# FINITE-DIMENSIONAL HILBERT SPACES

P. R. HALMOS, Indiana University

**Prologue.** I used to like linear algebra because it gave me a motivation for the study of operators on Hilbert space and because it gave me insight into the algebraic skeleton of operator theory, which made that study easier. Now I like what I learned about Hilbert space because it keeps shedding light on more and more new aspects of linear algebra and because it succeeds in keeping that classical subject alive and exciting. The purpose of this report is to illustrate the latter point by describing three non-trivial parts of finite-dimensional linear algebra, the original impetus for which came from operator theory on infinite-dimensional Hilbert space. The subjects are (1) an algebraic characterization of pairs of subspaces of a finite-dimensional Hilbert space, (2) a geometric characterization of linear transformations in terms of rotations and projections (dilation theory), and (3) a statement of some fragmentary results and challenging open problems about lattices of invariant subspaces.

A finite-dimensional Hilbert space is, by definition, a finite-dimensional unitary space (complex inner product space). The only prerequisite for an intelligent reading of this paper is acquaintance with the language, notation, and principal facts of finite-dimensional unitary geometry; see, for instance, [**6**]. As for the insistence on complex numbers: the geometric language of Hilbert space is motivated by the real case, but the algebraic hurdles are most easily overcome if complex numbers are allowed. The customary way out (followed here) is to use complex coefficients, and, at the same time, continue to use real language; this does not seem to lead to serious or permanent confusion.

**Two subspaces.** What are all the different ways in which a subspace can be placed in a finite-dimensional Hilbert space? The question is vague, but it has a reasonably definite answer. If $H=C^n$ (where $C$ is the complex number field), then one way to get an $r$-dimensional subspace of $H$ $(0\leqq r\leqq n)$ is to form the set of all those vectors whose last $n-r$ coordinates vanish. More to the point is that to within isomorphism this is the only way; every $r$-dimensional subspace of $C^n$ can be obtained from this one by a suitable rotation.

What are all the different ways in which two subspaces, or three, or any number can be placed in a finite-dimensional Hilbert space? The difficulty of the answer seems to increase with the number. The preceding paragraph shows that there is no difficulty about putting *one* subspace $M$ into a space $H$; just put it down, anywhere, let $M^\perp$ fall where it may, and there is nothing left to ask. Before the position of *two* subspaces $M$ and $N$ in $H$ can be said to be known, many questions must be asked and answered. Is $M$ included in $N$? Is $M$ orthogonal to $N$? Does $M^\perp$ have a nontrivial intersection with $N^\perp$? If the relation between $M$ and $N$ is not describable in the simple terms of inclusion and orthogonality, does it make sense to ask for the "angle" between them? Such questions were first raised by Dixmier [**3**]; the point of view described below is somewhat different and more recent [**11**]. As for three subspaces, or more, the mind boggles. There is, in fact, reason to believe that the problem of three subspaces will be out of human reach for a long time to come. A comment of Chandler Davis [**2**] indicates that if we knew all about three subspaces, then we could learn more about unitary equivalence than, apparently, we are meant to know.

In the study of pairs of subspaces there are four thoroughly uninteresting cases, the ones in which both $M$ and $N$ are either 0 or $H$. In the most general case the entire space is the direct sum of five subspaces:

$$M \cap N,\ M \cap N^{\perp},\ M^{\perp} \cap N,\ M^{\perp} \cap N^{\perp},$$

and the rest. The parts of $M$ and $N$ in the first four are "thoroughly uninteresting". In "the rest", the orthogonal complement of the span of the first four, $M$ and $N$ are in *generic position*, in the sense that all four of the special intersections listed above are equal to 0.

The simplest example of two subspaces in generic position consists of two distinct non-orthogonal lines in a plane, and there is no loss of generality in taking one of them to be the first coordinate axis. To get a useful generalization, suppose that $T$ is a non-singular linear transformation on a finite-dimensional Hilbert space $K$, write $H = K \oplus K$, let $M$ be the "horizontal axis" consisting of all vectors of the form $\langle f, 0\rangle$ in $H$, and let $N$ be the graph of $T$, i.e., the set of all vectors of the form $\langle f, Tf\rangle$ in $H$. The assertion that $M$ and $N$ are in generic position needs a little proof. The first step is to show that $M \cap N = 0$. Indeed, how can an $\langle f, 0\rangle$ be equal to a $\langle g, Tg\rangle$? Answer: only if $Tg = 0$, whence $g = 0$ (because $T$ is non-singular), and therefore $f = 0$. For the rest of the proof it is necessary to know $M^{\perp}$ (trivial: all $\langle 0, f\rangle$) and $N^{\perp}$ (easy and standard computation: all $\langle -T^*f, f\rangle$). From this it is easy to deduce that $M^{\perp} \cap N^{\perp} = 0$: since $T^*$ is just as non-singular as $T$, the proof just given applies again. The equations $M \cap N^{\perp} = 0$ and $M^{\perp} \cap N = 0$ are trivial.

