A Survey of Canonical Forms and Invariants for Unitary Similarity

Helene Shapiro
Department of Mathematics
Swarthmore College
Swarthmore, Pennsylvania 19081

Submitted by Roger A. Horn

ABSTRACT

Matrices A and B are said to be unitarily similar if $U^*AU = B$ for some unitary matrix U. This expository paper surveys results on canonical forms and invariants for unitary similarity. The first half gives a detailed description of methods developed by several authors (Brenner, Littlewood, Mitchell, McRae, Radjavi, Sergeĭchuk, and Benedetti and Cragnolini) using inductively defined reduction procedures to transform matrices to canonical form. The matrix is partitioned and successive unitary similarities applied to reduce the submatrices to some nice form. At each stage, one refines the partition and restricts the set of permissible unitary similarities to those that preserve the already reduced blocks. The process ends in a finite number of steps, producing both the canonical form and the subgroup of the unitary group that preserves that form. Depending on the initial step, various canonical forms may be defined. The method can also be used to define canonical forms relative to certain subgroups of the unitary group, and canonical forms for finite sets of matrices under simultaneous unitary similarity. The remainder of the paper surveys results on unitary invariants and other topics related to unitary similarity, such as the Specht-Pearcy trace invariants, the numerical range, and unitary reducibility.

1. INTRODUCTION

This paper is a survey of results on unitary similarity, particularly methods for reducing matrices to special canonical forms via unitary similarities, and properties or quantities that are invariant under unitary similarity. For the equivalence relation of similarity, these problems are fairly well understood and there are well-known standard solutions. This is not the case for the more specialized relation of unitary similarity, where the situation is

655 Avenue of the Americas, New York, NY 10010 0024-3795/91/$3.50

more complicated and much of the published work on canonical forms seems to be little known.

We work over $\mathbf{C}$, the field of complex numbers. We use $\mathbf{C}^n$ to denote the vector space of column vectors of length n over $\mathbf{C}$, with the usual inner product $\langle \mathbf{x},\mathbf{y}\rangle = \sum_{i=1}^{n} x_i\bar{y}_i$ and norm $\|\mathbf{x}\| = \langle \mathbf{x},\mathbf{x}\rangle^{1/2}$. The set of $m\times n$ complex matrices will be called $\mathbf{C}(m,n)$, while $\mathbf{C}(n)$ will denote the set of $n\times n$ complex matrices. If A is a matrix, A^T means the transpose of A, and A^* denotes the conjugate transpose. Two $n\times n$ matrices, A and B, are said to be similar if $S^{-1}AS = B$ for some nonsingular matrix S; if there is a unitary matrix U such that $U^{-1}AU = U^*AU = B$, then we say A and B are unitarily similar.

The algebra $\mathbf{C}(n)$ may be viewed as the set of linear transformations on the finite dimensional Hilbert space $\mathbf{C}^n$; it is also an example of a $\mathbf{C}^*$-algebra. While we deal mainly with the finite dimensional case, some of the results discussed have generalizations to bounded, linear operators on Hilbert spaces. The term operator will always mean a bounded, linear operator.

An $m\times n$ matrix A represents a linear transformation from $\mathbf{C}^n$ to $\mathbf{C}^m$ with respect to a choice of bases for $\mathbf{C}^n$ and $\mathbf{C}^m$. One can phrase the statements and proofs either in matrix language or in terms of vector spaces and choice of bases. Unless otherwise noted, we usually use the standard bases of unit coordinate vectors. If $m = n$, we may view the similar matrices A and $S^{-1}AS$ as representations of the same linear transformation with respect to two different bases; the nonsingular matrix S tells us how to change to the new basis. In the canonical form problem for the equivalence relation of similarity one tries to find a basis in which the linear transformation has a particularly nice matrix representation. If we restrict ourselves to unitary similarity, the new basis determined by the unitary matrix U in U^*AU must be orthonormal. Thus in the canonical form problem for unitary similarity, one wants to find an orthonormal basis in which the transformation has a nice matrix representation. Unitary transformations preserve the inner product and thus preserve important geometric information. They are also important in numerical methods for stability reasons [122]. Since the equivalence relation of unitary similarity is more restrictive than similarity, each similarity class of matrices will generally contain many different equivalence classes of unitarily similar matrices.

