An Exploratory Approach to Kaplansky's Lemma Leads to a Generalized Resultant

David Callan

Let $T : V \to V$ be a linear transformation on a vector space V. If the powers $I, T, T^2, \ldots, T^k$ of T are linearly dependent, then surely for every $v \in V$, the vectors $v, Tv, T^2v, \ldots, T^kv$ are also linearly dependent. Kaplansky noted that the converse assertion is also true.

Theorem 1 (Kaplansky's Lemma in finite dimensions) *Suppose $T : V \to V$ is a linear transformation on a finite-dimensional vector space V over an arbitrary field F and suppose the set of powers $\{T^i\}_{i=0}^k$ is linearly independent. Then there exists $v \in V$ such that the set $\{T^iv\}_{i=0}^k$ is linearly independent.*

This result actually holds in arbitrary dimensions and the extension from finite to infinite dimensions is fairly straightforward [6, p. 64]. Aupetit [1] gives an elegant proof when $F = \mathrm{C}$ (valid for any dimensions) using Liouville's theorem from complex analysis. His proof is reproduced, slightly polished, in Prasolov's recent intriguing survey of linear algebra emphasizing modern proofs of classical results [5, p. 46].

Henceforth, we assume a finite-dimensional space. So T has a minimal polynomial of degree n, say. The maximum k can be is $n-1$ and so we may as well assume $k = n-1$. Prasolov then gives a slick proof, valid for infinite fields, using minimal vector-annihilating polynomials and their subspaces [5, **13.1.2**, p. 73]. It relies on the result: a vector space cannot be expressed as a finite set-theoretic union of proper subspaces, and this of course requires the underlying field to be infinite. For arbitrary fields, one can readily construct a suitable vector v using the theory of the rational canonical form [4, p. 198]. This says that T acts as a block diagonal matrix, the diagonal blocks being the companion matrices of the invariant factors $d_1(x) \mid d_2(x) \mid \ldots \mid d_s(x)$ of T. The vector v consisting of 0's except for a 1 in the first position corresponding to the $d_s(x)$ block has minimal annihilating polynomial $d_s(x)$. This vector v fits the bill since $d_s(x)$ is also the minimal polynomial of T.

But suppose out of curiosity we look for a suitable v using the version of the rational canonical form involving elementary divisors [3, p. 262]. Now T acts as a block diagonal matrix whose diagonal blocks are the companion matrices of the elementary divisors $\{p_i(x)^{e_{ij}}\}$ of T. Here the polynomials $p_i(x)$ are irreducible and distinct. Let $f_i(x) = p_i(x)^{r_i}$, $1 \le i \le m$, be the elementary divisors of highest degree, so that $\prod_{i=1}^m f_i(x)$ is the minimal polynomial of T with degree $n = \sum_{i=1}^m \deg f_i$. In fact, it is sufficient to assume these are the *only* elementary divisors; after dealing with this case, v can be augmented with zeros to cover the general case.

Thus we assume T is the n-square matrix $\mathrm{diag}(A_1, A_2, \ldots, A_m)$ where A_i is the companion matrix of f_i, $1 \le i \le m$. Let $(v_1^T, v_2^T, \ldots, v_m^T)^T$ be the corresponding partition of our desired (column) vector v. For v_i, let's try the simplest possible nonzero vector: just one nonzero entry—a 1, say—which should be the first entry to maximize linear independence among $\{A_i^r v_i\}_{r \ge 0}$. This simple scheme would work if only we had a guarantee that the resulting n-square matrix $B := (v \,|\, Tv \,|\, \cdots \,|\, T^{n-1}v)$ has linearly independent columns—equivalently, $\det B \ne 0$—based solely on the premise of pairwise relative primeness of the polynomials f_i from which T is formed. But for polynomials, relative primeness means no common root (in an algebraic closure of F) and so a simple numerical quantity guaranteed to be nonzero precisely when the f_i are pairwise relatively prime is the product of all differences of roots of different f_i's. Call this product Δ. We are thus emboldened to conjecture that Δ and $\det B$ are closely related. In fact, as we show in Theorem 2, they are equal (up to sign).