The basic result in the theory of two subspaces is that this way of constructing pairs of subspaces in generic position is, to within unitary equivalence, the only way. More precisely: *if $M$ and $N$ are subspaces in generic position in a finite-dimensional Hilbert space $H$, then there exists a finite-dimensional Hilbert space $K$, and there exists a non-singular linear transformation $T$ on $K$, such that the pair $\langle M, N\rangle$ is unitarily equivalent to the pair* $\langle K \oplus 0, \text{graph } T\rangle$.

What follows is an outline of the proof; with suitable analytic caution the proof is generalizable to the infinite-dimensional case. Let $P$ be the projection with range $M$. Assertion: the restriction of $P$ to $N$ is a non-singular linear transformation from $N$ onto $M$. Suppose, indeed, that $Pg = 0$ for some $g$ in $N$. It follows that $g \in M^{\perp} \cap N$, and hence (generic position) that $g = 0$; this proves that the kernel of $P$ in $N$ is 0. To prove that the image $PN$ is equal to $M$, suppose that $f \in M$ and $f \perp PN$. This means that if $g \in N$, then $0 = (f, Pg) = (Pf, g) = (f, g)$, so that $f \in M \cap N^{\perp}$. It follows (generic position) that $f = 0$; the proof of the assertion is complete.

The existence of a non-singular linear transformation from $N$ onto $M$ implies that $M$ and $N$ have the same dimension. (This could have been proved more quickly, but the slower approach is needed for the rest of the proof anyway.) Since $M$ and $M^{\perp}$ on the one hand and $N$ and $N^{\perp}$ on the other hand enter the hypotheses with perfect symmetry, it follows that all four of these subspaces have the same dimension. Since this applies to $M$ and $M^{\perp}$ in particular, there exists an isometric linear mapping from $M$ onto $M^{\perp}$; the idea from now on is to identify each element of $M^{\perp}$ with the element of $M$ that it corresponds to.

Now put $K=M$. To define $T$ at an element $f$ of $K$, recall first that $f=Pg$ for a uniquely determined vector $g$ in $N$, project $g$ into $M^{\perp}$, and let $Tf$ be the element of $M$ that is identified with the element of $M$ so obtained. (A simple 2-dimensional picture should make that long-winded sentence crystal clear.) The verification that $K$ and $T$ do what is expected of them is straightforward.

Given a line in the plane, distinct from both the horizontal and the vertical axes, rotate the plane through the negative of half the angle of inclination. The given line and the horizontal axis become, after the rotation, a line and its reflection through the (new) horizontal axis. This half-angle rotation can be generalized to yield a different and useful representation for a pair of subspaces in generic position; the result is that any such pair is unitarily equivalent to a pair of the form $\langle \text{graph } T_0, \text{graph } (-T_0)\rangle$, for a suitable linear transformation $T_0$. From this representation, in turn, it is easy to recapture Dixmier's main theorem on pairs of subspaces in generic position: the result is that a *single* Hermitian transformation, namely the sum of the two projections whose ranges are the given subspaces, constitutes a complete set of unitary invariants for the pair.

**Unitary dilations.** Suppose that $H$ is a subspace of a finite-dimensional Hilbert space $K$, and let $P$ be the projection from $K$ onto $H$. Each linear transformation $B$ on $K$ induces in a natural way a linear transformation $A$ on $H$ defined for each $f$ in $H$ by

$$Af = PBf.$$

Under these conditions the transformation $A$ is called the *compression* of $B$ to $H$, and $B$ is called a *dilation* of $A$ to $K$. This geometric definition of compression and dilation is to be contrasted with the customary concepts of restriction and extension: if it happens that $H$ is invariant under $B$, then it is not necessary to project $Bf$ back into $H$ (it is already there), and, in that case, $A$ is the restriction of $B$ to $H$, and $B$ is an extension of $A$ to $K$. Restriction-extension is a special case of compression-dilation, the special case in which the linear transformation on the larger space leaves the smaller space invariant.

