Matrices with prescribed Ritz values

Beresford Parlett [a,*], Gilbert Strang [b]

[a] *Department of Mathematics, and Computer Science Division, of EECS Department, University of California at Berkeley, United States*

[b] *Department of Mathematics, Massachusetts Institute of Technology, United States*

Received 23 October 2006; accepted 12 October 2007

Submitted by R.A. Horn

Abstract

On the way to establishing a commutative analog to the Gelfand–Kirillov theorem in Lie theory, Kostant and Wallach produced a decomposition of *M(n)* which we will describe in the language of linear algebra. The "Ritz values" of a matrix are the eigenvalues of its leading principal submatrices of order $m = 1, 2, \dots, n$. There is a unique unit upper Hessenberg matrix H with those eigenvalues. For real symmetric matrices with interlacing Ritz values, we extend their analysis to allow eigenvalues at successive levels to be equal. We also decide whether given Ritz values can come from a tridiagonal matrix.

AMS classification: 15A18; 15A29

Keywords: Eigenvalues; Principal submatrices; Interlacing; Hessenberg

1. Introduction

This paper is concerned with the $1 + 2 + \cdots + n$ eigenvalues of the leading principal submatrices of an n by n matrix M. We call these $n(n + 1)/2$ numbers the *Ritz values* $\mathscr{R}$. They play an important role in numerical linear algebra (especially when M is tridiagonal), but our purpose here is different. We begin with two inverse questions, asked and answered by earlier authors, when these numbers $\mathscr{R}$ are prescribed:

* Corresponding author.

E-mail addresses: parlett@math.berkeley.edu (B. Parlett), gs@math.mit.edu (G. Strang).©

1. Does there exist a *real symmetric* matrix M with the Ritz values $\mathscr{R}$?
2. Does there exist a *complex* matrix M with the Ritz values $\mathscr{R}$?

In the symmetric case, the Ritz values must be real and *interlacing*: the ordered values d_i at level j and λ_i at the next level $j+1$ must satisfy

$$\lambda_1 \leqslant d_1 \leqslant \lambda_2 \leqslant d_2 \leqslant \cdots \leqslant d_j \leqslant \lambda_{j+1}. \tag{1}$$

This was proved by Cauchy [2]. It is known (the first published proof might be [3]) that $M = M^{\mathrm{T}}$ can be constructed with the given Ritz values.

The general complex case arises in two far-reaching papers by Kostant and Wallach [5,6]. In proving the Gelfand–Zeitlin conjecture in Lie theory, they describe the structure of the family (fiber, for Lie theorists) of all matrices M with given Ritz values $\mathscr{R}$. In the symmetric case, with strict interlacing, there are $N = n(n-1)/2$ sign choices in the construction of M, leading to 2^N symmetric matrices in the family.

The unsymmetric case places no restriction on $\mathscr{R}$. Any set of $n(n+1)/2$ complex numbers can be Ritz values. Assuming that the $2j+1$ d's and λ's at each step are distinct complex numbers, Kostant and Wallach construct the unique *unit Hessenberg matrix* H with given $\mathscr{R}$. (The lower triangular part of H is zero except for a diagonal of $n-1$ ones.) Again their purpose was to study the structure of the family of matrices M, headed by this H, sharing the given $\mathscr{R}$.

Now N choices of complex numbers, not just signs, will determine a specific matrix M in this family. Those $n(n-1)/2$ choices, together with the $n(n+1)/2$ numbers in $\mathscr{R}$, give n^2 parameters in a remarkable description by Kostant and Wallach of the space of n by n complex matrices.

We hope those authors will forgive us if we revisit these two matrix questions in a language we understand, matrix theory. Our approach has several rewards. In the symmetric case we do not need to invoke strict interlacing. The cardinality of the family with weakly interlacing Ritz values can sink below 2^N, or expand to a continuum. In that analysis we present a simple proof and extension of a lemma of Loewner on arrow matrices (Section 2). In addition we describe a procedure to decide whether a given $\mathscr{R}$, satisfying the interlacing condition, can come from a symmetric tridiagonal matrix.

The unsymmetric case also has extra questions when $\mathscr{R}$ is degenerate. The Ritz values $\mathscr{R} = \{1; 1, 1\}$ would allow

$$H = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and all } M = \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix} \text{ including } I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

All those matrices associated with $\mathscr{R}$ are similar to each other *except for* I.

Our terminology of Ritz values is dictated by brevity. In the symmetric case, the spectra of the leading principal submatrices are indeed the Rayleigh–Ritz approximations to the spectrum of the whole matrix from the subspaces spanned by successive columns of I. In the nonsymmetric case, the connection with Rayleigh and Ritz vanishes and we considered the longer name "principal values." The property of sharing $\mathscr{R}$ is a true equivalence relation on square matrices. We are tempted to say that two such matrices are spectrally aligned.

In the generic case treated by Kostant and Wallach, the Ritz values at consecutive levels have no values in common. Then all matrices that share a given $\mathscr{R}$ are "principally similar" – the leading principal submatrices of each order $j = 1, \ldots, n$ are similar (=conjugate). H serves as a canonical form for the equivalence class. Each matrix M in this class is non-derogatory, with only one Jordan block for each distinct eigenvalue. This property naturally holds for each leading principal submatrix.

We will show, for a non-degenerate $\mathscr{R}$, how each matrix M is uniquely determined by its N entries below the diagonal. We also consider briefly the non-generic derogatory case. In particular we describe the extreme case when $\mathscr{R}$ is all zeros and all leading submatrices are nilpotent.

In this unsymmetric case, Colarusso's important thesis [1] extends the full Kostant–Wallach theory to allow equal eigenvalues at a given level j (not overlapping with levels $j-1$ and $j+1$). He also studies $\mathscr{R}=0$. Our purpose is to bring key ideas of this theory to the attention of linear (but non-Lie) algebraists, with proofs that add insight when $\mathscr{R}$ is degenerate. A further paper by the first author will describe the "Ritz-preserving" similarity transformations that generate the equivalence class determined by $\mathscr{R}$.

2. The symmetric case

2.1. Loewner's lemma

We begin with the basic construction for arrow matrices. The first author learned the generic case as a student in a problem seminar taught by C. Loewner at Stanford University in 1958.

