormal Toeplitz operators (analytic and normal) are the only ones known; Problem 5 asks if they are the only ones that exist. There are many other problems about Toeplitz operators, big and small, solved and unsolved; here are two small ones, one old, one new. **Exercise 1:** is there a nonzero compact Toeplitz operator? **Exercise 2:** is every operator the sum of a Toeplitz operator and a compact one?

The opposite of compression is dilation. If, that is, A and B are operators on Hilbert spaces H and K respectively, where $H \subset K$, if P is the projection from K onto H , and if $Af = PBf$ for each f in H , then A is the compression of B to H , and B is a *dilation* of A to K . One reason for the analytic importance of Toeplitz operators is that they have easily treatable dilations.

Compressions and dilations can be usefully described in terms of matrices. If K is decomposed into H and H^\perp , and, correspondingly, operators on K are written as two-by-two matrices (whose entries are operators on H , operators on H^\perp , and transformations between the two), then a necessary and sufficient condition that B be a dilation of A is that the matrix of B have the form $\begin{pmatrix} A & X \\ Y & Z \end{pmatrix}$.

The purpose of dilation theory in general is to get information about difficult operators by finding their easy dilations. The program has been spectacularly successful. Unitary operators are among the easiest to deal with, and it turns out that, except for an easily adjusted normalization, every operator has a unitary dilation. Some normalization is clearly necessary: if B is unitary, then $\|Bf\| = \|f\|$ for every vector f , and it follows that $\|Af\| \leq \|f\|$ for every vector f ; in other words, if A has a unitary dilation, then A must be a contraction. That much normalization is sufficient: *every contraction has a unitary dilation*. The proof is explicit; given A on H , let B on $H \oplus H$ be defined by the matrix

$$\begin{pmatrix} A & S \\ T & -A^* \end{pmatrix},$$

where S and T are the positive square roots of $1 - AA^*$ and $1 - A^*A$ respectively. The verification that B is unitary is trivial but not obvious [20, Problem 177].

Projections, that is to say perpendicular projections, or projections onto a subspace along its orthogonal complement, are the right things to look at in Hilbert space, but, from a geometrical point of view, skew projections are just as good. In algebraic language skew projections are idempotent operators; perpendicular projections are the ones among them that happen to have unit norm. It is this normalization that forces every compression of a unitary operator to be a contraction; what happens if skew projections are allowed? **Exercise 3:** is it true that for every operator A on a Hilbert space H there exists a unitary operator B on a larger Hilbert space K and a skew projection P from K onto H such that $Af = PBf$ for each f in H ?

What makes the theory of unitary dilations useful is Nagy's theorem [20, Problem 178] on power dilations. The assertion is that for every contraction A on H there exists a unitary operator B on a larger K such that B^n is a dilation of A^n simultaneously for every positive integer n .

One of the most spectacular applications of unitary dilation theory is the proof of von Neumann's beautiful and powerful analytic theorem about contractions [20, Problem 180]. The assertion is that if $\|A\| \leq 1$ and if p is a polynomial such that $|p(z)| \leq 1$ whenever $|z| = 1$, then $\|p(A)\| \leq 1$. For the proof, let B be a unitary power dilation of A to, say, K , let P be the (perpendicular) projection from K onto H , and note that, for each f in H ,

$$\|p(A)f\| = \|Pp(U)f\| \leq \|p(U)\| \cdot \|f\| \leq \|f\|.$$

The crucial inequality, the second one, is an elementary consequence of the functional calculus for normal operators; all that is needed is the observation that, since U is unitary, the spectrum of U is on the unit circle, where $|p|$ is, by assumption, bounded by 1.

The success of this "one-variable" theory made it tempting to look for possible "several-variable" extensions. The first step was taken by Andô [1]; he proved that if A_1 and A_2 are commutative contractions on H , then there exist commutative unitary operators B_1 and B_2 on a larger K such that $B_1^m B_2^n$ is a dilation of $A_1^m A_2^n$ for every pair of positive integers m and n .

Much to everyone's surprise, Andô's theorem turned out to be a characterization of the integer 2; Parrott [37] proved that for three or more contractions the corresponding statement is false. In fact

even something much weaker is false: there exist three commutative contractions A_0, A_1, A_2 such that if B_0, B_1, B_2 respectively are isometric dilations (not necessarily unitary and not necessarily power dilations), then the set $\{B_0, B_1, B_2\}$ cannot be commutative. The following presentation of Parrott's idea is due to Chandler Davis.

Suppose that C_0, C_1, C_2 are isometries on some Hilbert space; write

$$A_i = \begin{pmatrix} 0 & 0 \\ C_i & 0 \end{pmatrix}.$$

It is clear that the A 's are contractions, and, since the product of any two of them is 0, it is clear that they commute. If $B_0, B_1,$ and B_2 are dilations of $A_0, A_1,$ and A_2 respectively, to the same enlarged space, then they can be written in the form

$$B_i = \begin{pmatrix} 0 & 0 & * \\ C_i & 0 & * \\ D_i & E_i & * \end{pmatrix},$$

where the entries indicated by * need not be known. If each B_i is an isometry, so that

$$\|B_i \langle f, 0, 0 \rangle\|^2 = \|\langle f, 0, 0 \rangle\|^2$$

for all f , then it follows that

$$\|C_i f\|^2 + \|D_i f\|^2 = \|f\|^2$$

for all f ; since C_i is an isometry, it follows that D_i must be 0. A similar glance at $\langle 0, f, 0 \rangle$ shows that E_i must be an isometry.

It is to be proved that the C 's can be chosen so that the B 's do not commute. Assertion: if $C_0 = 1$ and C_1 and C_2 do not commute (so that the Hilbert space they act on has dimension at least 2), then the desired result is achieved. Suppose, on the contrary, that the B 's do commute. Since the entry in row 3 and column 1 of the product $B_i B_j$ is $E_i C_j$, it follows that

$$E_1 C_2 = E_2 C_1, \quad E_0 C_1 = E_1, \quad \text{and} \quad E_0 C_2 = E_2.$$

This implies that

$$E_0 C_1 C_2 = E_1 C_2 = E_2 C_1 = E_0 C_2 C_1,$$

and therefore, since E_0 is an isometry, that $C_1 C_2 = C_2 C_1$; the contradiction has arrived and the proof is complete.

ANSWERS TO THE EXERCISES. (1) The only compact Toeplitz operator is 0. Indeed, if ϕ is a bounded measurable function on C , and if n and $n+k$ are non-negative integers, then $(\phi, e_k) = (T_\phi e_n, e_{n+k})$. If

T_ϕ is compact, then $\|T_\phi e_n\| \rightarrow 0$ (since $e_n \rightarrow 0$ weakly); it follows that $(\phi, e_k) = 0$ for all k (positive, negative, or zero), and hence that $\phi = 0$.

(2) If $A = T + C$, where T is a Toeplitz operator and C is compact, then $U^*AU = T + U^*CU$, where U is the unilateral shift, and therefore $U^*AU = A + K$, where K is compact. This implies that not every A has the form $T + C$. For an example, let A be the projection onto the span of $\{e_0, e_2, e_4, \dots\}$, so that $Ae_n = e_n$ when n is even and $Ae_n = 0$ when n is odd. If n is even, then $U^*AUe_n = 0$; if n is odd, then $U^*AUe_n = e_n$. Conclusion: $A - U^*AU$ is not compact.

(3) If skew projections are allowed, then every operator has a unitary dilation. Indeed, given A on H , write $K = H \oplus H$, identify H , as usual, with $H \oplus 0$, let B on K be given by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and let P the skew projection $\begin{pmatrix} 1 & A \\ 0 & 0 \end{pmatrix}$ from K onto H . Since PB has the matrix $\begin{pmatrix} A & 1 \\ 0 & 0 \end{pmatrix}$, the result follows. A generalization of this technique can be used to produce, for each positive integer n , a unitary operator B such that B^k is a skew dilation of A^k for each $k = 1, \dots, n$. For the typical case $n = 2$, write $K = H \oplus H \oplus H$,

$$B = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} 1 & A & A^2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

These results are due to L. J. Wallen and J. S. Johnson.

6. SIMILAR

Problem 6. *Is every polynomially bounded operator similar to a contraction?*

There is a sense in which questions of similarity are not "natural" in Hilbert space, but the few elegant results and interesting examples already known make it hard to resist the temptation to continue looking for an adequate general theory.

As is often true, the easiest operators to begin with are the unitary ones. If U is unitary and $A = S^{-1}US$, then, of course, $A^n = S^{-1}U^nS$, and therefore $\|A^n\| \leq c$ for every positive integer n (where $c = \|S^{-1}\| \cdot \|S\|$); in other words, each operator similar to a unitary operator is *power bounded*. Since A is invertible, the inequality makes sense and is true for negative exponents also. One of the earliest non-trivial results about similarity is Nagy's converse [31]: if A is invertible and both A and A^{-1} are power bounded, then A is similar to a unitary operator. (Caution: small diagonal matrices show that the inverse of a power bounded invertible operator may fail to be power bounded.) This is a satisfactory state of affairs; for similarity

to a unitary operator there is a simple and usable necessary and sufficient condition.

