The Weyr Characteristic

Helene Shapiro

1. INTRODUCTION. The Jordan canonical form is a well-known and standard topic in linear algebra. It is thoroughly covered in many texts on linear algebra and abstract algebra. The purpose of this article is to publicize a different approach to the canonical form problem introduced by Eduard Weyr in 1885 [**28**], [**29**]. Several older books ([**15**, pp. 73–74] and [**16**, pp. 117–118]) mention Weyr characteristics but it does not appear in recent linear algebra texts. The basic idea of Weyr's approach is useful in several areas, such as describing algorithms for computing the Jordan form in a stable manner ([**8**], [**13**], and [**18**]), and in developing canonical forms for matrices under unitary similarity ([**2**], [**14**], [**21**], and [**22**]), but Weyr's papers are rarely referenced and the sequence of numbers we call the Weyr characteristic is not named. Thus, while Weyr's work seems to be little known, his basic idea has been rediscovered and used several times. I first learned of the Weyr characteristic from Hans Schneider, when I was a post-doc at the University of Wisconsin in 1980. Schneider and others have studied the relationship between the Weyr characteristic and the singular graph of an M-matrix ([**9**], [**10**], [**17**], and [**19**]).

In this paper we define the Weyr characteristic and discuss its connection with the so-called "staircase" forms used in numerical linear algebra to determine the Jordan form in a stable manner. There is a simple relationship between the Weyr characteristic and the better known Segre characteristic, which is associated with the Jordan canonical form. This relationship leads to a quick derivation of Weyr's canonical form from the Jordan canonical form; we also present a proof that is independent of the Jordan canonical form, as Weyr did in his original paper.

The Jordan canonical form gives a canonical form for square matrices under the equivalence relation of similarity. It can be used whenever the field contains the eigenvalues of the matrix; typically, one is interested in matrices over the field of complex numbers. The Jordan canonical form of a square matrix A is a direct sum of square submatrices, called *Jordan blocks*. Each such block has an eigenvalue of A in the diagonal entries, a line of 1's along the superdiagonal, and zeros in all other entries, as shown in Figure 1.

$$\begin{pmatrix} \alpha & 1 & 0 & 0 & \cdots & 0 \\ 0 & \alpha & 1 & 0 & \cdots & 0 \\ 0 & 0 & \alpha & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \alpha & 1 \\ 0 & 0 & 0 & 0 & \cdots & \alpha \end{pmatrix}$$

Figure 1. A Jordan block with eigenvalue α.

There is at least one Jordan block for each eigenvalue of A and there may be several Jordan blocks for a single eigenvalue. The list of the non-increasingly ordered sizes of the blocks belonging to a given eigenvalue α is called the *Segre characteristic* of A relative to α. The Jordan canonical form displays all the information needed to know the algebraic structure of a linear transformation. The eigenvalues appear on the main diagonal, and the Segre characteristic reflects the action of the transformation on the generalized eigenspaces. To quote Golub and Wilkinson [**8**, p. 5768], "From the standpoint of classical algebra, the algebraic eigenvalue problem has been completely solved. The problem is the subject of classical similarity theory, and the fundamental result is embodied in the Jordan canonical form."

Weyr's canonical form is a block triangular matrix in which the diagonal blocks are scalar matrices (that is, scalar multiples of identity matrices), the superdiagonal blocks contain identity matrices augmented by rows of zeros, and all the other blocks are zero. The list of the non-increasingly ordered sizes of the diagonal blocks corresponding to an eigenvalue α is called the *Weyr characteristic* of A relative to α. These numbers are determined by the dimensions of the nullspaces of powers of $(A - \alpha I)$; we give precise definitions later. For example, if the Weyr characteristic of A corresponding to α is $(7, 5, 2, 2)$, then the block of the Weyr canonical form of A corresponding to α would have the form shown in Figure 2.

Weyr's approach is related to methods developed in numerical linear algebra for computing the complete eigenstructure of a matrix. While one can derive the Jordan canonical form using an algorithmic approach [**4**], there are numerical reasons to avoid direct computation of the Jordan form [**5**, p. 146]. In numerical computations, one must consider the effect of rounding errors and errors in the input. If the matrix is ill-conditioned with respect to the desired computation, or if the algorithm is not carefully designed, then small errors in the input or rounding errors may result in large errors in the output. Computing the inverse of a matrix that is close to being singular, or applying a similarity that is close to a singular matrix can lead to disaster. It is better to use algorithms that involve only orthogonal or unitary transformations. Algorithms developed by Kublanovskaya [**13**], Ruhe [**18**], and Golub and Wilkinson [**8**] for computing the Jordan canonical form of a matrix in an efficient and stable manner use unitary transformations to transform the matrix to a block triangular, or "staircase" form, in which the block sizes correspond to the Weyr characteristic. These algorithms are typically described in terms of QR factorizations, and/or singular value decompositions, but