Theorem 1 (Loewner's lemma). *Suppose $\lambda_1, \ldots, \lambda_{j+1}$ interlace $d_1, \ldots, d_j$:*

$$\lambda_1 \leqslant d_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_j \leqslant d_j \leqslant \lambda_{j+1}. \tag{2}$$

Then there exist real $c_1, \ldots, c_j, \delta$ such that the symmetric arrow matrix

$$A = \begin{bmatrix} d_1 & & & c_1 \\ & \ddots & & \vdots \\ & & d_j & c_j \\ c_1 & \cdots & c_j & \delta \end{bmatrix} \quad \textit{has eigenvalues } \lambda_1, \ldots, \lambda_{j+1}.$$

Each c_i^2 is unique except in case 2, and the trace determines $\delta = \sum \lambda_i - \sum d_i$.

1. *If $\lambda_i < d_i < \lambda_{i+1}$ then $c_i^2 > 0$ is uniquely determined.*
2. *Suppose, for some i, $\lambda_i < d_i = \lambda_{i+1} = \cdots = d_{i+p-1} < \lambda_{i+p}$ with $p > 1$. Then only the sum $g_i = c_i^2 + \cdots + c_{i+p-1}^2 > 0$ is determined. There is a continuum of arrow matrices A with the required eigenvalues.*
3. *If any inequality in 1 or 2 becomes an equality, then g_i is zero.*

Proof. Compute the characteristic polynomial $\det(\lambda I - A)$ by row operations that preserve the determinant:

$$\begin{bmatrix} \lambda - d_1 & & & -c_1 \\ & \ddots & & \vdots \\ & & \lambda - d_j & -c_j \\ -c_1 & \cdots & -c_j & \lambda - \delta \end{bmatrix} \longrightarrow \begin{bmatrix} \lambda - d_1 & & & -c_1 \\ & \ddots & & \vdots \\ & & \lambda - d_j & -c_j \\ 0 & \cdots & 0 & S(\lambda) \end{bmatrix}.$$

The determinant is seen to be $S(\lambda)$ times the product $\prod(\lambda - d_k)$ down the diagonal. This is a familiar computation in Gaussian elimination. Each row k (for $k \leqslant j$) was multiplied by $c_k/(\lambda - d_k)$ and added to row $j+1$, to give those zeros in the last row. Then the Schur complement $S(\lambda)$ is

$$S(\lambda) = \lambda - \delta - \sum_{k=1}^{j} c_k^2/(\lambda - d_k). \tag{3}$$

Multiplying by the product $\prod(\lambda - d_m)$ gives the determinant of $\lambda I - A$. This is required to equal $\prod(\lambda - \lambda_m)$. The kth term $c_k^2/(\lambda - d_k)$ in $S(\lambda)$ yields a product with the factor $(\lambda - d_k)$ excluded:

$$\det(\lambda I - A) = \prod_{m=1}^{j+1}(\lambda - \lambda_m) = (\lambda - \delta)\prod_{m=1}^{j}(\lambda - d_m) - \sum_{k=1}^{j}\left(c_k^2 \prod_{m \neq k}(\lambda - d_m)\right). \tag{4}$$

That is the cofactor expansion of $\det(\lambda I - A)$ with respect to the last row.

Case 1 ($\lambda_i < d_i < \lambda_{i+1}$). For this index i, set $\lambda = d_i$ in (4) to find

$$\prod_{m=1}^{j+1}(d_i - \lambda_m) = -c_i^2 \prod_{m \neq i}(d_i - d_m). \tag{5}$$

The product on the left has $j + 1 - i$ negative terms. The product on the right has $j - i$ negative terms. So $c_i^2 > 0$ in this case of strict interlacing.

This argument seems to us simpler than the proof in [3], which omits the multiplication by $\prod(\lambda - d_m)$. That approach leads to a linear system for the new entries c_i^2. The difficulty is to prove that the solution is positive, which came immediately from (5).

Case 2 ($\lambda_i < d_i = \lambda_{i+1} = d_{i+1} = \cdots = d_{i+p-1} < \lambda_{i+p}$ *with* $p > 1$). This case cannot arise unless $j > 1$. The factor $(\lambda - d_i)$ is repeated $p - 1$ times in all products in equation (4). Divide each product by $(\lambda - d_i)^{p-1}$. Then set $\lambda = d_i$ and replace $c_i^2 + \cdots + c_{i+p-1}^2$ by g_i:

$$\prod_{m \leqslant i}(d_i - \lambda_m) \prod_{m \geqslant i+p}(d_i - \lambda_m) = -g_i \prod_{m<i}(d_i - d_m) \prod_{m \geqslant i+p}(d_i - d_m). \tag{6}$$

On each side, the first product contains only positive factors and the second product contains only negative factors. Then $g_i > 0$ because the left side has one extra negative factor $d_i - \lambda_{j+1}$.

Case 1 could be included in case 2 by allowing $p = 1$. But the essential point is that $p > 1$ introduces a continuum of permissible c_i^2.

Case 3 (*at least one inequality in* 1 *or* 2 *becomes an equality*). The left side of equation (5) or (6) is now zero. Therefore $c_i = 0$ in the first case and $g_i = 0$ in the second case. The latter implies that all of $c_i, c_{i+1}, \ldots, c_{i+p-1}$ are zero. The proof of Theorem 1 is complete. □

Thus a continuum of arrow matrices A will have the eigenvalues λ_i when there is a drop from multiplicity $p > 1$ for d_i at stage j to multiplicity $p - 1$ at stage $j + 1$.

As an extreme example of cases 2 and 3, *suppose all d's are zero*. If $\lambda_1 = -1$ and $\lambda_{j+1} = 1$, the trace gives $\delta = 0$. In this case 2, the column vector c is constrained only by $c^{\mathrm{T}}c = \sum c_i^2 = g = 1$, with a continuum of solutions:

$$A^2 = \begin{bmatrix} 0 & c \\ c^{\mathrm{T}} & 0 \end{bmatrix}^2 = \begin{bmatrix} cc^{\mathrm{T}} & 0 \\ 0 & c^{\mathrm{T}}c \end{bmatrix} \quad \text{has double eigenvalue 1 when } c^{\mathrm{T}}c = 1.$$

Case 3 occurs if $\lambda_1 = 0$, instead of -1. The trace is required to be $\delta = \lambda_{j+1}$, the only possible nonzero entry of A. All c's are zero and A becomes diagonal.

2.2. The size of the class

Given Theorem 1, the existence and structure of the equivalence class determined by $\mathscr{R}$ is established by induction.

