

# Chapter 8

## The Invariant Subspace Problem

LOMONOSOV'S FAMOUS THEOREM

**Overview.** This chapter is about the most vexing problem in the theory of linear operators on Hilbert space:

THE INVARIANT SUBSPACE PROBLEM. *Does every operator on Hilbert space have a nontrivial invariant subspace?*

Here “operator” means “continuous linear transformation,” and “invariant subspace” means “closed (linear) subspace that the operator takes into itself.” To say that a subspace is “nontrivial” means that it is neither the zero subspace nor the whole space. Examples constructed toward the end of the last century show that in the generality of Banach spaces there do exist operators with only trivial invariant subspaces. For *Hilbert space*, however, the Invariant Subspace Problem remains open, and is the subject of much research. In this chapter we'll see why invariant subspaces are of interest and then will prove one of the subject's landmark theorems: Victor Lomonosov's 1973 result, a special case of which states:

*If an operator  $T$  on a Banach space commutes with a non-zero compact operator, then  $T$  has a nontrivial invariant subspace.*

This result, which far surpassed anything that seemed attainable at the time, is only part of what Lomonosov proved in an astonishing two-page paper [71] that introduced nonlinear methods—in particular the Schauder Fixed-Point Theorem—into this supposedly hard-core-linear area of mathematics.

**Prerequisites.** Basics of inner-product spaces, Hilbert and Banach spaces.

## 8.1 Invariant Subspaces

For linear transformations on vector spaces for which no topology is assumed, “invariant subspace” will simply mean “subspace taken into itself by the transformation.” Eigenvalues give rise to an important class of nontrivial invariant subspaces.

**Theorem 8.1** (Invariance of eigenspaces). *Suppose  $T$  is a linear transformation on a vector space  $V$ , and that  $T$  is not a scalar multiple of the identity. Let  $\lambda$  be an eigenvalue of  $T$ . Then the subspace  $\ker(T - \lambda I)$  is nontrivial and invariant for every linear transformation on  $V$  that commutes with  $T$ .*

*Proof.* Let  $E = \ker T - \lambda I$ . By hypothesis there is a vector  $v \in V \setminus \{0\}$  with  $Tv = \lambda v$ . Thus  $v \in E$ , so  $E \neq \{0\}$ . Since  $T \neq \lambda I$  we know  $E \neq V$ . Thus  $E$  is nontrivial.

Now suppose  $S$  is a linear transformation on  $V$  that commutes with  $T$ . Suppose  $v \in E$ . We wish to show that  $Sv \in E$ , i.e., that  $TSv - \lambda Sv = 0$ . This follows right away from the commutativity of  $S$  and  $T$ :

$$TSv - \lambda Sv = STv - \lambda Sv = S(\lambda v) - \lambda Sv = \lambda Sv - \lambda Sv = 0. \quad \square$$

*Exercise 8.1* (Invariant subspaces without eigenvalues). Let  $C([0, 1])$  denote the Banach space of complex-valued continuous functions on the unit interval  $[0, 1]$ , endowed with the “max-norm”

$$\|f\| = \max\{|f(x)| : 0 \leq x \leq 1\} \quad (f \in C([0, 1])).$$

Show that the *Volterra operator*, defined by

$$Vf(x) = \int_0^x f(t) dt \quad (f \in C([0, 1]), x \in [0, 1]),$$

is an operator that takes the Banach space  $C([0, 1])$  into itself, that has no eigenvalue, but that nonetheless has nontrivial invariant subspaces.

*Hyperinvariant subspaces.* If a subspace of a Banach space is invariant for every operator that commutes with a given operator  $T$ , we’ll call that subspace *hyperinvariant* for  $T$ . Thus Theorem 8.1 shows that every operator on  $\mathbb{C}^N$  that’s not a scalar multiple of the identity has a nontrivial hyperinvariant subspace. It’s not known, however, if this is true for infinite dimensional Hilbert spaces. In other words, the “Hyperinvariant Subspace Problem” is just as open as is the “Invariant Subspace Problem.”

**Why the Invariant Subspace Problem?** In studying the Invariant Subspace Problem one is searching for two things: simplicity and approximation.

*Simplicity.* One hopes that restriction of an operator to an invariant subspace will result in a simpler operator that provides insight into the workings of the original one. This is just what happens in the finite dimensional setting where the study of invariant subspaces leads to Schur’s Theorem (Theorem 8.3 below), which asserts that every operator on  $\mathbb{C}^N$  has—relative to an appropriately chosen orthonormal basis—an upper-triangular matrix. Schur’s Theorem in turn leads to the Jordan Canonical

Form (see, e.g., [51, Chap. 3]), which tells us that every operator on  $\mathbb{C}^N$  is similar to either an operator of the form  $\lambda I + N$ , where  $\lambda$  is a scalar and  $N$  is nilpotent (possibly the zero-operator), or to a direct sum of such operators.

For the infinite dimensional situation, suppose we have an operator  $T$  on a separable Hilbert space and that  $T$  has a nontrivial invariant subspace  $M$ . Upon choosing an orthonormal basis for  $M$  and completing it to one for the whole space we can write—just as in the finite dimensional case—a matrix (an infinite one this time) representing  $T$  with respect to this basis. This matrix will have a “block upper triangular” form  $\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$ , where the matrix  $A$  represents the restriction of  $T$  to  $M$ ,  $B$  the restriction of  $PT$  to  $M^\perp$  ( $P$  being the orthogonal projection of our Hilbert space onto  $M$ ), and  $C$  the restriction to  $M^\perp$  of  $(I - P)T$ . In fact the existence of a nontrivial invariant subspace is *equivalent* to  $T$  having such a matrix representation.