Another familiar equivalence relation on the set of $n\times n$ matrices is congruence—we say A and B are congruent if there is some nonsingular P such that $B = P^TAP$. This arises naturally in the study of real quadratic and bilinear forms; symmetric matrices A and B represent the same bilinear form, relative to different bases, if and only if they are congruent. The transforming matrix P represents the change of basis. The complex analogue of this relation is conjunctivity—we say complex matrices A and B are conjunctive if there exists a nonsingular P such that $B = P^*AP$. This arises in the theory of conjugate, bilinear forms; Hermitian matrices A and B represent the same conjugate-symmetric, conjugate-bilinear form, relative to different bases, if and only if they are conjunctive. If U is unitary, then $U^* = U^{-1}$, so $U^*AU = U^{-1}AU$ is both a similarity and a conjunctive transformation.

The equivalence relation of unitary congruence, where A and B are said to be unitarily congruent if $U^TAU = B$ for some unitary matrix U, has also been studied [43–46, 124].

In general, one can study an equivalence relation in several ways. One approach is to find a canonical form—i.e., determine a "nice" set of canonical matrices such that each equivalence class corresponds to exactly one canonical matrix. Two matrices are then equivalent if and only if they have the same canonical form. For the equivalence relation of similarity, the rational canonical form, which applies over any field, and the Jordan canonical form, which applies over an algebraically closed field, are two standard solutions to this problem. Alternatively, one might seek a set of invariants that completely specify the equivalence class, so that two matrices will be equivalent if and only if they have the same set of invariants. For the similarity relation, one may use determinantal divisors, invariant factors, or elementary divisors. Note that if one knows any one of these sets of invariants, one can find either of the other sets of invariants. Furthermore, the invariants tell one the Jordan and rational canonical forms of the matrix, and conversely, one can read off the invariants from the canonical form.

These canonical forms and invariants for similarity classes of matrices are well known. Many algebra and linear algebra texts [33, 39, 48, 50, 61, 65, 121] discuss the theory of determinantal divisors, elementary divisors, invariant factors, Jordan canonical form, and rational canonical form; the subject also arises as an application of the structure theorem for finitely generated modules over principal ideal domains. For symmetric and Hermitian matrices, congruence and conjunctivity are also well understood, and many texts discuss diagonalization, reduction to sums of squares, rank and signature, and Sylvester's inertia theorem. However, the equivalence relation of unitary similarity is less well understood. More information is needed to specify a matrix up to unitary similarity than to determine it simply up to similarity, and more invariants are needed to completely specify an equivalence class. Thus, any canonical form for unitary similarity must display more information than the Jordan form, and hence can be expected to be more complicated.

A normal matrix can be diagonalized with a unitary similarity, but in the general case the form is much more complicated. Except for Schur's result [97] that any complex matrix can be put in triangular form via a unitary similarity, and the trace invariants found by Specht [110], much of the literature on this subject seems to be little known. Many authors [9, 12, 18, 52, 59, 62, 69, 81, 89, 90, 92, 99] have studied the problem of finding a canonical form for unitary similarity and proposed methods for reducing a matrix to a canonical form under unitary similarity. As we shall see, these reduction methods are based on inductive procedures that involve partitioning the matrix into blocks and successively applying unitary transformations to reduce the blocks to some nice form. The final "canonical form" is usually not easily visualized. Recent workers seem to be unaware of some of the work done by earlier authors, and the basic ideas used in the reduction procedures seem to have been rediscovered several times. Several approaches have been proposed, but they share several common ideas. In this survey we try to give a unified presentation of the known results.