Theorem 2. *If $\mathscr{R}$ is interlacing, there is a real symmetric matrix M_n with Ritz values $\mathscr{R}$. If $\mathscr{R}$ is strictly interlacing, the step from order j to $j+1$ gives j sign choices for $c_1, \ldots, c_j$. There is a total of $N = n(n-1)/2$ sign choices and 2^N different matrices M_n.*

Proof. We concentrate on the step from order j to $j+1$. The effects of weak interlacing are seen at that step (Theorem 1). If there are equalities between the j eigenvalues d_i, then infinitely many matrices M_{j+1} have the required Ritz values. If there are extra equalities between those d_i and the $j+1$ Ritz values λ_i at the new level, then the number of choices for M_{j+1} is diminished.

This induction step begins with a similarity transformation to diagonalize the matrix M_j produced at the previous step:

$$M_j = Z_j D_j Z_j^{-1} = Z_j D_j Z_j^{\mathrm{T}} \quad \text{with } D_j = \operatorname{diag}(d_1, \ldots, d_j). \tag{7}$$

Z_j is an orthogonal matrix, with the eigenvectors in its columns. The eigenvalues $d_1, \ldots, d_j$ lie in between the required eigenvalues $\lambda_1, \ldots, \lambda_{j+1}$ of M_{j+1}. To obtain M_{j+1} append a column c and a row c^{T} to D_j to make an *arrow matrix*. By Loewner's lemma there are choices for c which give the required eigenvalues λ_i to the arrow matrix. Reversing the similarity yields M_{j+1}:

$$M_{j+1} = \begin{bmatrix} Z_j & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} D_j & c \\ c^{\mathrm{T}} & \delta \end{bmatrix} \begin{bmatrix} Z_j^{\mathrm{T}} & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} M_j & Z_j c \\ (Z_j c)^{\mathrm{T}} & \delta \end{bmatrix}. \tag{8}$$

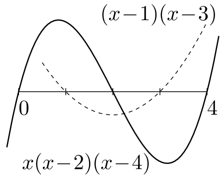
With strict interlacing, the structure of the N candidates M_n is described in Section 2.5 (see also [5,6]). □

2.3. Examples

It may be helpful to illustrate each case in Loewner's lemma by an example with $j+1 = 3$. The given diagonal entries will be $d_1 = 1$ and $d_2 = 3$ in case 1 and $d_1 = d_2 = 2$ in case 2. The eigenvalues $\lambda_1, \lambda_2, \lambda_3$ may interlace the d's strictly or weakly.

The excluded product at the end of (4) is the derivative of $\prod(\lambda - d_m)$ at $\lambda = d_i$. In the graphs below, this is the slope of the dotted curve at d_i.

Case **1** $\quad A = \begin{bmatrix} 1 & 0 & c_1 \\ 0 & 3 & c_2 \\ c_1 & c_2 & 2 \end{bmatrix}$ has $\lambda = 0, 2, 4$

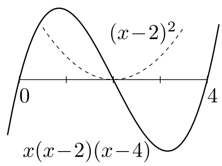


Each c_i allows two choices

$$c_1^2 = \frac{-(1-0)(1-2)(1-4)}{(1-3)} = \frac{3}{2}$$

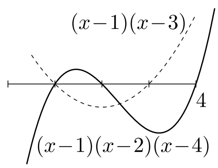
$$c_2^2 = \frac{-(3-0)(3-2)(3-4)}{(3-1)} = \frac{3}{2}$$

Case **2** $\quad A = \begin{bmatrix} 2 & 0 & c_1 \\ 0 & 2 & c_2 \\ c_1 & c_2 & 2 \end{bmatrix}$ has $\lambda = 0, 2, 4$

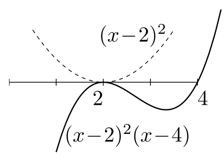


Continuum $\quad s_1 = c_1^2 + c_2^2 = \dfrac{-(2-0)(2-4)}{1} = 4$

Case **3** from **1**
One choice $\quad A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & \pm 1 \\ 0 & \pm 1 & 3 \end{bmatrix}$ has $\lambda = 1, 2, 4$



Case **3** from **2**
No choice $\quad A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ has $\lambda = 2, 2, 4$



2.4. Eight matrices with $\mathscr{R} = \{0; -1, 4; -2, 1, 5\}$.

These Ritz values are strictly interlacing, so the construction will allow $N = n(n-1)/2 = 3$ sign choices. Then $2^3 = 8$ symmetric matrices share these Ritz values $\mathscr{R}$, and all eight must begin with $M_{11} = 0$. The next step has two possibilities:

$$\begin{bmatrix} 0 & 2 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & -2 \\ -2 & 3 \end{bmatrix} \text{ have eigenvalues } -1 \text{ and } 4.$$

The final step diagonalizes those matrices, adds a row and column by Loewner's lemma, and reverses the diagonalization as in (8). Four new numbers enter the 3 by 3 matrices:

$$c_1 = \frac{4\sqrt{3} + 3\sqrt{2}}{5} \approx 2.2, \quad c_2 = \frac{-2\sqrt{3} + 6\sqrt{2}}{5} \approx 1.0,$$
$$d_1 = \frac{4\sqrt{3} - 3\sqrt{2}}{5} \approx .54, \quad d_2 = \frac{2\sqrt{3} + 6\sqrt{2}}{5} \approx 2.4.$$

It is instructive to see the eight matrices! We write them as B_1–B_4 and C_1–C_4, and find the similarities that preserve $\mathscr{R}$.

$$\overset{B_1}{\begin{bmatrix} 0 & 2 & c_1 \\ 2 & 3 & c_2 \\ c_1 & c_2 & 1 \end{bmatrix}}, \overset{B_2}{\begin{bmatrix} 0 & 2 & d_1 \\ 2 & 3 & -d_2 \\ d_1 & -d_2 & 1 \end{bmatrix}}, \overset{B_3}{\begin{bmatrix} 0 & -2 & d_1 \\ -2 & 3 & d_2 \\ d_1 & d_2 & 1 \end{bmatrix}}, \overset{B_4}{\begin{bmatrix} 0 & -2 & c_1 \\ -2 & 3 & -c_2 \\ c_1 & -c_2 & 1 \end{bmatrix}},$$

$$\overset{C_1}{\begin{bmatrix} 0 & 2 & -c_1 \\ 2 & 3 & -c_2 \\ -c_1 & -c_2 & 1 \end{bmatrix}}, \overset{C_2}{\begin{bmatrix} 0 & 2 & -d_1 \\ 2 & 3 & d_2 \\ -d_1 & d_2 & 1 \end{bmatrix}}, \overset{C_3}{\begin{bmatrix} 0 & -2 & -d_1 \\ -2 & 3 & -d_2 \\ -d_1 & -d_2 & 1 \end{bmatrix}}, \overset{C_4}{\begin{bmatrix} 0 & -2 & -c_1 \\ -2 & 3 & c_2 \\ -c_1 & c_2 & 1 \end{bmatrix}}.$$