Exercise 1: is every idempotent operator similar to a projection?

What happens if the invertibility part of Nagy's condition is dropped? In other words: is there a simply describable class of operators such that power boundedness is a necessary and sufficient condition for being similar to one of them? Superficially it might seem that the class of isometries is the answer (drop the invertibility condition from the definition of unitary operators), but that is not right. (A quick way to see that it is not right is to recall that on finite-dimensional spaces all isometries are unitary.) A reasonable second guess might lead to the question: is every power bounded operator similar to a contraction? For some classes of operators the answer is yes. It is, for instance, an easy exercise in analysis to prove that every power bounded weighted shift is similar to a contraction.

It is clear that every eigenvalue of a power bounded operator must have modulus less than or equal to 1. A finite Jordan block with eigenvalue of modulus 1 (and size greater than 1) is not power bounded; if the eigenvalue has modulus less than 1, then the block is power bounded and it is easily seen to be similar to a contraction. These considerations imply that in the finite-dimensional case every power bounded operator is similar to a contraction, and it follows, of course, that the same is true for operators of finite rank on spaces of arbitrary dimension. Nagy [32] extended the result to compact operators. For not necessarily compact operators the answer is no; a counterexample was constructed by Foguel [15], [19]. **Exercise 2:** every power bounded subnormal operator is similar to a contraction.

At this point it is reasonable to begin to think that the whole approach is wrong. The preceding two paragraphs encouraged a search for a class of operators whose conjugates (i.e., transforms by similarities) were characterized by a prescribed condition (power boundedness). It might be better to prescribe an interesting class of operators (e.g., contractions) and search for a characterization of their conjugates.

With hindsight it might be said that two-sided power boundedness is a natural characterization of the conjugates of unitary operators in that it is a condition on the *group* of invertible operators. To characterize the conjugates of contractions (which may, of course, fail to be invertible) it seems reasonable to put a condition on operators considered as elements of an *algebra*. To find a good candidate for a characteristic condition, invert the problem: what can be said

about the elements of the algebra generated by an operator similar to a contraction? In other words: if C is a contraction, $A = S^{-1}CS$, and p is a polynomial, what can be said about $p(A)$? Since $p(A) = S^{-1}p(C)S$, and since (by the von Neumann theorem mentioned in §5) $\|p(C)\| \leq \|p\|_\infty$ ($= \sup\{|p(z)| : |z| \leq 1\}$), one answer is the inequality $\|p(A)\| \leq c\|p\|_\infty$ (where $c = \|S^{-1}\| \cdot \|S\|$). Operators satisfying this condition for all p are called *polynomially bounded*; Problem 6 asks whether they are exactly the conjugates of contractions.

Foguel's work provides an operator that is power bounded but not similar to a contraction; if that operator were polynomially bounded also, Problem 6 would be solved (in the negative). Lebow [30] examined Foguel's operator and proved that it is not polynomially bounded; Problem 6 is still unsolved.

Some interesting special cases of Problem 6 are solved.

(i) If $r(A) < 1$ (r is the spectral radius), then A is similar to a contraction. The result is due to Rota [20, Problem 122]. It takes more than a casual glance to see that it is a special case of Problem 6. The easiest proof uses the result itself. That is: if $r(A) < 1$, then A is polynomially bounded because it is similar to a contraction. There is an elegant quantitative version of Rota's theorem; it asserts that $r(A)$ is always equal to the infimum of the norms of all conjugates of A .

(ii) If A^2 is a contraction, then A is similar to a contraction. (This was proved in collaboration by A. L. Shields, L. J. Wallen, and myself; simultaneously and independently it was proved by J. P. Williams.) For the proof, define a new inner product by $((f, g)) = (f, g) + (Af, Ag)$. Since $(f, f) \leq ((f, f)) \leq (1 + \|A\|^2)(f, f)$, the new inner product is equivalent to the old one. The assumption that A^2 is a contraction with respect to the old inner product implies that $((Af, Af)) \leq ((f, f))$. This means that A is a contraction with respect to the new inner product, and hence that A was similar to a contraction in the first place. For an alternative approach to essentially the same proof, let S be the positive square root of $1 + A^*A$ and verify that SAS^{-1} is a contraction. Generalization: if A^2 is similar to a contraction, then so is A . Proof: if $A^2 = S^{-1}CS$, where C is a contraction, then write $B = SAS^{-1}$. Since $B^2 = SA^2S^{-1} = C$, the result just proved implies that B is similar to a contraction and hence so is A . Both the result and the generalization extend automatically to A^n in place of A^2 .

What does (ii) have to do with Problem 6? Answer: if A^2 is a contraction, then A is polynomially bounded. For the proof, given a polynomial p , express it in terms of its even and odd parts, i.e., let

q^+ and q^- be the polynomials defined by

$$q^+(z^2) = \frac{1}{2} (p(z) + p(-z)) \quad \text{and} \quad q^-(z^2) = \frac{1}{2z} (p(z) - p(-z)),$$

so that $p(z) = q^+(z^2) + zq^-(z^2)$. Since A^2 is a contraction, the von Neumann theorem on contractions applies and proves that

$$\|q^\pm(A^2)\| \leq \|q^\pm\|_\infty.$$

Consequence:

$$\|p(A)\| \leq \|q^+\|_\infty + \|A\| \cdot \|q^-\|_\infty \leq (1 + \|A\|) \|p\|_\infty.$$

(The last inequality follows from the definitions of q^+ and q^- .) What the argument really proves is that if A^2 is polynomially bounded, then so is A . The extension to A^n is routine.

A final small comment about (ii) must be made before the subject can be abandoned: it is not the general case. That is: there exists a polynomially bounded operator A such that no power of A is a contraction. Example (A. L. Shields): a unilateral weighted shift with weights $1 + 1/2^n$, $n = 0, 1, 2, \dots$. The norm of A^n is $\prod_{k=0}^{n-1} (1 + 1/2^k)$, which is always strictly greater than 1; to prove that A is polynomially bounded, apply the general theory of similarity for weighted shifts [20, Problem 76] and infer that A is similar to the (unweighted) unilateral shift.

(iii) **Exercise 3:** every polynomially bounded operator is the limit (in the norm) of operators that are similar to contractions. That is: granted that equality is not yet attainable, at least asymptotic equality can be achieved.

Unitary operators are too special and contractions are too general, there are several interesting classes between. What, for instance, can be said about the conjugates of isometries? Answer: A is similar to an isometry if and only if the powers of A are uniformly bounded from both above and below. To say that the powers of A are uniformly bounded from above means, of course, that the supremum of the set of numbers

$$\{\|A^n f\| : n = 0, 1, 2, \dots, \|f\| = 1\}$$

is finite (i.e., that A is power bounded); the corresponding condition from below is that the infimum of the same set of numbers is (strictly) positive. The proof of the assertion is a routine imitation of Nagy's proof for the unitary case.

Although this characterization of the conjugates of isometries looks as if it might be awkward to apply, it does have some pleasant

consequences. These consequences are minor, and each one can be proved directly, but it is more efficient to derive them from a common source. Sample: every square root of an isometry is similar to an isometry. (Isometries can have non-isometric square roots. A trivial example is $\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$; a non-trivial infinite-dimensional example is a unilateral weighted shift with alternating weights 2 and $\frac{1}{2}$.) For the proof, observe that if A^2 is an isometry, then A has a left inverse and, consequently, A is bounded from below; since A^n is alternately an isometry and A times an isometry, the desired result follows. The extension to higher powers is obvious.

Another sample: if a normal operator A is similar to an isometry, then A is an isometry (and hence A is unitary). Proof: use the power characterization of the conjugates of isometries and the spectral theorem. Here is a small surprise: for subnormal operators the conclusion is false; a counterexample was constructed by Sarason [20, Problem 156].

The problem of characterizing the conjugates of partial isometries has apparently not been studied.

ANSWERS TO THE EXERCISES. (1) If $E^2 = E$, form $A = 2E - 1$, note that $A^n = 1$ or A according as n is even or odd, and conclude that A is similar to a unitary operator U . Since $U^2 = 1$ (because $A^2 = 1$), it follows that U is Hermitian, and hence so is $\frac{1}{2}(U+1)$. Conclusion: $\frac{1}{2}(U+1)$ is a projection similar to E ($=\frac{1}{2}(A+1)$). This is an elegant application of Nagy's similarity theorem, but it has a flaw; the trouble is that the result is easier to prove directly. Indeed: decompose the space into $\text{ran} E$ ($=\ker(1-E)$) and its orthogonal complement; the matrix corresponding to E then takes the form $\begin{pmatrix} 1 & A \\ 0 & 0 \end{pmatrix}$. If $S = \begin{pmatrix} 1 & A \\ 0 & 1 \end{pmatrix}$, then S is invertible and $S^{-1}ES = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. No harm done: the conclusion is worth mentioning in any discussion of similarity. The second proof is due to J. G. Stampfli.