$$\begin{pmatrix}
\alpha & & & & & & & 1 & & & & & & & & \\
& \alpha & & & & & & & 1 & & & & & & & \\
& & \alpha & & & & & & & 1 & & & & & & \\
& & & \alpha & & & & & & & 1 & & & & & \\
& & & & \alpha & & & & & & & 1 & & & & \\
& & & & & \alpha & & 0 & 0 & 0 & 0 & 0 & & & & \\
& & & & & & \alpha & 0 & 0 & 0 & 0 & 0 & & & & \\
\hline
& & & & & & & \alpha & & & & & 1 & & & \\
& & & & & & & & \alpha & & & & & 1 & & \\
& & & & & & & & & \alpha & & & 0 & 0 & & \\
& & & & & & & & & & \alpha & & 0 & 0 & & \\
& & & & & & & & & & & \alpha & 0 & 0 & & \\
\hline
& & & & & & & & & & & & \alpha & & 1 & \\
& & & & & & & & & & & & & \alpha & & 1 \\
\hline
& & & & & & & & & & & & & & \alpha & \\
& & & & & & & & & & & & & & & \alpha
\end{pmatrix}$$

Figure 2. The block of the Weyr canonical form corresponding to an eigenvalue α with Weyr characteristic $(7, 5, 2, 2)$.

in theoretical terms, these computations find the null spaces of powers of $(A - \alpha I)$, for each eigenvalue α. Related ideas also appear in Van Dooren's work ([**1**], [**25**], [**26**], and [**27**]) on computing the Kronecker normal form of a matrix pencil, $A + \lambda B$. We do not describe these methods here and refer the reader to the original sources for specific algorithms and an analysis of their stability and efficiency. Our aim is to present Weyr's basic theory and give some proofs that are motivated both by the methods used in the numerical algorithms and by Weyr's original presentation.

2. PRELIMINARIES. We work over an algebraically closed field F. The vector space $V = F^n$ is the space of column vectors of length n over F. If T is a linear operator on V, that is, a linear transformation from V to V, then T can be represented by an $n \times n$ matrix over F, relative to a choice of basis for V; the matrix representation depends on the choice of basis. If A and B are two $n \times n$ matrices that represent T, relative to two choices of basis, then A and B are related by the equation $B = P^{-1}AP$, where the nonsingular matrix P is the change of basis matrix. We say A and B are *similar*.

If F is the field of complex numbers C, we have the usual inner product on C^n. A square, complex matrix U is said to be unitary if $U^{-1} = U^*$ (the star denotes the conjugate transpose); this is equivalent to saying that the columns of U form an orthonormal basis for C^n with respect to the usual inner product. Applying a unitary similarity to A is equivalent to a unitary change of basis.

We frequently deal with matrices that are partitioned into submatrices that have special forms. If A is an $n \times n$ matrix, we may partition the rows of A into t sets consisting of the first n_1 rows, the next n_2 rows, and so on, finishing with the last n_t rows, where $n_1 + n_2 + \cdots + n_t = n$. Partitioning the columns of A in the same way breaks the matrix up into t^2 blocks, A_{ij}, where A_{ij} denotes the block formed from the ith set of rows and the jth set of columns. Note that A_{ij} is $n_i \times n_j$ and the diagonal blocks are square. If all blocks below the diagonal blocks are zero ($A_{ij} = 0$ for $i > j$) then we say A is *block (upper) triangular*. One can visualize the form of such a block triangular matrix as a staircase. If A_i denotes the ith diagonal block (A_{ii}) then we also say that A is $\mathcal{T}(A_1, A_2, \ldots, A_t)$ or write $A = \mathcal{T}(A_1, A_2, \ldots, A_t)$.

$$A = \mathcal{T}(A_1, A_2, \ldots, A_t) = \begin{pmatrix} A_1 & A_{12} & A_{13} & \cdots & A_{1t} \\ 0 & A_2 & A_{23} & \cdots & A_{2t} \\ 0 & 0 & A_3 & \cdots & A_{3t} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & A_t \end{pmatrix}.$$

If A_i and B_i have the same size for each i, then the product of $A = \mathcal{T}(A_1, A_2, \ldots, A_t)$ with $B = \mathcal{T}(B_1, B_2, \ldots, B_t)$ has the form $\mathcal{T}(A_1B_1, A_2B_2, \ldots, A_tB_t)$. When all the off-diagonal blocks are zero ($A_{ij} = 0$ for $i \neq j$) then we say A is *block diagonal*, and say A is $\mathcal{D}(A_1, A_2, \ldots, A_t)$ or write $A = \mathcal{D}(A_1, A_2, \ldots, A_t)$. We also say A is the *direct sum* of $A_1, A_2, \ldots, A_t$.

We use $N(A)$ to denote the null space of A and null(A) for the nullity of A, i.e., the dimension of $N(A)$. The range space of A is denoted by $R(A)$ and rank(A) denotes the rank of A, i.e., the dimension of $R(A)$.

We use I_k to denote the $k \times k$ identity matrix and 0_k for the $k \times k$ zero matrix. For $r > s$, the notation $I_{r,s}$ means a matrix with r rows and s columns in which the first s rows are I_s and the remaining $r - s$ rows are rows of zeroes. For example,

$$I_{5,3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

A matrix with linearly independent columns is said to have *full column rank*; for example $I_{r,s}$ has full column rank. Note that if $r > s$ and A is an $r \times s$ matrix with full column rank, then there exists a nonsingular $r \times r$ matrix B such that $BA = I_{r,s}$.