Each B_j is connected to C_j by a diagonal similarity:

$$\Omega_3 = \operatorname{diag}(1, 1, -1) = \Omega_3^{-1} \quad \text{gives } C_j = \Omega_3 B_j \Omega_3.$$

Another diagonal similarity connects two pairs of B's:

$$\Omega_2 = \operatorname{diag}(1, -1, 1) = \Omega_2^{-1} \quad \text{has } B_4 = \Omega_2 B_1 \Omega_2 \text{ and } B_3 = \Omega_2 B_2 \Omega_2.$$

The key question is to relate B_1 to B_2. They are symmetric, with the same eigenvalues $-2, 1, 5$, so they are orthogonally similar:

$$Q_1 B_1 Q_1^{\mathrm{T}} = \operatorname{diag}(-2, 1, 5) = Q_2 B_2 Q_2^{\mathrm{T}} \quad \text{yields } B_2 = (Q_2^{\mathrm{T}} Q_1) B_1 (Q_1^{\mathrm{T}} Q_2).$$

But there are crucial sign choices in the columns of Q_1 and Q_2 (the eigenvectors of B_1 and B_2). The right choice involves Ω_2 and the mapping into arrow form. It leads to the similarity that preserves all Ritz values:

$$B_2 = Q B_1 Q^{\mathrm{T}} \quad \text{with } Q = \frac{1}{5}\begin{bmatrix} 3 & -4 & 0 \\ -4 & -3 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

The upper left 2 by 2 submatrices of Q and B_1 commute.

The first author plans to study similarities that preserve Ritz values in a future paper.

2.5. Symmetric tridiagonal case

Ritz values are of special importance when they come from a *symmetric tridiagonal matrix* T. Starting with a symmetric M, the "Lanczos method" constructs T so that its Ritz values quickly approach the extreme eigenvalues of M. This is an important algorithm and it is natural to ask the inverse question: *Which Ritz values* $\mathscr{R}$ *can come from a tridiagonal* T?

Construct the sequence of monic polynomials $p_1, p_2, \ldots$ with (interlacing) zeros given by $\mathscr{R}$. The coefficient s_j in $p_j(t) = t^j - s_j t^{j-1} + \cdots$ will be the sum of the zeros of p_j (s_j is the trace of T's leading $j \times j$ submatrix). Define $\delta_1 = s_1$ and $\delta_j = s_j - s_{j-1}$ for $j = 2, \ldots, n$.

In (9), if the polynomials p_j do come from a symmetric tridiagonal matrix T, they will satisfy a three-term recurrence with coefficients of known sign. That will be our test. The key is to proceed in reverse order $j = n-1, n-2, \ldots, 2$ and introduce the appropriate difference polynomial

$$q_{j-1}(t) = p_{j+1}(t) - (t - \delta_{j+1}) p_j(t) = r_{j-1} t^{j-1} + \cdots \tag{9}$$

It may be verified that the definition of δ_{j+1} causes the coefficient of t^j in the two terms above in q_{j-1} to cancel yielding the desired degree $j-1$ for q_{j-1}.

Here is the criterion. *In* (9) *if* $r_{j-1} \leqslant 0$ *or if* $q_{j-1}(t) \neq r_{j-1}p_{j-1}(t)$, *for any* j, *then the required tridiagonal matrix* T *does not exist. Otherwise a tridiagonal* T *can be constructed from the roots* $\mathscr{R}$ *as described above.*

2.6. The family of symmetric M_n.

The eigenvector matrices of members of the fiber for a given $\mathscr{R}$ are built up from the eigenvector matrices of arrow matrices. Strict interlacing (case **1** in Theorem 1) has distinct d's and λ's. Then the eigenvectors of the arrow matrix A, with different sign choices for the c's, have a simple structure. *First choose positive* c_i *in* A:

$$\begin{bmatrix} A-\lambda I \end{bmatrix}\begin{bmatrix} q \\ z \end{bmatrix} = \begin{bmatrix} D-\lambda I & c \\ c^{\mathrm{T}} & \delta-\lambda \end{bmatrix}\begin{bmatrix} q \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{for } \lambda = \lambda_1, \ldots, \lambda_{j+1}. \tag{10}$$

Certainly $z = 0$ is impossible, because the first block row would become $(D - \lambda I)q = 0$ which implies $q = 0$ by strict interlacing. Thus $q = (\lambda I - D)^{-1}cz$. We normalize to eigenvectors of unit length with $z > 0$. The eigenvectors for distinct $\lambda_1, \ldots, \lambda_{j+1}$ are orthonormal because A is symmetric. Then the matrix with those eigenvectors in its columns is orthogonal: $Q^{-1} = Q^{\mathrm{T}}$.

Write that eigenvector matrix as Q^+_{j+1}, to indicate that all $c_i > 0$. Any choice of signs is given by Ωc, with $\Omega = \mathrm{diag}(+/-1)$. Such matrices are called *signature* matrices. Changing c to Ωc in (10) just changes q to Ωq, because Ω commutes with the diagonal matrix $D - \lambda I$ (and $\Omega^{\mathrm{T}}\Omega = I$ since each diagonal entry is $+1$ or -1):

$$\begin{bmatrix} D-\lambda I & \Omega c \\ (\Omega c)^{\mathrm{T}} & \delta-\lambda \end{bmatrix}\begin{bmatrix} \Omega q \\ z \end{bmatrix} = \begin{bmatrix} \Omega(D-\lambda I)q + \Omega c z \\ c^{\mathrm{T}}q + (\delta-\lambda)z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{11}$$

Extend Ω (of size j) to the signature matrix Ω_{j+1}, in which *the last sign is always* $+1$. For the sign choice Ω, the eigenvector matrix is $Q^{\Omega}_{j+1} = \Omega_{j+1}Q^+_{j+1}$.