(2) Suppose that A is subnormal and power bounded, and let B be its minimal normal extension. If Λ denotes spectrum, then $\Lambda(B^n) = (\Lambda(B))^n$ (spectral mapping theorem) $\subset (\Lambda(A))^n$ (spectral inclusion theorem for subnormal operators) $= \Lambda(A^n)$. If r denotes spectral radius, then it follows that

$$\|B^n\| = r(B^n) \leq r(A^n) \leq \|A^n\|.$$

Consequence: B is power bounded. Since the only power bounded complex numbers are the ones of modulus less than or equal to 1, the spectral theorem implies that B must be a contraction, and it follows that so is A . Conclusion: a power bounded subnormal operator is not only similar to a contraction, but actually is one.

(3) Every operator A with $r(A) \leq 1$ is the limit of operators similar to contractions. (Note that if A is polynomially bounded, then it is power bounded, and therefore $r(A) \leq 1$.) Indeed: if $A_n = (1 - 1/n)A$, then $r(A_n) < 1$, so that Rota's theorem (i) applies, and, clearly, $A_n \rightarrow A$. Since there are easy examples of operators A with $r(A) \leq 1$ that are not polynomially bounded (a hard example is the Foguel operator), this shows incidentally that the set of polynomially bounded operators is not closed.

7. NILPOTENT

Problem 7. *Is every quasinilpotent operator the norm limit of nilpotent ones?*

The similarity theory of linear transformations on finite-dimensional spaces reduces to that of nilpotent ones, and the theory of nilpotent transformations turns out to be algebraically tractable. In the infinite-dimensional case nilpotence no longer plays the same central role, and the theory of the appropriately generalized concept is both algebraically and topologically refractory.

If A is nilpotent, say $A^n = 0$, then, by the spectral mapping theorem, the spectrum of A consists of 0 alone. In the finite-dimensional case the converse is true; in the infinite-dimensional case it is not. The standard example [20, Problem 80] is any unilateral weighted shift whose weights tend to 0. If A is such a shift, then A is not nilpotent (in fact if $A^n f = 0$ for some positive integer n , then $f = 0$). The operator A is, however, *quasinilpotent* in the sense that $\|A^n\|^{1/n} \rightarrow 0$. Since $\lim_n \|A^n\|^{1/n}$ is always equal to the spectral radius of A [20, Problem 74], the spectrum of a quasinilpotent operator is the singleton $\{0\}$.

Problem 7 is important because it calls for the discovery of new techniques; the question itself is "wrong". What is wrong is that the condition whose necessity is in question is already known to be *not* sufficient; a limit of nilpotent operators may fail to be quasinilpotent. The basic example is due to Kakutani [20, Problem 87]. It is a unilateral weighted shift A whose weight sequence $\{\alpha_0, \alpha_1, \alpha_2, \dots\}$ is obtained as follows: every second α is equal to 1, (i.e., $\alpha_0 = 1, \alpha_2 = 1, \alpha_4 = 1, \dots$); every second one of the remaining α 's is equal to $\frac{1}{2}$ ($\alpha_1 = \frac{1}{2}, \alpha_3 = \frac{1}{2}, \alpha_5 = \frac{1}{2}, \dots$); every second one of the remaining α 's is $\frac{1}{4}$, etc., etc. If A_n is the "weighted shift" obtained from A by replacing $1/2^n$ by 0, then A_n is nilpotent of index 2^{n+1} and $A_n \rightarrow A$. The spectral radius of A can be obtained by a mildly onerous computation; the result is that $r(A) = 1$. Problem 7 is specific but mis-

directed. The more honest formulation has to be more vague; it should be something like “what is the closure of the set of nilpotent operators?”. As long as the subject has been raised: what is the closure of the set of quasinilpotent operators?

There is an elegant partial solution of Problem 7 due to R. G. Douglas: every compact quasinilpotent operator is the limit of nilpotent ones. The main tool in the proof is the upper semicontinuity of the spectrum [20, Problem 86]. According to that result if A is quasinilpotent, then corresponding to each positive integer n there exists a positive number ϵ_n such that every operator within ϵ_n of A has a spectrum smaller than $1/n$. Assume, with no loss of generality, that $\epsilon_n \rightarrow 0$. If now A is compact (as well as quasinilpotent), then for each n there exists an operator B_n of finite rank such that $\|A - B_n\| < \epsilon_n$. Since all the eigenvalues of B_n are less than $1/n$ in modulus, it follows (triangular form) that B_n has the form $C_n + D_n$, where C_n is nilpotent and D_n is diagonal, with $\|D_n\| < 1/n$. Conclusion: $C_n \rightarrow A$.

Exercise 1: is every quasinilpotent operator either nilpotent or compact? (It is clear from the context that the answer must be no; otherwise Problem 7 would be solved in the affirmative.)

The general structure of quasinilpotent operators is not known at all. There are some large questions, and there are a few fragmentary theorems. Two questions have to do with the invariant subspace problem. (i) Does every quasinilpotent operator have nontrivial invariant subspaces? What makes the question interesting is not only that the answer is not known, but also that, possibly, it is equivalent to the general invariant subspace problem. No convincing reduction is known, but meditation on the finite-dimensional case, and on the connectedness of the spectrum of a transitive operator, has led some people sometimes to hope that one exists. (ii) An operator is called *unicellular* in case all its invariant subspaces are comparable (i.e., if M and N are invariant subspaces, then $M \subset N$ or $N \subset M$). The terminology comes from the finite-dimensional case: there the phenomenon occurs if and only if the Jordan form has just one block (cell). Since one block has just one eigenvalue, it is natural to raise the general question [44]: is every unicellular operator equal to the sum of a quasinilpotent operator and a scalar? For compact operators the answer is yes.

What follows is a discussion of some of the fragmentary theorems about quasinilpotence that are known. The first two are more fragmentary than the rest; their discussion is left to the reader.

Exercise 2: An analytic quasinilpotent operator is nilpotent. (Here A is called “analytic” in case there exists a function f analytic

in a neighborhood of 0 such that $f(A) = 0$.)

Exercise 3: If U is the unilateral shift and A is quasinilpotent, then $\|U - A\| \geq 1$.

Perhaps the most important place where quasinilpotent operators enter functional analysis is in Dunford's theory of spectral operators [13]. The concept belongs to Banach spaces, but in Hilbert spaces, by virtue of an elegant theorem of Wermer [52], it has an especially simple characterization. An operator A (on a Hilbert space) is *spectral* if and only if it can be represented as a sum, $A = S + Q$, where S is similar to a normal operator, Q is quasinilpotent, and S and Q commute; the representation is unique. The theory of the classical Jordan form shows that on a finite-dimensional Hilbert space every operator is spectral.

A basic result about spectral operators is that if $A = S + Q$, as above, then the spectra of A and S are the same. To prove this, observe first that if T is an operator that commutes with Q , then QT is quasinilpotent (because $\|(QT)^n\| = \|Q^n T^n\|$), and it follows that if, moreover, T is invertible, then $T + Q$ is invertible (because $T + Q = T(1 + T^{-1}Q)$). Suppose now that λ is not in the spectrum of S and write $T = S - \lambda$; by what was just said, $T + Q$ is invertible, so that λ is not in the spectrum of $S + Q$. Since S and $S + Q$ play symmetric roles, it follows that, indeed, their spectra are the same. This theorem has an ingenious generalization due to Colojoară and Foiaş [8]. They define two operators (which might as well be called A and S) to be *quasinilpotent equivalent* in case

$$\left\| \sum_{k=0}^n (-1)^k \binom{n}{k} A^k S^{n-k} \right\|^{1/n} \rightarrow 0$$

and in case the same is true with A and S interchanged. The point is that if A and S happen to commute (which happens exactly when $A - S$ and S commute), then the sum inside the norm is equal to $(A - S)^n$. In the commutative case quasinilpotent equivalence reduces to having a quasinilpotent difference; the non-commutative case is a proper generalization. The Colojoară-Foiaş theorem is that any two operators that are quasinilpotent equivalent have equal spectra.

If the imaginary part of a quasinilpotent operator is compact, does it follow that the operator itself is compact? The answer is yes, and the result plays an important role in the theory of invariant subspaces. The fact was discovered almost simultaneously by Ringrose [42] and Schwartz [47]. Schwartz's proof is sophisticated but short; it goes as follows. Consider the algebra \mathfrak{B} of all operators, the

ideal \mathcal{C} of compact operators, the quotient algebra \mathfrak{B}/\mathcal{C} , and the canonical homomorphism π from \mathfrak{B} to \mathfrak{B}/\mathcal{C} . If Q is quasinilpotent (in \mathfrak{B}), then $\pi(Q)$ is quasinilpotent (in \mathfrak{B}/\mathcal{C}); if $\text{Im } Q (= (1/2i)(Q - Q^*))$ is compact, then $\text{Im } \pi(Q) = 0$. Since \mathfrak{B}/\mathcal{C} is representable as an operator algebra, and since the image of Q in such a representation has now been proved to be quasinilpotent and Hermitian, it follows that $\pi(Q) = 0$, so that Q is compact.

