ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 22, Number 2, Spring 1992

TRANSCENDENTAL OPERATORS ON A BANACH SPACE

FRED RICHMAN

ABSTRACT. Let A be a bounded operator on a normed linear space V. If $p(A) \neq 0$ for each nonzero polynomial p of degree less than n, then there exists $x \in V$ such that $p(A)x \neq 0$ for each nonzero polynomial p of degree less than n. We give a proof of this theorem that is constructive in the sense of Errett Bishop.

A classical theorem [**4**, 3.3.15] states that if the minimal polynomial of a matrix M is equal to its characteristic polynomial p, then M is similar to the companion matrix of p. Another way of saying this is that, given a linear transformation A on an n-dimensional vector space V, then

(1) *if the transformations $I, A, A^2, \ldots, A^{n-1}$ are linearly independent, then there exists $x \in V$ such that $x, Ax, A^2x, \ldots, A^{n-1}x$ are linearly independent.*

In fact, (1) holds whether or not n is the dimension of the space V; it holds even when the dimension of V is infinite. The hypothesis of (1) is equivalent to the condition that $p(A) \neq 0$ for any nonzero polynomial p of degree less than n. The purpose of this note is to give a constructive proof (in the sense of Bishop [**1**]) of (1) when A is a bounded operator on a Banach space. We shall show that the set of vectors x that work for (1) is open and dense, so that if *all* powers of A are independent, that is, if A is *transcendental*, then the Baire category theorem [**2**, 3.9 Chapter 4] constructs a single x that works for all powers of A.

Except for this last appeal to the Baire category theorem, the completeness of the Banach space plays no role, so we state our results for a normed linear space The field of scalars can be either the real numbers or the complex numbers.

Although (1) holds in complete generality from a classical point of view, there are serious barriers to a constructive proof. In fact, in [**5**] it

Received by the editors on March 20, 1989, and in revised form on August 31, 1989.
1980 *Mathematics subject classifications.* Primary 03F65, 45C05.
Research supported by NSF grant DMS-8802062.

is shown that (1) does not admit a constructive proof even for finitely generated vector spaces over a discrete field (a field with decidable equality), although the natural constructive contrapositive

(2) *if for each $x \in V$ the vectors $x, Ax, A^2x, \ldots, A^{n-1}x$ are linearly dependent, then the transformations $I, A, A^2, \ldots, A^{n-1}$ are linearly dependent.*

holds in this case. It is important to remember here that a constructive proof that vectors are linearly dependent requires that you construct the coefficients of a dependence relation.

Classically we address (2) and assume that for each $x \in V$ there is a monic polynomial p_x of least degree such that $p_x(A)x = 0$. Given x and y in V, we show that there is z in V such that p_z is the least common multiple of p_x and p_y; this follows, for instance, from the general theory of finitely generated modules over a Euclidean ring [**3**, Theorem 4.5.1]. So if $\deg p_x < n$ for all x, then there is x such that $p_x(A) = 0$.

Constructively, we need not be able to compute the polynomials p_x, or the least common multiple of p_x and p_y, or the vector z. And even if we could do all this, we would need to be able to decide whether $p_x(A) = 0$, or whether $p_x(A)y \neq 0$ for some $y \in V$, in order to complete the proof of the contrapositive. Moreover, there would still remain the problem of converting our knowledge that $\deg p_x < n$ cannot hold for all x, into a construction of x such that $\deg p_x \geq n$.

The reader should keep in mind the following facts about working in a constructive context. You can get arbitrarily close rational approximations to a given real number r, but you may not be able to tell whether or not $r = 0$. A real number r is *nonzero* if you can find a positive integer n such that $|r| \geq 1/n$. A vector in a normed linear space is nonzero if its norm is nonzero. If the sum of two vectors in a normed linear space is nonzero, then one or the other of the vectors is nonzero (and you can tell which). The vectors $v_1, \ldots, v_n$ are *linearly independent* if each c_i is small whenever $\sum_{i=1}^n c_i v_i$ is small (rather than the usual definition with 'small' replaced by 'zero').

Lemma 1. *Let A be a bounded operator on a normed linear space V, and suppose $x \in V$ is such that $x, Ax, \ldots, A^{n-1}x$ are linearly independent. Then there is a monic polynomial p_x of degree n such*

that if $p_x(A)x \neq 0$, then $q(A)x$ is bounded away from 0 as q ranges over all monic polynomials of degree n.

