A linear algebra proof of the fundamental theorem of algebra

Andrés E. Caicedo

May 18, 2010

Abstract

We present a recent proof due to Harm Derksen, that any linear operator in a complex finite dimensional vector space admits eigenvectors. The argument avoids the use of the fundamental theorem of algebra, which can then be deduced from it. Our presentation avoids any appeal to the theory of determinants.

2010 Mathematics Subject Classification. 15-01, 15A18.

Keywords and phrases. Fundamental theorem of algebra, eigenvectors.

1 Introduction

The goal of this note is to present a reasonably self-contained proof of the fundamental theorem of algebra:

Theorem 1.1 (Gauß)**.** *Any non-constant polynomial $p(x)$ with complex coefficients has a complex root.*

The proof described here is due to Derksen [Der03] and uses linear algebra. The main point is that linear operators admit eigenvectors, and this can be proved without the use of the fundamental theorem:

Theorem 1.2. *Every linear operator in a finite dimensional complex vector space admits an eigenvector.*

Theorem 1.1 is an immediate corollary of Theorem 1.2, once one shows that any non-constant polynomial is the minimal polynomial of some operator.

Keeping with the spirit of Axler's book [Axl97], I organize the presentation in a way that avoids the use of determinants. This adds a couple of

slight complications to the argument in Derksen [Der03], since we need to reprove its Lemma 4 and Corollary 8. They correspond here to Lemma 4.1, the nice proof of which was suggested by David Milovich, and Lemma 2.1. Let me emphasize that nothing here is new or due to me.

It is well known that the fundamental theorem of algebra is really a consequence of the following two facts:

Fact 1.3. *Odd degree polynomials with real coefficients have a real root.* □

Fact 1.4. *Positive real numbers have real square roots.* □

What this means, that these two facts suffice, is that given any real closed field R, the extension $C = R(\sqrt{-1})$ is algebraically closed. Recall:

Definition 1.5. *R is a real closed field iff*

1. *R is a field.*
2. *There is a linear order $<$ of R that makes R into an ordered ring, i.e., for all $a, b, c \in R$, if $a \leq b$ then $a + c \leq b + c$, and if also $0 \leq c$, then $a \cdot c \leq b \cdot c$.*
3. *Odd degree polynomials $p \in R[x]$ admit a root in R.*
4. *Positive elements of R have a a square root.*

We do not need this level of generality, and work with $\mathbb{R}$ and $\mathbb{C}$ explicitly, but the reader may want to keep in mind that this is really all that we are using, and therefore the argument in this note actually proves:

Theorem 1.6. *Let R be a real closed field. Then $C = R(\sqrt{-1})$ is algebraically closed.* □

On the other hand, one can give examples of fields R satisfying *1.–3.* or *1., 2., 4.* of the definition above and such that $R(\sqrt{-1})$ is not algebraically closed, so Facts 1.3 and 1.4 are necessary as well. Naturally, the argument below uses Fact 1.3 and 1.4: Fact 1.3 is used explicitly in the proof of Lemma 4.1. Fact 1.4 is used in the proof of Lemma 4.4, as follows: Fact 1.4 implies that any complex number admits a complex square root. In effect, if $a, b \in \mathbb{R}$ then, letting $c = \sqrt{a^2 + b^2}$, we have that $\sqrt{\frac{c+a}{2}} + i\sqrt{\frac{c-a}{2}}$ is a square root of $a + ib$. This allows us to use the familiar quadratic formula, to conclude that any quadratic polynomial factors over $\mathbb{C}$.

Our notation is standard; we refer to Axler [Axl97].

1.1 Acknowledgements

Thanks are in order to the National Science Foundation for partial support through grant DMS-0801189. I also want to thank David Milovich for his argument for Lemma 4.1.

2 The fundamental theorem of algebra

Theorem 1.1 follows from Theorem 1.2 by a well-known procedure.

First, we need to recall some standard facts: Let V be a finite dimensional complex vector space and let $T : V \to V$ be linear. We can define as usual the **minimal** polynomial m_T of T. This is a monic polynomial with the property that, for any polynomial q, $m_T|q$ iff $q(T) = 0$, see Axler [Axl97, Theorem 8.34].