Schur's theorem [97], which appears in a 1909 paper, seems to be the most widely known result on unitary similarity. Röseler's 1933 Ph.D. dissertation [92] studies the problem of finding normal forms for matrices under unitary similarity and uses Schur's theorem to deal with some special cases. Currie's 1950 abstract [18] describes a triangular canonical form, but Currie does not seem to have published these results. Brenner [12] gives an inductive definition of a canonical form in his 1951 paper on this problem; the basic ideas of Brenner's induction argument reappear in later work. Brenner suggests that the nonuniqueness of the triangular form established in Schur's theorem makes this an unsuitable starting point, and proposes an alternative method based on the structure of A^*A. However, Littlewood [59], in his 1953 paper, does begin with Schur's theorem, first using it to describe a canonical form for matrices with distinct eigenvalues, and then developing methods to deal with the general case. Although Littlewood's description of the actual reduction procedure seems a bit sketchy, his method of treating the triangular form reappears in more recent work. Mitchell's 1954 paper [69] deals with the nonderogatory case and also proposes a triangular canonical form; McRae's 1955 dissertation [62], among other things, proposes a general scheme for defining various canonical forms under unitary similarity, including the approach proposed by Brenner. McRae also discusses triangular forms, but seems to be unaware of the Littlewood and Mitchell papers. In 1962, Radjavi [89] described an algorithm for constructing a canonical form; Radjavi's approach is related to Brenner's work but develops a detailed reduction process reducing a matrix to canonical form. Radjavi gives a different reduction method in a 1968 paper [90]. Recent papers (1984) by Sergeĭchuk [99] and Benedetti and Cragnolini [9] propose schemes for reducing a general matrix to a triangular canonical form. Some of the key ideas of these 1984 papers can actually be found in Littlewood's paper, as well as in the work of Röseler, McRae, and Mitchell. These earlier papers, however, seem to be unknown to the recent authors, as they do not appear in the references of either of the 1984 papers. Sergeĭchuk [99] and Benedetti and Cragnolini [9] do precisely specify the actual reduction procedure.

Sections 2–5 contain a fairly detailed account of the known results on the canonical form problem. Since the key ideas and arguments appear in several versions and notations in the literature, and seem to have been rediscovered several times by authors who were unaware of earlier versions, it seems worthwhile to give a fairly complete account of the methods and proofs in this survey. Sections 6–8 deal with issues related to unitary similarity such as unitary invariants, the numerical range, and unitary reducibility. These sections have fewer details and proofs. Unlike the situation for similarity, where one can determine the canonical form from the invariants and vice versa, there seems to be little connection between the known canonical forms and sets of invariants for matrices under unitary similarity. Only in the 2×2 case does one get a completely satisfying picture of the situation.

In Section 2, we discuss matrices with distinct eigenvalues and the 2×2 case. In these special cases, the triangular canonical form is fairly easy to describe and understand, and also helps to introduce some of the basic problems and ideas involved in the general case. Section 3 contains preliminary results that will provide the basic tools for describing and developing canonical forms. These results also show that the method used for matrices with distinct eigenvalues applies to nonderogatory matrices, as shown by Mitchell in [69]. In Section 4, we develop the technical details needed to deal with the general case and describe the reduction procedure for the triangular type canonical forms of Littlewood [59], Sergeĭchuk [99], and Benedetti and Cragnolini [9]. The general treatment is somewhat complicated. Section 5 deals with nontriangular forms and the alternative approaches proposed by Brenner [12] and Radjavi [89, 90]. We then move on to the question of invariants in Section 6. The main result here is a theorem of Specht [110], which gives a complete set of unitary invariants in terms of traces of certain matrices. The Specht theorem has been refined by Pearcy [83], and also generalized to certain classes of operators on Hilbert space [83, 85, 86, 19, 22]. Section 7 discusses the field of values, including a result of Arveson [2, 4], which shows that a generalization of the field of values, called the matrix range, classifies compact operators up to unitary similarity. Section 8 gives some results on unitary reducibility.

2. SPECIAL CASES

Before discussing canonical forms for general matrices, we use Schur's theorem to examine some special cases. As often happens with such problems, matrices with distinct eigenvalues are easier to handle. While the arguments are simpler for this case, they serve to introduce the basic ideas used to analyze the general situation. We also look at 2×2 matrices, where one can give a fully understood and satisfying canonical form.

We say an $m \times n$ matrix A is *upper triangular* if $a_{ij} = 0$ whenever $i > j$, that is, if all entries below the main diagonal are zero. If all entries above the main diagonal are zero, we say A is *lower triangular*. Results about triangular matrices can be stated and proved in either an upper or lower triangular version. As we mainly use the upper triangular form, we shall use the term "triangular" to mean "upper triangular."

Schur's theorem [97], which says that any square matrix is unitarily similar to a triangular matrix, is probably the best-known result concerning unitary similarity. Many texts [33, 39, 48, 61, 65] include this result.