*Approximation.* For an operator  $T$  on a Banach space  $X$ , here’s a natural way to construct an invariant subspace. Fix a non-zero vector  $x_0 \in X$  and take the linear span of its iterate sequence under  $T$ , i.e., look at the linear subspace of  $X$  consisting of all vectors  $p(T)x_0$ , where  $p$  is a polynomial with complex coefficients. This linear subspace is taken into itself by  $T$ , hence so is its closure  $\mathcal{V} = \mathcal{V}(T, x_0)$ . Since  $\mathcal{V}$  contains  $x_0$ , it is not the zero subspace; in fact  $\mathcal{V}$  is the smallest  $T$ -invariant subspace containing  $x_0$ . If  $\mathcal{V} \neq X$  then we’ve produced a nontrivial invariant subspace for  $T$ . On the other hand, if  $\mathcal{V} = X$  then we have an approximation theorem: every vector in  $X$  is the limit of a sequence of polynomials in  $T$  applied to the *cyclic vector*  $x_0$ .

*Example.* Let  $T$  denote the linear transformation of “multiplication by  $x$ ” on the Banach space  $C([0, 1])$ . More precisely,

$$(Tf)(x) = xf(x) \quad (f \in C([0, 1]), 0 \leq x \leq 1).$$

It’s easy to see that  $T$  is a bounded operator on  $C([0, 1])$ .

*Claim:* *The constant function 1 is a cyclic vector for  $T$ .*

*Proof of Claim.* For  $p$  a polynomial with complex coefficients, the vector  $p(T)1$  is just  $p$ , now viewed as a function on  $[0, 1]$ . Thus  $\mathcal{V}(T, 1)$  is the closure in  $C([0, 1])$  of the polynomials. Now convergence in  $C([0, 1])$  is uniform convergence on  $[0, 1]$  so by the Weierstrass Approximation Theorem [101, Theorem 7.26, p. 159],  $\mathcal{V}(T, 1) = C([0, 1])$ .  $\square$

The operator of “multiplication by  $x$ ” also makes sense for the Hilbert space  $L^2([0, 1])$ , and since the continuous functions are dense therein, the function 1 is a cyclic vector in that setting too. This is not to say that our operator  $T$  is devoid of nontrivial invariant subspaces; it has non-cyclic vectors, too. For example, in the setting of  $C([0, 1])$  each function  $f$  that takes the value zero somewhere on  $[0, 1]$  is a non-cyclic vector (exercise), so  $\mathcal{V}(T, f)$  is a nontrivial invariant subspace for  $T$ .

*Exercise 8.2.* Characterize the cyclic vectors for the operator of “multiplication by  $x$ ” when setting is

- (a) The Banach space  $C([0, 1])$ .
- (b) The Hilbert space  $L^2([0, 1])$ .

*Exercise 8.3.* Show that every operator on a *non-separable* Banach space has a nontrivial invariant subspace. Thus the invariant subspace problem really concerns only *separable* Banach spaces.

*Exercise 8.4 (Reducing subspaces).* A subspace is said to *reduce* an operator if it’s invariant and has an invariant complement, i.e., if the whole space can be decomposed as the direct sum of the original invariant subspace and another one. Not every operator, even in finitely many dimensions, has a nontrivial reducing subspace; show that the operator induced on  $\mathbb{C}^2$  by the matrix  $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$  does not have such a subspace. More generally the same is true for every  $N \times N$  matrix whose  $N$ -th power is the zero-matrix, but whose  $(N - 1)$ -st power is not.

**Invariant subspaces and projections.** Suppose  $X$  is a vector space,  $V$  a linear subspace, and  $P$  a projection taking  $X$  onto  $V$ , i.e.,  $P$  is a linear transformation with  $P(X) = V$  whose restriction to  $V$  is the identity operator.<sup>1</sup> The fact that  $P$  is the identity map when restricted to its image can be expressed by the equation  $P^2 = P$ . Clearly the linear transformation  $Q = I - P$  is also a projection with  $PQ = QP = 0$ . Since  $P + Q = I$  these equations tell us that the projections  $P$  and  $Q$  decompose  $X$  into the direct sum of  $V = P(X)$  and  $W = Q(X)$ .

**Proposition 8.2.** *Suppose  $X$  is a vector space,  $V$  a linear subspace,  $P$  a projection taking  $X$  onto  $V$ , and  $T: X \rightarrow X$  a linear transformation on  $X$ . Then the following three statements are equivalent:*

- (a)  $T(V) \subset V$ .
- (b)  $PTP = TP$ .
- (c)  $QTQ = QT$ , where  $Q = I - P$ .

*Proof.* Statements (a) and (b) both assert that the restriction of  $P$  to  $T(V)$  is the identity map on  $T(V)$ . As for the equivalence of (b) and (c): note that since  $Q = I - P$  we have

$$QTQ = T - TP - PT + PTP = QT + (PTP - TP)$$

so  $QTQ = QT$  if and only  $PTP - TP = 0$ . □

## 8.2 Invariant Subspaces in $\mathbb{C}^N$

Invariant subspaces are important even for finite dimensional operators. For example, the following 1909 result of Issai Schur is a fundamental result in matrix theory.

---

<sup>1</sup> If  $X$  were a normed linear space with  $P$  continuous we could use the language introduced in Sect. 4.1 and call  $P$  a *retraction* of  $X$  onto  $V$ .