Compressions and dilations can be usefully described in terms of matrices. If $K$ is decomposed into $H$ and $H^{\perp}$, and, correspondingly, transformations on $K$ are written in terms of matrices (whose entries are transformations on $H$, and on $H^{\perp}$, and 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 is to get information about difficult transformations by finding their easy dilations. The program is spectacularly successful. Unitary transformations (rotations) are among the easiest to deal with, and it turns out that, except for an easily adjusted normalization, every transformation has a unitary dilation. Some normalization is clearly necessary: if $B$ is unitary, then $\|Bf\| = \|f\|$ for every vector $f$, and it follows that $\|Af\| \leqq \|f\|$ for every vector $f$; in other words, if $A$ has a unitary dilation, then $A$ must not increase the norm of any vector. (In the appropriate geometrical technical term, $A$ must be a *contraction.*) That much normalization is sufficient: *every contraction has a unitary dilation.*

As a heuristic guide to the proof, consider the very special case in which the given Hilbert space is 1-dimensional real Euclidean space and the dilation space $K$ is the plane. In that case the given contraction is a scalar $\alpha$ (with $|\alpha| \leqq 1$), and, in geometric terms, the assertion is that multiplication by $\alpha$ (on the line) can be achieved by a suitable rotation (in the plane), followed by projection (back to the line). A picture makes all this crystal clear again.

The proof in the general case can be obtained by first transcribing the synthetic proof just outlined to analytic form, and then imitating the analytic geometry with matrices in the place of numbers. The conceptual problems that the program encounters are familiar ones, and so are their solutions; see [5] for the details.

The least unitary looking contraction is 0, but, of course, even it has a unitary dilation; one such is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This dilation does not have many useful algebraic properties. It is not necessarily true, for instance, that the square of a dilation is a dilation of the square; indeed, the square of the dilation of 0 exhibited above is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which is not a dilation of the square of 0. Is there a unitary dilation of 0 that is fair to squares? The answer is yes:

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

is an example. The square of this dilation is

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

which is a dilation of the square of 0. Unfortunately, however, this dilation is not perfect either; its cube is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

which is not a dilation of the cube of 0. That is (in self-explanatory language): the $3\times 3$ matrix is a 2-dilation of 0, but not a 3-dilation.

Question: does every contraction on an $n$-dimensional Hilbert space have a unitary $k$-dilation for every $k$? Answer: yes, on a Hilbert space of dimension $n(k+1)$; an elementary proof was given by Egerváry [**4**]. More is true: Nagy proved [**15**] that every contraction $A$ has a unitary dilation $B$ such that $B^k$ is a dilation of $A^k$ simultaneously for every positive integer $k$. (Nagy's paper came between [**5**] and [**4**].) The result is true for infinite-dimensional Hilbert spaces too, but its impressive generality has a price: even if the given space $H$ is finite-dimensional, the dilation space $K$ may have to be infinite-dimensional.

To show how dilation theory can be used, consider the finite-dimensional special case of a beautiful and powerful analytic theorem of von Neumann [**18**]. The assertion is that if $A$ is a contraction and if $q$ is a polynomial such that $|q(z)| \leqq 1$ whenever $|z| = 1$, then $q(A)$ is a contraction. The proof via dilation theory goes as follows: If the degree of $q$ is $k$, find a unitary $k$-dilation $B$ of $A$. It follows then that $q(B)$ is a dilation of $q(A)$, and hence it is sufficient to prove that $q(B)$ is a contraction. (In other words, dilation theory reduces the problem to the consideration of unitary transformations only.) But that is trivial: a unitary $B$ has a diagonal matrix, whose diagonal entries are complex numbers of modulus 1; since the corresponding matrix of $q(B)$ has diagonal entries whose moduli are not greater than 1, the desired conclusion becomes obvious.