Looking back at the whole sequence of sign choices for the c's at previous levels, those choices are specified by a sequence of diagonal signature matrices:

$$\Omega = \{\Omega_1, \Omega_2, \ldots, \Omega_n\}.$$

Every Ω_j of order j ends with sign $+1$. There are $0 + 1 + \cdots + (n-1) = n(n-1)/2 = N$ sequences Ω. Each Ω determines a symmetric matrix M^{Ω} of order n whose leading principal submatrices have the required (interlacing) eigenvalues from $\mathscr{R}$. We have seen that the arrow matrices in that construction have the eigenvector matrices

$$Q^{\Omega} = \{\Omega_1 Q^+_1, \Omega_2 Q^+_2, \ldots, \Omega_n Q^+_n\}. \tag{12}$$

The (Abelian) group of Ω's has $\Omega \times \Omega' = \{\Omega_1\Omega'_1, \Omega_2\Omega'_2, \ldots, \Omega_n\Omega'_n\}$. This group is clearly isomorphic to the group of all Q^{Ω}, when the operation is defined by

$$Q^{\Omega} \times Q^{\Omega'} = \{\Omega_1\Omega'_1 Q^+_1, \Omega_2\Omega'_2 Q^+_2, \ldots, \Omega_n\Omega'_n Q^+_n\}. \tag{13}$$

It remains to recognize the eigenvector matrix Z^{Ω}_n that diagonalizes the final matrix M^{Ω}_n. This Z^{Ω}_n is built recursively from the eigenvector matrices in Q^{Ω} for the sequence of arrow matrices $A_1, \ldots, A_n$ and from the eigenvector matrices $Z_1, \ldots, Z_{n-1}$ in (7) for the leading submatrices.

First we find the recursion for Z^+_j (eigenvectors of M^+_j) when the c's are chosen positive: $M^+_j = Z^+_j D_j (Z^+_j)^{\mathrm{T}}$ leads to

$$M^+_{j+1} = (Z^+_j \oplus 1)A^+_{j+1}(Z^+_j \oplus 1)^{\mathrm{T}} = (Z^+_j \oplus 1)Q^+_{j+1}D_{j+1}(Q^+_{j+1})^{\mathrm{T}}(Z^+_j \oplus 1)^{\mathrm{T}}. \tag{14}$$

Thus the recursion is

$$Z^+_{j+1} = (Z^+_j \oplus 1)Q^+_{j+1}. \tag{15}$$

For different sign choices of the c's at the previous step, a sign matrix Ω_j precedes Q^+_j as in (12). The whole sequence of sign choices will enter the eigenvector matrices Z_j at successive steps of the recursion. Then the $Z^\Omega = \{Z_1, \dots, Z_n\}$ have a group structure isomorphic to the group of sign choices $\{\Omega_1, \dots, \Omega_n\}$. The recurrence is

$$Z^{\Omega_{j+1}}_{j+1} = (Z^{\Omega_j}_j \oplus 1)\Omega_{j+1}Q^+_{j+1}. \tag{16}$$

3. The non-symmetric case

3.1. A canonical form

In the non-symmetric case, the interlacing requirement disappears. We are given one (complex) number, then two, and eventually n numbers $\lambda_1, \dots, \lambda_n$ that make up $\mathscr{R}$. We want to prove that there is a *unique unit Hessenberg matrix* H_n whose leading submatrices have these numbers as their eigenvalues. Thus, when $n = 4$

$$H_4 = N_4 + U_4 = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 1 & u_{22} & u_{23} & u_{24} \\ 0 & 1 & u_{33} & u_{34} \\ 0 & 0 & 1 & u_{44} \end{bmatrix},$$

where U is (complex) upper triangular and N has 1's down the subdiagonal. N makes the matrix "unit Hessenberg." The 4 diagonal entries are determined by $\mathscr{R}$, and the $1 + 2 + 3$ entries above the diagonal can be chosen (uniquely) to give the required eigenvalues of the leading principal submatrices H_1, H_2, H_3.

Theorem 3. *There is a unique unit* (*upper*) *Hessenberg matrix H whose leading principal submatrices of orders* $1, \dots, n$ *have arbitrarily prescribed eigenvalues* (*the Ritz values* $\mathscr{R}$).

The proof is by induction. Clearly u_{11} must equal the prescribed eigenvalue for the 1 by 1 submatrix. We write e^{T}_j for the j-component row vector $[0 \quad \cdots \quad 0 \quad 1]$. Allow us to isolate the inductive step from j to $j + 1$:

> Given any unit Hessenberg matrix $H_j = N_j + U_j$, and any complex numbers $\lambda_1, \dots, \lambda_{j+1}$, there is a unique choice of the last column of H_{j+1} so that
>
> $$H_{j+1} = N_{j+1} + U_{j+1} = \begin{bmatrix} N_j + U_j & d \\ e^{\mathrm{T}}_j & \delta \end{bmatrix} \quad \text{has eigenvalues } \lambda_1, \dots, \lambda_{j+1}.$$

First proof. This is almost a standard example of the "pole assignment problem" in control theory. The text by Wonham [9, pp. 45 and 50] gives the necessary and sufficient condition on the first n columns of H_{j+1} so that any λ's can be achieved by a correct choice of d and δ in the last column. In our notation, that controllability condition is

$$\operatorname{rank}\left[e_j \quad H^{\mathrm{T}}_j e_j \quad \cdots \quad (H^{\mathrm{T}}_j)^{j-1}e_j\right] = j. \tag{17}$$

For our e_j and $H_j = N_j + U_j$ this matrix does have full rank j:

$$\operatorname{rank}\begin{bmatrix} 0 & 0 & \cdot & 0 & 1 \\ 0 & 0 & \cdot & 1 & x \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & \cdot & x & x \\ 1 & x & \cdot & x & x \end{bmatrix} = j.$$

The articles [4,7,8] resolve the general question of adding a row and column to achieve a desired spectrum. A second more explicit proof will bring out the uniqueness of H_{j+1}. □

Second proof. For an explicit construction of d and δ in the last column of H_{j+1}, we put the first j columns into a simple form. Every unit Hessenberg matrix H_j is similar to a *companion matrix* C_j with last column $c = (c_1, \dots, c_j)$:

$$S^{-1} H_j S = C_j = N_j - c e_j^{\mathrm{T}}. \tag{18}$$

Reason: A companion matrix is unit Hessenberg with zeros on and above the diagonal, *except in its last column*. We can choose that last column c so that C_j has the same characteristic polynomial as H_j. More than that, *C_j and H_j are similar*. We noted earlier that in all unit Hessenberg matrices, the Jordan form has only one block for each eigenvalue. For any λ, the ranks of $\lambda I - H_j$ and $\lambda I - C_j$ cannot go below $j - 1$ because of the 1's from N_j.

