The Cyclic Decomposition of a Nilpotent Operator

J.H. Shapiro

1 Introduction.

Suppose T is a linear transformation on a vector space V. Recall Exercise #3 of Chapter 8 of our text, which we restate here as:

1.1 Key Lemma. *If, for some vector x in V and some positive integer m, we have:*

$$T^{m-1}x \neq 0 \quad \text{but} \quad T^m x = 0$$

then the list $(x, Tx, \ldots, T^{m-1}x)$ is linearly independent.

Suppose in particular that V is finite dimensional with dimension n, and T is nilpotent of index n. Then there is an $x \in V$ with $T^{n-1}x \neq 0$, and $T^n x = 0$ (the first condition arising from the definition of "index of nilpotence" and the second from the fact that T^n is the zero-operator). By the Key Lemma, the list $(x, Tx, \ldots, T^{n-1}x)$ forms a *basis* for V.

Recall that in this case the matrix of T with respect to the basis

$$(x, Tx, \ldots, T^{n-1}x)$$

has zeros everywhere except on the first subdiagonal, where it has ones. It is more common to write the basis in reverse order, in which case the matrix of T has zeros everywhere but on the first *superdiagonal*, where it has ones. We will say more about such matrices later.

1.2 Cyclic terminology. Let us call a linear transformation $T \in \mathcal{L}(V)$ a *cyclic operator* if there is a vector $x \in V$ for which the list $(x, Tx, \ldots, T^{n-1}x)$ spans (and is therefore a basis) for V. We call x a *cyclic vector* for T, and sometimes, when overcome by exuberance, we say that the space V is *cyclic for T*, or just *T-cyclic*. If x is a cyclic vector for T we call the corresponding basis $(x, Tx, \ldots, T^{n-1}x)$ a *cyclic basis* for V.

Using this terminology we can restate the Key Lemma like this:

1.3 Proposition. *If V is a finite dimensional vector space and T a nilpotent operator with index of nilpotence equal to* $\dim V$*, then T is cyclic.*

1.4 Exercise. (a) Suppose T is a linear transformation of a finite dimensional vector space V, and that x is a cyclic vector for T.

(i) What is the matrix of T with respect to the corresponding cyclic basis of V?

(ii) In terms of this matrix, what is the minimal polynomial of T?

(b) Give an example of a cyclic linear transformation that is *not* nilpotent.

(c) Give an example of a nilpotent linear transformation that is not cyclic. What's the most extreme example of this situation?

1.5 The Fundamental Question. *What can be said of T if it is nilpotent with index of nilpotence $< \dim V$?*

To get a handle on this question, suppose $n = \dim V$ (finite) and m is the index of nilpotence of T, with $m < n$. By the definition of "index of nilpotence" we know that there exists a vector $x_1 \in V$ such that $T^{m-1}x_1 \neq 0$. Since $T^m = 0$, the Key Lemma guarantees that the list $(x_1, Tx_1, \ldots, T^{m-1}x_1)$ is linearly independent, and clearly its linear span V_1 is T-invariant. We are going to prove that:

(*) $\quad V_1$ *has a T-invariant complement,*

i.e. that there exists a T-invariant subspace W_1 of V such that $V = V_1 \oplus W_1$. Once this is done, we'll be able to restrict T to W_1, where it's still nilpotent of some index $m_1 \leq m$, and find a vector $x_2 \in W_1$ such that $T^{m_1-1}x_2 \neq 0$, whereupon, as above, the list $(x_2, Tx_2, \ldots, T^{m_1-1}x_2)$ is linearly independent, and its linear span V_2 is T-invariant. Upon applying (*) to the restriction of T to W_1 we obtain a splitting of W_1 into two subspaces, $V_2 = \text{span}\,(x_2, Tx_2, \ldots, T^{m-1}x_2)$ and a T-invariant complement W_2.

Thus far $V = V_1 \oplus V_2 \oplus W_2$, where all three subspaces are T-invariant, and the restrictions of T to the first two are cyclic. Since V is finite dimensional, we arrive, after a finite number of repetitions of this argument, at a decomposition of V into a finite direct sum of invariant subspaces, with T restricted to each subspace being cyclic.

In summary: once we have established (*), we will have established that V splits into the direct sum of finitely many T-invariant subspaces with the property that the restriction of T to each is cyclic. Put more succinctly, we will have proved:

1.6 The Cyclic Nilpotent Theorem. *Every nilpotent linear transformation of a finite dimensional vector space splits into a direct sum of cyclic nilpotent transformations.*

We are also interested in the matrix interpretation of this result. It asserts that if T is nilpotent then V has a basis with respect to which the matrix of T is block diagonal, each block being zero except for ones on the first superdiagonal.

If, instead, it is $T - \lambda I$ that is nilpotent for some scalar λ, then each of these cyclic subspaces is T-invariant, and the corresponding matrices have λ down the main diagonal and ones down the first superdiagonal. Such matrices are called λ-*Jordan* matrices. These observations, along with the Fundamental Decomposition Theorem and the Cyclic Nilpotent Theorem combine to establish:

1.7 Jordan's Theorem. *Suppose T is any linear transformation on a finite dimensional complex vector space V, and suppose $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T. Then V has a basis with respect to which the matrix of T is block diagonal, where each block is a Jordan λ_j-matrix, and every eigenvalue λ_j is represented by at least one such block.*

The matrix produced in the last theorem is called the *Jordan canonical matrix* for T. Up to the order in which the Jordan λ_j-blocks occur, it is uniquely determined by T.