3. REDUCTION TO THE NILPOTENT CASE. As with the Jordan form, deriving the Weyr form boils down to analyzing the action of the linear transformation on its generalized eigenspaces, and ultimately to analyzing nilpotent transformations.

Let T be a linear operator on V, and let $\operatorname{spec}(T) = \{\alpha_1, \alpha_2, \ldots, \alpha_t\}$ denote the set of distinct eigenvalues, or spectrum, of T. The generalized eigenspace for each eigenvalue α_i of T is the subspace

$$V_{\alpha_i} = \left\{x \in V | (T - \alpha_i I)^k x = 0 \quad \text{for some nonnegative integer } k\right\}.$$

The space V_{α_i} is invariant under T and contains the eigenspace $U_{\alpha_i} = \{x \in V | (T - \alpha_i I)x = 0\}$. Furthermore, V is the direct sum of the generalized eigenspaces V_{α_i}. Thus, setting $V_i = V_{\alpha_i}$, we have $V = V_1 \oplus V_2 \oplus \cdots \oplus V_t$. Now let n_i be the dimension of V_i and let T_i denote the action of T on the subspace V_i. Choose a basis for each V_i and form a basis B for V by taking the union of these bases. Then the matrix of T with respect to B is $\mathscr{D}(n_1, \ldots, n_t)$, where the ith diagonal block represents T_i. Thus, we can describe a canonical form for T by describing a form for the blocks, or for each T_i. Now let $N_i = T - \alpha_i I_{n_i}$. Then N_i is a nilpotent linear operator on V_i and we have reduced the problem to analyzing the action of a nilpotent linear operator or matrix.

4. THE WEYR CHARACTERISTIC FOR THE NILPOTENT CASE. Suppose A is an $n \times n$ nilpotent matrix. The smallest positive integer k such that $A^k = 0$ is called the *index* of A. Then

$$N(A) \subsetneqq N(A^2) \subsetneqq N(A^3) \subsetneqq \cdots \subsetneqq N(A^k) = V$$

and so $0 < \operatorname{null}(A) < \operatorname{null}(A^2) < \cdots < \operatorname{null}(A^k) = n$. For $i = 1, \ldots, k$, set $\omega_i = \operatorname{null}(A^i) - \operatorname{null}(A^{i-1})$. The sequence of positive numbers $\omega_1, \omega_2, \ldots, \omega_k$ is called the *Weyr characteristic* of A; in Lemma 2 we show that the sequence $\omega_1, \omega_2, \ldots, \omega_k$ is non-increasing. We write $\omega(A) = (\omega_1, \omega_2, \ldots, \omega_k)$. Note that $\omega_1 = \operatorname{null}(A)$.

We begin by showing how to compute $\omega(A)$ via a recursive process that avoids computing the powers of A; lemmas 1 and 2 are based on work of Kublanovskaya **[13]**. If $k = 1$, then A is the zero matrix, so we may safely assume that $k \geq 2$. Since $\omega_1 = \operatorname{null}(A)$, the matrix A is similar to a matrix with zeros in the first ω_1 columns and thus we can assume A is in the block form

$$\begin{pmatrix} 0 & A_{12} \\ 0 & A_2 \end{pmatrix}$$

where A_{12} is $\omega_1 \times (n - \omega_1)$ and A_2 is square of size $n - \omega_1$. If we are working over the complex numbers, A can be transformed to this block form with a unitary similarity, because we can choose an orthonormal basis for $N(A)$ and can then extend it to an orthonormal basis for the whole space. Since $\text{rank}(A) = n - \omega_1$, the matrix

$$\begin{pmatrix} A_{12} \\ A_2 \end{pmatrix}$$

has linear independent columns.

Lemma 1. *Suppose A is an $n \times n$ matrix in the form $\mathscr{T}(0_{\omega_1}, A_2)$, where $\omega_1 = \text{null}(A)$. Partition X in F^n as*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

where $X_1 \in F^{\omega_1}$ and $X_2 \in F^{n-\omega_1}$. Then for any given positive integer r, we have $A^r X = 0$ if and only if $A_2^{r-1} X_2 = 0$.

Proof: Since

$$A^r = \begin{pmatrix} 0 & A_{12} A_2^{r-1} \\ 0 & A_2^r \end{pmatrix},$$

we have

$$A^r X = \begin{pmatrix} A_{12} \\ A_2 \end{pmatrix} \left(A_2^{r-1} X_2 \right).$$

Since the rank of A is $n - \omega_1$, the matrix

$$\begin{pmatrix} A_{12} \\ A_2 \end{pmatrix}$$

has linearly independent columns, and so

$$\begin{pmatrix} A_{12} \\ A_2 \end{pmatrix} Y = 0$$

if and only if $Y = 0$. Putting $Y = A_2^{r-1} X_2$ we see that $A^r X = 0$ if and only if $A_2^{r-1} X_2 = 0$. ■

Lemma 2. *Let $A = \mathscr{T}(0_{\omega_1}, A_2)$ be an $n \times n$, nonzero, nilpotent matrix with Weyr characteristic $\omega(A) = (\omega_1, \omega_2, \ldots, \omega_k)$. Then $\omega(A_2) = (\omega_2, \ldots, \omega_k)$. Furthermore, $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_k$.*

Proof: Lemma 1 ensures that $\text{null}(A^i) = \omega_1 + \text{null}(A_2^{i-1})$, so for each $i \geq 2$ we have $\text{null}(A_2^{i-1}) - \text{null}(A_2^{i-2}) = \text{null}(A^i) - \text{null}(A^{i-1}) = \omega_i$. Thus, $\omega(A_2) = (\omega_2, \ldots, \omega_k)$.

