THE FUNDAMENTAL THEOREM OF ALGEBRA VIA LINEAR ALGEBRA

KEITH CONRAD

Our goal is to use abstract linear algebra to prove the following result, which is called the fundamental theorem of algebra.

Theorem 1. *Any nonconstant polynomial with complex coefficients has a complex root.*

We will prove this theorem by reformulating it in terms of eigenvectors of linear operators. Let

$$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$

have degree $n \geq 1$, with $a_j \in \mathbf{C}$. By induction on n, the matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -a_{n-2} \\ 0 & 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}$$

satisfies $\det(\lambda I_n - A) = f(\lambda)$. Therefore Theorem 1 is a consequence of

Theorem 2. *For each $n \geq 1$, every $n \times n$ square matrix over $\mathbf{C}$ has an eigenvector. Equivalently, for each $n \geq 1$, every linear operator on an n-dimensional complex vector space has an eigenvector.*

Theorem 2 is also consequence of Theorem 1, so the two theorems are equivalent. In fact, the implication Theorem 1 $\Rightarrow$ Theorem 2 is usually how one first meets the fundamental theorem of algebra in a linear algebra course: it assures us that any complex square matrix has an eigenvector because the characteristic polynomial of the matrix has a complex root. But here, we will prove Theorem 2 without assuming Theorem 1, so we can deduce Theorem 1 as a consequence of Theorem 2. Our argument is a modification of a proof by H. Derksen [1]. It uses an interesting induction. Our starting point is the following lemma.

Lemma 3. *Fix an integer $m > 1$ and a field F. Suppose that, for every F-vector space V whose dimension is not divisible by m, every linear operator on V has an eigenvector. Then for every F-vector space V whose dimension is not divisible by m, any pair of commuting linear operators on V has a common eigenvector.*

The hypothesis of Lemma 3 is quite restrictive, so Lemma 3 does not apply in too many examples. For instance, it does not apply for any $m > 1$ when $F = \mathbf{Q}$. We will use Lemma 3 when m is a power of 2. I know no worthwhile applications of Lemma 3 for other values of m.

Proof. We induct on the dimension d which runs through integers not divisible by m. The case $d = 1$ is trivial: any nonzero vector in a one-dimensional space is an eigenvector of any linear operator on the space (any two such operators also commute).

Assume now that $d > 1$ is not divisible by m and we have settled all dimensions less than d which are not divisible by m. Let A_1 and A_2 be commuting linear operators on an

F-vector space V of dimension d. Since d is not divisible by n, the hypothesis of the lemma tells us A_1 has an eigenvalue in F, say λ. Let

$$U = \mathrm{im}(A_1 - \lambda I_V), \quad W = \ker(A_1 - \lambda I_V).$$

Each of these subspaces of V is A_1-stable (that is, $u \in U \Rightarrow A_1(u) \in U$, and $w \in W \Rightarrow A_1(w) \in W$), and $\dim_F W \geq 1$ since λ is an eigenvalue of A_1. Each of U and W is also A_2-stable since A_1 and A_2 commute. (For instance, if $u \in U$, write $u = A_1(v) - \lambda v$. Then

$$A_2(u) = A_2(A_1(v)) - A_2(\lambda v) = A_1(A_2(v)) - \lambda(A_2(v)) = (A_1 - \lambda I_V)(A_2(v))$$

is also in U.) Note $\dim_F U + \dim_F W = d$ is not divisible by m, so one of U or W has dimension not divisible by m. If the subspace U or W with dimension not divisible by m is a proper subspace of V, then A_1 and A_2 have a common eigenvector in that subspace (and thus in V) by induction. The other case is that U or W is all of V and the other subspace is $\{0\}$ since the dimensions add up to d, in $W = V$ since we already noted that $\dim_F W$ is positive. From the equation $W = V$ we see every vector in V is an eigenvector for A_1, and one of them is an eigenvector for A_2 since V has dimension not divisible by m. □

Corollary 4. *For every real vector space V whose dimension is odd, any pair of commuting linear operators on V has a common eigenvector.*

Proof. In Lemma 3, use $F = \mathbf{R}$ and $m = 2$. Any linear operator on an odd-dimensional real vector space has an eigenvector since the characteristic polynomial has odd degree and therefore has a real root, which is a real eigenvalue. Any real eigenvalue leads to a real eigenvector. □

Note Lemma 3 and Corollary 4 are *not* saying commuting linear operators have a common eigenvalue, but rather a common eigenvector. A common eigenvector does not have to occur with the same eigenvalue for A_1 and A_2.

Example 5. Let

$$A_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 2 \end{pmatrix},$$

viewed as linear operators on $\mathbf{R}^3$. A direct calculation shows $A_1A_2 = A_2A_1$, so there must be a common eigenvector for A_1 and A_2 in $\mathbf{R}^3$. One common eigenvector is

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

with eigenvalue 1 for A_1 and 3 for A_2. In fact, this is the only common eigenvector for A_1 and A_2 in $\mathbf{R}^3$, up to scaling.

Now we turn to the proof of the Fundamental Theorem of Algebra.