If T is the left-multiplication operator on $\mathbf{F}^n$ associated with an n by n $\mathbf{F}$-matrix A, then the Jordan canonical matrix for T is called the *Jordan canonical form of A*.

2 Toward the proof of (*).

Here, in detail, is the statement of the result we need to prove.

2.1 The Nilpotent-Splitting Theorem. *Suppose V is a real or complex vector space (not necessarily finite dimensional), and $T \in \mathcal{L}(V)$ is nilpotent of index m. Let x be a vector in V with $T^{m-1}x \neq 0$. Let V_1 be the span of the list $(x, Tx, \dots, T^{m-1}x)$ (so V_1 is T-invariant). Then there is a subspace W_1 of V that is T-invariant, such that $V = V_1 \oplus W_1$.*

We will get serious about the proof of this result in the next section. The task will be made considerably easier if we first negotiate some easy preliminaries about inverse images. For the rest of this section V is any vector space and $T \in \mathcal{L}(V)$. We begin with a familiar definition.

2.2 Definition. If S is a subset of V, then

$$T^{-1}(S) = \{x \in V : Tx \in S\}.$$

In words, $T^{-1}(S)$ is the set of vectors that T sends into S.

We will need some special properties of the inverse image of T when it acts on *subspaces* of V. We begin with a simple exercise.

2.3 Exercise. Suppose W is a subspace of V. Then

(a) $T^{-1}(W)$ is a subspace of V that contains $\operatorname{null} T$.

(b) $T(T^{-1}(W)) = W$. (Does this require that W be a subspace?)

(c) W is T-invariant if and only if $W \subset T^{-1}(W)$.

2.4 Proposition. $T^{-1}(T(W)) = W + \mathrm{null}\,(T)$.

PROOF. $T(W)$ is a subspace of V, so by part (a) of the last Exercise, so is $T^{-1}(T(W))$. Clearly $T^{-1}(T(W))$ contains W, and by part (a) of Exercise 2.3 it also contains $\mathrm{null}\,T$. Because it is a subspace, it therefore contains $W + \mathrm{null}\,T$.

Conversely, suppose $x \in T^{-1}(T(W))$. then $Tx \in T(W)$, so $Tx = Tw$ for some $w \in W$. By linearity, $T(x - w) = 0$, i.e. $x - w \in \mathrm{null}\,T$, so

$$x = w + (x - w) \in W + \mathrm{null}\,T,$$

which completes the proof. ///

2.5 Proposition. If W is a T-invariant subspace of V, then so is $T^{-1}(W)$.

PROOF. This follows from the containments:

$$T(T^{-1}(W)) = W \subset W + \mathrm{null}\,T \subset T^{-1}(W),$$

in which the first is just part (b) of Exercise 2.3, the second is obvious, and the third follows from parts (a) and (c) of Exercise 2.3. ///

3 Proof of the Nilpotent Splitting Theorem.

THE SITUATION: We are given a nilpotent linear transformation T on a vector space V, which need not be finite dimensional. Since m is the index of nilpotence of T, there exists a vector $x_1 \in V$ such that $T^{m-1}x_1 \neq 0$. We know from the Key Lemma (Lemma 1.1 that the list $(x_1, \ldots, T^{m-1}x_1)$ is linearly independent, and that its linear span V_1 is T-invariant.

TO FIND: A T-invariant subspace W_1 of V such that $V = V_1 \oplus W_1$.

The proof proceeds by induction on m, the index of nilpotence.

THE CASE $m = 1$. In this case T is the zero-operator on V, so x_1 is any non-zero vector in V, $V_1 = \mathrm{span}\,(x_1)$, and we can take W_1 to be any subspace of V that is complementary to V_1 (every subspace of V is invariant for the zero-operator).

THE INDUCTION HYPOTHESIS. Suppose $m > 1$ and suppose the result is true for all operators that are nilpotent of index $m - 1$. The proof now proceeds in four steps.

STEP I: *Pushing Down.* We focus on $\mathrm{ran}\,T$, which you can easily check is an invariant subspace for T on which T is nilpotent of index $m-1$, and which is clearly the span of the linearly independent list $(Tx_1, \ldots, T^{m-1}x_1) = (y_1, \ldots, T^{m-2}y_1)$, where $y_1 = Tx_1$. So we may apply the induction hypothesis to the restriction of T to $\mathrm{ran}\,T$, with V_1 replaced by the subspace

$$Y_1 = \mathrm{span}\,(y_1, \ldots, T^{m-2}y_1). \tag{1}$$

The result is a T-invariant subspace Y_2 of $\mathrm{ran}\,T$ such that

$$\mathrm{ran}\,T = Y_1 \oplus Y_2 \tag{2}$$

STEP II. *Pulling Back.* We claim that:

$$V = V_1 + T^{-1}(Y_2). \tag{3}$$

To prove this we write equation (2) as $T(V) = T(V_1) \oplus Y_2$, and distribute the inverse image of T over the sum (Exercise: prove that this is legal) to get:

$$V = T^{-1}(T(V_1)) + T^{-1}(Y_2).$$

Now by Proposition 2.4 we have $T^{-1}(T(V_1)) = V_1 + \operatorname{null} T$, hence

$$V = V_1 + \operatorname{null} T + T^{-1}(Y_2).$$

Since $\operatorname{null} T$ is a subspace of $T^{-1}(Y_2)$, the right-hand side of the last equation is just $V_1 + T^{-1}(Y_2)$, which proves (3).

Equation (3) is a step in the right direction, since it splits V into the sum of V_1 and the subspace $T^{-1}(Y_2)$ which is also T-invariant (by the T-invariance of Y_2 and Proposition 2.5). Unfortunately $T^{-1}(Y_2)$ may have nontrivial intersection with V_1, so it is not, in general, a complement for V_1. The rest of the argument seeks to remedy this deficiency by "cutting out the excess" from $T^{-1}(Y_2)$.

STEP III: *An important trivial intersection.* We claim that, even though V_1 and $T^{-1}(Y_2)$ may have nontrivial intersection, there is still *this* bit of good news:

$$V_1 \cap Y_2 = \{0\}. \tag{4}$$

To prove this, note that:

$$\begin{aligned} T(V_1 \cap Y_2) &= T(V_1) \cap T(Y_2) \\ &\subset Y_1 \cap Y_2 \\ &= \{0\}, \end{aligned}$$

where the first equality is a general property of mappings (can you prove it?), and the second follows from the fact that $T(V_1) = Y_1$ (definition of Y_1) and the T-invariance of Y_2. This shows that $V_1 \cap Y_2 \subset \operatorname{null} T$, or more precisely:

$$V_1 \cap Y_2 \subset \operatorname{null} T \cap V_1 \tag{5}$$

Now recall the definition of V_1:

$$V_1 = \operatorname{span}(x_1, \ldots, T^{m-1}x_1)$$

where $T^{m-1}x_1 \neq 0$, but $T^m x_1 = 0$ (recall that T^m is the zero-operator). From this it's immediate that the only vectors in V_1 that T annihilates are the scalar multiples of $T^{m-1}x$, in other words:

$$\operatorname{null}(T) \cap V_1 = \operatorname{span}(T^{m-1}x) = \operatorname{span}(T^{m-2}y) \subset Y_1,$$

where the last containment follows from our assumption that $m > 1$.

Summarizing:

$$V_1 \cap Y_2 \subset \operatorname{null} T \cap V_1 \subset Y_1,$$

hence

$$V_1 \cap Y_2 \subset Y_1 \cap Y_2 = \{0\},$$

as promised.

STEP IV: *Excising the excess.* Let's summarize what we have so far. We began with a T-invariant subspace V_1 on which our nilpotent operator T is cyclic, and have:

(a) Produced a T-invariant subspace Y_2 of $T(V)$ that is complementary to $Y_1 = T(V_1)$.

(b) Proved that $V = V_1 + T^{-1}(Y_2)$, where $T^{-1}(Y_2)$ is T-invariant, but may unfortunately intersect V_1.

(c) Proved that nevertheless $V_1 \cap Y_2 = \{0\}$.

Recall that $Y_2 \subset T^{-1}(Y_2)$ (just a restatement of the T-invariance of Y_2). By (c) above, $V_1 \cap T^{-1}(Y_2)$ intersects Y_2 in the zero-subspace, so if we let Z be any complement of the sum of these two subspaces in $T^{-1}(Y_2)$ we have

$$T^{-1}(Y_2) = Y_2 \oplus (V_1 \cap T^{-1}(Y_2)) \oplus Z. \tag{6}$$

Thus:

$$\begin{aligned} V &= V_1 + T^{-1}(Y_2), && \text{restating (b) above} \\ &= V_1 + Y_2 + (V_1 \cap T^{-1}(Y_2)) + Z, && \text{by (6)} \\ &= V_1 + Y_2 + Z, && \text{since } V_1 \cap T^{-1}(Y_2) \subset V_1. \end{aligned}$$

Now by its definition, $Y_2 + Z$ $(= Y_2 \oplus Z)$ lies in $T^{-1}(Y_2)$, and it has only trivial intersection with $V_1 \cap T^{-1}(Y_2)$. Thus it has only trivial intersection with V_1, hence $V = V_1 \oplus Y_2 \oplus Z$.

We claim that $V_2 = Y_2 \oplus Z$, is the subspace we seek. We have just shown that $V = V_1 \oplus V_2$, so it only remains to show that V_2 is T-invariant. But this is easy: we already know Y_2 is T-invariant, and although our choice of Z might have seemed arbitrary, it is a subspace of $T^{-1}(Y_2)$ so its image under T lies in Y_2. Thus $T(V_2) \subset Y_2 \subset V_2$, which finishes the proof. ///