To prove that $\omega_{i+1} \leq \omega_i$ we use induction on k, starting with $k = 2$. Now, $\text{rank}(A) \leq \text{rank}(A_{12}) + \text{rank}(A_2)$. Substituting $\text{rank}(A) = n - \omega_1$ and $\text{rank}(A_2) = (n - \omega_1) - \text{null}(A_2)$ gives $\text{null}(A_2) \leq \text{rank}(A_{12})$. But $\omega_2 = \text{null}(A_2)$ and $\text{rank}(A_{12}) \leq \omega_1$, so $\omega_2 \leq \omega_1$. By the induction hypothesis, the result holds for the matrix A_2 and so we have $\omega_{i+1} \leq \omega_i$ for all $i \geq 2$. ■

Lemma 2 leads to a recursive process for computing the Weyr characteristic of a nilpotent matrix. First one applies a similarity to put A in the form $\mathscr{T}(0_{\omega_1}, A_2)$, where $\omega_1 = \text{null}(A)$. This is equivalent to finding the null space of A and choosing a basis, B, for V in which the first ω_1 vectors of B are a basis for $N(A)$. When $F = C$, this can be done with a unitary similarity by choosing B to be an orthonormal basis. Lemma 2 tells us that we have now reduced the problem to finding the Weyr characteristic of the smaller matrix A_2. Repeated application of Lemma 2 leads to a block triangular form in which the diagonal blocks are zero blocks of sizes $\omega_1, \omega_2, \ldots, \omega_k$. In Section 5 we examine this form more carefully and show that the superdiagonal blocks have full column rank; this leads to the Weyr canonical form.

We now look at the relationship between the Weyr and Segre characteristics of A. Let S_r denote the $r \times r$ matrix with a 1 in each superdiagonal position and zeros elsewhere; S_r is a nilpotent matrix of index r. Observe that as we form powers of S_r, the superdiagonal line of ones moves upwards, and for $1 \leq m \leq r$, the power S_r^m has rank $r - m$ and nullity m. The Jordan canonical form of A is $J = \mathscr{D}(S_{\sigma_1}, S_{\sigma_2}, \ldots, S_{\sigma_t})$ where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_t$. The list $(\sigma_1, \sigma_2, \ldots, \sigma_t)$ is the Segre characteristic of A. Since each block S_{σ_i} has nullity one, $\text{null}(A) = t$. Hence, if $\omega(A) = (\omega_1, \omega_2, \ldots, \omega_k)$, then $\omega_1 = t$ is the number of blocks in the Jordan form of A. The nullity of J^2 exceeds $\text{null}(J)$ by exactly the number of blocks of size at least two, so $\text{null}(J^2) = t +$ (the number of blocks of size 2 or more). But $\text{null}(A^2) = \omega_1 + \omega_2$, so ω_2 is the number of blocks in the Jordan form that have size at least 2. Now if we look at J^3, we see that $\text{null}(J^3)$ exceeds $\text{null}(J^2)$ by exactly the number of blocks in J with size greater than or equal to 3, so ω_3 is the number of blocks in the Jordan form that have size at least 3. In general, computing $\text{null}(J^m)$ shows that ω_m is the number of blocks in the Jordan form that have size at least m. If we regard the Weyr and Segre characteristics as partitions of n, then the Weyr characteristic is the *conjugate partition* of the Segre characteristic, and we can easily derive one from the other. Using a *Ferrers diagram* to represent the partition $\omega(A) = (\omega_1, \omega_2, \ldots, \omega_t)$, the number of dots in row i is ω_i, while σ_i is the number of dots in column i. For example, if $\omega(A) = (4,3,3,2,2,2,1)$, then the Segre characteristic for A is $(7,6,3,1)$ and the corresponding Ferrers diagram is shown in Figure 3.