**Theorem 8.3** (Schur’s Triangularization Theorem). *Suppose  $V$  is a finite dimensional complex inner-product space and  $T$  is a linear transformation on  $V$ . Then  $V$  has an orthonormal basis relative to which the matrix of  $T$  is upper triangular.*

Schur’s Theorem is really a statement about invariant subspaces. Suppose  $\dim V = N$ , and let  $\mathcal{V} = (v_j : 1 \leq j \leq N)$  be the orthonormal basis it promises for the operator  $T$  (it’s important to note here that “basis” means: “linearly independent spanning set, written as an ordered list”). Let  $[T]$  denote the matrix of  $T$  with respect to this basis, i.e., for each index  $j$ , the  $j$ -th column of  $[T]$  is the column vector of coefficients of  $Tv_j$  with respect to the basis  $\mathcal{V}$ . Thus the upper-triangularity of  $[T]$  asserts that  $Tv_j$  belongs to the linear span  $V_j$  of the basis vectors  $(v_1, v_2, \dots, v_j)$ , so for each  $j$  between 1 and  $N$ :

*$V_j$  is a nontrivial invariant subspace for  $T$ .*

Schur’s Theorem therefore promises, for each operator  $T$  on  $V$ , the existence of a descending chain of invariant subspaces

$$V = V_N \supset V_{N-1} \supset \dots \supset V_1 \supset V_0 = \{0\}, \quad (8.1)$$

each of which has codimension one in the preceding one. It’s an easy exercise to see that the existence of such a chain is equivalent to that of the basis promised by Schur’s Theorem.

*Proof of Schur’s Theorem.* This proceeds by induction on the dimension  $N$ . For  $N = 1$  the theorem is trivial, so suppose  $N > 1$  and the result is true for  $N - 1$ . The transformation  $T$  has an eigenvalue; let  $v_1$  be a unit eigenvector for this eigenvalue, let  $V_1$  be the (one dimensional) linear span of the singleton  $\{v_1\}$ , and let  $W = V_1^\perp$ , the orthogonal complement in  $V$  of  $V_1$ . The subspace  $W$  has dimension  $N - 1$ , but unfortunately it need not be invariant under  $T$ . To remedy this, let  $P$  denote the orthogonal projection of  $V$  onto  $W$  and consider the operator  $R = PT$ , for which  $W$  is invariant. Our induction hypothesis applies to the restriction  $R|_W$  of  $R$  to  $W$ , and produces an orthonormal basis  $(v_2, v_3, \dots, v_N)$  for  $W$  relative to which the matrix of  $R|_W$  is upper triangular.

Thus  $(v_1, v_2, v_3, \dots, v_N)$  is an orthonormal basis for  $V$ . We aim to show that the matrix of  $T$  with respect to this basis is upper triangular, i.e., that  $Tv_j$  lies in the linear span of the vectors  $v_1, v_2, \dots, v_j$  for each index  $1 \leq j \leq N$ . We already know  $Tv_1 \in V_1$ , so suppose  $j > 1$ . We have

$$Tv_j = (I - P)Tv_j + PTv_j = (I - P)Tv_j + Rv_j$$

with  $I - P$  the orthogonal projection of  $V$  onto  $V_1$ . Now  $R$  takes  $v_j$  into the subspace spanned by the vectors  $v_k$  for  $2 \leq k \leq j$ . Thus  $Tv_j$  belongs to the linear span of the vectors  $(v_1, v_2, \dots, v_j)$ , as we wished to prove.  $\square$

**Applications of Schur’s Theorem.** Before moving on let’s see how Schur’s Theorem makes short work of several fundamental results of linear algebra.

*Hermitian Operators.* Let  $V$  be a finite dimensional inner-product space, with inner product  $\langle \cdot, \cdot \rangle$ . Then to each operator  $T$  on  $V$  we can attach another one called the *adjoint*  $T^*$  of  $T$ , defined by

$$\langle Tx, y \rangle = \langle x, T^*y \rangle \quad (x, y \in V). \quad (8.2)$$

To say an operator  $T$  on  $V$  is *hermitian* means that  $T = T^*$ . If  $(v_1, v_2, \dots, v_N)$  is an orthonormal basis for  $V$ , then an operator  $T$  on  $V$  is hermitian if and only if (8.2) holds with  $T = T^*$  when  $x$  and  $y$  run through the elements of this basis, i.e., when

$$\langle Tv_i, v_j \rangle = \langle v_i, Tv_j \rangle = \langle Tv_j, v_i \rangle^*, \quad (1 \leq i, j \leq N),$$

where the notation  $\lambda^*$ , when applied to a complex scalar  $\lambda$ , denotes “complex conjugate.” Thus:

*An operator  $T$  on  $V$  is hermitian if and only if, with respect to every (or even “some”) orthonormal basis, its matrix and the conjugate-transpose of this matrix are the same.*

With these preliminaries in hand we obtain from Schur’s Theorem—almost trivially—one of the most important theorems of linear algebra:

**Corollary 8.4** (The Spectral Theorem for hermitian operators). *Suppose  $T$  is a hermitian operator on a finite dimensional inner-product space. Then the space has an orthonormal basis relative to which the matrix of  $T$  is diagonal.*

*Proof.* Schur’s Theorem promises an orthonormal basis for the space, relative to which  $T$  has an upper-triangular matrix. With respect to this basis, the matrix of the adjoint  $T^*$  has all entries *above* the main diagonal equal to zero. But  $T = T^*$ , so the matrix of  $T$  has all entries off the main diagonal equal to zero.  $\square$

Why is this result called a “spectral theorem?” For a finite dimensional operator, the set of eigenvalues is often called the “spectrum,” and for each diagonal matrix this is precisely the set of diagonal entries. With this in mind, it’s an easy exercise to check that the above Corollary can be restated:

*If  $T$  is a hermitian operator on a finite dimensional inner-product space  $V$  then there is an orthonormal basis for  $V$  consisting of eigenvectors of  $T$ .*

*Normal Operators.* To say an operator on a finite dimensional inner-product space, or even a Hilbert space, is *normal* means that the operator commutes with its adjoint. Hermitian operators are normal, but not all normal operators are hermitian (Example: a diagonal matrix with at least one non-real entry.). It turns out that the spectral theorem for hermitian operators holds as well for normal operators. The proof follows the hermitian model, once we have the following surprisingly simple generalization of Schur’s Theorem.