To state this result in appropriate generality, let λ_{ij}, $1 \le i \le m$, $1 \le j \le n_i$ be algebraically independent commuting indeterminates over F (the "roots") and for $1 \le i \le m$ let σ_{ij}, $1 \le j \le n_i$, denote the elementary symmetric polynomials in $\{\lambda_{ij}\}_{j=1}^{n_i}$ (the "coefficients"). Then let $f_i(x) = x^{n_i} - \sigma_{i1}x^{n_i-1} + \sigma_{i2}x^{n_i-2} - \cdots + (-1)^{n_i}\sigma_{in_i}$, $1 \le i \le m$, so that the roots of f_i are $\{\lambda_{ij}\}_{j=1}^{n_i}$. Set $n = \sum_{i=1}^m n_i$. The companion matrix of f_i is

$$A_i = \begin{pmatrix} 0 & & & & (-1)^{n_i-1}\sigma_{in_i} \\ 1 & 0 & & & \vdots \\ & 1 & \ddots & & \\ & & \ddots & 0 & -\sigma_{i2} \\ & & & 1 & \sigma_{i1} \end{pmatrix}$$

and the characteristic polynomial of A_i is f_i. Now let A be the n-square block diagonal matrix with A_i as its ith diagonal block, $1 \le i \le m$, let e_i denote the n_i-length (column) vector with 1 in the first position and 0's elsewhere, and let e be the n-vector obtained by concatenating the e_i. Finally, let B denote the n-square matrix whose jth column is $A^{j-1}e$, $1 \le j \le n$. For example, when $m = 3$ and $n_1 = 3$, $n_2 = 2$, $n_3 = 1$,

$$B = \begin{pmatrix} 1 & 0 & 0 & \sigma_{13} & \sigma_{11}\sigma_{13} & \sigma_{11}^2\sigma_{13} - \sigma_{12}\sigma_{13} \\ 0 & 1 & 0 & -\sigma_{12} & -\sigma_{11}\sigma_{12} + \sigma_{13} & -\sigma_{11}^2\sigma_{12} + \sigma_{11}\sigma_{13} + \sigma_{12}^2 \\ 0 & 0 & 1 & \sigma_{11} & \sigma_{11}^2 - \sigma_{12} & \sigma_{11}^3 - 2\sigma_{11}\sigma_{12} + \sigma_{13} \\ 1 & 0 & -\sigma_{22} & -\sigma_{21}\sigma_{22} & -\sigma_{21}^2\sigma_{22} + \sigma_{22}^2 & -\sigma_{21}^3\sigma_{22} + 2\sigma_{21}\sigma_{22}^2 \\ 0 & 1 & \sigma_{21} & \sigma_{21}^2 - \sigma_{22} & \sigma_{21}^3 - 2\sigma_{21}\sigma_{22} & \sigma_{21}^4 - 3\sigma_{21}^2\sigma_{22} + \sigma_{22}^2 \\ 1 & \sigma_{31} & \sigma_{31}^2 & \sigma_{31}^3 & \sigma_{31}^4 & \sigma_{31}^5 \end{pmatrix} \tag{1}$$

Theorem 2 *For the matrix B formed from the λ_{ij} through their symmetric functions σ_{ij} as in (1),*

$$\det B = \prod_{1 \le i < j \le m} \prod_{k=1}^{n_i} \prod_{\ell=1}^{n_j} (\lambda_{j\ell} - \lambda_{ik}) \tag{2}$$

Before proving Theorem 2 (and thereby also Kaplansky's Lemma) we make some remarks. The classical resultant of two polynomials expresses the right side of (2) (for $m = 2$) as the determinant of the *Sylvester matrix* [7, p. 102] also formed from the polynomials' coefficients. Thus Theorem 2 gives a form of generalized resultant, though the matrix B does not coincide with the Sylvester matrix for $m = 2$. Further information about resultants (including a generalization in a different direction) and their applications in finding greatest common divisors of polynomials is given in [2, §2.1].