Proof. Let H be the space spanned by $x, Ax, \ldots, A^{n-1}x$. As H is finite-dimensional, H is complete and *located* [**2**, 6.2 Chapter 4], that is, we can compute the distance from any point v in V to H, and we can find points in $h \in H$ such that $||v-h||$ approximates that distance. By [**2**, 3.8 Chapter 4], for each $v \in V$, we can find $h \in H$ such that if $v \neq h$, then the distance from v to H is nonzero; if V is an inner product space, we can simply take h to be the projection of v onto H, which has the virtue of avoiding appeal to the axiom of dependent choices.

For $v = A^n x$, we get $h = c_0 x + c_1 Ax + \cdots + c_{n-1}A^{n-1}x$ in H, and we can let $p_x(X) = X^n - c_{n-1}X^{n-1} - \cdots - c_1 X - c_0$. □

We will need the following, rather trivial, result.

Lemma 2. *If A is a nonzero operator on a normed linear space V, then $\{x : Ax \neq 0\}$ is dense in V.*

Proof. Suppose $Ay \neq 0$ and z is an arbitrary element of V. Then either $Az \neq 0$ or $A(z+y) \neq 0$. As we can take y as small as we please, we can find x arbitrarily close to z such that $Ax \neq 0$. □

The division algorithm works for polynomials with coefficients in an arbitrary ring, if the divisor is monic. We need to observe the exact form of the quotient when the divisor is $X - r$.

Lemma 3. *Let $p(X) = p_n X^n + p_{n-1}X^{n-1} + \cdots + p_1 X + p_0$, and write $p(X) = (X-r)q(X) + p(r)$ by the division algorithm. Then*

$$q(X) = p_n X^{n-1} + (rp_n + p_{n-1})X^{n-2} + (r^2 p_n + rp_{n-1} + p_{n-2})X^{n-3} + \cdots + (r^{n-1}p_n + r^{n-2}p_{n-1} + \cdots + p_1).$$

Proof. Compute. □

By the *norm* $||p||$ of a polynomial p we mean the supremum of the absolute values of its coefficients.

If f is a monic polynomial with complex coefficients and $f(z_0)$ is small, then z_0 is near a root of f; indeed, it is easy to see that if $f(X) = \prod_{i=1}^{n}(X - a_i)$, then for each $\delta > |f(z_0)|$ there exists i such that $|z_0 - a_i|^n < \delta$. As $f(X) = q(X)(X - z_0) + f(z_0)$, this is a special case of the more general fact that the set of roots of f is a continuous function of f in the following sense.

Lemma 4. *Let M and n be positive integers. Then there exists C such that if $f(X) = \prod_{i=1}^{n}(X - a_i)$ and $g(X) = \prod_{i=1}^{n}(X - b_i)$ are monic polynomials with complex coefficients, whose roots are bounded by M, and $||f - g|| < \delta$, then there is a permutation σ of $\{1, \dots, n\}$ such that, for all $i \in \{1, \dots, n\}$,*

$$|a_i - b_{\sigma(i)}| < C(\delta^{1/n!} + \delta)$$

Proof. We may assume that $M \geq 1$. As $|g(a_1)| = |f(a_1) - g(a_1)| < nM^n\delta$, there is j so that $|a_1 - b_j| < n^{1/n}M\delta^{1/n}$. Write $f(X) = (X - a_1)f_0(X)$ and $g(X) = (X - b_j)g_0(X)$, and consider

$$(X - b_j)(f_0(X) - g_0(X)) = (a_1 - b_j)f_0(X) + f(X) - g(X).$$

We see that the norm of the right hand side is bounded by a constant (depending only on M and n) times $\delta^{1/n} + \delta$. So Lemma 3 says that $||f_0(X) - g_0(X)|| < K(\delta^{1/n} + \delta)$ for some constant K, which we may take to be greater than 1. By induction on n we can find a permutation σ of $\{1, \dots, n\}$ such that $\sigma(1) = j$ and, for $i \neq 1$,

$$\begin{aligned} |a_i - b_{\sigma(i)}| &< C_0((K\delta^{1/n} + K\delta)^{1/(n-1)!} + K(\delta^{1/n} + \delta)) \\ &< C_0K((\delta^{1/n} + \delta)^{1/(n-1)!} + \delta^{1/n} + \delta) \\ &< 4C_0K(\delta^{1/n!} + \delta), \end{aligned}$$

where the last inequality comes from considering the two cases $\delta \leq 1$ and $\delta \geq 1$ and using continuity. Set $C = \sup(4C_0K, nM)$. □