**Theorem 8.5** (Schur’s Theorem for Commuting Pairs of Operators). *If two operators commute on a finite dimensional inner-product space then the space has an orthonormal basis with respect to which each operator has upper-triangular matrix.*

*Proof.* This one is a minor modification of the induction proof of Theorem 8.3. Let  $V$  be the inner-product space in question, with  $N = \dim V$ , and let  $S$  and  $T$  be operators on  $V$  that commute. The result we want to prove is trivially true for  $N = 1$ , so suppose it holds for dimension  $N - 1$ , where  $N > 1$ . We want to prove it for dimension  $N$ . Once again we observe that  $T$  has an eigenvalue—call it  $\mu$ , but now, instead of choosing just one unit  $T$ -eigenvector for  $\mu$ , we look at the full eigenspace  $E = \ker(T - \mu I)$ , and note that since  $S$  commutes with  $T$ , Theorem 8.1 guarantees that this eigenspace is invariant for  $S$ . Thus the restriction of  $S$  to  $E$  has an eigenvalue  $\lambda$ , hence a corresponding unit eigenvector  $v_1$ , which by design is a  $\lambda$ -eigenvector for  $T$ . As before, let  $V_1$  be the span of the single vector  $v_1$ , let  $W = V_1^\perp$ , and let  $P$  be the orthogonal projection of  $\mathbb{C}^N$  onto  $W$ . Let  $A = PS$  and  $B = PT$ . Both operators take  $W$  into itself, so if we can show that their restrictions to  $W$  commute, our induction hypothesis will supply an orthonormal basis for  $W$  relative to which the matrices of these restrictions are both upper triangular. Upon adjoining  $v_1$  to this basis, then applying to both  $S$  and  $T$  the argument that finished off the proof of Theorem 8.3, we’ll be done.

In fact, it’s easy to see that  $A$  commutes with  $B$  on all of  $V$ . Since  $W^\perp = V_1$  is invariant for both  $S$  and  $T$ , we know from the equivalence of (a) and (c) in Proposition 8.2 (with the roles of  $P$  and  $Q$  reversed) that  $PTP = PT$  and  $PSP = PS$ . Thus

$$AB = PSPT = PST = PTS = PTPS = BA,$$

where the third inequality uses the commutativity of  $S$  and  $T$ . □

**Corollary 8.6** (The Spectral Theorem for normal operators on  $\mathbb{C}^N$ ). *If  $T$  is a normal operator on  $\mathbb{C}^N$  then there exists an orthonormal basis for  $\mathbb{C}^N$  relative to which the matrix of  $T$  is diagonal.*

The above proof of Schur’s Theorem for commuting pairs of operators can easily be extended to finite collections of commuting operators. The following exercise shows that this proof extends even further:

*Exercise 8.5* (Triangularization for commuting families). *Show that: If  $\mathcal{C}$  is a family of commuting operators on a finite dimensional inner-product space  $V$ , then there exists an orthonormal basis of  $V$  relative to which each operator in  $\mathcal{C}$  has upper-triangular matrix.*

In particular, if the commuting family  $\mathcal{C}$  consists of normal operators, there’s an orthonormal basis for  $V$  relative to which each operator in the family has a diagonal matrix.

*Outline of proof:* The key is to prove that the family  $\mathcal{C}$  has a common eigenvector; then the proof can proceed like that of Theorem 8.5. Note first that there are nontrivial subspaces of  $V$  that are  $\mathcal{C}$ -invariant (meaning: “invariant for every operator in  $\mathcal{C}$ ”). Example: the eigenspace of any operator in  $\mathcal{C}$ . Let  $m$  be the minimum of the dimensions of all the eigenspaces of operators in  $\mathcal{C}$ , so  $m \geq 1$ . Choose a  $\mathcal{C}$ -invariant subspace of  $\mathbb{C}^N$  having this minimum dimension  $m$ . Show that every operator in  $\mathcal{C}$  is, when restricted to that subspace, a scalar multiple of the identity.

### 8.3 Compact Operators

The result we seek to understand, Lomonosov’s Theorem, deals with two concepts: invariant subspaces and compact operators. Having spent some time getting a feeling for the former, let’s now take a moment to review some of the fundamental facts about the latter.

A linear transformation on a normed linear space is said to be *compact* if it takes the closed unit ball into a relatively compact set. Since relatively compact sets are bounded it follows from Proposition C.8 (Appendix C, p. 196) that: *Every compact linear transformation is continuous, and so is an “operator.”*

*Exercise 8.6* (Basic Facts About Compact Transformations). Here all linear transformations act on a normed linear space  $X$ .

- (a) If  $\dim X < \infty$  then every linear transformation on  $X$  is compact.
- (b) For operators  $A$  and  $K$  on  $X$ : if  $K$  is compact then so are  $AK$  and  $KA$  (i.e., the compact operators on  $X$  form a closed ideal in the algebra of all operators).

The following exercise gives some feeling for the concept of compactness for a natural class of concrete operators on the Hilbert space  $\ell^2$ .

*Exercise 8.7.* For a bounded sequence  $\Lambda := (\lambda_k)$  of complex numbers, define the linear “diagonal map”  $D_\Lambda$  on  $\ell^2$  by  $D_\Lambda(x) = (\lambda_k \xi_k)$  for each vector  $x = (\xi_k) \in \ell^2$ . Show that  $D_\Lambda$  is continuous on  $\ell^2$ , and compact if and only if  $\lambda_k \rightarrow 0$ .