5. THE WEYR CANONICAL FORM FOR THE NILPOTENT CASE. We now obtain Weyr's canonical form for the nilpotent case. Since two nilpotent matrices have the same Weyr characteristic if and only if they have the same Segre characteristic, we see that two nilpotent matrices are similar if and only if they have the same Weyr characteristic. Now let $W = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \ldots, 0_{\omega_k})$ be the block

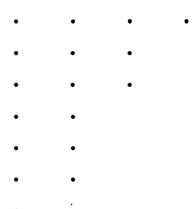

Figure 3. Ferrers diagram for $(4,3,3,2,2,2,1)$.

triangular matrix in which each superdiagonal block is $W_{i,i+1} = I_{\omega_i, \omega_{i+1}}$ and all other blocks are zero. Thus,

$$W = \begin{pmatrix} 0_{\omega_1} & I_{\omega_1, \omega_2} & 0 & \cdots & 0 \\ & 0_{\omega_2} & I_{\omega_2, \omega_3} & \cdots & 0 \\ & & \ddots & & \\ & & & 0_{\omega_{k-1}} & I_{\omega_{k-1}, \omega_k} \\ 0 & & & & 0_{\omega_k} \end{pmatrix}.$$

Direct calculation of the powers of W shows that W has Weyr characteristic $(\omega_1, \omega_2, \dots, \omega_k)$. Hence, W is a canonical form for all nilpotent matrices with Weyr characteristic $(\omega_1, \omega_2, \dots, \omega_k)$.

This approach is quick and easy, but it depends on the Jordan canonical form. Weyr, of course, developed his theory independently. The remainder of this section presents a derivation of Weyr's form that does not depend on the Jordan canonical form. We use Lemma 2 to obtain a block triangular form $\mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$, show that the superdiagonal blocks have full column rank, and then show how to further reduce this form to obtain the Weyr canonical form. The proofs of the main results are by induction; to get started we need the following lemma.

Lemma 3. *Let T be a nilpotent linear operator on V with $\omega(T) = (\omega_1, \omega_2, \dots, \omega_k)$. Then T can be represented by a matrix $A = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \hat{A})$, where $\operatorname{rank}(A_{12}) = \omega_2$ and so A_{12} has full column rank.*

Proof: Since $\omega_1 = \operatorname{null}(T)$, we can represent T by a matrix $B = \mathscr{T}(0_{\omega_1}, B_2)$. Lemma 2 ensures that $\omega_2 = \operatorname{null}(B_2)$ so there is a square matrix Q of size $n - \omega_1$ such that $Q^{-1}B_2Q = \mathscr{T}(0_{\omega_2}, \tilde{A})$. Now let $P = \mathscr{D}(I_{\omega_1}, Q)$; then $P^{-1}BP = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \tilde{A})$, so $A = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \tilde{A})$ is a matrix representation for T:

$$\mathscr{T}\left(0_{\omega_1}, 0_{\omega_2}, \tilde{A}\right) = \begin{pmatrix} 0_{\omega_1} & A_{12} & A_{13} \\ 0 & 0_{\omega_2} & A_{23} \\ 0 & 0 & \tilde{A} \end{pmatrix}.$$

Since A has rank $n - \omega_1$, the last $n - \omega_1$ columns of A must be linearly independent, and hence the block A_{12} (which is $\omega_1 \times \omega_2$) must have full column rank. ∎

When $k = 2$, Lemma 3 tells us that T can be represented by a block triangular matrix $\mathscr{T}(0_{\omega_1}, 0_{\omega_2})$, where the $\omega_1 \times \omega_2$ block A_{12} has full column rank, i.e., $\operatorname{rank}(A_{12}) = \omega_2$.

Remark 1. If $F = C$, then in the proof of Lemma 3, we can use an orthonormal basis for C^n in which the first ω_1 vectors are a basis for $N(T)$ and can use a unitary matrix for Q. Hence, we can obtain a representation for T in the form given in Lemma 3 by using an appropriate orthonormal basis.

Theorem 1. *Let T be a nilpotent linear operator on V. Then $\omega(T) = (\omega_1, \omega_2, \dots, \omega_k)$ if and only if T can be represented by a block triangular matrix $A = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$*

in which each superdiagonal block $A_{i,i+1}$ has full column rank, i.e., rank$(A_{i,i+1}) = \omega_{i+1}$.

Proof: We use induction on k. Assume $\omega(T) = (\omega_1, \omega_2, \dots, \omega_k)$. If $k = 1$, then T is the zero matrix. If $k = 2$, then Lemma 3 gives the result. For the general case, we apply Lemma 3 to see that T has a matrix representation $B = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \tilde{B})$, where B_{12} has full column rank. Let B_2 denote the square submatrix in the last $n - \omega_1$ rows and columns; then B_2 is $\mathscr{T}(0_{\omega_2}, \tilde{B})$. Lemma 2 tells us that $\omega(B_2) = (\omega_2, \dots, \omega_k)$, so by the induction hypothesis, there is a nonsingular matrix Q, of size $n - \omega_1$, such that $Q^{-1}B_2Q = \mathscr{T}(0_{\omega_2}, 0_{\omega_3}, \dots, 0_{\omega_k})$ with each superdiagonal block having full column rank. Apply the similarity $P = \mathscr{D}(I_{\omega_1}, Q)$ to B to get a matrix, A, in the desired form.