THEOREM 2.1 (Schur [97]). *If A is in $\mathbf{C}(n)$, then there is an $n \times n$ unitary matrix U such that U^*AU is upper triangular.*

The typical proof of this uses induction on n. Let α_1 be an eigenvalue of A with corresponding unit eigenvector $\mathbf{u}_1$. Use $\mathbf{u}_1$ as the first vector of an orthonormal basis. Equivalently, let U_1 be a unitary matrix with $\mathbf{u}_1$ as the first column. Then the first column of $U_1^*AU_1$ has α_1 in the first position and zeros in the remaining $n-1$ positions. Now apply the induction hypothesis to the lower right hand block of size $n-1$ to complete the proof. One can, of course, obtain a similar result for lower triangular matrices. Note that the diagonal entries of the resulting triangular form must be the eigenvalues of A, and that the unitary matrix U may be chosen so that the eigenvalues appear in any desired order along the diagonal of A.

Murnaghan and Wintner [75] give a version of Schur's theorem for real matrices under orthogonal similarity; they show that if A is a real matrix, then there is a real, orthogonal matrix O such that O^TAO is block triangular, with 1×1 and 2×2 diagonal blocks. The real eigenvalues of A appear on the diagonal. The nonreal eigenvalues of a real matrix must occur in conjugate pairs; such a conjugate pair gives a 2×2 block on the diagonal.

The triangular matrix $T = U^*AU$ of Theorem 2.1 will be diagonal if and only if A is normal, as can be shown by noting that T is normal if and only if A is normal, and then comparing the entries in the matrix equation $T^*T = TT^*$. Alternatively, one can apply the following result of Schur.

THEOREM 2.2 (Schur [97]). *Let A be an $n \times n$ matrix with entries a_{ij} and eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_n$. Then $\sum_{i=1}^n |\alpha_i|^2 \leqslant \sum_{j=1}^n \sum_{i=1}^n |a_{ij}|^2$, and equality holds if and only if A is normal.*

This follows from Theorem 2.1, the fact that $\Sigma_{j=1}^{n}\Sigma_{i=1}^{n}|a_{ij}|^2$ is the trace of A^*A, and the fact that if $B = U^*AU$, where U is unitary, then B^*B and A^*A have the same trace.

If A is normal, then A is unitarily similar to a diagonal matrix (Toeplitz, [119]), and this will be the canonical form. In general, however, A is unitarily similar only to a triangular matrix, and while the diagonal entries must be the eigenvalues of A, the entries above the diagonal are not uniquely determined by A. The basic approach of Littlewood [59], Sergeĭchuk [99], and Benedetti and Cragnolini [9] is to specify further the entries above the diagonal in order to obtain a unique canonical form. While this analysis is somewhat complicated for a matrix with repeated eigenvalues, a simpler treatment, described by Röseler [92] and Littlewood [59], applies if A has distinct eigenvalues. Mitchell [69] also showed that nonderogatory matrices may be handled the same way; we discuss this in Section 4.

Let A be an $n \times n$ matrix with n distinct eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_n$, in some fixed order. Schur's theorem tells us that A is unitarily similar to a triangular matrix T with diagonal entries $t_{ii} = \alpha_i$. While T is not unique, the following theorem tells us that it is determined up to transformation by a diagonal unitary matrix. Theorem 2.3 has been observed by Röseler [92] and Littlewood [59], and is a special case of a more general result (Theorem 3.7) stated in Section 3.

THEOREM 2.3. *Let S and T be $n \times n$ triangular matrices with the same diagonal entries $s_{ii} = t_{ii} = \alpha_i$ and with $\alpha_i \neq \alpha_j$ for $i \neq j$. Suppose U is a unitary matrix such that $U^*TU = S$. Then U must be diagonal.*

Proof. Suppose U is unitary and $U^*TU = S$. Then $TU = US$. Now compare the entries on the two sides. From the $n,1$ entry, we have $\alpha_n u_{n1} = \alpha_1 u_{n1}$. Since $\alpha_n \neq \alpha_1$, we have $u_{n1} = 0$. Now move to the $n-1,1$ entry to obtain $\alpha_{n-1}u_{n-1,1} = \alpha_1 u_{n-1,1}$, so $u_{n-1,1} = 0$. In a similar fashion, one shows that $u_{i1} = 0$ for $i > 1$. Applying the same argument to entries in columns $2, 3, \ldots, n-1$, or using an induction proof, shows that U is triangular. But U is unitary, so U must actually be diagonal. ∎