*Suggestion.* For compactness: first show that a subset  $S$  of  $\ell^2$  is relatively compact if and only if it is “equicontinuous at  $\infty$ ” in the sense that

$$\limsup_{n \rightarrow \infty} \sum_{f \in S} \sum_{k \geq n} |f(k)|^2 = 0.$$

As noted in Exercise 8.6(a), every operator on a finite dimensional normed linear space is compact. By contrast we pointed out at the beginning of Sect. 7.2 that the unit ball of an infinite dimensional Hilbert space is *not* compact; according to Exercise 7.2 the same is true for  $C([0, 1])$ . Thus the identity operator is not compact on either of these spaces. More is true:

**Theorem 8.7.** *If a normed linear space is infinite dimensional then its closed unit ball is not compact.*

This result, along with Proposition C.9 of Appendix C, shows that for normed linear spaces, compactness of the closed unit ball *characterizes* finite dimensionality. The key to its proof is the following lemma:

**Lemma 8.8.** *Suppose  $X$  is a normed linear space,  $Y$  a finite dimensional proper subspace, and  $0 < r < 1$ . Then there exists a unit vector  $x \in X$  whose distance to  $Y$  is greater than  $r$ .*

*Proof.* Fix a vector  $x_0 \in X$  that is not in  $Y$ , and let  $d$  denote the distance from  $x_0$  to  $Y$ , i.e.,



$$d = \inf\{\|x_0 - y\| : y \in Y\}.$$

According to Corollary C.10 of Appendix C, the subspace  $Y$  is complete in the norm-induced metric on  $X$ , thus  $Y$  is closed in  $X$ . It follows that  $d > 0$ , hence there exists  $y_0 \in Y$  with  $\|x_0 - y_0\| < d/r$ .

*Claim:* The unit vector  $x = \frac{x_0 - y_0}{\|x_0 - y_0\|}$  is the one we seek.

*Proof of Claim.* If  $y \in Y$  then

$$x - y = \frac{1}{\|x_0 - y_0\|} \left[ x_0 - \underbrace{(y_0 + \|x_0 - y_0\|y)}_{\in Y} \right].$$

Thus the term on the right in square brackets has norm  $\geq d$ , so

$$\|x - y\| \geq d/\|x - x_0\| > d/(d/r) = r,$$

hence the distance from  $x$  to  $Y$  is  $> r$ , as desired. □

*Proof of Theorem 8.7.* Let  $X$  be an infinite dimensional normed linear space. Fix a countable linearly independent set  $\{x_n\}_1^\infty$  in  $X$  and let  $Y_n$  be the linear span of the vectors  $\{x_1, \dots, x_n\}$ . There results the strictly increasing chain

$$Y_1 \subset Y_2 \subset Y_3 \subset \dots$$

of subspaces of  $X$ , each of which is finite dimensional hence closed in its successor. By Lemma 8.8 there is, for each index  $n > 1$ , a unit vector  $y_n \in Y_n$  at distance  $\geq 1/2$  to  $y_{n-1}$ . Let  $y_1 = x_1/\|x_1\|$ . Suppose the indices  $i$  and  $j$  are different, say  $i < j$ . Then  $y_i \in Y_{j-1}$ , so  $\|y_i - y_j\| \geq 1/2$ . Thus  $(y_n)$  is a sequence of vectors in the closed unit ball of  $X$  that has no convergent subsequence. □

**Corollary 8.9.** *The identity operator on a normed linear space is compact if and only if the space is finite dimensional.*

This suggests that for operators, compactness should be intertwined with finite dimensionality. The following result gives one important way in which this is true; it's the beginning of what's known as "The Riesz Theory of Compact Operators."

**Proposition 8.10.** *Suppose  $K$  is a compact operator on a Banach space. If  $\lambda \neq 0$  is an eigenvalue of  $K$  then the eigenspace  $\ker(K - \lambda I)$  is finite dimensional.*

*Proof.* We may suppose without loss of generality that  $\lambda = 1$  (exercise). Thus  $M := \ker(K - I)$  is an invariant subspace for  $K$  and the restriction of  $K$  to  $M$  is a compact operator on  $M$ . Since this restriction equals the identity operator on  $M$ , the closed unit ball of  $M$  must be compact, hence  $M$  is finite dimensional by Theorem 8.7. □

On infinite dimensional Banach spaces, compact operators need not have eigenvalues. The exercise below provides an example: the Volterra operator, which was shown in Exercise 8.1 to have no eigenvalues.

*Exercise 8.8 (Compactness Without Eigenvalues).* Use the Arzela–Ascoli Theorem (Appendix B, Theorem B.8) to show that the Volterra operator is compact on  $C([0, 1])$ .

## 8.4 Lomonosov's Theorem

We now turn to what is easily the most celebrated result on the existence of invariant subspaces. Here's a special case:

**Theorem 8.11** (Lomonosov 1973). *Every non-zero compact operator on an infinite dimensional Banach space has a nontrivial hyperinvariant subspace.*

This result says that not only does every operator commuting with a non-zero compact have a nontrivial invariant subspace—already far more than was previously known—but also that there's even a nontrivial subspace invariant for *all* the operators commuting with the given compact. We'll devote the rest of this section to proving this remarkable result; the method of proof will provide an even more remarkable generalization.

The key to Theorem 8.11 is the following Lemma which, although Lomonosov did not state it explicitly, is in fact the crucial step in his argument.

**Lemma 8.12.** *Suppose  $X$  is an infinite dimensional Banach space and  $K$  is a non-zero compact operator on  $X$ . If  $K$  does not have a hyperinvariant subspace then there is an operator  $A$  on  $X$  that commutes with  $K$  and for which  $KA$  has a fixed point in  $X \setminus \{0\}$ .*

*Proof that Lemma 8.12 implies Theorem 8.11.* Suppose  $K$  is a non-zero compact operator on  $X$  that has no hyperinvariant subspace. Let  $A$  be as in the Lemma. Thus  $M = \ker(AK - I)$  is not the zero subspace, and since  $AK$  is compact (by Exercise 8.6(b)) its eigenspace  $M$  is finite dimensional (Proposition 8.10), hence not equal to  $X$ . Now  $K$  commutes with  $A$ , hence it commutes with  $AK$ . Theorem 8.1 guarantees that  $M$  is invariant for every operator that commutes with  $AK$ , hence  $M$  is invariant for  $K$ . Since  $M$  is finite dimensional the restriction of  $K$  to  $M$ —hence  $K$  itself—has an eigenvalue; call it  $\lambda$ .