To prove the converse, it suffices to show that a matrix $A = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$ with superdiagonal blocks of full column rank has Weyr characteristic $(\omega_1, \omega_2, \dots, \omega_k)$. We again use induction on k. Observe that the last $n - \omega_1$ columns of such a matrix are linearly independent, so null$(A) = \omega_1$. If $k = 1$, then A is the zero matrix and we are done. Otherwise, A has the form $\mathscr{T}(0_{\omega_1}, A_2)$ given in Lemma 1, and Lemma 2 tells us that the Weyr characteristic of A is $(\omega_1, \omega_2', \dots, \omega_k')$, where $(\omega_2', \dots, \omega_k') = \omega(A_2)$. But the induction hypothesis then tells us that $\omega_i = \omega_i'$ for $i \geq 2$ and we are done. ■

Using Remark 1 and a unitary matrix for the matrix Q in the proof of Theorem 1, we obtain the following unitary version of Theorem 1.

Theorem 1′. *Let A be an $n \times n$ nilpotent complex matrix. Then $\omega(A) = (\omega_1, \omega_2, \dots, \omega_k)$ if and only if there is a unitary matrix U such that U^*AU is a block triangular matrix of the form $\mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$ in which each superdiagonal block $A_{i,i+1}$ has full column rank, i.e., rank$(A_{i,i+1}) = \omega_{i+1}$.*

It is also possible to apply further unitary similarities to reduce the superdiagonal blocks to special forms; see **[2]**, **[21]**, and **[22]**.

For purposes of computing the Weyr characteristic, one would stop with the staircase form of Theorem 1′, which can be reached via a unitary similarity. However, this block triangular form is not unique; for a canonical form we must go further and use non-unitary similarities.

Theorem 2. *Let T be a nilpotent linear operator on V. Then $\omega(T) = (\omega_1, \omega_2, \dots, \omega_k)$ if and only if T can be represented by the block triangular matrix $W = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$, in which the only nonzero blocks are the superdiagonal blocks $W_{i,i+1} = I_{\omega_i, \omega_{i+1}}$, $i = 1, \dots, k-1$.*

Proof: Using Theorem 1, it suffices to show that a matrix $B = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$ in which each superdiagonal block has full column rank is similar to W. We use induction on k. When $k = 1$, we have $B = 0$ and there is nothing to do. Assume $k > 1$. The matrix occupying the last $n - \omega_1$ rows and columns of B has Weyr characteristic $(\omega_2, \omega_3, \dots, \omega_k)$, so the induction hypothesis ensures that it is similar to a matrix in the desired form. Thus, there is a nonsingular matrix Q, of size $n - \omega_1$, such that $C = \mathscr{D}(I_{\omega_1}, Q^{-1})B\mathscr{D}(I_{\omega_1}, Q)$ has the desired form except possibly in the first row of blocks, $(0_{\omega_1}, C_{12}, C_{13}, \ldots, C_{1k})$. Thus,

$$C = \begin{pmatrix} 0_{\omega_1} & C_{12} & C_{13} & C_{14} & \cdots & C_{1k} \\ & 0_{\omega_2} & I_{\omega_2,\,\omega_3} & 0 & \cdots & 0 \\ & & 0_{\omega_3} & I_{\omega_3,\,\omega_4} & \cdots & 0 \\ & & & \ddots & & I_{\omega_{k-1},\,\omega_k} \\ 0 & & & & & 0_{\omega_k} \end{pmatrix}.$$

Now, $\text{null}(B) = \text{null}(C) = \omega_1$ so C_{12} has full column rank. We now reduce C to the desired form in two steps. First, we clear out the blocks $C_{13}, \ldots, C_{1k}$, and then we reduce C_{12} to the form $I_{\omega_1,\,\omega_2}$.

The block C_{1r} is $\omega_1 \times \omega_r$; let $\tilde{C}_{1r}$ denote the $\omega_1 \times \omega_{r-1}$ matrix obtained by adjoining $\omega_{r-1} - \omega_r$ columns of zeros to C_{1r}. Thus, we have

$$\tilde{C}_{1r} = \begin{pmatrix} C_{1r} & 0_{\omega_1 \times (\omega_{r-1} - \omega_\omega)} \end{pmatrix},$$

and $\tilde{C}_{1r} I_{\omega_{r-1},\,\omega_r} = C_{1r}$. Now let P be the matrix of the form $\mathscr{T}(I_{\omega_1}, I_{n-\omega_1})$ in which the first ω_1 rows are the blocks $(I_{\omega_1}, \tilde{C}_{13}, \tilde{C}_{14}, \ldots, \tilde{C}_{1k}, 0_{\omega_1 \times \omega_k})$, that is,

$$P = \left(\begin{array}{c|c} I_{\omega_1} & \tilde{C}_{13}\tilde{C}_{14} \cdots \tilde{C}_{1k}\, 0_{\omega_1 \times \omega_k} \\ \hline & I_{n-\omega_1} \end{array}\right).$$

Then P^{-1} has the same form, but its first ω_1 rows are the blocks $(I_{\omega_1}, -\tilde{C}_{13}, -\tilde{C}_{14}, \ldots, -\tilde{C}_{1k}, 0_{\omega_1 \times \omega_k})$. A computation using block multiplication shows that $P^{-1}CP$ has C_{12} in its $1,2$ block, but otherwise has the desired form.