A nilpotent operator is *locally nilpotent* also; that is, if $A^n = 0$, then $A^n f = 0$ for each f . The converse is true: if for each f there is an $n (= n(f))$ such that $A^n f = 0$, then there is an n (independent of f) such that $A^n = 0$. Indeed, if A is locally nilpotent, then the space is the union of the subspaces $\ker A^n$, $n = 1, 2, 3, \dots$. The Baire category theorem implies that $\ker A^n$ has a non-empty interior for at least one n , and a subspace with a non-empty interior equals the whole space. Note: local nilpotence on a dense set is not enough to guarantee nilpotence.

The preceding paragraph extends elegantly to quasinilpotence, as follows. A quasinilpotent operator is *locally quasinilpotent* also; that is, if $\|A^n\|^{1/n} \rightarrow 0$, then $\|A^n f\|^{1/n} \rightarrow 0$ for each f . To prove the converse, observe that if $\|A^n f\| < \epsilon^n$ whenever n is sufficiently large (for each fixed f and each fixed positive number ϵ), then $\{\|A^n f/\epsilon^n\| : n = 1, 2, 3, \dots\}$ is bounded. The principle of uniform boundedness [20, Problem 40] implies that $\{\|A^n/\epsilon^n\| : n = 1, 2, 3, \dots\}$ is bounded, so that there exists a constant c such that $\|A^n\| \leq c\epsilon^n$ for all n and all ϵ . Conclusion: $\|A^n\|^{1/n} \rightarrow 0$, as promised. This smooth proof is due to Colojoară and Foiaş [9]. Local quasinilpotence on a dense set is not enough to guarantee quasinilpotence. Indeed, if U is the unilateral shift, and if f is a polynomial (i.e., a finite linear combination of the basis vectors being shifted), then $U^n f = 0$ for n sufficiently large; U^* , however, is obviously not quasinilpotent.

ANSWERS TO THE EXERCISES. (1) Begin with a quasinilpotent operator A with $\ker A = 0$, and put $B = \begin{pmatrix} 0 & A \\ 1 & 0 \end{pmatrix}$. It follows that $\ker B = 0$, and hence that B is not nilpotent. Since $B^2 = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}$, it follows that B^2 is quasinilpotent, and therefore so is B . A simple special case is a unilateral weighted shift with every other weight equal to 1 and the in-between ones tending to 0. For another example, start with an operator that is quasinilpotent and not nilpotent, but possibly compact, and form the direct sum of infinitely many copies of it.

(2) Write $f(z) = z^n g(z)$, where g is analytic in a neighborhood of 0 and $g(0) \neq 0$. The function g has an analytic reciprocal; i.e., there exists a function h analytic in a neighborhood of 0 such that $g(z)h(z) = 1$. It follows that $0 = f(A) = f(A)h(A) = A^n g(A)h(A) = A^n$. (This

proof is due to D. E. Sarason.)

(3) If $\|U - A\| < 1$, then $\|1 - U^*A\| < 1$, so that U^*A is invertible, and therefore A is left invertible. But if $BA = 1$, then $B^n A^n = 1$ (discard BA 's from the middle), and therefore $1 \leq \|B^n\|^{1/n} \|A^n\|^{1/n}$ for all n . It follows that $1 \leq r(B)r(A)$, which cannot happen when A is quasinilpotent. Corollary: the set of quasinilpotent operators is not dense. As far as the corollary is concerned, however, Exercise 3 is not needed. For that purpose it is enough to note that if A is quasinilpotent, then $\|1 - A\| \geq 1$, and that is obvious: a quasinilpotent operator surely cannot be invertible. The corollary should be contrasted with the result that the set of all nilpotent operators of index 2 is strongly dense [20, Problem 91]. The statement of Exercise 3 is due to R. A. Hirschfeld.

8. REDUCIBLE

Problem 8. *Is every operator the norm limit of reducible ones?*

Recall that a subspace M of a Hilbert space H *reduces* an operator A on H in case both M and M^\perp are invariant under A ; equivalently, M reduces A in case the projection with range M commutes with A . The operator A is reducible in case it has a non-trivial reducing subspace.

On a finite-dimensional Hilbert space the answer to the question is no: the set \mathcal{R} of reducible operators is closed. Suppose, indeed, that A_n is reducible and $A_n \rightarrow A$, and, for each n , let P_n be a non-trivial projection that commutes with A_n . Finite-dimensionality implies the compactness of the unit ball in the space of operators. There is, therefore, no loss of generality in assuming that $P_n \rightarrow P$, where, of course, P is a projection, and, clearly, $AP = PA$. If $\dim H = k$, then $1 \leq \text{trace } P_n \leq k - 1$; since trace is continuous, it follows that $P \neq 0, 1$. This proves that \mathcal{R} is closed; since not every operator is reducible (witness $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$), it follows that \mathcal{R} cannot be dense.

Exercise 1: on a non-separable Hilbert space every operator is reducible.

In view of the preceding comments, the solution of Problem 8 is negative for small spaces and trivially affirmative for large ones. The scope of the problem has been reduced to medium-sized Hilbert spaces (infinite-dimensional but separable), and there it is unsolved. Irreducible operators on such spaces certainly do exist; the unilateral shift is one of them. The unilateral shift, however, does not yield a negative solution of Problem 8; it can be approximated by reducible operators.

Another pertinent comment concerns bilateral weighted shifts. Such a shift is reducible if and only if its weight sequence is periodic [20, Problem 129]. If the weight sequence is not periodic but is the uniform limit of periodic ones, then, of course, the corresponding shift is the norm limit of reducible ones. If, however, the weight sequence is not the uniform limit of periodic ones (if, for instance, all the weights are 1 except the one with index 0, which is 2), then it does not follow that the shift is not reducibly approximable; all that follows is that the obvious approximation breaks down. The facts are not known.

Irreducible operators not only exist, they exist in profusion: the set \mathfrak{I} of irreducible operators (on a separable Hilbert space) is dense. To prove this, consider an arbitrary operator A and write $A = B + iC$ with B and C Hermitian. Represent B as a multiplication on L^2 over a finite measure space. (This is one place where separability comes in.) The (real) multiplier can be uniformly approximated by (real) simple functions. Multiplication by a real simple function is the direct sum of a finite set of real scalars, and consequently it is a diagonal operator; a Hermitian diagonal operator can obviously be approximated by one with no repeated eigenvalues. Call such an approximant B_0 , and consider the matrix of C with respect to a basis formed by the eigenvectors of B_0 . That matrix might have some entries equal to 0, but in any event it can be approximated by a (Hermitian) matrix with all non-zero entries; let C_0 be the operator corresponding to such a matrix. The operator $A_0 = B_0 + iC_0$ approximates A ; it remains to prove that A_0 is irreducible. If a subspace M reduces A_0 , then M is invariant under both B_0 and C_0 . A subspace invariant under B_0 is spanned by the subset of the eigenvectors of B_0 that it contains; this is a consequence of the distinctness of the eigenvalues of B_0 and is proved by a standard and elementary computation. Such a subspace, however, cannot be invariant under C_0 , unless it is either 0 or the whole space; this is a consequence of the non-zerosness of the matrix entries of C_0 . Conclusion: A_0 is irreducible, and the proof of the density of \mathfrak{I} is complete. The first appearance of the theorem is in [21]; the proof here presented is due to Radjavi and Rosenthal [39].

The proof works for all separable spaces, and, in particular, for the finite-dimensional ones. Since in the finite-dimensional case \mathfrak{I} is open (because its complement \mathfrak{R} is closed), and since the complement of a dense open set is nowhere dense, it follows that in the finite-dimensional case \mathfrak{R} is, in the sense of topological size, very small indeed. It is of interest to note that for separable spaces \mathfrak{R} is always topologically small (meager, set of the first category). Explicitly: if

the space is separable, then \mathfrak{R} is an F_σ [21].

The proof that \mathfrak{R} is an F_σ is tricky. Let \mathcal{O} be the set of all those Hermitian operators P for which $0 \leq P \leq 1$. Recall that \mathcal{O} is exactly the weak closure of the set of projections. Let \mathcal{O}_0 be the subset of those elements of \mathcal{O} that are *not* scalars. Since \mathcal{O} is a weakly closed subset of the unit ball, it is weakly compact, and hence the weak topology for \mathcal{O} is metrizable. Since the set of scalars is weakly closed, it follows that \mathcal{O}_0 is weakly locally compact. Since the weak topology for \mathcal{O} has a countable base, the same is true for \mathcal{O}_0 , and therefore \mathcal{O}_0 is weakly σ -compact. Let $\mathcal{O}_1, \mathcal{O}_2, \dots$ be weakly compact subsets of \mathcal{O}_0 whose union is \mathcal{O}_0 , and, for each n , let $\hat{\mathcal{O}}_n$ be the set of all those operators A that commute with some element of \mathcal{O}_n . The spectral theorem implies that $\bigcup_{n=1}^{\infty} \hat{\mathcal{O}}_n = \mathfrak{R}$.