The corresponding eigenspace  $E := \ker(K - \lambda I)$  is a non-zero subspace of  $X$  that is, by Theorem 8.1, invariant for every operator that commutes with  $K$ . Also,  $E \neq X$ ; if  $\lambda = 0$  this follows from the fact that  $K \neq 0$ , while if  $\lambda \neq 0$  then it follows from the finite dimensionality of  $E$ . Thus  $E$  is a nontrivial hyperinvariant subspace for  $K$ , contradicting our assumption that  $K$  had no such subspace. Conclusion:  $K$  does have a nontrivial hyperinvariant subspace.  $\square$

*Proof of Lemma 8.12.* We're given a non-zero compact operator  $K$  on an infinite dimensional Banach space  $X$  and are assuming that  $K$  has *only trivial* hyperinvariant subspaces. Our goal is to produce an operator  $A$  that commutes with  $K$  such that  $KA$  has a non-zero fixed point (i.e., has 1 as an eigenvalue).

*Step I. An Algebra of Operators.* Let  $\mathcal{A}$  denote the collection of operators on  $X$  that commute with  $K$ , the notation reflecting the fact that  $\mathcal{A}$  is an *algebra* of operators, i.e., closed under addition, scalar multiplication and multiplication (= composition) of operators. In particular: for each  $x \in X$  the set of vectors  $\mathcal{A}x = \{Ax : A \in \mathcal{A}\}$  is a linear subspace of  $X$  (since  $\mathcal{A}$  is closed under addition and scalar multiplication

of operators) that's taken into itself by each operator in  $\mathcal{A}$  (since  $\mathcal{A}$  is closed under operator multiplication). Furthermore  $\mathcal{A}$  contains the identity operator on  $X$ , so if  $x \neq 0$  then  $\mathcal{A}x \neq \{0\}$ . Since we're assuming  $K$  has only trivial hyperinvariant subspaces,  $\mathcal{A}x$  has to be dense for each  $0 \neq x \in X$ ; otherwise its closure would be a nontrivial hyperinvariant subspace for  $K$ .

*Step II. Some Sets.* Since multiplication of  $K$  by a non-zero scalar changes neither its compactness, its commutation properties, nor its hyperinvariant subspace structure, we may without loss of generality assume that  $\|K\| = 1$ . Thus  $K$  is *contractive*:  $\|Kx\| \leq \|x\|$  for every  $x \in X$ . Choose a vector  $x_0 \in X$  for which  $\|Kx_0\| > 1$ . Because  $\|K\| = 1$  this implies  $\|x_0\| > 1$ , so the closed ball

$$B = \{x \in X : \|x - x_0\| \leq 1\}$$

does not contain the origin. Let  $C$  denote the closure in  $X$  of  $K(B)$ . Since  $K$  is a compact operator and  $B$  is a bounded subset of  $X$ , the set  $C$  is compact. In addition, since  $B$  is convex and  $K$  linear,  $C$  is convex. Finally (and crucially), as the calculation below shows,  $C$  does not contain the origin. Indeed, for each  $x \in X$ :

$$\|Kx\| = \|K(x - x_0) + Kx_0\| \geq \|Kx_0\| - \|K(x - x_0)\| \geq \|Kx_0\| - \|x - x_0\|,$$

the last inequality arising from the contractivity of  $K$ . Thus for each  $x \in B$  we have  $\|Kx\| \geq \|Kx_0\| - 1 := \delta > 0$ , hence every vector in  $K(B)$ , so also in its closure  $C$ , has norm at least  $\delta$ .

*Some wishful thinking.* If we could produce an operator  $A \in \mathcal{A}$  for which  $A(C) \subset B$ , then  $KA$ , which also belongs to the algebra  $\mathcal{A}$ , would map the compact, convex set  $C$  continuously into itself, so by Schauder's theorem would have the desired fixed point. This is not quite what's going to happen, but it's still worth keeping in mind as we proceed.

*Step III. A Map with a Fixed Point.* Let  $B^\circ$  denote the interior of the closed ball  $B$ . Suppose  $0 \neq y \in X$ . Since  $\mathcal{A}y$  is dense in  $X$  there exists  $A \in \mathcal{A}$  for which  $y \in A^{-1}(B^\circ)$ . Thus  $\{A^{-1}(B^\circ) : A \in \mathcal{A}\}$  is an open cover of  $X \setminus \{0\}$ , hence an open cover of  $C$ . Since  $C$  is compact, it has a finite subcover  $\mathcal{U} = \{U_j\}_1^N$ , where  $U_j := A_j^{-1}(B^\circ)$ .

While we haven't produced a map  $A \in \mathcal{A}$  with  $A(C) \subset B$ , we have produced a finite collection  $\{A_1, A_2, \dots, A_N\}$  of operators in  $\mathcal{A}$ , each of which takes a piece of  $C$  into  $B$ , as shown by the right-hand side of Fig. 8.1.