Similarly, for any vector v, we can define the minimal polynomial $m_{T,v}$ of T for v. This is a monic polynomial with the property that, for any polynomial q, $m_{T,v}|q$ iff $q(T)v = 0$. In particular, $m_{T,v}|m_T$.

Thanks to Theorem 1.2, we can now define the **characteristic** polynomial c_T of T and prove the Cayley-Hamilton theorem that $c_T(T) = 0$. Recall that c_T is monic and $\deg(c_T) = \dim(V)$. None of this requires the use of determinants or the fundamental theorem of algebra; note that the use of the fundamental theorem in this argument, as described in Axler [Axl97] (namely, the conjunction of Theorems 5.13, 8.10, and 8.20), is really a use of Theorem 1.2.

Since $m_T|c_T$, it follows that $\deg(m_T) \leq \dim(V)$.

Lemma 2.1. *Let p be a monic polynomial of degree n with complex coefficients. Let V be a complex vector space of dimension n. Then there is a linear operator $T : V \to V$ such that $m_T = p$.*

Proof. Let

$$p(x) = x^n + \sum_{k=0}^{n-1} a_k x^k.$$

We define a linear operator $T : V \to V$ with $m_T = p$. Let $v_1, \ldots, v_n$ be a basis for V. By linearity, it suffices to specify the values Tv_j for $1 \leq j \leq n$.

First, set $Tv_j = v_{j+1}$ for $1 \leq j \leq n-1$, and note that for any such T, we have $T^j v_1 = Tv_j$ for $1 \leq j \leq n$. In particular, since the v_j are linearly independent, it follows that $\deg(m_{T,v_1}) \geq n$ and, therefore, $\deg(m_{T,v_1}) = n$ since $m_{T,v_1}|m_T$ and $\deg(m_T) \leq n$. In fact, $m_{T,v_1} = m_T$, since both are monic.

Now set $T^n v_1 = Tv_n = -\sum_{k=0}^{n-1} a_k v_{k+1}$, so

$$T^n v_1 = -\sum_{k=0}^{n-1} a_k T^k v_1,$$

or

$$p(T)v_1 = 0.$$

It follows that $p = m_{T,v_1} = m_T$. □

For example, taking $V = \mathbb{C}^n$ and letting the v_j be the standard vectors e_j, the matrix representation of T in terms of the standard basis is

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & 0 & \dots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & -a_{n-2} \\ 0 & 0 & 0 & \dots & 1 & -a_{n-1} \end{pmatrix}.$$

Proof of Theorem 1.1. Let p be a non-constant polynomial of degree n with complex coefficients. We need to show that p admits a complex root. Without loss, p is monic. By Lemma 2.1, there is a linear operator $T : \mathbb{C}^n \to \mathbb{C}^n$ such that $p = m_T$, so $p(T) = 0$. Using Theorem 1.2, let v be an eigenvector of T with eigenvalue λ. Then

$$0 = p(T)v = p(\lambda)v,$$

and it follows that λ is a root of p. □

3 Commuting operators

Theorem 1.2 shows that if V is a complex finite dimensional vector space and $T : V \to V$ is linear, then T admits an eigenvector. It turns out that this statement can be easily strengthened as follows:

Corollary 3.1. *Let V be a complex finite dimensional vector space. Let $\mathcal{F}$ be a (possibly infinite) family of pairwise commuting linear operators on V, i.e., for any $T, S \in \mathcal{F}$, $TS = ST$. Then the operators in $\mathcal{F}$ admit a common eigenvector, i.e., there is a vector $v \neq 0$ that is an eigenvector of each $T \in \mathcal{F}$.*

The proof of Theorem 1.2 actually requires the use of an appropriate (local) version of Corollary 3.1 where $|\mathcal{F}| = 2$. Accordingly, we begin by showing this. Since the proof for arbitrary $\mathcal{F}$ is only microscopically longer than the argument for $|\mathcal{F}| = 2$, we present the general version. This corresponds to Derksen [Der03, Lemma 3], although Derksen only considers finite families $\mathcal{F}$.

Let $E(K, d)$ be the following statement about an arbitrary field K and a positive integer d:

> If V is a vector space over K of finite dimension, and $d \nmid \dim(V)$, then any family $\mathcal{F}$ of pairwise commuting linear maps from V to itself admits a common eigenvector.