Suppose A has distinct eigenvalues and $U^*AU = S$ and $V^*AV = T$, where U and V are unitary and S and T satisfy the hypothesis of Theorem 2.3. Then $S = U^*(VTV^*)U = (V^*U)^*T(V^*U)$ and V^*U is unitary, so V^*U must be diagonal. Thus, once we specify the order in which the eigenvalues appear on the diagonal of the triangular T, the only unitary similarities preserving this form are diagonal. Since the diagonal entries of a diagonal unitary matrix must be of the form $e^{i\theta}$ for some angle $0 \leqslant \theta < 2\pi$, we have

$$U = \operatorname{diag}[e^{i\theta_1}, e^{i\theta_2}, e^{i\theta_3}, \dots, e^{i\theta_n}].$$

For $B = U^*AU$ we have $b_{jk} = e^{i(\theta_k - \theta_j)}a_{jk}$ and so $|b_{jk}| = |a_{jk}|$. Thus, when A has distinct eigenvalues, the magnitudes of the entries in the triangular matrix $T = U^*AU$ are uniquely determined. In particular, note that the positions of the zero entries are uniquely determined. We can now choose the angles θ_j in order to make certain nonzero entries of T positive real numbers; this will lead to a uniquely determined T. There are several ways to do this. Since we can remove a scalar factor from U without affecting anything, we can assume the first diagonal entry of U is 1, or $\theta_1 = 0$. Put $D = \operatorname{diag}[1, e^{i\theta_2}, e^{i\theta_3}, \dots, e^{i\theta_n}]$; then the j,k entry of D^*TD is $e^{i(\theta_k - \theta_j)}t_{jk}$, and we are free to choose the $n-1$ angles $\theta_2, \theta_3, \dots, \theta_n$. For example, one might choose the θ_j's in order to transform the entries $t_{12}, t_{13}, \dots, t_{1n}$ in the first row of T into nonnegative reals. If $t_{12}, t_{13}, \dots, t_{1n}$ are all nonzero, then this will completely specify the D, and thus give a unique canonical form. However, if $t_{1j} = 0$ for some j, then θ_j will still be undetermined and we can thus choose θ_j to make some other entry positive. Another approach might be to change the $n-1$ superdiagonal entries into nonnegative real numbers, and again, if these are nonzero, then D is uniquely determined and we are done. However, if some of these entries are zero, then it will be possible to make other entries real by choosing the as yet unspecified entries of D. The basic idea is to create as many positive entries as possible.

To specify a unique canonical form, one can order the positions above the diagonal and then proceed to successively create real entries in that order. Littlewood chooses to order the entries by row: $t_{12}, t_{13}, \dots, t_{1n}, t_{23}, t_{24}, \dots, t_{2n}, t_{34}, t_{35}, \dots, t_{3n}, \dots, t_{n-2,n-1}, t_{n-2,n}, t_{n-1,n}$. In general one expects to make $n-1$ of the entries positive by specifying the $n-1$ angles $\theta_2, \theta_3, \dots, \theta_n$ of D. However, the zero-nonzero pattern of the entries above the diagonal of T will play an important role here. The following example shows that it is not always possible to create positive entries in $n-1$ positions, even if T has $n-1$ nonzero entries above the diagonal.

Example 2.1. Let $n = 4$, and suppose T has the form below, where an asterisk signifies a nonzero entry:

$$\begin{pmatrix} 1 & * & * & 0 \\ 0 & 2 & * & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

We can then choose θ_2 and θ_3 to make the 1,2 and 1,3 entries positive. However, the argument of the 2,3 entry will then be fixed, since it depends only on θ_2 and θ_3 and is not affected by θ_4. Alternatively, we could choose θ_2 and θ_3 to make the 1,2 and 2,3 entries positive, but then the argument of the 1,3 entry would be determined.