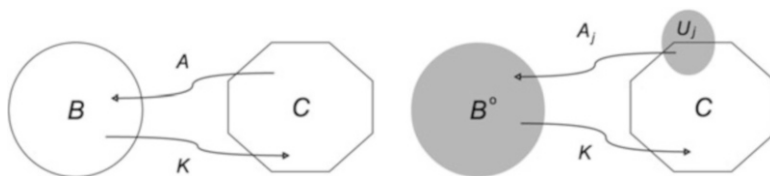


Fig. 8.1 What we want (left) vs. what we've got (right)

Lomonosov’s great insight was to use a standard “nonlinear” argument to glue the operators  $A_j$  together into a continuous map that takes  $C$  into  $B$ . By Proposition B.6 of Appendix B there is a partition of unity subordinate to the open covering  $\mathcal{U}$  of  $C$ , i.e., a set  $\{p_j : 1 \leq j \leq N\}$  of continuous functions taking  $C$  into the unit interval  $[0, 1]$  that sum to 1 at each point of  $C$ , and have the property that for each index  $j$  the function  $p_j$  is  $\equiv 0$  on  $C \setminus U_j$ . Define  $\Phi : C \rightarrow X$  by

$$\Phi(y) = \sum_{j=1}^N p_j(y) A_j y \quad (y \in C).$$

Being a finite sum of continuous maps,  $\Phi$  is continuous. Moreover  $\Phi(y)$  is, for each  $y \in C$ , a convex combination of vectors in the convex ball  $B$ , so it, too, belongs to  $B$ . Thus  $\Phi$  is a continuous map taking  $C$  into  $B$ , hence  $K \circ \Phi$  takes  $C$  continuously into itself. Since  $C$  is a compact, convex subset of a Banach space, the *Schauder Fixed-Point Theorem* (Theorem 7.1) guarantees that  $K \circ \Phi$  has a fixed point  $y_0 \in C$ .

*Step IV. Linearization.* Let  $A = \sum_{j=1}^N p_j(y_0) A_j$ , a linear combination of operators in  $\mathcal{A}$  and therefore also an operator in  $\mathcal{A}$ . Moreover

$$(KA)y_0 = K \left( \sum_{j=1}^N p_j(y_0) A_j y_0 \right) = K(\Phi(y_0)) = y_0.$$

Thus  $A \in \mathcal{A}$  and  $y_0 \in X \setminus \{0\}$  are the operator and vector we seek. This establishes Lemma 8.12 and with it, Lomonosov’s Theorem 8.11. □

*Exercise 8.9.* The hypothesis of Theorem 8.11 does not hold for every operator; there exist operators that commute with no non-zero compact operator. For  $\varphi \in C([0, 1])$  not identically zero, let  $M_\varphi$  denote the operator on  $C([0, 1])$  of “multiplication by  $\varphi$ ,” i.e.,

$$(M_\varphi f)(x) = \varphi(x)f(x) \quad (0 \leq x \leq 1; f \in C([0, 1])).$$

If  $\varphi(x) \equiv x$  we’ll write  $M_x$  instead of  $M_\varphi$ . Show that the operators  $M_\varphi$  are the only ones that commute with  $M_x$ , and that none of these is compact. *Hint:* If  $T = M_\varphi$  then  $\varphi = T1$ .

We mentioned earlier that there are Banach space operators with no nontrivial invariant subspace, but that the problem is still open for Hilbert space (see the *Notes* at the end of this chapter for references and more details). Thus Exercise 8.9 would have more significance if it were set in a Hilbert space. The following modification does just that, replacing  $C[0, 1]$  with the Hilbert space  $L^2 = L^2([0, 1])$  consisting of (a.e.-equivalence classes of) measurable complex-valued functions on  $[0, 1]$  whose moduli are square integrable with respect to Lebesgue measure. The arguments are similar to those of the exercise above, but they require a bit more work.

*Exercise 8.10.* Let  $L^\infty$  denote the space of (a.e.-equivalence classes of) essentially bounded complex-valued functions on  $[0, 1]$ . Define multiplication operators  $M_\varphi$  for  $\varphi \in L^\infty$ , and  $M_x$ , as above. Show that if  $\varphi \in L^\infty \setminus \{0\}$  then  $M_\varphi$  is an operator on  $L^2$  that is not compact. Show that if an operator  $T$  on  $L^2$  commutes with  $M_x$ , then  $T = M_\varphi$  for some  $\varphi \in L^\infty$ .

## 8.5 What Lomonosov Really Proved

According to Exercise 8.10 there exist operators on  $L^2$  that commute with no non-zero compact operator. Consequently Lomonosov's Theorem 8.11, spectacular as it is, does not solve the Invariant Subspace Problem for Hilbert space. However the story does not end here. At the very end of his paper [71], Lomonosov notes that the reasoning he used to prove Theorem 8.11 yields more. In what follows, let's agree to call an operator "nonscalar" if it is not a scalar multiple of the identity operator.

**Theorem 8.13** (Lomonosov). *If a nonscalar operator  $T$  on an infinite dimensional Banach space commutes with a non-zero compact operator, then  $T$  has a nontrivial hyperinvariant subspace.*

Our original Lomonosov Theorem implies that, on an infinite dimensional Banach space, every operator that commutes with a non-zero compact operator has a nontrivial invariant subspace. This one implies that a nontrivial invariant subspace exists for every operator that commutes with a nonscalar operator that commutes with a compact one.

*Proof of Theorem 8.13.* Let  $X$  be our infinite dimensional Banach space. The proof of Lemma 8.12 goes through word-for-word to establish this:

**Lemma 8.12, Enhanced.** *Suppose  $\mathcal{A}$  is an algebra of operators on  $X$ , and  $K$  is a non-zero compact operator on  $X$ . Suppose there is no nontrivial closed subspace invariant for every member of  $\mathcal{A}$ . Then there exists an operator  $A \in \mathcal{A}$  for which  $KA$  has a fixed point in  $X \setminus \{0\}$ .*