What does the matrix B look like when expressed in terms of the λ's? The following two propositions give some useful information.

Proposition 1 *All nonzero terms in the determinant of* any *submatrix of B have the same total degree in the λ's (depending only on the submatrix).*

Proof. The matrix B contains various size identity matrices flush left; the other entries are polynomials in the σ's that are easily seen to be homogeneous polynomials in the λ's whose (total) degrees are as illustrated in Figure 1 (for $m = 3$; $n_1 = 3$, $n_2 = 2$, $n_3 = 1$). Here the blanks represent zero entries in B.

$$\begin{pmatrix} 0 & - & - & 3 & 4 & 5 \\ - & 0 & - & 2 & 3 & 4 \\ - & - & 0 & 1 & 2 & 3 \\ 0 & - & 2 & 3 & 4 & 5 \\ - & 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ -1 & 0 & 1 & 2 & 3 & 4 \\ -2 & -1 & 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ -1 & 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{pmatrix}$$

Fig. 1 Fig. 2

It is possible to fill in the blanks so that each row is a sequence of consecutive integers, yielding a matrix C as in Figure 2. The degree of the product of given nonzero entries in B is the sum of the corresponding entries in C. Thus it certainly suffices to show: for any square submatrix D of C, the sum of the entries in a diagonal is constant over all diagonals of D (a diagonal is a maximal set of entries, no two in the same row or column). Now the entries in C are of the form $c_{ij} = d_i + e_j$ for some sequences $\{d_i\}, \{e_j\}$ (e_j may be taken $= j$). Thus the sum of the entries in any diagonal of D is the sum of the d_i corresponding to the rows of D and the e_j corresponding to the columns of D, independent of the diagonal.

Proposition 2 *Suppose B is partitioned into m^2 submatrices B_{ij} of size n_i-by-n_j in the obvious way. Fix i in $[1, m]$. Then*

(i) $B_{ii} = A_i^{n_1+n_2+\cdots+n_{i-1}}$,

(ii) $\det B_{ii} = (\lambda_{i1}\lambda_{i2}\cdots\lambda_{in_i})^{n_1+n_2+\cdots+n_{i-1}}$,

(iii) The degree (in the λ's) of each term in $\det B_{ii}$ is $n_i(n_1 + n_2 + \cdots + n_{i-1})$,

(iv) If B_i' is any n_i-square submatrix of B involving the same rows as B_{ii} but no column to the right of the rightmost column of B_{ii}, then the degree of each term in $\det B_i'$ is strictly less than $n_i(n_1 + n_2 + \cdots + n_{i-1})$ unless $B_i' = B_{ii}$,

(v) The degree of each term in $\det B$ is $\sum_{1\le i<j\le m} n_i n_j$.

Proof. (i) If an n_i-square window slides from left to right along the ith row of the partitioned

B, the successive powers $I, A_i, A_i^2, \ldots$ are obtained.

(ii) This is clear since $\det A_i = \sigma_{in_i}$.

(iii) follows from (ii) and Proposition 1.

(iv) follows from (iii) and Proposition 1 since degrees are strictly increasing along each row of B.

(v) also follows from (iii) and Proposition 1.

$\square$

Now to the proof of Theorem 2 in three steps: a) the right side of (2), Δ, divides $\det B$, b) Δ and $\det B$ have the same degree and hence a) implies they can differ only by a scalar factor, c) Δ and $\det B$ contain an identical term and hence this scalar factor is 1. At this point only step a) requires some work since the degree of Δ is $\sum_{1 \le i < j \le m} n_i n_j$, as is the degree of $\det B$ by Proposition 2 (v). Also, the product of the first terms of the factors comprising Δ is $X := (\lambda_{21} \cdots \lambda_{2n_2})^{n_1} (\lambda_{31} \cdots \lambda_{3n_3})^{n_1+n_2} \cdots (\lambda_{m1} \cdots \lambda_{mn_m})^{n_1+n_2+\cdots+n_{m-1}}$ (and this monomial does not occur elsewhere in Δ). On the other hand, by Proposition 2 (iv) and a Laplace expansion of $\det B$, only the entries in the main diagonal blocks $B_{11}, B_{22}, \ldots, B_{mm}$ can produce the monomial X and their contribution is precisely $\det B_{11} \det B_{22} \cdots \det B_{mm} = X$ by Proposition 2 (ii).