The proof can be completed by showing that each $\hat{\mathcal{O}}_n$ is norm closed. Suppose therefore that A^k is in $\hat{\mathcal{O}}_n$ and that $\|A_k - A\| \rightarrow 0$. For each k , find P_k in \mathcal{O}_n so that it commutes with A_k . Since \mathcal{O}_n is weakly compact and metrizable, there is no loss of generality in assuming that the sequence $\{P_k\}$ is weakly convergent to P , say. (This is the point where it is advantageous to consider all the operators in \mathcal{O} , and not just the projections; there is no guarantee that P is a projection even if the P_k 's are. Note that P is in \mathcal{O}_n , so that, in particular, P is not a scalar.)

Since $A_k \rightarrow A$ (norm) and $P_k \rightarrow P$ (weak), it follows that $A_k P_k \rightarrow AP$ and $P_k A_k \rightarrow PA$ (weak). (The proof of this continuity assertion is elementary; all that it needs is that the sequence $\{P_k\}$ is bounded.) Conclusion: A commutes with P , so that A is in $\hat{\mathcal{O}}_n$.

Category arguments are sometimes used for existence proofs. That is certainly not the point here; the assertion that \mathfrak{I} is not empty was obvious long before the proof that \mathfrak{I} is a dense G_δ . Consider, however, invariance instead of reducibility. That is, let \mathfrak{I} be the set of transitive operators; one way to try to prove that \mathfrak{I} is not empty might be to prove that it is (or includes) a dense G_δ . As it stands, this is doomed to failure; since every transitive operator is cyclic, \mathfrak{I} cannot be dense (cf. §1).

How far is \mathfrak{I} from being topologically large? Could it be that it is so small that its complement is (or includes) a dense G_δ ? The answer is not known, but the evidence is toward the affirmative: the complement of \mathfrak{I} is at least dense. A much stronger statement is true (**Exercise 2**): the set of all operators with an eigenvalue is dense.

There is not much more to the theory of irreducible operators, but there is a little; it would be good to have more. Here, in conclusion for now, are two known results. (i) The set \mathfrak{R}_1 of all operators (on a

separable infinite-dimensional Hilbert space) that have a reducing subspace of dimension 1 is not dense, but (**Exercise 3**): its closure contains every isometry. (ii) Every operator (on a separable Hilbert space) is the sum of two irreducible ones [14], [38].

ANSWERS TO THE EXERCISES. (1) If A is an operator on H and f is a non-zero vector in H , then the smallest subspace of H that contains f and reduces A is spanned by f together with the set of all vectors of the form $A_1 \cdots A_n f$, where n is an arbitrary positive integer and each A_j is either A or A^* . It follows that that (non-zero) subspace is separable; if H is not separable, then the construction has yielded a non-trivial subspace that reduces A .

(2) Given A , find an approximate eigenvector for A , use it as the first element of a basis, form the matrix of A , and then find an approximant by replacing by 0 all but the first entry in the first column. Conclusion: every operator can be approximated by operators with eigenvalues.

(3) It is to be proved that if U is an isometry, then U can be approximated by operators with reducing eigenvectors. The proof is similar to the one in (2). Let λ be an approximate eigenvalue of U with modulus 1. It follows, by definition, that corresponding to each positive number ϵ there exists a unit vector e such that $\|Ue - \lambda e\| < \epsilon$, and hence such that $\|U^*e - \lambda^*e\| = \|- \lambda^*U^*(Ue - \lambda e)\| < \epsilon$. Use e as the first element of a basis, form the matrix of U , and then find an approximant by replacing by 0 all but the first entry in both the first row and the first column.

9. REFLEXIVE

Problem 9. *Is every complete Boolean algebra reflexive?*

The invariant subspace problem asks whether the set of all invariant subspaces of an operator can consist of the two extremes only, but what people would really like to know is what the set of all invariant subspaces of an operator, or, for that matter, of any set of operators can look like. A few necessary conditions are easy to obtain. If, for instance, \mathcal{G} is a set of operators (on H) and \mathcal{L} is the set of all those subspaces (of H) that are invariant under every operator in \mathcal{G} , then it is clear that \mathcal{L} is a lattice (i.e., that \mathcal{L} is closed under the formation of intersections and spans), and it is even clearer that \mathcal{L} contains 0 and H . In addition to these algebraic conditions, \mathcal{L} satisfies a somewhat less obvious topological condition, namely that it is strongly closed. (Explanation: the set of all projections whose ranges are in \mathcal{L} is strongly closed. Proof: if A is in \mathcal{G} and $\{P_n\}$ is a net

of projections strongly convergent to P and satisfying $AP_n = P_nAP_n$ for each n , then, because multiplication is jointly continuous on bounded sets, $AP = PAP$.) The topological condition has algebraic reverberations. **Exercise 1:** a strongly closed lattice is complete. (Caution: the converse is false. Example: discard a 1-dimensional space from the lattice of all subspaces of a 2-dimensional space; the remainder is a complete but non-closed lattice.) It is easy to show that the necessary conditions listed so far are nowhere near sufficient; the characterization problem for the set of all invariant subspaces of a set of operators is still open.

There is a kind of Galois theory connecting sets of operators and sets of subspaces. Half of it was just described: to each set \mathcal{A} of operators there corresponds the set of those subspaces that are invariant under the elements of \mathcal{A} ; call that set $\text{Lat}\mathcal{A}$. The other half goes backwards: to each set \mathcal{L} of subspaces there corresponds the set $\text{Alg}\mathcal{L}$ of those operators that leave invariant each element of \mathcal{L} . It is clear that $\text{Alg}\mathcal{L}$ is always an algebra (closed under the formation of sums, products, and scalar multiples); it is even clearer that $\text{Alg}\mathcal{L}$ contains 1; and the one-sided weak continuity of multiplication implies that $\text{Alg}\mathcal{L}$ is weakly closed.

The basic properties of Alg and Lat are easy; their proofs belong to universal algebra. The facts are that both Alg and Lat are order-reversing, and that if \mathcal{L} is any set of subspaces and \mathcal{A} is any set of operators, then

$$\mathcal{L} \subset \text{Lat Alg}\mathcal{L} \quad \text{and} \quad \mathcal{A} \subset \text{Alg Lat}\mathcal{A}.$$

(**Exercise 2:** $\text{Alg}\mathcal{L} = \text{Alg Lat Alg}\mathcal{L}$ and $\text{Lat}\mathcal{A} = \text{Lat Alg Lat}\mathcal{A}$. Analogy with linear algebra (when is an object equal to its own second dual?) suggests that if $\mathcal{L} = \text{Lat Alg}\mathcal{L}$, then the lattice \mathcal{L} be called *reflexive*; the same word is used for an algebra \mathcal{A} such that $\mathcal{A} = \text{Alg Lat}\mathcal{A}$. Word-saving convention: every lattice contains 0 and H and is strongly closed, every algebra contains 1 and is weakly closed.)

The theory of reflexive algebras is relatively new and, judging from the analytic techniques it has been using, quite deep. A typical result is Sarason's [45]: a commutative algebra of normal operators is reflexive. Example: the set of all matrices of the form $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$. In a different direction, Radjavi and Rosenthal [40] generalized a theorem of Arveson's [3] as follows: if \mathcal{A} includes a maximal abelian self-adjoint algebra and if $\text{Lat}\mathcal{A}$ is a chain (totally ordered), then \mathcal{A} is reflexive. Example: the set of all matrices of the form $\begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix}$. A typical example of a non-reflexive algebra is the set of all matrices of

the form $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$. A complete characterization of reflexive algebras is unknown even for finite-dimensional spaces.

To ask about invariant subspace lattices of sets of operators is the same as to ask about reflexive lattices. Here are two trivial but illuminating examples.

(i) If H is 2-dimensional and \mathcal{L} consists of 0 , H , and two distinct 1-dimensional subspaces of H , then \mathcal{L} is reflexive. To say of an operator that it leaves invariant each subspace in \mathcal{L} is to say just that it has two prescribed eigenvectors, and hence that its matrix with respect to the basis they form is diagonal. Since the only subspaces simultaneously invariant under all such diagonal operators are the ones in \mathcal{L} , the lattice \mathcal{L} is reflexive. (ii) Suppose again that H is 2-dimensional, and let \mathcal{L} consist of 0 , H , and three distinct 1-dimensional subspaces. It is very hard for an operator on H to have three distinct eigenvectors; the only operators that can do it are the scalars. Scalars, on the other hand, leave invariant every subspace of H ; the lattice \mathcal{L} is not reflexive.

The first non-trivial theorem about reflexive lattices is a sharpening of some related work of Kadison and Singer [27]; the statement, proved in complete generality by Ringrose [43], is that every complete chain is reflexive. Here again "chain" means a lattice whose order (in general partial) is in fact total. No topological assumption is needed: it is easy to prove that a complete chain of subspaces is necessarily strongly closed.