It is important to notice that S is upper triangular with ones on the diagonal. We now have to choose δ and $b = S^{-1} d$ in the last column of B_{j+1} (similar to H_{j+1}) so that B_{j+1} has eigenvalues $\lambda_1, \dots, \lambda_{j+1}$:

$$B_{j+1} = \begin{bmatrix} S^{-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} H_j & d \\ e_j^{\mathrm{T}} & \delta \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} C_j & b \\ e_j^{\mathrm{T}} & \delta \end{bmatrix}. \tag{19}$$

As before, δ is chosen to give the required trace. Then in parallel with Loewner's lemma in the symmetric case (where we had a diagonal matrix D instead of the companion matrix C), we need an expression for the characteristic polynomial of B_{j+1}. □

Companion lemma. *The characteristic polynomial of the companion matrix B_{j+1}, required to equal $\prod(\lambda - \lambda_i)$, is*

$$\det(\lambda I_{j+1} - B_{j+1}) = (\lambda - \delta) \det(\lambda I_j - C_j) - \sum_{1}^{j} b_i \lambda^{i-1}. \tag{20}$$

The b_i can be chosen (uniquely) to give the required eigenvalues $\lambda_1, \dots, \lambda_{j+1}$, since (by choice of δ) the coefficient of λ^j already gives the correct trace.

Proof of (20). Expand $\det(\lambda I_{j+1} - B_{j+1})$ in cofactors of $(\lambda - \delta)$ and -1 in its last row. For $j + 1 = 4$,

$$\det \begin{bmatrix} \lambda & 0 & -c_1 & -b_1 \\ -1 & \lambda & -c_2 & -b_2 \\ 0 & -1 & \lambda - c_3 & -b_3 \\ 0 & 0 & -1 & \lambda - \delta \end{bmatrix} = (\lambda - \delta)\det(\lambda I - C) + \det \begin{bmatrix} \lambda & 0 & -b_1 \\ -1 & \lambda & -b_2 \\ 0 & -1 & -b_3 \end{bmatrix}. \tag{21}$$

The 3 by 3 determinant is $-b_1 - b_2\lambda - b_3\lambda^2$ as the lemma requires. It is the characteristic polynomial of a typical companion matrix, without the leading term λ^3.

With this lemma the recursive construction of a unique unit Hessenberg H with Ritz values $\mathscr{R}$ is complete. Reversing the similarity in (19),

$$H_{j+1} = \begin{bmatrix} S & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} B_{j+1} \end{bmatrix} \begin{bmatrix} S^{-1} & 0 \\ 0 & 1 \end{bmatrix}. \tag{22}$$

Since S is unit upper triangular the (2, 1) block of H_{j+1} is $e_j^{\mathrm{T}} = e_j^{\mathrm{T}} S^{-1}$. Next we consider all matrices M_n that have the required (non-degenerate) Ritz values $\mathscr{R}$. There are N sign choices in the symmetric case (Section 2.2) and N complex parameters in the non-symmetric case (Section 3.2).

3.2. *The family of non-symmetric M_n.*

The Ritz values $\mathscr{R}$ are *generic* if at each stage, the eigenvalues $d_1, \ldots, d_j, \lambda_1, \ldots, \lambda_{j+1}$ are $2j + 1$ distinct numbers. We want to construct M_{j+1} from M_j, allowing any j complex numbers $y_1, \ldots, y_j$ in the new last row:

$$M_{j+1} = \begin{bmatrix} M_j & x \\ y^{\mathrm{T}} & \delta \end{bmatrix} \quad \text{with } \delta = \sum_1^{j+1} \lambda_i - \sum_1^{j} d_i. \tag{23}$$

We are now looking at the whole equivalence class of matrices M with given Ritz values $\mathscr{R}$, not at the unit Hessenberg matrix H in that family. The problem is to choose x so that M_{j+1} has the eigenvalues λ_i. This was solved in Section 3.1 when $M_j = H_j$ and $y = e_j$.

Let $P(\lambda)$ be the characteristic polynomial of M_j. The "generic assumption" means that each $P(\lambda_i) \neq 0$. In this case block elimination of y^{T} in M_{j+1} leads for each λ_i to

$$\det(M_{j+1} - \lambda_i I) = P(\lambda_i)(\delta - \lambda_i - \ell_i^{\mathrm{T}} x), \tag{24}$$

with $\ell_i^{\mathrm{T}} = y^{\mathrm{T}}(M_j - \lambda_i I)^{-1}$. Form the j by $j + 1$ matrix L and the row vector g^{T}:

$$L = \begin{bmatrix} \ell_1 & \ldots & \ell_{j+1} \end{bmatrix} \quad \text{and} \quad g^{\mathrm{T}} = \begin{bmatrix} \delta - \lambda_1 & \ldots & \delta - \lambda_{j+1} \end{bmatrix}.$$

Then Eq. (24) gives $\det(M_{j+1} - \lambda_i I) = 0$ for $i = 1, \ldots, j + 1$ when

$$L^{\mathrm{T}} x = g. \tag{25}$$

Provided L has full rank j, this has the unique solution

$$x = (LL^{\mathrm{T}})^{-1} L g.$$

The matrix L depends heavily on y^{T}, and the full rank condition shows that there are constraints on viable rows y^{T}. The condition is not easily recognized and has different names in different fields. In linear algebra one would say that the minimal polynomial of y^{T} for the matrix M_j must have maximal degree j. In systems theory one would say that the SISO (Single Input, Single Output) linear system presented by M_{j+1} in (23) must be "observable":

The observability matrix $[y, M^{\mathrm{T}} y, (M^{\mathrm{T}})^2 y, \ldots]$ must have full rank j.

Theorem 4. *If each leading principal submatrix of M_n defines an observable SISO system then, given $\mathscr{R}$, the strictly lower triangular part of M_n determines uniquely the strictly upper triangular part. If the systems are "controllable" then the strictly upper triangular part determines uniquely the strictly lower triangular part.*

Let us summarize the new description introduced by Kostant and Wallach. Each set of Ritz values $\mathscr{R}$ determines an equivalence class of matrices M_n. In the generic case, each member of an equivalence class is determined by its strictly lower triangular part. When this part is the identity matrix, we have the canonical Hessenberg matrix H in the class, and all members of a class are related by similarity transformations. These transformations (of a particular form, to preserve $\mathscr{R}$) will be discussed in another paper.