For a sharper bound in Lemma 4, and a longer proof, see [**6**, Appendix A]. It might be worthwhile to point out that, from a constructive point

of view, Lemma 4 says something even when $f = g$. In that case we cannot necessarily find a permutation σ such that $a_i = b_{\sigma(i)}$ for all i. Indeed, consider $(X-x)(X-y) = (X-x\wedge y)(X-x\vee y)$. If $x = x\wedge y$, then $x \le y$, while if $x = x\vee y$, then $y \le x$. But it is well known that we cannot, in general, decide which of $x \le y$ and $y \le x$ holds (you must decide on the basis of some arbitrarily good rational approximations to x and y, and no matter how accurate these are, you might not have enough information to decide).

Our last lemma is a classical triviality but is needed because we cannot necessarily decide whether or not a given real number is zero.

Lemma 5. *Let $x_1, \dots, x_n$ be elements of a metric space (X, d). Then for any $r > 0$, and positive integer N, we can find $\varepsilon > 0$ and a partition of $\{1, \dots, n\}$, such that*

(1) $\varepsilon < r$,

(2) $d(x_i, x_j) < \varepsilon$ *if i and j are in the same element of the partition,*

(3) $d(x_i, x_j) > N\varepsilon$ *if i and j are in different elements of the partition.*

Proof. Let S be a finite subset of $\{1, \dots, n\}^2$ such that if $(i, j) \in S$, then $d(x_i, x_j) \neq 0$. We may think of S as the set of pairs (i, j) such that $d(x_i, x_j)$ is known to be nonzero—initially we can take S to be empty. We may assume that $N \ge 2$.

We proceed by induction on $m = n^2 - n - \#S$. Let $\varepsilon > 0$ be less than r and less than each $d(x_i, x_j)/N$ for $(i, j) \in S$. Either $d(x_i, x_j) < \varepsilon$ whenever $(i, j) \notin S$ or there exists $(i, j) \notin S$ such that $d(x_i, x_j) > 0$. In the former case the desired partition is induced by the equivalence relation $i \equiv j$ if $(i, j) \notin S$; in the latter case we can increase the size of S, and we are done by induction. □

It will be convenient to have a quantitative measure of linear independence. Define the *modulus of linear independence* of a finite sequence $x_1, \dots, x_n$ of vectors in a normed linear space V by

$$\omega(x_1, \dots, x_n) = \inf \left\{ \left\| \sum_{i=1}^{n} c_i x_i \right\| : \sup_{i=1}^{n} |c_i| = 1 \right\}.$$

Note that ω is defined as being the infimum of a uniformly continuous function on a totally bounded set and ω is uniformly continuous on bounded subsets of V^n. Moreover, by linearity we have

$$\left\| \sum_{i=1}^{n} c_i x_i \right\| \geq \omega(x_1, \ldots, x_n) \sup_{i=1}^{n} |c_i|,$$

so $x_1, \ldots, x_n$ are linearly independent if and only if $\omega(x_1, \ldots, x_n) > 0$.

Theorem. *Let A be a bounded operator on a normed linear space V. If $I, A, A^2, \ldots, A^n$ are linearly independent operators on V, then*

$$S_n = \{x \in V : x, Ax, A^2x, \ldots, A^nx \text{ are linearly independent}\}$$

is dense and open in V.

Proof. For each $x \in S_{n-1}$ we let p_x denote the (not necessarily unique) polynomial of Lemma 1. We may assume, by induction, that S_{n-1} is dense and open in V. Note that if $x \in S_{n-1}$, then Lemma 1 says that $x \in S_n$ if and only if $p_x(A)x \neq 0$. Because the modulus of linear independence $\omega(u, Au, A^2u, \ldots, A^nu)$ is bounded away from zero in some neighborhood of any vector in S_n, we see that S_n is open.