**Reflexive lattices.** The set of all subspaces of a finite-dimensional Hilbert space $H$ is a lattice (with respect to the operations of intersection and span) with zero element 0 and unit element $H$. Certain of its sublattices, the ones called *reflexive*, are of interest in linear algebra. The definition of reflexivity requires of a lattice $\mathcal{L}$ that a two-step process performed on $\mathcal{L}$, which always yields a lattice of subspaces at least as large as $\mathcal{L}$, should, in fact, yield exactly $\mathcal{L}$, and nothing more. The two steps are these: (1) form all linear transformations that leave invariant each subspace of $\mathcal{L}$, and then (2) form all subspaces of $H$ that are invariant under all those linear transformations.

Here is an example. Suppose that $H$ is 2-dimensional, and let $\mathcal{L}$ consist of 0, $H$, and two distinct lines (i.e., 1-dimensional subspaces of $H$). To say of a linear transformation 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 transformations are the ones in $\mathcal{L}$, the lattice $\mathcal{L}$ is reflexive indeed.

Here is a non-example. Suppose again that $H$ is 2-dimensional, and let $\mathcal{L}$ consist of 0, $H$, and three distinct lines. It is very hard for a linear transformation on $H$ to have three distinct eigenvectors; the only linear transformations that can do it are the scalar multiples of the identity. Such scalar multiples, on the other hand, leave invariant every subspace of $H$. The two-step process applied to this $\mathcal{L}$ drastically enlarges $\mathcal{L}$; instead of the 5-element lattice $\mathcal{L}$, the enlargement is the infinite lattice of all subspaces of $H$.

The non-example of the preceding paragraph fails to be reflexive the worst way anything can; the enlargement it effects is maximal. Another way of saying the same thing is that a linear transformation that leaves invariant every subspace of the lattice is necessarily a scalar. A lattice that is non-reflexive in this extreme way is called *transitive.*

A basic open problem about operator theory on infinite-dimensional Hilbert spaces is to characterize all reflexive lattices and all transitive (extremally non-reflexive) lattices. A little progress has been made, but not very much. The finite-dimensional specialization of what is known amounts to two statements: (1) every chain (totally ordered set) of subspaces is reflexive, and (2) every Boolean algebra of subspaces is reflexive. The infinite-dimensional case of (1) is Ringrose's generalization [**17**] of a result of Kadison and Singer [**14**]. For a statement of the appropriate infinite-dimensional formulation of (2) see [**10**]; the proof has not been published yet. In the finite-dimensional case both results become almost trivial.

As for transitive lattices, even less is known. The example above (the one with three lines) is in a certain sense degenerate. From the point of view of projective geometry, which is quite appropriate here, the space of that example has dimension 1, not 2, and the example does not help to answer the question whether higher-dimensional examples exist at all.

An interesting unpublished observation of J. E. McLaughlin shows that they do exist; one such, in $C^n$, consists of all those subspaces that are invariant under the formation of complex conjugates. More explicitly: call a subspace $M$ of $C^n$ *symmetric* in case $M$ contains, along with each of its vectors, the vector whose coordinates are obtained from the given one by complex conjugation. Assertion: the set of all symmetric subspaces is a transitive lattice. The proof requires a moment's thought, but there is nothing profound about it.

The example of the preceding paragraph yields many examples. Given a finite-dimensional Hilbert space, coordinatize it (i.e., establish an isomorphism between it and $C^n$); the lattice of symmetric subspaces with respect to that coordinatization is a transitive lattice. Question: are all transitive lattices obtainable in this way?

The answer is no for two reasons, both trivial, and the heart of the question is still unanswered. *First*, a topological distinction arises: are the lattices under discussion closed or not? (There is only one reasonable topology for the space of subspaces; the question makes unambiguous sense.) If a lattice is dense in a transitive lattice, then it itself is transitive; the question loses no vigor at all if attention is restricted to closed lattices only. As long as restrictions are in order, here is one spot where the complex field makes life more complicated, not less; the question retains all its interest if attention is restricted to real spaces only. *Second*, for spaces of even dimension $2n$ there is a construction that yields a transitive lattice whose non-trivial elements are all of dimension $n$. Such a lattice imitates an already observed misbehavior; it is isomorphic to a sublattice of the lattice of the projective line.

The result of the indicated specializations is the following question: is every closed transitive lattice of subspaces of an odd-dimensional Euclidean space (real inner product space) equal to the lattice of all subspaces? The answer does not seem to be known.

*Note added in proof.* K. J Harrison (Monash University) has recently discovered a new transitive lattice of 18 elements that shows that the answer to the question as it stands is no. A modification, however, restores the question: just add the hypothesis that the atoms of the lattice span the whole space. Harrison's discovery makes the problem of determining all transitive lattices even more challenging than before.