The smallest lattice that is not a chain is a 4-element Boolean algebra, i.e., a lattice of the form $\mathcal{L} = \{0, M, N, H\}$, where the subspaces M and N are such that $M \cap N = 0$ and $M \vee N = H$. It is true but not obvious that each such \mathcal{L} is reflexive. To prove this it is sufficient (and necessary) to exhibit a set \mathcal{Q} of operators such that $\text{Lat } \mathcal{Q} = \mathcal{L}$. Let P and Q be the projections onto M and N respectively and let \mathcal{Q} be the set of all operators that are either of the form $PA(1-Q)$ or else of the form $QA(1-P)$. Since $PA(1-Q)$ annihilates N and maps everything into M , and, similarly, $QA(1-P)$ annihilates M and maps everything into N , it follows that $\mathcal{L} \subset \text{Lat } \mathcal{Q}$. It remains to prove the reverse inclusion.

A useful lemma is that if K is in $\text{Lat } \mathcal{Q}$ and $K \not\subset M$, then $K \supset N$. (Similarly, of course, if K is in $\text{Lat } \mathcal{Q}$ and $K \not\subset N$, then $K \supset M$.) For the proof, take a vector f in K but not in M , and take an arbitrary vector g in N . Since f is not in M , it follows that $(1-P)f \neq 0$, and hence that there exists an operator A such that $A(1-P)f = g$. Since $QA(1-P)f$ is in K (because K is in $\text{Lat } \mathcal{Q}$) and $QA(1-P)f = g$ (because g is in N), it follows that g is in K ; this proves the lemma.

Suppose now that K is in $\text{Lat } \mathcal{A}$, $K \neq 0$, M , N ; it is to be proved that $K = H$. The subspace K cannot be included in both M and N (because $K \neq 0$). If, say, $K \not\subset M$, then, by the lemma, $K \supset N$. Since, however, $K \neq N$, it is not the case that $K \subset N$; the lemma applies again and yields $K \supset M$. The conclusion follows from the assumption that $M \vee N = H$.

The result just proved is contained in a much more general theorem: every complete atomic Boolean algebra is reflexive. The complementation implicit in the phrase "Boolean algebra" is not, of course, assumed to be orthogonal; compare the \mathcal{L} just treated. To say of a Boolean algebra of subspaces that it is atomic means that every subspace in the algebra is the span of all the atoms it includes; an *atom* is a non-zero subspace in the algebra that includes no subspace in the algebra other than 0 and itself. Problem 9 asks whether the assumption of atomicity can be omitted. The atomic special case was first announced in [23].

There are at least two ways of making new reflexive lattices out of old ones: direct sums and ordinal sums. The latter is more useful. A lattice \mathcal{L} of subspaces of a Hilbert space H is the ordinal sum of two lattices in case it contains a comparable element H_0 . That means that, for each M in \mathcal{L} , either $M \subset H_0$ or $H_0 \subset M$. If that is the case, then there are, indeed, two other lattices associated with \mathcal{L} ; (i) the lattice \mathcal{L}^- of all those subspaces of H_0 that happen to be in \mathcal{L} , and (ii) the lattice \mathcal{L}^+ of all those subspaces of H_0^\perp that are of the form $M \cap H_0^\perp$ for some M in \mathcal{L} . (This ad hoc definition is not elegant, but it avoids a long digression, and suggests the correct picture.) Theorem: if both \mathcal{L}^- and \mathcal{L}^+ are reflexive, then so is \mathcal{L} . Sample application: if $\dim H = 3$ and \mathcal{L} (the "pendulum") consists of 0, two 1-dimensional spaces, their span, and H , then \mathcal{L} is reflexive. This is of interest because (Exercise 3) \mathcal{L} is not the Lat of any single operator.

The theorems and techniques mentioned so far yield many reflexive lattices and a few non-reflexive ones, and they are sufficient to decide the status of all sufficiently small lattices. The following comments exhaust all that is known, and show therefore how much remains to be done.

If a lattice has four elements, or fewer, then it is reflexive. There are five isomorphism types of lattices with five elements: the chain, the pendulum, the pendulum upside down, the non-modular pentagon, and the double triangle. Every lattice of subspaces that belongs to one of the first three types is reflexive. The pentagon ($= \{0, L, M, N, H\}$, with $M \subset N$, $L \cap M = L \cap N = 0$ and $L \vee M = L \vee N = H$) is not realizable as a lattice of subspaces of a finite-dimensional space. If $\dim H$ is

infinite, the pentagon can occur, and in the one manifestation that has been studied it turns out to be reflexive. (Can it have a non-reflexive occurrence? The answer is not known.) As for the double triangle ($= \{0, L, M, N, H\}$, with $L \cap M = L \cap N = M \cap N = 0$ and $L \vee M = L \vee N = M \vee N = H$), all its known manifestations are non-reflexive. The known manifestations include all possible finite-dimensional ones; the unknown cases are all infinite-dimensional. The finite-dimensional facts are covered by a theorem of R. E. Johnson [25]: a finite lattice of subspaces of a finite-dimensional space is reflexive if and only if it is distributive. The "if" remains true for infinite-dimensional spaces (K. J. Harrison); the reflexive realizability of the pentagon shows that the "only if" is false for them.

ANSWERS TO THE EXERCISES. (1) Suppose that \mathcal{L} is a strongly closed lattice and that \mathcal{L}_0 is a subset of \mathcal{L} . Consider the directed set of all finite subsets of \mathcal{L}_0 ordered by inclusion. If n is such a finite subset, let M_n be the intersection of the subspaces in n . Since $n \rightarrow M_n$ is a decreasing net of subspaces, the corresponding net of projections is strongly convergent to the projection whose range is $\bigcap \mathcal{L}_0$; this proves that $\bigcap \mathcal{L}_0$ is in \mathcal{L} . The proof for spans is similar.

(2) Apply Lat to the relation $\mathcal{Q} \subset \text{Alg Lat } \mathcal{Q}$, and apply the relation $\mathcal{L} \subset \text{Lat Alg } \mathcal{L}$ to $\text{Lat } \mathcal{Q}$ in place of \mathcal{L} . This proves that $\text{Lat } \mathcal{Q} = \text{Lat Alg Lat } \mathcal{Q}$; the equation $\text{Alg } \mathcal{L} = \text{Alg Lat Alg } \mathcal{L}$ is proved dually. Corollary: every lattice that is the Lat of something is the Lat of its own Alg .

(3) If $\dim H < \infty$, and if $\mathcal{L} = \text{Lat } A$ for some operator A on H , then \mathcal{L} is self-dual, in the sense that \mathcal{L} is isomorphic to a lattice in which all the order relations of \mathcal{L} are reversed. Proof: the formation of orthogonal complements proves that $\text{Lat } A$ is anti-isomorphic to $\text{Lat } A^*$; since every matrix is similar to its transpose, and similar operators have isomorphic lattices, the proof can be completed by recalling that the matrix of A^* is the conjugate transpose of that of A [5].

10. TRANSITIVE

Problem 10. *Is every non-trivial strongly closed transitive atomic lattice either medial or self-conjugate?*

This is the most awkward of the ten problems here proposed; the reason is that it points to the darkest area of ignorance.

The crucial word is "transitive". A set \mathcal{Q} of operators is called transitive in case $\text{Lat } \mathcal{Q} = \mathcal{L}_{\min}$ (=the smallest lattice, the one consisting of 0 and H only); a set \mathcal{L} of subspaces is transitive in case $\text{Alg } \mathcal{L} = \mathcal{A}_{\min}$ (=the smallest algebra, the one consisting of scalars only).

If a set \mathcal{Q} of operators is transitive, then so is every subset of \mathcal{Q} that generates the same weakly closed algebra. It follows that the search for transitive sets might as well be restricted to weakly closed algebras. (Word-saving convention: every algebra contains 1, every lattice contains 0 and H .) Problem 10 has a dual, a problem for algebras, that is easier to state: is every weakly closed transitive algebra equal to \mathcal{Q}_{\max} (=the largest algebra, the one consisting of all operators)? It is implicit in the question that \mathcal{Q}_{\max} itself is transitive; if a subspace M of H is invariant under every operator on H , then $M=0$ or $M=H$. Conjecture: the solution of this dual of Problem 10 is negative. In fact, there probably exists a commutative transitive algebra. This conjecture, in turn, is a special case of one already on record (in §3): if a transitive operator exists (in present language an operator A is transitive in case $\text{Lat } A = \mathcal{L}_{\min}$), then the weakly closed algebra it generates is a transitive algebra. All these conjectures seem to be out of reach at present. If H is small (finite-dimensional), then Burnside's classical theorem applies [24, p. 276] and says that the only transitive algebra is \mathcal{Q}_{\max} . If H is very large (non-separable), then an easy cardinality argument shows that no countably generated algebra can be transitive. The cases between these two extremes remain shrouded in mystery.

The corresponding questions for lattices are newer, and, although the principal characterization problem is still unsolved, they promise to be somewhat more accessible.

The first thing to settle is that the definition is not vacuous: are there any transitive lattices? Since a lattice larger than a transitive one is also transitive, the question reduces to a known projective geometric fact (**Exercise 1**): if $\mathcal{L} = \mathcal{L}_{\max}$ (=the largest lattice, the one consisting of all subspaces), then \mathcal{L} is transitive.