Let $E(K, d, r)$ be the particular case of the statement above where we add the requirement that $|\mathcal{F}| = r$.

Lemma 3.2. *For any field K and any d, $E(K, d, 1)$ implies $E(K, d)$.*

Proof. First we argue by induction on $r = |\mathcal{F}|$ that the result holds when $\mathcal{F}$ is finite. Accordingly, assume $E(K, d, r)$. We prove $E(K, d, r + 1)$ by induction on the dimension of the vector space under consideration. Let V be a K-vector space of dimension n not divisible by d, let $\mathcal{F}$ be a family of $r + 1$ commuting linear maps from V to itself, and suppose that whenever X is a K-vector space of dimension $m < n$ such that $d \nmid m$, then any family of $r+1$ commuting linear maps of X to itself admits a common eigenvector.

Fix $T \in \mathcal{F}$. By $E(K, d, 1)$, the map T admits an eigenvalue λ. Let

$$W = \operatorname{null}(T - \lambda I)$$

and

$$U = \operatorname{ran}(T - \lambda I).$$

Then both W and U are S-invariant for all $S \in \mathcal{F}$, by the commutativity assumption. Moreover, W is nontrivial and therefore $U \neq V$.

Recall that $\dim(W) + \dim(U) = n$. It follows that either $d \nmid \dim(W)$ or $d \nmid \dim(U)$. If $W = V$, every non-zero vector is an eigenvector of T, and we are done by the inductive assumption on r. Otherwise, we are done by the inductive assumption on n, since either W or U has dimension strictly less than n and not divisible by d.

Now we deduce the general case: Let $\mathcal{F}$ be a commuting family of operators on the finite dimensional K-vector space V such that $d \nmid \dim(V)$, and suppose that they do not have a common eigenvector. Let $T_1 \in \mathcal{F}$, and

consider a maximal linearly independent family $v_1, \ldots, v_n$ of eigenvectors of T_1. For $1 \leq i \leq n$, let $T_{i+1} \in \mathcal{F}$ be an operator for which v_i is not an eigenvector. Then $T_1, \ldots, T_{n+1}$ is a finite family of commuting operators on V without a common eigenvector, contradiction. □

4 Eigenvectors for operators in finite dimensional complex vector spaces

Now we present the argument in Derksen [Der03] for Theorem 1.2. The proof proceeds by a "thinning out" process: We will prove Theorem 1.2 by showing by induction that $E(\mathbb{C}, 2^k)$ holds for all k, although (as mentioned above) only the case of two commuting operators needs to be considered. This clearly implies Theorem 1.2 and Corollary 3.1, since any finite dimension is eventually covered by these statements as k increases.

Lemma 4.1. *$E(\mathbb{R}, 2)$ holds.*

Explicitly, Lemma 4.1 asserts that if V is an odd dimensional real vector space, then any family of commuting linear operators $T : V \to V$ admits a common eigenvector.

The proof that follows was found by David Milovich (personal communication); it is in the spirit of Axler [Axl97]. However, I feel it is easier than the argument there, Theorem 5.26 (and it has the additional advantage of not depending on the fundamental theorem of algebra, of course).

Proof. By Lemma 3.2, it suffices to establish $E(\mathbb{R}, 2, 1)$. The argument is by induction on $n = \dim(V)$, with the case $n = 1$ being obvious.

Assume then that n is odd, that the result holds for all positive odd dimensions smaller than n, and that $T : V \to V$ is linear. We may also assume that the result is false for n, and argue by contradiction. It follows that $T - \lambda I$ is invertible for all $\lambda \in \mathbb{R}$ (where I is the identity operator), and that if W is a proper T-invariant subspace of V, then $\dim(W)$ is even.