For the distinct eigenvalue case, the canonical form is then an upper triangular matrix in which we have fixed the ordering of the diagonal entries (the eigenvalues) and have then created as many positive entries above the diagonal as possible by choosing the entries of the diagonal D. While this does give a unique canonical matrix in each equivalence class, the final result has some drawbacks. The canonical form is specified by describing a procedure for reducing a given matrix to canonical form, but we don't know exactly what it will look like until we obtain the triangular matrix T and then apply the diagonal, unitary similarity. In general, we hope to have positive entries in the first row, or in the "first" $n-1$ nonzero positions in the ordered list, but this will not always be possible. One must examine the positions of the zero entries in order to know exactly which entries can be made positive. Another drawback is that the canonical form depends on an ordering imposed on the superdiagonal positions. Ordering them by row, by column, or along successive diagonals is possible, but one could use any well-defined ordering. As we shall see, these problems persist and become more complicated in the general case. All of the proposed canonical forms depend on inductive definitions, and on some sort of imposed ordering scheme.

One case with a completely satisfying solution is when $n = 2$.

THEOREM 2.4. *Let A be a 2×2 matrix with eigenvalues α_1 and α_2, which may or may not be distinct. Let*

$$r = \sqrt{\operatorname{trace}(A^*A) - |\alpha_1|^2 - |\alpha_2|^2}, \qquad \textit{where} \quad r \geqslant 0.$$

Then A is unitarily similar to a triangular matrix T with α_1 and α_2 on the main diagonal and r in the superdiagonal position. Furthermore, if A is unitarily similar to any triangular matrix S with α_1 and α_2 on the main diagonal, then $|s_{12}| = r$.

Proof. By Schur's theorem, A is unitarily similar to a triangular matrix S with α_1 and α_2 on the main diagonal; let $s_{12} = re^{i\theta}$ be the superdiagonal entry of S. Since A and S are unitarily similar, $\operatorname{trace}(A^*A) = \operatorname{trace}(S^*S) = |\alpha_1|^2 + |\alpha_2|^2 + r^2$. For $U = \operatorname{diag}[1, e^{-i\theta}]$, we have $U^*SU = T$. ∎

The number r in Theorem 2.4 can also be described in terms of the field of values, or numerical range. The numerical range of a 2×2 matrix is an ellipse with foci corresponding to the eigenvalues [119], where we identify an eigenvalue $\lambda = a + ib$ with the real point (a, b). The number r is the length of the minor axis of this ellipse [73].

Paulsen [81] gives a different, nontriangular canonical form for 2×2 matrices. Paulsen studies the problem of whether or not there exist continuous canonical forms for unitary similarity. He shows that the answer is no for $n \geqslant 3$, but gives a continuous canonical form for the case $n = 2$.

As a special case of the canonical form problem one might ask when a matrix is unitarily similar to its Jordan canonical form. Vesilic [123] has solved this problem for nilpotent matrices.

3. PRELIMINARY RESULTS

This section presents some of the results needed to describe and establish the canonical forms developed by Brenner [12], Littlewood [59], Mitchell [69], McRae [62], Radjavi [89], Sergeĭchuk [99], and Benedetti and Cragnolini [9]. These results are of independent interest; we also hope to clarify the exposition by presenting this preliminary material first and later focusing on the mechanics of the reduction to canonical form. Many of the theorems in this section are well known and can be found in standard texts such as [33, 39, 48, 61, 65, 121] as well as in [9, 12, 59, 62, 69, 89, 92, 99]. We include them here for completeness and for the convenience of the reader. There are also some less well-known theorems showing that a matrix can be transformed to a special triangular form [9, 59, 62, 89, 99], determined by the Weyr characteristic, via a unitary similarity. One can present and prove these results in a variety of ways, and we shall try to indicate these different points of view. Some proofs will be included, others merely outlined.

We frequently deal with matrices that are partitioned into submatrices. Thus, 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 manner breaks up the matrix A into blocks A_{ij}, where A_{ij} is the submatrix specified by the ith set of rows and the jth set of columns. Note that A_{ij} has size $n_i \times n_j$ and that the diagonal blocks are square and have sizes $n_i \times n_i$. If all blocks below the diagonal blocks are zero ($A_{ij} = 0$ whenever $i > j$), then we say A is *block triangular*.