Suppose  $T$  is a nonscalar operator on  $X$  that commutes with our non-zero compact operator  $K$ . Let  $\mathcal{A}$  denote the algebra of all operators that commute with  $T$ . We wish to show that there is a closed subspace, neither the zero subspace nor the whole space, that is invariant under every operator in  $\mathcal{A}$ . Suppose this is not the case. Then by the enhanced Lemma 8.12 we know that there exists  $A \in \mathcal{A}$  such that  $KA$  has a fixed point in  $X \setminus \{0\}$ . The eigenspace  $M := \ker(KA - I)$  is, just as before:  $\neq \{0\}$ , finite dimensional so  $\neq X$ , and invariant for every operator that commutes with  $KA$ . Since  $T$  commutes with both  $K$  and  $A$ , it commutes with  $KA$ , hence  $M$  is invariant for  $T$ . The restriction of  $T$  to the finite dimensional invariant subspace  $M$  therefore has an eigenvalue  $\lambda$ . The eigenspace  $M_\lambda := \ker(T - \lambda I)$  is a closed subspace of  $X$  that is: not the zero subspace, not  $X$  (because  $T$  is nonscalar), and invariant for every operator that commutes with  $T$ . But we've assumed that  $\mathcal{A}$  has no such subspace. Contradiction! Therefore  $\mathcal{A}$  does have such a subspace.  $\square$

## Notes

*Schur's Triangularization Theorem.* This occurs in Schur's paper [109, p. 490], where it's applied to the study of integral equations. According to Horn and Johnson [51, p. 101], Schur's Theorem is "perhaps the most fundamentally useful fact of elementary matrix theory." Exercise 8.5 is from [51], see in particular Theorems 1.3.19, pp. 63–63 and 2.3.3, p. 103.

*Compact operators.* Lemma 8.8 is due to F. Riesz; it's Lemma 1 on p. 218 of his book [98] with Sz.-Nagy. Sections 76–80 of this book contain a nice exposition, set in the Hilbert space  $L^2$ , of the Riesz Theory of Compact Operators, a fundamental piece of which is—as we noted above—Proposition 8.10. The Riesz theory shows that compact operators behave “spectrally” very much like operators on finite dimensional spaces. For a modern exposition set in Banach spaces, see [103, Sects. 4.16–4.25, pp. 103–111]. J. H. Williamson showed in [124] that with the proper definition of “compact operator” the Riesz theory carries over to arbitrary (Hausdorff, but not necessarily locally convex) topological vector spaces.

*Lomonosov's Theorem: prehistory.* In the early 1930s von Neumann proved that every compact operator on Hilbert space has a nontrivial invariant subspace. He never published this result, and it was rediscovered about thirty years later by Aronszajn who, along with K. T. Smith, simplified the proof and in [4] generalized the result to Banach spaces.

The work of Aronszajn and Smith suggested the question of whether or not every operator whose *square* is compact has a nontrivial invariant subspace. This remained open until 1966 when Bernstein and Robinson in [11] showed, using non-standard analysis, that an operator has a nontrivial invariant subspace whenever some *polynomial* (not  $\equiv 0$ ) in it is compact.

Various authors refined the Bernstein–Robinson proof, replacing their polynomial hypotheses with one of the form: “Some of limit of polynomials or rational functions in the operator is compact.” Lomonosov's results superseded most of this earlier work. The version presented here of Lomonosov's work closely follows his original paper [71], as well as the exposition [92] of Percy and Shields.

*Chains of commutation.* For operators  $S$  and  $T$  on some Banach space, let's write  $T \leftrightarrow S$  whenever  $S$  commutes with  $T$ , and let's write  $K$  for a generic non-zero compact operator. Theorem 8.11 implies that:

$$T \leftrightarrow K \implies T \text{ has a nontrivial invariant subspace.}$$

We've observed that, thanks to Exercise 8.10, the above consequence of Lomonosov's theorem doesn't solve the Invariant Subspace Problem for Hilbert space. However Theorem 8.13, the “real” Lomonosov Theorem, tells us that:

$$T \leftrightarrow S \text{ (nonscalar)} \leftrightarrow K \implies T \text{ has a nontrivial invariant subspace,}$$

so it makes sense to ask if *this* might solve the Invariant Subspace Problem for Hilbert space, i.e., “Does every operator on Hilbert space commute with a nonscalar operator that commutes with a non-zero compact?” This hope was destroyed in 1980 by Hadwin et al. [44].

One might still hope to solve the Invariant Subspace Problem by extending Lomonosov's method to get a result for longer “commutation chains.” Unfortunately Troitsky in [119] showed that at least for the Banach space  $\ell^1$  there's no hope for such a result (see below for more details).

*Counterexamples for Banach spaces.* Per Enflo produced the first example of an operator on a Banach space possessing no nontrivial invariant subspace. Enflo's paper is [38, 1987], but his result was already circulating in preprint form over a decade earlier. A few years after Enflo released his preprint, Charles Read produced a much simpler counterexample, and then went on to find one set in the sequence space  $\ell^1$  [95, 1986]. Read later gave examples of Banach-space operators having no closed invariant *subset* [96, 1988].

In [119] Troitsky showed for Read's operator  $T$  on  $\ell^1$  that there exist non-scalar operators  $S_1$  and  $S_2$  on  $\ell^1$  such that  $T \leftrightarrow S_1 \leftrightarrow S_2 \leftrightarrow K$ , thus showing that Lomonosov's arguments cannot be extended to handle longer commutation chains.

In a totally different direction Argyros and Haydon [3] recently showed that there exist Banach spaces on which every bounded operator has the form "compact plus scalar multiple of the identity." Thus every bounded operator on such a space has a nontrivial invariant subspace (by the Aronszajn-Smith theorem), and even one that is *hyperinvariant* (by Lomonosov's theorem). Needless to say, such Banach spaces do not occur in the course of every-day mathematical life.

*The current state of affairs.* It's impossible to summarize quickly the many research efforts currently under way related to the Invariant Subspace Problem. The book [24, 2011] is an up-to-date exposition of the subject, while [94, 2003] is the standard reference for the state of the art circa 1973, and contains an outline, along with extensive references, of subsequent results up to 2003.