3.3. *The case* $\mathscr{R} = 0$

This case $\mathscr{R} = 0$, when the principal submatrices are all nilpotent, is studied by Colarusso [1] with important new results. The first question to ask of any nilpotent matrix is its *index of nilpotency* which we abbreviate by nilp and define by

$$\mathrm{nilp}[B] := \max\{j : B^j \neq 0\}.$$

By convention $\mathrm{nilp}[0] = 0$.

Each leading principal submatrix B_k of a matrix $B = B_n$ with $\mathscr{R} = 0$ is nilpotent and B_n is characterized by the sequence $\mathrm{nilp}[B_k]$, $k = 1, 2, \ldots, n$ which we call its *skeleton*:

$$\mathrm{skeleton}(B) := (\mathrm{nilp}_1, \mathrm{nilp}_2, \ldots, \mathrm{nilp}_n).$$

Clearly $\mathrm{nilp}_1 = 0$ for $\mathscr{R} = 0$. Almost as clearly, $\mathrm{nilp}_{k+1} \geqslant \mathrm{nilp}_k$. The maximal skeleton is $(0, 1, 2, \ldots, n-1)$ and the corresponding non-derogatory matrices are the easiest to describe.

We need an extra piece of notation $[k : j]$ to denote the product of entries $(k, k-1), (k-1, k-2), \ldots, (j+1, j)$ of a given matrix with $j < k$. Also, if P is a permutation matrix then we call the conjugation $M \longrightarrow PMP^{\mathrm{T}}$ a symmetric permutation of M in order to distinguish it from $M \longrightarrow PMQ$ with Q another permutation.

Lemma 1. *Each $n \times n$ matrix with $\mathscr{R} = 0$ and a maximal skeleton is a symmetric permutation of a strictly lower triangular matrix with $[n : 1] \neq 0$. Its canonical form, under preservation of $\mathscr{R}$, coincides with its Jordan form N_n, the push-down matrix introduced at the start of Section* 3.1.

Proof. For $n = 1$, Lemma 1 is vacuous. For $n = 2$, it holds because the class consists of all matrices $B_2 = \begin{bmatrix} 0 & 0 \\ b & 0 \end{bmatrix}$, $b \neq 0$, and their transposes. Note that $b = 0$ gives nilp 0, not 1.

The induction assumption is that Lemma 1 holds for $n = k \geqslant 2$. Designate a matrix of order $k+1$ with $\mathscr{R} = 0$ by

$$B_{k+1} = \begin{bmatrix} B_k & x \\ y^{\mathrm{T}} & 0 \end{bmatrix}. \tag{26}$$

By the induction assumption there is no loss of generality in taking B_k to be strictly lower triangular. Its nilp is $k-1$ by assumption, and it then follows that $[k : 1] \neq 0$.

Gaussian elimination by blocks shows that, for any y, x, and scalar t

$$\det[tI_{k+1} - B_{k+1}] = \det[tI_k - B_k](t - 0 - y^{\mathrm{T}}(tI_k - B_k)^{-1}x). \tag{27}$$

$\mathscr{R} = 0$ implies that $\det[tI_k - B_k] = t^k$ and $\det[tI_{k+1} - B_{k+1}] = t^{k+1}$. Then

$$y^{\mathrm{T}}(tI_k - B_k)^{-1}x = 0 \quad \text{for all } t. \tag{28}$$

Since B_k is nilpotent, the power series for $(tI_k - B_k)^{-1}$ ends after k terms:

$$(tI_k - B_k)^{-1} = \sum_{j=0}^{k-1} B_k^j / t^{j+1}. \tag{29}$$

The powers of $1/t$ are linearly independent, so that

$$y^{\mathrm{T}} B_k^j x = 0, \quad j = 0, 1, \ldots, k-1. \tag{30}$$

We can condense these equations in two ways. Define Krylov matrices with $B = B_k$ and columns and rows constructed from x and y^{T}:

$$K(x) := [x, Bx, \ldots, B^{k-1}x] \quad \text{and} \quad K^{\mathrm{T}}(y) := [y, B^{\mathrm{T}}y, \ldots, (B^{\mathrm{T}})^{k-1}y]^{\mathrm{T}}. \tag{31}$$

The conditions (30) for $\mathscr{R} = 0$ become

$$y^{\mathrm{T}} K(x) = 0 \quad \text{and} \quad K^{\mathrm{T}}(y)x = 0. \tag{32}$$

Case 1 ($x_1 \neq 0$). In this case $K(x)$ is a lower triangular matrix. Its diagonal entries, from top to bottom, are given by $x_1(1, [2:1], [3:1], \ldots, [k:1])$. Thus $K(x)$ is invertible and $y^{\mathrm{T}} = 0$. Let P be the cyclic permutation matrix that puts the last entry of a column vector to the top. Then $PB_{k+1}P^{\mathrm{T}}$ is strictly lower triangular and has nilp $= k$ because x_1 does not vanish.

Case 2 ($y_k \neq 0$). In this case $K^{\mathrm{T}}(y)$ is a "northwest" triangular matrix (reverse columns of an upper triangle). Its antidiagonal entries, from top to bottom, are given by $y_1(1, [k:k-1], [k:k-2], \ldots, [k:1])$. Thus $K(y)$ is invertible and $x = 0$. Thus B_{k+1} is strictly lower triangular and has nilp $= k$ because y_k does not vanish.

Case 3 ($y_k = 0$ *and* $x_1 = 0$). The case $x = 0$ would not have maximal skeleton because $y_k = 0$. So let m be the smallest index with $x_m \neq 0$.Clearly $2 \leqslant m \leqslant k$. In (31), $K(x)$ is zero except for an invertible lower triangular matrix L of order $k - m + 1$ in the lower left corner. The nonzero diagonal entries of L are $x_m(1, [m+1:m], [m+2:m], \ldots, [k:m])$. Condition (32) for $\mathscr{R} = 0$ reduces to $0 = (y_m, y_{m+1}, \ldots, y_k)L$. Since L is invertible these entries $y_m, \ldots, y_k$ must all vanish. No constraints are imposed on $y_1, y_2, \ldots, y_{m-1}$.