Since C_{12} has full column rank, there is a nonsingular $\omega_1 \times \omega_1$ matrix W such that $WC_{12} = I_{\omega_1,\,\omega_2}$. Let $S = \mathscr{D}(W^{-1}, I_{\omega_2}, I_{\omega_3}, \ldots, I_{\omega_k})$; then $S^{-1}P^{-1}CPS$ has the desired form. ∎

6. THE GENERAL CASE. We can now use our form for the nilpotent case to deal with a general linear operator T. As described in Section 2, we can decompose T into a direct sum $T_1 \oplus T_2 \oplus \cdots T_t$, where each T_i is the action of T on the generalized eigenspace V_i. Then $T_i - \alpha_i I$ is a nilpotent transformation on V_i. We say that $\omega(T_i - \alpha_i I)$ is the *Weyr characteristic of T, relative to the eigenvalue α_i*. Let W_i be the Weyr canonical form of N_i; then T can be represented by the block diagonal matrix $\mathscr{D}(\alpha_1 I + W_1, \alpha_2 I + W_2, \ldots, \alpha_t I + W_t)$. This is the canonical form described by Weyr [**28**]; we call it the *Weyr canonical form* of T. For each eigenvalue, α_i, the Weyr characteristic, $\omega(T_i - \alpha_i I)$ is related to the Segre characteristic for α_i as described in Section 4, and so the Jordan canonical form of a matrix can be read off from the Weyr canonical form, and vice versa.

7. OBTAINING THE WEYR CHARACTERISTIC BY UNITARY SIMILARITY. Two $n \times n$ complex matrices, A and B, are *unitarily similar* if there is a unitary matrix U such that $B = U^*AU$. In general, a matrix is not unitarily similar to its Jordan or Weyr canonical form. However, in numerical computations, it is desirable to obtain the information needed to specify the canonical form by using only unitary similarities. We briefly outline, in theory, why the Weyr characteristic can be found using only unitary similarities.

The process begins with Schur's result that a square complex matrix can be triangularized with a unitary similarity [**11**, pp. 79–81].

Theorem (Schur [**20**]). *If A is an $n \times n$ complex matrix, then there is a unitary matrix U such that U^*AU is triangular.*

Proof: Start with an eigenvalue, α_1, of A and an associated eigenvector x, where x has length one. Then construct an orthonormal basis for C^n in which x is the first basis element. Let U_1 be the unitary matrix that has the basis vectors in its columns. Then $U_1^*AU_1$ has the form $\mathscr{T}(\alpha_1, A_1)$ where A_1 is $(n-1) \times (n-1)$. Using induction, let U_2 be a unitary matrix of size $n-1$ that puts A_1 in triangular form and let $U_2 = \mathscr{D}(1, \tilde{U}_2)$. Then if $U = U_1U_2$, the matrix U^*AU is triangular. ■

Note that we can obtain a triangular form for A with the eigenvalues in any given order along the diagonal. Thus, if $\text{spec}(A) = \{\alpha_1, \alpha_2, \dots, \alpha_t\}$, where α_i has multiplicity n_i, we can unitarily put A into the form $\mathscr{T}(A_1, A_2, \dots, A_t)$ where A_i is an $n_i \times n_i$ triangular matrix with α_i along its diagonal.

The next step is to show that $\mathscr{T}(A_1, A_2, \dots, A_t)$ is similar to $\mathscr{D}(A_1, A_2, \dots, A_t)$, for this will tell us that the Weyr characteristic of A, relative to the eigenvalue α_i is simply the Weyr characteristic of the nilpotent matrix $A_i - \alpha_i I$. To show that $\mathscr{T}(A_1, A_2, \dots, A_t)$ and $\mathscr{D}(a_1, A_2, \dots, A_t)$ are similar, we use a well-known theorem of Sylvester, which may be found in many sources, e.g., [**3**], [**6**, Vol 1, p. 225], [**11**, Section 2.4, Problems 9 and 13], and [**12**, Theorem 4.4.6].

Theorem (Sylvester) [**23**]. *Let A be $m \times m$ and B be $n \times n$. Then the matrix equation $AX - XB = C$ has a unique solution for every $m \times n$ matrix C if and only if $\text{spec}(A) \cap \text{spec}(B) = \emptyset$.*

Lemma 4. *If $A = \mathscr{T}(A_1, A_2)$ and $\text{spec}(A_1) \cap \text{spec}(A_2) = \emptyset$ then A is similar to $\mathscr{D}(A_1, A_2)$.*

Proof: Let A_i be size $n_i \times n_i$ for $i = 1, 2$. Let X be the unique $n_1 \times n_2$ matrix that satisfies $A_1X - XA_2 = -A_{12}$. Let S be of the form $\mathscr{T}(I_{n_1}, I_{n_2})$ with X in the 1,2 block. Then S^{-1} is $\mathscr{T}(I_{n_1}, I_{n_2})$ with $-X$ in the 1,2 block. A computation then shows that $S^{-1}AS$ is $\mathscr{D}(A_1, A_2)$.

Using Lemma 4 with an induction argument proves the following result.