Proof. (Derksen) Our proof will be by induction on the highest power of 2 dividing n. That is, writing $n = 2^k n'$, where $k \geq 0$ and n' is odd, we will prove the theorem by induction on k. Let's be concrete about how the induction proceeds, before we carry it out. First we will treat the case $k = 0$, which means n is odd.

For the base case ($k = 0$), let n be odd. Pick $A \in \mathrm{M}_n(\mathbf{C})$. To show A has an eigenvector in $\mathbf{C}^n$, we will associate to A some linear operators on the space

$$\mathcal{H}_n = \{M \in \mathrm{M}_n(\mathbf{C}) : M^* = M\},$$

where $M^* = \overline{M}^\top$ denotes the conjugate-transpose of M. (In coordinates, $(m_{ij})^* = (\overline{m}_{ji})$. Note $(MN)^* = N^*M^*$.) Matrices in $\mathcal{H}_n$ are called Hermitian. They form a real vector space, with dimension n^2 over $\mathbf{R}$. In particular, $\mathcal{H}_n$ has odd dimension over $\mathbf{R}$.

We can decompose any $C \in \mathrm{M}_n(\mathbf{C})$ as

$$C = \frac{C + C^*}{2} + i\frac{C - C^*}{2i},$$

where $(C+C^*)/2$ and $(C-C^*)/2i$ are Hermitian. (Symbolically, $\mathrm{M}_n(\mathbf{C}) = \mathcal{H}_n + i\mathcal{H}_n$.) For $B \in \mathrm{M}_n(\mathbf{C})$, take $C = AB$:

$$AB = \frac{AB + B^*A^*}{2} + i\frac{AB - B^*A^*}{2i}. \tag{0.1}$$

When $B \in \mathcal{H}_n$, (0.1) becomes

$$AB = \frac{AB + BA^*}{2} + i\frac{AB - BA^*}{2i}. \tag{0.2}$$

Looking at (0.2), we are inspired to consider the $\mathbf{R}$-linear operators $L_1, L_2 \colon \mathcal{H}_n \to \mathcal{H}_n$ defined by

$$L_1(B) = \frac{AB + BA^*}{2}, \quad L_2(B) = \frac{AB - BA^*}{2i}.$$

The operators L_1 and L_2 commute with each other (check!). Since $\mathcal{H}_n$ has odd (real) dimension, Corollary 4 implies L_1 and L_2 have a common eigenvector in $\mathcal{H}_n$. That is, there is some nonzero $B \in \mathcal{H}_n$ such that

$$L_1(B) = \lambda_1 B, \quad L_2(B) = \lambda_2 B,$$

where $\lambda_1, \lambda_2 \in \mathbf{R}$. Therefore

$$AB = L_1(B) + iL_2(B) = (\lambda_1 + i\lambda_2)B.$$

Any nonzero column vector of B is an eigenvector in $\mathbf{C}^n$ for A. That concludes the proof of the Fundamental Theorem of Algebra for odd n, which was the base case. (Notice that the base case of a proof by induction is not always trivial!)

Now we treat the inductive step. The following is our *inductive hypothesis* for $k \geq 1$: for any $d \geq 1$ where the highest power of 2 dividing d is less than 2^k (that is, d is not divisible by 2^k), any linear operator on any d-dimensional complex vector space has an eigenvector. It follows, by Lemma 3 with $F = \mathbf{C}$, that any pair of commuting linear operators on a vector space of such a dimension has a common eigenvector. The hypothesis of Lemma 3 is precisely our inductive hypothesis.

(Derksen's paper [1] has a more involved inductive hypothesis, because he is trying to prove not only the Fundamental Theorem of Algebra, but also that any finite set of commuting linear operators on a complex vector space – not just two commuting operators as in Lemma 3 – has a common eigenvector.)

The case to consider now is linear operators on complex vector spaces with dimension divisible by 2^k but not by a higher power of 2. By choosing bases to convert operators into matrices, it will suffice to treat the case of matrices acting on concrete spaces of the form $\mathbf{C}^n$ instead of linear operators acting on abstract complex vector spaces. However, in the course of our argument for matrix operators, we will be using linear operators on vector spaces other than $\mathbf{C}^n$ (specifically, we will be using linear operators on the vector space $\mathrm{Sym}_n(\mathbf{C})$), so it is important to have the abstract point of view in mind.

Let n be a positive integer such that the highest power of 2 dividing n is 2^k. Let A be any $n \times n$ complex matrix. We want to show A has an eigenvector in $\mathbf{C}^n$.

Consider the space of $n \times n$ complex symmetric matrices,

$$\mathrm{Sym}_n(\mathbf{C}) = \{M \in \mathrm{M}_n(\mathbf{C}) : M^\top = M\}.$$

This is a complex vector space with dimension $\frac{n(n+1)}{2}$. Notice that the highest power of 2 dividing $\frac{n(n+1)}{2}$ is 2^{k-1}. Since $\frac{n(n+1)}{2}$ is not divisible by 2^k, any pair of commuting linear operators on $\mathrm{Sym}_n(\mathbf{C})$ has a common eigenvector. (This is the application of Lemma 3 which we made right after formulating the inductive hypothesis.) We are going to apply this result to a pair of operators on $\mathrm{Sym}_n(\mathbf{C})$ built from the matrix A.