Once that's settled, it is natural to regard \mathcal{L}_{\max} as a trivial example and to ask whether there exist any non-trivial strongly closed transitive lattices. The answer is yes (J. E. McLaughlin), but that's not obvious. To get an interesting class of examples, consider an arbitrary conjugation J on a Hilbert space H . (A *conjugation* is an involutory semilinear isometry, i.e., a mapping J of H into itself such that $J^2=1$, $J(\alpha f + \beta g) = \alpha^* Jf + \beta^* Jg$, and $\|Jf\| = \|f\|$. If H is an L^2 space, then an example of a conjugation can be obtained by writing $Jf = f^*$, and, conversely, every conjugation can be represented in this way. For a detailed discussion see [50, p. 357].) Call a subspace M of H self-conjugate (with respect to J) in case Jf is in M whenever f is in M , and let $\mathcal{L} (= \mathcal{L}_J)$ be the set of all self-conjugate subspaces. It is easy to verify that \mathcal{L} is a strongly closed transitive lattice. (The proof

of transitivity uses the same technique as is needed to prove the transitivity of \mathcal{L}_{\max} .) Call every \mathcal{L} obtained in this way a *self-conjugate* lattice; there are as many of them as there are J 's.

Once that's settled, it is again natural to ask whether every transitive lattice has been found by now. In other words, do there exist non-trivial strongly closed transitive lattices other than the self-conjugate ones? The answer is yes again, but that's even less obvious than anything that led up to it.

To construct an example, let K be a Hilbert space, write $H = K \oplus K$, and let \mathcal{L} be the 7-element lattice of subspaces of H consisting of 0 , H , the two axes ($K \oplus 0$ and $0 \oplus K$), the diagonal (the set of all $\langle f, f \rangle$'s with f in K), and the graphs of two operators S and T on K . Assertion: for suitable choice of S and T , the lattice \mathcal{L} is transitive.

Since H is given as a direct sum of two copies of K , every operator on H is a two-by-two matrix of operators on K . The assumption that \mathcal{L} contains the two axes implies that every operator in $\text{Alg } \mathcal{L}$ has the form $\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}$; the assumption that \mathcal{L} contains the diagonal implies that $Y = X$.

The operators S and T are to be chosen so that their graphs are disjoint from one another and from the other non-trivial elements of \mathcal{L} (i.e., from the axes and the diagonal). When can $\langle f, Sf \rangle$ be equal to $\langle g, Tg \rangle$? Answer: exactly when $f = g$ and $(S - T)f = 0$. Consequence: $\text{graph } S \cap \text{graph } T = 0$ if and only if $\ker(S - T) = 0$. Since the horizontal axis ($K \oplus 0$) and the diagonal are graphs, and since every graph is disjoint from the vertical axis, it follows that the five non-trivial elements of \mathcal{L} are pairwise disjoint if and only if

$$\ker S = \ker T = \ker(1 - S) = \ker(1 - T) = \ker(S - T) = 0.$$

The consideration of the orthogonal complements of graphs shows that the five non-trivial elements of \mathcal{L} pairwise span H if and only if their adjoints satisfy the vanishing kernel conditions just found. (If K is finite-dimensional, the resulting conditions are equivalent to the original ones; in the infinite-dimensional case they are not. Note that the orthogonal complement of the graph of, say, S is the "co-graph" consisting of all vectors of the form $\langle -S^*f, f \rangle$.)

When does $\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}$ leave invariant the graph of S ? An obvious computation shows that a necessary and sufficient condition is that $SX = XS$. Consequence: \mathcal{L} is transitive if and only if

$$\text{Com } S \cap \text{Com } T = \mathfrak{A}_{\min},$$

where "Com" means commutant.

The problem of finding a transitive lattice of the kind promised

above comes down to this: find two operators S and T satisfying the commutant condition just stated and such that both they and their adjoints satisfy the vanishing kernel condition. That is not difficult, but it does seem to lead to some matrix computation. One example (with $\dim K < \infty$) is obtained by letting S be the sum of a nilpotent Jordan block and a scalar, say, 2 (what is important about 2 is that it is neither 0 nor 1), and letting T be a diagonal matrix none of whose eigenvalues is 0, 1, or 2. The only condition whose verification is not obvious is the one involving $\ker(S - T)$. To prove it, compute $T^{-1}S$ and note that it is a lower triangular matrix with all diagonal entries distinct from 1; it follows that $T^{-1}S - 1$ is invertible, and hence so is $S - T$.

A similar example can be constructed on infinite-dimensional spaces. Suppose, for instance, that K is L^2 of the circle, let S be the bilateral shift ($Se_n = e_{n+1}$ for all n), and let T be a diagonal operator ($Te_n = \lambda_n e_n$ for all n). If the λ 's are bounded, if none of them is 0 or 1, and if they are such that

$$\sum_{n=1}^{\infty} |1/\lambda_1 \cdots \lambda_n|^2 < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} |1/\lambda_{-1} \cdots \lambda_{-n}|^2 < \infty,$$

then S and T satisfy all the requirements. The proof is straightforward.

The lattices \mathcal{L} just constructed have the property that any two of their non-trivial elements are complements. Since in the lattice diagram of such an \mathcal{L} all the non-trivial elements occur on the same level, halfway between 0 and H , the word *medial* might serve as an adequate description. P. Rosenthal reports that in an infinite-dimensional space a transitive medial lattice can be constructed with only four non-trivial elements. The construction uses, as above, the two axes and the diagonal, but only one graph, namely the graph of a suitably chosen unbounded transformation. It is not known whether three non-trivial elements can do the job. In finite-dimensional spaces (of dimension greater than 2) a medial lattice with four (or fewer) non-trivial elements can never be transitive; the proof proceeds by observing that any such lattice can be represented, as above, by two axes and two (or fewer) graphs, and then proving that there are too many degrees of freedom to make transitivity possible. **Exercise 2:** is every medial lattice with five (or more) non-trivial elements transitive?

Each new construction brings with it the question whether it's the last. Do there exist non-trivial strongly closed transitive lattices

other than the medial or the self-conjugate ones? Once again the answer is yes, but that yes is at the boundary of what is known today. K. J. Harrison has constructed a transitive lattice of 16 non-trivial elements (the dimension of the space is at least 8), of which 5 are atoms, 5 are co-atoms, and 6 are in the middle. It seems hard to describe a general class of which Harrison's example is an instance and which, together with the classes described before, has a chance of catching all transitive lattices. A reasonable guess is that atoms play a crucial role. Recall that a lattice is called *atomic* in case every element is the span of all the atoms it includes. (**Exercise 3:** if $2 \leq \dim H < \infty$ and if a lattice \mathcal{L} of subspaces of H has exactly one atom, then \mathcal{L} is not transitive.) The trouble with Harrison's 18-element lattice is that it is not atomic. A fussy examination of low-dimensional cases indicates that transitive lattices want to be atomic. In the infinite-dimensional case atomicity still makes sense but it may be too special to play an important role. In any event, since no intelligent guess at the structure of all (strongly closed) transitive lattices is at hand, the best that can be done is to grasp at the atomic straw; that is what Problem 10 does.

ANSWERS TO THE EXERCISES. (1) It is to be proved that if an operator A on H leaves invariant every subspace of H , then A is a scalar. The assumption implies that to every non-zero vector f there corresponds a unique scalar $\lambda(f)$ such that $Af = \lambda(f)f$. Since $A(f+g) = Af + Ag$, it follows that $(\lambda(f) - \lambda(f+g))f + (\lambda(g) - \lambda(f+g))g = 0$; this implies that if f and g are linearly independent, then $\lambda(f) = \lambda(g)$ ($= \lambda(f+g)$). If, on the other hand, $f = \alpha g$ (with $\alpha \neq 0$), then $Af = \alpha Ag$ and therefore, again, $\lambda(f) = \lambda(g)$, i.e., the function λ is a constant.

(2) There exist intransitive medial lattices with exactly five non-trivial elements. For a simple example in $H = K \oplus K$ (where K is separable and $\dim K \geq 2$), let \mathcal{L} consist of the two axes, the diagonal, and the graphs of two operators S and T . If, for instance, both S and T are diagonal matrices, such that the set of all diagonal entries in both of them together is a set of distinct numbers all distinct from 0 and 1, then the vanishing kernel conditions are satisfied. It follows that \mathcal{L} is indeed a medial lattice. Since both $\text{Com } S$ and $\text{Com } T$ consist of all diagonal matrices, it follows that $\text{Alg } \mathcal{L}$ consists of all $\begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}$, where X is diagonal; since $\dim K \geq 2$, this implies that $\text{Alg } \mathcal{L} \neq \mathcal{O}_{\text{min}}$.