Claim 4.2. *Any $v \in V$ is contained in a proper T-invariant subspace.*

Proof. This is clear if $v = 0$. If $v \neq 0$, since $v, Tv, \ldots, T^n v$ are linearly dependent, there is a least $k \leq n$ for which there is a polynomial $p \in \mathbb{R}[x]$ such that $p(T)v = 0$ and p is non-constant and of degree k. If $k < n$ we are done, because then $\operatorname{span}\{v, Tv, \ldots, T^{k-1}v\}$ is T-invariant and of dimension at most k. If $k = n$, we use that p has a real root, to factor

$$p(T) = (T - \lambda I)q(T)$$

for some $\lambda \in \mathbb{R}$ and $q \in \mathbb{R}[x]$ of degree $n-1$. Since $T - \lambda I$ is invertible, we have that $q(T)v = 0$, contradicting the minimality of k. □

Let W be a maximal proper T-invariant subspace. Let $v \notin W$, and let U be a proper T-invariant subspace with $v \in U$. Note that $W + U$ is also T-invariant, and strictly larger than W, since it contains v. By maximality of W, we must have $W + U = V$. Since W and U have even dimension, it follows that $\dim(W \cap U)$ is odd. But $W \cap U$ is T-invariant and proper, contradiction. □

Lemma 4.3. *$E(\mathbb{C}, 2)$ holds.*

Proof. Again, it suffices to establish $E(\mathbb{C}, 2, 1)$. We use Lemma 4.1. Let n be odd, consider a complex vector space V of dimension n, and let $T : V \to V$ be linear. Derksen's first key idea here is the following:

> Let $\mathcal{L}(V)$ denote the space of $\mathbb{C}$-linear transformations of V to itself. Consider the linear map $f_T : \mathcal{L}(V) \to \mathcal{L}(V)$ given by $f_T(S) = TS$. If f_T admits an eigenvector S, we are done: Let $v \in \mathrm{ran}(S)$. Then v is an eigenvector of T.

In order to ensure that f_T has an eigenvector, one possible approach is to try to identify an f_T-invariant subset W of $\mathcal{L}(V)$ that is a *real* vector space of *odd* dimension, and deduce the existence of the eigenvector S from $E(\mathbb{R}, 2, 1)$. This approach, however, cannot succeed directly, since T may not have any real eigenvalues. It is here that Derksen's second key idea appears.

Rather than considering f_T directly, Derksen considers V as an inner product space, and looks separately at what would correspond to the real and imaginary parts of f_T,

$$R_T(S) = \frac{1}{2}(TS + S^*T^*),$$

and

$$I_T(S) = \frac{1}{2i}(TS - S^*T^*).$$

Here, as usual, S^* is the **adjoint** of S, the unique map in $\mathcal{L}(V)$ such that

$$\langle Su, v\rangle = \langle u, S^*v\rangle$$

for all vectors $u, v \in V$, see Axler [Axl97, Chapter 6].

Note that $TS = R_T(S) + iI_T(S)$. Also, $\mathrm{ran}(R_T)$ and $\mathrm{ran}(I_T)$ are both contained in $\mathcal{H}(V)$, the set of self-adjoint maps from V to itself, i.e., those

linear maps S such that $S = S^*$. Although $\mathcal{H}(V)$ is not a subspace of $\mathcal{L}(V)$ since it is not closed under scalar multiplication (because $\langle Su, u\rangle \in \mathbb{R}$ for all $u \in V$ whenever $S \in \mathcal{H}(V)$), this suggests us to take $W = \mathcal{H}(V)$, considered as a **real** vector space.

Note that $\dim_{\mathbb{R}}(W) = n^2$, an odd number. In effect, fixing an orthonormal basis $v_1, \ldots, v_n$ for V, a basis for W is obtained by taking:

- For $1 \leq j \leq n$, the unique linear map T_j such that $T_j v_j = v_j$ and $T_j v_k = 0$ if $j \neq k$.
- For $1 \leq j < k \leq n$, the unique linear map $T_{j,k}$ such that $T_{j,k} v_j = v_k$, $T_{j,k} v_k = v_j$, and $T_{j,k} v_l = 0$ for $l \notin \{j, k\}$.
- For $1 \leq j < k \leq n$, the unique linear map $T'_{j,k}$ such that $T'_{j,k} v_j = \sqrt{-1} v_k$, $T'_{j,k} v_k = -\sqrt{-1} v_j$, and $T'_{j,k} v_l = 0$ for $l \notin \{j, k\}$.