Since $x_1, x_2, \ldots, x_{m-1}$ all vanish too, we may permute B_{k+1} into $PB_{k+1}P^{\mathrm{T}}$ using the cyclic permutation $P = (m, m+1, \ldots, k+1)$ that brings row $k+1$ into row m. It may be verified that the result is strictly lower triangular:

$$PB_{k+1}P^{\mathrm{T}} = \begin{bmatrix} 0 & & & & & & \\ \vdots & \ddots & & & & & \\ & & 0 & & & & \\ y_1 & \cdots & y_{m-1} & 0 & & & \\ & & & x_m & 0 & & \\ \vdots & & & \vdots & & \ddots & \\ & & & x_k & & & 0 \end{bmatrix}.$$

The product of the $(i+1, i)$ entries is, in our notation,

$$[m-1:1]y_{m-1}x_m[k:m], \quad x_m \neq 0.$$

By our induction assumptions nilp$[B_{k+1}] = k$ if and only if $y_{m-1} \neq 0$.

In all three cases Lemma 1 holds for $n = k+1$ if it holds for $n = k$. We know it holds for $n = 2$ and so, by the principle of finite induction, it holds for all finite orders. □

Warning. The class with maximal skeleton is not closed under *all* symmetric permutations. Thus the Jordan form does not characterize a class with a given skeleton.

As an introduction to the analysis that follows we display without comment the skeletons and canonical forms, all upper Hessenberg matrices, for $n = 3$.

Skeleton	(0, 0, 0)	(0, 0, 1)	(0, 0, 2)	(0, 1, 1)	(0, 1, 2)
Canonical form	0	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

The canonical forms for (0, 0, 2) and (0, 1, 2) are symmetric permutations of each other. The whole family with skeleton (0, 0, 2) is given by

$$B_3 = \begin{bmatrix} 0 & 0 & f \\ 0 & 0 & g \\ -g & f & 0 \end{bmatrix}. \tag{33}$$

Note that the number of degrees of "freedom" in B is the sum of the nilp in the skeleton.

Trivial extensions. Suppose that $\text{nilp}(B_k) > 0$. Then we can create larger matrices $B_{k+1}, B_{k+2}, \ldots$ with the same nilp by simply appending zero columns and rows, $x = 0$, $y = 0$. To obtain nontrivial B_{k+1} without increasing nilp we shall therefore insist that $x \neq 0$, allowing $y = 0$. However as the results given below reveal we may allow initial zero blocks of arbitrary order:

	(0, 1, 2, 2)	(0, 0, 2, 2)
Nontrivial extensions	$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

It turns out that there is a constraint on how much nilp_{k+1} can exceed nilp_k.

Lemma 2. *Given nilpotent B and nilpotent*

$$C = \begin{pmatrix} B & x \\ y^{\mathrm{T}} & 0 \end{pmatrix}$$

then

$$\text{nilp}[C] \leqslant \min\{2\text{nilp}[B] + 2, \text{order}(B)\}.$$

Proof. From the proof of Lemma 1 the constraints on x and y so that C be nilpotent are $y^{\mathrm{T}} B^j x = 0$ for $j = 0, 1, \ldots, \text{nilp}[B]$ (see (30)). Using these relations repeatedly it may be verified that, for $j \geqslant 2$,

$$C^j = \begin{pmatrix} B^j + \sum_{i=0}^{j-2} B^i x y^{\mathrm{T}} B^{j-2-i} & B^{j-1}x \\ y^{\mathrm{T}} B^{j-1} & 0 \end{pmatrix}$$

The terms in the sum with the smallest powers of B, when j is odd, are $B^{(j-3)/2} x y^{\mathrm{T}} B^{(j-1)/2} + B^{(j-1)/2} x y^{\mathrm{T}} B^{(j-3)/2}$, and $B^m x y^{\mathrm{T}} B^m$ when $j = 2m + 2$.

Now we draw the consequences of these expressions. Take $j = 2\text{nilp}[B] + 3$ and observe that even the smallest powers of B in the (1, 1) block of C^j involve $B^{(j-1)/2}$ and thus vanish because $(j-1)/2 > \text{nilp}[B]$. Thus $\text{nilp}[C] < j = 2\text{nilp}[B] + 3$.

To get a lower bound observe that the (1, 1) block of C^{j-1} is $B^{\text{nilp}[B]} x y^{\mathrm{T}} B^{\text{nilp}[B]}$ and there exist x and y for which this outer product does not vanish. For such x and y, $\text{nilp}[C] = j - 1 = 2\text{nilp}[B] + 2$.

On the other hand, since C is nilpotent, $\mathrm{nilp}[C] \leqslant \mathrm{order}[C] - 1 = \mathrm{order}[B]$. The proof is complete. $\square$

We give an example of maximal increase because it demonstrates that we cannot have a canonical form for non-maximal skeletons which is both upper Hessenberg and has $y = 0$ whenever there is an increase in nilp.

$$\begin{array}{ll} \text{Skeleton} & \begin{pmatrix} 0 & 1 & 1 & 1 & 4 \end{pmatrix} \\ \text{Canonical form?} & \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \end{array}$$

Within the equivalence class $\mathscr{R} = 0$ there is an equivalence relation of sharing the same skeleton. We leave open the task of finding a canonical form for each class under this equivalence.

References

[1] Mark Colarusso, The Gelfand–Zeitlin Algebra and Polarizations of Coadjoint Orbits for Classical Groups, Ph.D. Thesis, University of California at San Diego, 2007.

[2] Augustin Cauchy, Sur l'Équation à l'Aide de Laquelle on Détermine les Inegalités Séculaires des Mouvements des Planètes,Oeuvres Complètes, Séconde Série, 9 (1929) 174–195.

[3] Ky Fan, Gordon Pall, Imbedding conditions for Hermitian and normal matrices, Canad. J. Math. 9 (1957) 298–304.

[4] H. Farahat, W. Ledermann, Matrices with prescribed characteristic polynomial,Proc. Edinburgh Math. Soc. 11 (1959) 143–146.

[5] Bertram Kostant, Nolan Wallach, Gelfand–Zeitlin theory from the perspective of classical mechanics I, Prog. Math. 243 (2006) 319–364.

[6] Bertram Kostant, Nolan Wallach, Gelfand–Zeitlin theory from the perspective of classical mechanics II, Prog. Math. 244 (2006) 387–420.

[7] Graciano N. de Oliveira, Matrices with prescribed characteristic polynomial and prescribed submatrix I–II, Pacific J. Math. 29 (1969) 653–667.

[8] Graciano N. de Oliveira, Matrices with prescribed characteristic polynomial and prescribed submatrix III, Monatsh. Math. 75 (1971) 441–446.

[9] W. Murray Wonham, Linear Multivariable Control, third ed., Springer, New York, 1985.