(3) If \mathcal{L} has exactly one atom M_0 , then, since finite-dimensionality implies that every element of \mathcal{L} includes at least one atom, it follows that every element of \mathcal{L} includes M_0 . If $M_0 = H$, then \mathcal{L} is not transitive (because $\dim H \geq 2$). If $M_0 \neq H$, then there exists a non-scalar

operator A on H with range M_0 . Such an A leaves invariant every M in \mathfrak{L} ; indeed $AM \subset AH = M_0 \subset M$.

BIBLIOGRAPHY

1. T. Andô, *On a pair of commutative contractions*, Acta Sci. Math. (Szeged) **24** (1963), 88–90. MR **27** #5132.
2. N. Aronszajn and K. T. Smith, *Invariant subspaces of completely continuous operators*, Ann of Math. (2) **60** (1954), 345–350. MR **16**, 488.
3. W. B. Arveson, *A density theorem for operator algebras*, Duke Math. J. **34** (1967), 635–647. MR **36** #4345.
4. E. Bishop, *Spectral theory for operators on a Banach space*, Trans. Amer. Math. Soc. **86** (1957), 414–445. MR **20** #7217.
5. L. Brickman and P. A. Fillmore, *The invariant subspace lattice of a linear transformation*, Canad. J. Math. **19** (1967), 810–822. MR **35** #4242.
6. A. Brown and P. R. Halmos, *Algebraic properties of Toeplitz operators*, J. Reine Angew. Math. **213** (1963/64), 89–102. MR **28** #3350; errata MR **30**, 1205.
7. A. Brown, P. R. Halmos, and A. L. Shields, *Cesàro operators*, Acta Sci. Math. (Szeged) **26** (1965), 125–137. MR **32** #4539.
8. I. Colojoară and C. Foiaş, *Quasi-nilpotent equivalence of not necessarily commuting operators*, J. Math Mech. **15** (1966), 521–540. MR **33** #570.
9. ———, *Theory of generalized spectral operators*, Mathematics and its Applications, vol. 9, Gordon and Breach, New York, 1968.
10. D. Deckard, R. G. Douglas, and C. Pearcy, *On invariant subspaces of quasi-triangular operators*, Amer. J. Math. **91** (1969), 637–647.
11. R. G. Douglas and C. Pearcy, *A characterization of thin operators*, Acta Sci. Math. (Szeged) **29** (1968), 295–297. MR **38** #2628.
12. ———, *A note on quasitriangular operators*, Duke Math. J. **37** (1970), 177–188.
13. N. Dunford, *A survey of the theory of spectral operators*, Bull. Amer. Math. Soc. **64** (1958), 217–274. MR **21** #3616.
14. P. A. Fillmore and D. M. Topping, *Sums of irreducible operators*, Proc. Amer. Math. Soc. **20** (1969), 131–133. MR **38** #1549.
15. S. Foguel, *A counterexample to a problem of Sz.-Nagy*, Proc. Amer. Math. Soc. **15** (1964), 788–790. MR **29** #2646.
16. R. Gellar, *Cyclic vectors and parts of the spectrum of a weighted shift*, Trans. Amer. Math. Soc. **146** (1969), 69–85.
17. P. R. Halmos, *Measure theory*, Van Nostrand, Princeton, N. J., 1950. MR **11**, 504.
18. ———, *Normal dilations and extensions of operators*, Summa Brasil. Math. **2** (1950), 125–134. MR **13**, 359.
19. ———, *On Foguel's answer to Nagy's question*, Proc. Amer. Math. Soc. **15** (1964), 791–793. MR **29** #2647.
20. ———, *A Hilbert space problem book*, Van Nostrand, Princeton, N. J., 1967. MR **34** #8178.
21. ———, *Irreducible operators*, Michigan Math. J. **15** (1968), 215–223. MR **37** #6788.
22. ———, *Quasitriangular operators*, Acta Sci. Math. (Szeged) **29** (1968), 283–293. MR **38** #2627.
23. ———, *Invariant subspaces*, Abstract Spaces and Approximation, Proc. M. R. I. Oberwolfach, Birkhäuser, Basel, 1968, pp. 26–30.

24. N. Jacobson, *Lectures in abstract algebra*. Vol. II: *Linear algebra*, Van Nostrand, Princeton, N. J., 1953. MR 14, 837.
25. R. E. Johnson, *Distinguished rings of linear transformations*, Trans. Amer. Math. Soc. 111 (1964), 400–412. MR 28 #5088.
26. R. V. Kadison, *Strong continuity of operator functions*, Pacific J. Math. 26 (1968), 121–129. MR 37 #6766.
27. R. V. Kadison and I. M. Singer, *Triangular operator algebras. Fundamentals and hyperreducible theory*, Amer. J. Math. 82 (1960), 227–259. MR 22 #12409.
28. G. K. Kalisch, *On similarity, reducing manifolds, and unitary equivalence of certain Volterra operators*, Ann. of Math. (2) 66 (1957), 481–494. MR 19, 970.
29. R. L. Kelley, *Weighted shifts on Hilbert space*, Dissertation, University of Michigan, Ann Arbor, Mich., 1966.
30. A. Lebow, *A power-bounded operator that is not polynomially bounded*, Michigan Math. J. 15 (1968), 397–399. MR 38 #5047.
31. B. Sz.-Nagy, *On uniformly bounded linear transformations in Hilbert space*, Acta Sci. Math. (Szeged) 11 (1947), 152–157. MR 9, 191.
32. ———, *Completely continuous operators with uniformly bounded iterates*, Magyar Tud. Akad. Mat. Kutató Int. Közl. 4 (1959), 89–93. MR 21 #7436.
33. B. Sz.-Nagy and C. Foias, *Opérateurs sans multiplicité*, Acta Sci. Math. (Szeged) 30 (1969), 1–18.
34. N. K. Nikol'skiĭ, *Invariant subspaces of weighted shift operators*, Mat. Sb. 74 (116) (1967), 171–190. Math. USSR Sb. 3 (1967), 159–176. MR 37 #4659.
35. R. S. Palais, *Seminar on the Atiyah-Singer index theorem*, Ann. of Math. Studies, no. 57, Princeton Univ. Press, Princeton, N. J., 1965. MR 33 #6649.
36. S. K. Parrott, *Weighted translation operators*, Dissertation, University of Michigan, Ann Arbor, Mich., 1965.
37. ———, *Unitary dilations for commuting contractions*, Pacific J. Math. (to appear).
38. H. Radjavi, *Every operator is the sum of two irreducible ones*, Proc. Amer. Math. Soc. 21 (1969), 251–252. MR 38 #6388.
39. H. Radjavi and P. M. Rosenthal, *The set of irreducible operators is dense*, Proc. Amer. Math. Soc. 21 (1969), 256. MR 38 #5042.
40. ———, *On invariant subspaces and reflexive algebras*, Amer. J. Math. 91 (1969), 683–692.
41. W. C. Ridge, *Approximate point spectrum of a weighted shift*, Trans. Amer. Math. Soc. 147 (1970), 349–356.
42. J. R. Ringrose, *On the triangular representation of integral operators*, Proc. London. Math. Soc. (3) 12 (1962), 385–399. MR 25 #3372.
43. ———, *On some algebras of operators*, Proc. London Math. Soc. (3) 15 (1965), 61–83. MR 30 #1405.
44. P. M. Rosenthal, *On lattices of invariant subspaces*, Dissertation, University of Michigan, Ann Arbor, Mich., 1967.
45. D. E. Sarason, *Invariant subspaces and unstarred operator algebras*, Pacific J. Math. 17 (1966), 511–517. MR 33 #590.
46. H. H. Schaefer, *Eine Bemerkung zur Existenz invarianter Teilräume linearer Abbildungen*, Math. Z. 82 (1963), 90. MR 27 #4815.
47. J. Schwartz, *Subdiagonalization of operators in Hilbert space with compact imaginary part*, Comm. Pure Appl. Math. 15 (1962), 159–172. MR 26 #1759.
48. A. L. Shields, *A note on invariant subspaces*, Michigan Math. J. (to appear).

49. J. G. Stampfli, *Which weighted shifts are subnormal?* Pacific J. Math. **17** (1966), 367–379. MR **33** #1740.

50. M. H. Stone, *Linear transformations in Hilbert space*, Amer. Math. Soc. Colloq. Publ., vol. 15, Amer. Math. Soc., Providence, R. I., 1932.

51. N. Suzuki, *On the irreducibility of weighted shifts*, Proc. Amer. Math. Soc. **22** (1969), 579–581.

52. J. Wermer, *Commuting spectral measures on Hilbert space*, Pacific J. Math. **4** (1954), 355–361. MR **16**, 143.

53. H. Weyl, *Über beschränkte quadratische Formen deren Differenz vollstetig ist*, Rend. Circ. Mat. Palermo **27** (1909), 373–392.

INDIANA UNIVERSITY, BLOOMINGTON, INDIANA 47401

NOTE ADDED IN PROOF. One of the consequences of the preprint system of scientific communication is that some part of almost everything that is printed is superseded by the time it is published. The present paper is no exception apparently. Solutions have been reported to two of the ten problems above: Problem 1, negative, J. G. Stampfli; Problem 4, affirmative, I. D. Berg.