Let x be an arbitrary element of S_{n-1}, and let B be some ball around x, contained in S_{n-1} so that $\omega(u, Au, A^2u, \ldots, A^{n-1}u)$ is bounded away from zero for $u \in B$. As $I, A, A^2, \ldots, A^n$ are linearly independent, $p_x(A) \neq 0$, so Lemma 2 says that we can choose y in B so that $p_x(A)y \neq 0$. As $p_x(A)y = (p_x(A) - p_y(A))y + p_y(A)y$, either $p_x \neq p_y$ or $p_y(A)y \neq 0$. In the latter case $y \in S_n$, so we may assume that $p_x \neq p_y$.

Factor p_x and p_y into monic linear factors over the complex numbers, and let $\{q_i\}_{i \in I}$ be the finite family of n-fold products of these $2n$ linear factors. By Lemma 5 we can find $\varepsilon > 0$ and a partition of I such that $3\varepsilon < ||p_x - p_y||$, and if i and j are in the same element of the partition, then $||q_i - q_j|| < \varepsilon$, while if i and j are in different elements of the partition, then $||q_i - q_j|| > 3\varepsilon$.

As $\omega(u, Au, A^2u, \ldots, A^{n-1}u)$ is bounded away from zero for $u \in B$, there exists $\delta > 0$ such that for each $u \in B$ and each polynomial p of degree less than n, if $||p(A)u|| < 2\delta$, then $||p|| < \varepsilon$. There exists C such

that if p is a polynomial of degree less than n, then $C||p||$ is a bound on the operator $p(A)$; in fact, we can take $C = nC_0^{n-1}$ where $C_0 \geq 1$ is a bound on A.

Let μ be the supremum of $||q_i||$ for $i \in I$, and let U be a finite set of equally spaced elements of the interval connecting x to y so that if u and v are adjacent elements of U, then $||u - v|| < \delta/C(\mu + \varepsilon)$. For each u in U, use the division algorithm to write $p_x p_y = qp_u + r$, where r has degree less than n. We consider three cases.

(1) $p_x(A)p_y(A)u$ and $p_u(A)u$ are small,

(2) $p_x(A)p_y(A)u \neq 0$,

(3) $p_u(A)u \neq 0$.

In case (2), as u is a convex combination of x and y, either $p_x(A)x \neq 0$ or $p_y(A)y \neq 0$, so in either case (2) or case (3) we find a vector in $U \cap S_n$, and we are done. Thus, we may assume that case (1) holds for each $u \in U$. As $u \in S_{n-1}$ and r has degree less than n, then r is small. Apply Lemma 4 with $f = p_x p_y$ and $g = qp_u$; if $r = f - g$ is small enough, then the roots of p_u are close to the roots of some $q_{i(u)}$, so $||p_u - q_{i(u)}|| < \varepsilon$.

If u and v are adjacent elements of U, then

$$(*) \qquad (p_u(A) - p_v(A))u = p_u(A)u - p_v(A)v - p_v(A)(u - v).$$

We may assume that $||p_u(A)u||$ and $||p_v(A)v||$ are less than $\delta/2$, as otherwise one of them would be nonzero and we would be done. As p_v is within ε of (the real part of) $q_{i(v)}$, the operator $p_v(A)$ is bounded by $C(\mu + \varepsilon)$, so the right hand side of $(*)$ has norm less than 2δ, so $||p_u - p_v|| < \varepsilon$.

Thus, if u and v are adjacent elements of U, then $i(u)$ and $i(v)$ are in the same element of the partition of I. But $i(x)$ and $i(y)$ cannot be in the same element of the partition, a contradiction which shows that we must have found $u \in U \cap S_n$. □

REFERENCES

1. E. Bishop, *Foundations of constructive analysis*, McGraw-Hill, New York, 1967.

2. E. Bishop and D.S. Bridges, *Constructive analysis*, Springer-Verlag, Berlin, Heidelberg, 1985.

3. I.N. Herstein, *Topics in algebra*, Xerox College Publishing, Lexington, MA, 1975.

4. R.A. Horn and C.A. Johnson, *Matrix analysis*, Cambridge University Press, 1985.

5. F. Richman, *Polynomials and linear transformations*, Linear Algebra Appl. **131** (1990), 131–137

6. A.M. Ostrowski, *Solution of equations and systems of equations*, Academic Press, New York, 1966.

Florida Atlantic University, Boca Raton, FL 33431