By construction, each of these operators is self-adjoint, and there are

$$n + 2\left(\frac{n(n-1)}{2}\right) = n^2$$

of them. Using that $v = \sum_{j=1}^{n} \langle v, v_j\rangle v_j$ for any v, it is straightforward to verify that they are linearly independent and span W.[1]

It is here that Lemma 3.2 is first used: A direct computation verifies that R_T and I_T commute. Since W is R_T- and I_T-invariant, Lemma 4.1 implies that they admit a common eigenvector. This gives the result. □

Lemma 4.4. *For any positive integer k, $E(\mathbb{C}, 2^k)$ holds.*

Proof. Once again, it suffices to show $E(\mathbb{C}, 2^k, 1)$. The proof is by induction on k, with the case $k = 1$ being Lemma 4.3. Assume the result for k. Let V be a space whose dimension n is divisible by 2^k but not 2^{k+1}. Let $T : V \to V$ be linear.

Before proceeding to the argument, note that we must definitely use here Fact 1.4, since we have not appealed to it yet. As explained above, we will use it by showing that, for some nonzero vector v, there is a quadratic polynomial $p(x) = x^2 + \alpha x + \beta \in \mathbb{C}[x]$ such that $p(T)v = 0$. By factoring p

[1] Being even more explicit, recall the well-known and easily verifiable fact that if $v_1, \ldots, v_n$ is an orthonormal basis of V, and $T \in \mathcal{L}(V)$ then, if the matrix $A = (a_{i,j})_{i,j=1}^n$ represents T with respect to this basis, the matrix $A^* = (b_{i,j})_{i,j=1}^n$ representing T^* with respect to the same basis satisfies $b_{i,j} = \overline{a_{j,i}}$, see Axler [Axl97, Proposition 6.47]. This means that if $T \in \mathcal{H}$, then T is completely determined by the numbers $a_{i,j}$ for $1 \leq i \leq j \leq n$, and the basis described above is then the obvious one.

into linear terms, we conclude that T admits an eigenvector. To find v, we proceed as in Lemma 4.3, and instead find a non-zero operator S such that $p(T)S = 0$. We can then take as v any vector in $\mathrm{ran}(S)$.

To find S and p, we use the inductive assumption, and consider a complex subspace W of $\mathcal{L}(V)$ of dimension not divisible by 2^k, and exhibit two commuting linear operators $L_1, L_2 : W \to W$. The desired S will then be a common eigenvector.

To see how this could work, note we need that

$$(T^2 + \alpha T + \beta I)S = 0,$$

or $\beta S = -T^2 S - \alpha TS$, which we can rewrite as

$$\beta S = T(-\alpha S - TS).$$

If L_1 and L_2 are the two operators on W we are looking for, it would be reasonable to have $-\alpha$ to be the eigenvalue of L_1 corresponding to S,

$$L_1(S) = -\alpha S,$$

and to set

$$L_2(U) = T(L_1(U) - TU),$$

so β would be the eigenvalue of L_2 corresponding to S.

Derksen's final trick makes this work: Let $W = \mathcal{S}(V)$ be the complex vector space of **skew-symmetric** maps from V to itself. Perhaps the easiest way to define W consists in fixing an inner product and an orthonormal basis for V, and taking as W the set of all those S whose matrix $A = (a_{i,j})_{i,j=1}^n$ with respect to this basis satisfies $A^\intercal = -A$, i.e., $a_{i,j} = -a_{j,i}$ whenever $1 \le i \le j \le n$. We write $S^\intercal$ for the operator whose matrix is $A^\intercal$.[2]

A straightforward argument as in the previous lemma shows that this is a (complex) space of dimension $\dfrac{n(n-1)}{2}$. Note that this number is not divisible by 2^k, by our assumption on n.

Now let $L_1, L_2 \in W$ be given by $L_1(S) = TS + ST^\intercal$ and $L_2(S) = T(L_1(S) - TS) = TST^\intercal$. Note L_1, L_2 commute, and we are done. □

As explained in Section 3, Theorem 1.2 follows immediately from Lemma 4.4, concluding the proof.

[2]Of course, symmetry ($S^\intercal = S$) and skew-symmetry could also be defined without reference to matrices, but we do not need this here. Just as well, we could have adopted as W the space of symmetric maps, which has complex dimension $\dfrac{n(n+1)}{2}$.

References

[Axl97] Axler, Sheldon. **Linear algebra done right**, Springer, second edition (1997).

[Der03] Derksen, Harm. *The fundamental theorem of algebra and linear algebra*, American Mathematical Monthly, **110 (7)** (2003), 620–623.