Theorem 3. *If $A = \mathscr{T}(A_1, A_2, \dots, A_t)$, where each $\text{spec}(A_i) = \{\alpha_i\}$ and $\alpha_i \neq \alpha_j$ when $i \neq j$, the A is similar to $\mathscr{D}(A_1, A_2, \dots, A_t)$.*

Thus, once we have A in the triangular form $\mathscr{T}(A_1, A_2, \dots, A_t)$, we can find the Weyr characteristic of each eigenvalue of A by finding the Weyr characteristic of each nilpotent block $A_i - \alpha_i I$. As pointed out in Section 4, this can be done with a recursive procedure and can be done with unitary transformations. We refer the reader to references [**8**], [**13**], and [**18**] for detailed information on numerical algorithms and the stability issues involved.

REFERENCES

1. Th. Beelen and P. Van Dooren, An improved algorithm for the computation of Kronecker's canonical form for a singular pencil, *Linear Algebra Appl.* **105** (1988) 9–65.
2. R. Benedetti and P. Cragnolini, Versal families of matrices with respect to unitary conjugation, *Adv. Math.* **54** (1984) 314–335.
3. R. Bhatia and P. Rosenthal, How and why to solve the operator equation $AX - XB = Y$, *Bull. London Math. Soc.* **29** (1997) 1–21.

4. Richard Brualdi, The Jordan Canonical Form: an Old Proof, *Amer. Math. Monthly* **94** (1987) 257–267.
5. J. W. Demmel, *Applied Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, 1997.
6. F. R. Gantmacher, *The Theory of Matrices*, Vols. 1, 2, Chelsea, New York, 1959.
7. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd edition, The Johns Hopkins University Press, Baltimore and London, 1989.
8. G. H. Golub and J. H. Wilkinson, Ill-conditioned eigensystems and the computation of the Jordan canonical form, *SIAM Review* **18** (1976) 578–619.
9. D. Hershkowitz and H. Schneider, On the existence of matrices with prescribed height and level characteristics, *Israel J. Math.* **75** (1991) 105–117.
10. D. Hershkowitz and H. Schneider, Height bases, level bases, and the equality of the height and the level characteristics of an M-matrix, *Linear and Multilinear Algebra*, **25** (1989) 149–171.
11. R. Horn and C, Johnson, *Matrix Analysis*, Cambridge U. P., Cambridge, 1985.
12. R. Horn and C. Johnson, *Topics in Matrix Analysis*, Cambridge U. P., Cambridge, 1990.
13. V. N. Kublanovskaya, On a method of solving the complete eigenvalue problem for a degenerate matrix, *U.S.S.R. Comput. Math. and Math. Physics* **6** (1966) 1–14.
14. D. E. Littlewood, On unitary equivalence, *J. London Math. Soc.* **28** (1953) 314–322.
15. C. C. MacDuffee, *The Theory of Matrices*, Springer Verlag, Berlin, 1933.
16. A. I. Mal'cev, *Foundations of Linear Algebra*, W. H. Freeman and Company, San Francisco and London, 1963.
17. D. J. Richman and H. Schneider, On the singular graph and the Weyr characteristic of an M-matrix, *Aequationes Math.* **17** (1978) 208–234.
18. A. Ruhe, An algorithm for numerical determination of the structure of a general matrix, *BIT* 10 (1970) 196–216.
19. H. Schneider, The influence of the marked reduced graph of a nonnegative matrix on the Jordan form and related properties: A survey, *Linear Algebra Appl.* **84** (1986) 161–189.
20. I. Schur, Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen, *Math. Ann.* **66** (1909) 488–510.
21. V. V. Sergeichuk, Classification of linear operators in a finite-dimensional unitary space, *Functional Anal. Appl.* **18** (1984) 224–230.
22. H. Shapiro, A survey of canonical forms and invariants for unitary similarity, *Linear Algebra Appl.* **147** (1991) 101–167.
23. J. J. Sylvester, Sur l'equation en matrices $px = xq$, *C. R. Acad. Sci. Paris* **99** (1884) 67–71 and 115–116.
24. H. W. Turnbull and A. C. Aitken, *An Introduction to the theory of Canonical Matrices*, Blackie & Son Limited, London and Glasgow, 1932.
25. P. Van Dooren, *The Generalized Eigenstructure Problem*; *Applications in Linear System Theory*, Ph.D. Thesis, Katholieke Universiteit Leuven, May, 1979.
26. P. Van Dooren, The computation of Kronecker's canonical form of a singular pencil, *Linear Algebra Appl.* **27** (1979) 103–140.
27. P. Van Dooren, The generalized eigenstructure problem in linear system theory, *IEEE Trans. Automatic Control* **26** (1981) 111–129.
28. E. Weyr, Zur Theorie der bilinearen Formen, *Monatsh. Math. und Physik* **1** (1890) 163–236.
29. E. Weyr, Répartition des matrices en espèces et formation de toutes les espèces, *C. R. Acad. Sci. Paris* **100** (1885) 966–969.

HELENE SHAPIRO received her B.A. degree from Kenyon College in 1975 and her Ph.D. from the California Institute of Technology in 1979, where she was a student of Olga Taussky Todd. She spent one year at the University of Wisconsin and then joined the Department of Mathematics and Statistics at Swarthmore College, where she has been teaching since 1980.
Department of Mathematics and Statistics, Swarthmore College, Swarthmore, PA 19081
hshapir1@swarthmore.edu