It remains to establish a). Clearly, $\det B$ lies in the polynomial ring $F[\lambda_{11}, \ldots, \lambda_{mn_m}]$. First, we show by a standard technique that each factor in Δ is a divisor of $\det B$ in this ring. To do so, suppose $\lambda_{ik} = \lambda_{j\ell}$ for some i, j, k, ℓ with $i \neq j$. This means that A_i and A_j share a common eigenvalue $\lambda_{ik} = \lambda_{j\ell} \equiv \lambda$. Let u_i and u_j be *left* (row) eigenvectors of A_i and A_j, respectively, corresponding to λ. Then for all $r \geq 0$,

$$(u_i \ u_j) \begin{pmatrix} A_i & 0 \\ 0 & A_j \end{pmatrix}^r = \lambda^r (u_i \ u_j).$$

We claim the eigenvector u_i can be scaled so that the inner product

$$(u_i \ u_j) \begin{pmatrix} e_i \\ e_j \end{pmatrix} = 0.$$

Since only the first components of e_i, e_j are nonzero, it clearly suffices to show that both inner products u_ie_i and u_je_j are nonzero. But if, say, $u_ie_i = 0$ then for all $r \geq 0$, $u_iA_i^re_i = \lambda^ru_ie_i = 0$ while $\{A_i^re_i\}_{r\geq 0}$ spans all of F^{n_i} (its first n_i vectors are the standard basis for F^{n_i}). Thus $u_i \in F^{n_i}$ would be perpendicular to F^{n_i}, hence 0, and eigenvectors are not allowed to be 0. Thus the claim is established. It follows that

$$(u_i \ u_j)\begin{pmatrix} A_i & 0 \\ 0 & A_j \end{pmatrix}^r \begin{pmatrix} e_i \\ e_j \end{pmatrix} = 0 \qquad \text{for all } r \geq 0.$$

In other words, $(u_i \ u_j)$ provides a nontrivial, yet vanishing, linear combination of the rows of B indexed by its i and j blocks. Thus B is singular and $\det B = 0$. Hence $\lambda_{j\ell} - \lambda_{ik}$ divides $\det B$ in the ring $F[\lambda_{11}, \ldots, \lambda_{mn_m}]$. Since each $\lambda_{j\ell} - \lambda_{ik}$ is prime in this ring, their product Δ also divides $\det B$. Thus Theorem 2 holds and our generalized resultant yields yet another proof of Kaplansky's Lemma.

$\square$

References

[1] B. Aupetit, An improvement of Kaplansky's lemma on locally algebraic operators, *Studia Math.* 88 (1988), 275-278.

[2] S. Barnett, *Matrices in Control Theory,* Robert E. Krieger Publ. Co., Malabar, Florida, 1984.

[3] I. N. Herstein, *Topics in Algebra*, Blaisdell, Waltham, Massachusetts, 1964.

[4] N. Jacobson, *Basic Algebra I*, W. H. Freeman, New York, 1985.

[5] V. V. Prasolov, *Problems and Theorems in Linear Algebra*, Translations of Mathematical Monographs, Vol. 134, American Mathematical Society, Providence, Rhode Island, 1994.

[6] H. Radjavi and P. Rosenthal, *Invariant Subspaces,* Ergeb. Math. Grenzgeb. 77, Springer, Berlin, 1973.

[7] B. L. van der Waerden, *Algebra*, Springer-Verlag, New York, 1991.

Department of Statistics
University of Wisconsin-Madison
1210 W. Dayton Street
Madison, WI 53706-1693
callan@stat.wisc.edu