**Epilogue.** Three topics were discussed above to illustrate the thesis that operator theory on Hilbert space yields non-trivial questions and answers about finite matrices. Choosing the examples was not an easy task; many more are available than can be included in one lecture, or one paper, of reasonable length.

I could have chosen the theorem about "near" projections (if the projections onto two subspaces are near enough in norm, then the subspaces have the same dimension) [**8**, Problem 43]. I could have discussed the "power inequality" ($w(A^n) \leqq (w(A))^n$) for the numerical range [**8**, Problem 176], which started in the theory of partial differential equations and ended as a problem about matrices, a problem that refused to become trivial even in the $2\times 2$ case. The topological properties of sets of reducible and irreducible operators were of recent research interest [**9**, **16**], and so also were partial isometries [**7**, **12**]; in both these cases new and interesting facts about the finite-dimensional case emerged. The theory of matrices whose entries are matrices still has some life in it; thus, for instance, the "ultra-invariant" subspaces of binormal operators ($2\times 2$ matrices whose entries are commutative normal operators; see [**1**]) are still being studied (by R. G. Douglas and C. M. Pearcy), and there are small, amusing, and until recently unnoticed questions even about determinants. (Sample: if all four ways of forming the formal determinant of a $2\times 2$ matrix whose entries are matrices yield an invertible matrix, does the matrix itself have to be invertible? See [**13**].)

The subjects just given honorable mention, as well as the three actually discussed in detail, have been receiving serious research attention in the course of the last twenty years (and many still are), and they have all led to questions that could and should have been asked in the finite-dimensional case long before, but as a matter of historical fact they were not. The reason perhaps is that the powerful tools of finite-dimensional linear algebra are too good; they sometimes conceal the elegant and intricate structure that the difficulties of the infinite-dimensional theory bring out. Most of the problems of operator theory can be formulated in the finite-dimensional case, and there are two reasons why it is good to do so. The old reason is that the finite can suggest what should and should not be tried with the infinite; the new reason is the joy of seeing the infinite inspire and guide the finite and contribute to a new flowering of an old subject.

### References

1. A. Brown, The unitary equivalence of binormal operators, Amer. J. Math., 76 (1954) 414–434.

2. C. Davis, Generators of the ring of bounded operators, Proc. Amer. Math. Soc., 6 (1955) 970–972.

3. J. Dixmier, Position relative de deux variétés linéaires fermées dans un espace de Hilbert, Rev. Sci., 86 (1948) 387–399.

4. E. Egerváry, On the contractive linear transformations of $n$-dimensional vector space, Acta Sci. Math. (Szeged) 15 (1953) 178–182.

5. P. R. Halmos, Normal dilations and extensions of operators, Summa Brasil. Math., 2 (1950) 125–134.

6. ———, Finite-dimensional Vector Spaces, Van Nostrand, Princeton, 1958.

7. ——— and J. E. McLaughlin, Partial isometries, Pacific J. Math., 13 (1963) 585–596.

8. ———, A Hilbert Space Problem Book, Van Nostrand, Princeton, 1967.

9. ———, Irreducible operators, Michigan Math. J., 15 (1968) 215–223.

10. ———, Invariant subspaces, Abstract spaces and approximation, Proc. MRI Oberwolbach (1968).

11. ———, Two subspaces, Trans. Amer. Math. Soc., 144(1969) 381–389.

12. ——— and L. J. Wallen, Powers of partial isometries, J. Math. Mech., 19 (1970) 657–663.

13. ———, Advanced problem 5727, this MONTHLY, 77 (1970) 409.

14. R. V. Kadison and I. M. Singer, Triangular operator algebras, Amer. J. Math., 82 (1960) 227–259.

15. B. Sz.-Nagy, Sur les contractions de l'espace de Hilbert, Acta Sci. Math. (Szeged) 15 (1953) 87–92.

16. H. Radjavi and P. Rosenthal, The set of irreducible operators is dense, Proc. Amer. Math. Soc., 21 (1969) 256.

17. J. R. Ringrose, On some algebras of operators, Proc. London Math. Soc., 15 (1965) 61–83.

18. J. von Neumann, Eine Spektraltheorie für allgemeine Operatoren eines unitären Raumes. Math. Nachr., 4 (1951) 258–281.