Define $\mathcal{L}_1, \mathcal{L}_2 \colon \mathrm{Sym}_n(\mathbf{C}) \to \mathrm{Sym}_n(\mathbf{C})$ by

$$\mathcal{L}_1(B) = AB + BA^\top, \quad \mathcal{L}_2(B) = ABA^\top.$$

The maps $\mathcal{L}_1$ and $\mathcal{L}_2$ are $\mathbf{C}$-linear operators and commute (check!). Therefore, by the application of Lemma 3 in the previous paragraph, $\mathcal{L}_1$ and $\mathcal{L}_2$ have a common eigenvector: some nonzero $B \in \mathrm{Sym}_n(\mathbf{C})$ satisfies

$$AB + BA^\top = \lambda B, \quad ABA^\top = \mu B,$$

where $\lambda, \mu \in \mathbf{C}$. Applying A to the first equation on the left, and using the second equation to simplify, we find

$$A^2B + \mu B = \lambda AB,$$

so

$$(A^2 - \lambda A + \mu)B = 0.$$

Since any complex number has a complex square root, the quadratic formula shows any quadratic polynomial over $\mathbf{C}$ splits into linear factors. (This is the first time we have used something about $\mathbf{C}$ which is not true for $\mathbf{R}$. Up until this point, we could have been trying to prove the Fundamental Theorem of Algebra for operators on real vector spaces, and even invoked Lemma 3 with $F = \mathbf{R}$ as part of the induction. But the proof would break down now, because real quadratics generally don't have real roots.) Factor the complex polynomial $z^2 - \lambda z + \mu$ as $(z - \alpha)(z - \beta)$. Then the matrix product $(A - \alpha I)(A - \beta I)$ is O, so

$$(A - \alpha I)((A - \beta I)B) = O.$$

If $(A - \beta I)B = O$, then any nonzero column vector of B is an eigenvector of A (with eigenvalue β). If $(A - \beta I)B \neq O$, then any nonzero column vector of $(A - \beta I)B$ is an eigenvector of A (with eigenvalue α).

That concludes the Fundamental Theorem of Algebra. □

Derksen's base case made essential use of complex conjugation on $\mathbf{C}$, but the only place where Derksen's inductive argument uses something about $\mathbf{C}$ which is not true for $\mathbf{R}$ is near the end: every complex quadratic polynomial splits into linear factors. Up until that point, we could have been trying to prove the Fundamental Theorem of Algebra for $\mathbf{R}$. The base case of odd degree would be trivial since real polynomials of odd degree have a real root and in the inductive step we would use Lemma 3 with $F = \mathbf{R}$. But the proof of the inductive step for $\mathbf{R}$ instead of $\mathbf{C}$ would break down at the end when we try to factor a real quadratic over $\mathbf{R}$, because real quadratics generally don't have real roots.

While some motivation for L_1 and L_2 comes from (0.2), and the definition $\mathcal{L}_1(B) = AB + BA^\top$ is, at least up to a factor of $1/2$, the symmetric part of the product AB (and thus resembles the definition of L_1 and L_2 as the "real" and "imaginary" parts of a matrix product in $\mathcal{H}_n$), the definition of $\mathcal{L}_2$ seems to come out of nowhere. One rationale for the definition of $\mathcal{L}_2$ is the following. Given a square matrix C, two ways of producing a symmetric matrix from C are addition and multiplication with the transpose: $C + C^\top$ and $CC^\top$. Applying this to the matrix product $C = AB$ (with B symmetric, as in the proof) gives rise to the expressions

$$AB + (AB)^\top = AB + B^\top A^\top = AB + BA^\top$$

and

$$(AB)(AB)^\top = AB^2A^\top.$$

While $AB + BA^\top$ is linear in B (and defines $\mathcal{L}_1(B)$), $AB^2A^\top$ is not linear in B. However, if we work instead with $ABA^\top$, we do get a linear function of B, and thus we are led to $\mathcal{L}_2(B)$.

Two extended treatments of the fundamental theorem of algebra are in [2] and [3]. Chapter 4 of [2] gives an historical treatment of this result, while [3] proves the result in twelve different ways (six in the main text and six in appendices). Derksen's use of 2-power divisibility to view the passage from n to $n(n+1)/2$ as a *reduction* step is reminiscent of Laplace's proof of the fundamental theorem of algebra, which is in the appendix to [2, Chap. 4].

References

[1] H. Derksen, *The Fundamental Theorem of Algebra and Linear Algebra*, Amer. Math. Monthly **110** (2003), 620–623.

[2] H-D. Ebbinghaus *et al.*, "Numbers," Springer–Verlag, New York, 1991.

[3] B. Fine and G. Rosenberger, "The Fundamental Theorem of Algebra," Springer–Verlag, New York, 1997.