DEFINITION 3.1. If A is an $n \times n$ block triangular matrix with t square blocks on the diagonal of sizes $n_1, n_2, \ldots, n_t$, where $t \geqslant 1$ and

$n_1 + n_2 + \cdots + n_t = n$, then we say A is $\mathbf{T}(n_1, \ldots, n_t)$. If A_i denotes the ith diagonal block, we shall also say A is $\mathbf{T}(A_1, A_2, \ldots, A_t)$ and write $A = \mathbf{T}(A_1, A_2, \ldots, A_t)$.

A square matrix in which both the blocks above and those below the diagonal blocks are zero is said to be *block diagonal*; thus a block diagonal matrix is just a direct sum of its diagonal blocks.

DEFINITION 3.2. If A is an $n \times n$ block diagonal matrix with t square blocks on the diagonal of sizes $n_1, n_2, \ldots, n_t$, where $t \geqslant 1$ and $n_1 + n_2 + \cdots + n_t = n$, then we say A is $\mathbf{D}(n_1, \ldots, n_t)$. If A_i denotes the ith diagonal block, we shall also say A is $\mathbf{D}(A_1, A_2, \ldots, A_t)$ and write $A = \mathbf{D}(A_1, A_2, \ldots, A_t)$.

Essentially the same argument used to prove Schur's theorem (Theorem 2.1) establishes the following result about sets of matrices.

THEOREM 3.1 (Mitchell [69]). *Suppose $\mathscr{A}$ is a nonempty set of $n \times n$ matrices that can be put simultaneously in triangular form by a similarity S. Then there is a unitary similarity U that simultaneously triangularizes the set $\mathscr{A}$.*

The fact that $\mathscr{A}$ can be simultaneously triangularized implies that the matrices in $\mathscr{A}$ have a common unit eigenvector $\mathbf{u}_1$; the rest of the proof is essentially the same as the proof of Schur's theorem. Mitchell's proof [69] uses the following result, attributed to Schmidt [86].

THEOREM 3.2. *If S is a nonsingular matrix, then there exists a triangular matrix T such that $U = ST$ is unitary.*

Thus, if $S^{-1}AS$ is triangular, so is $T^{-1}(S^{-1}AS)T$, which is equal to U^*AU. Since the matrix equation $U = ST$ may be viewed as the Gram-Schmidt orthogonalization process applied to the columns of S, we see that this proof is really based on the same idea as the other.

Sets of matrices that can be simultaneously triangularized by a similarity transformation were characterized by McCoy [60]; see also [24, 114, 70, 44].

Theorem 3.1 extends to block triangular matrices.

THEOREM 3.3 (Specht [109]). *A nonempty set of n × n matrices that can be put simultaneously in block triangular form $\mathbf{T}(n_1, n_2, \ldots, n_t)$ by a similarity transformation S can be put simultaneously in this form by a unitary similarity U.*

Again, one can prove this by induction. Since the matrices can be simultaneously put in block triangular form, they have a common invariant subspace of dimension n_1. Now use an orthonormal basis for this subspace as the first n_1 columns of U, and then apply the induction hypothesis. Alternatively, one can use Theorem 3.2 and note that since T is triangular, $T^{-1}(S^{-1}AS)T$ will still be $\mathbf{T}(n_1, n_2, \dots, n_t)$.

Sets of matrices that can be put simultaneously in block triangular form via a similarity or unitary similarity are studied in [7, 31, 55, 57, 100, 102, 109, 127].

Mitchell [69], Littlewood [59], Sergeĭchuk [99], and Benedetti and Cragnolini [9] all base their canonical forms on the Schur theorem. Although the diagonal entries of the triangular matrix found in Theorem 2.1 must be the eigenvalues of the matrix, and hence are uniquely determined, up to ordering, by the original matrix, the entries above the main diagonal are not uniquely determined. The basic idea in [59, 69, 99, 9] is to specify more precisely the entries above the main diagonal in the triangular form by applying further unitary similarities that still preserve the triangular form. In preparation for this analysis, we need information about similarity transformations that preserve certain block triangular forms. We first need the following well-known theorem [33, 61, 65].

THEOREM 3.4 (Sylvester [113]). *Let A be in $\mathbf{C}(n)$ and let B be in $\mathbf{C}(m)$. Then the matrix equation $AX - XB = 0$ has a nontrivial solution X in $\mathbf{C}(n, m)$ if and only if A and B have a common eigenvalue.*

Various proofs of Theorem 3.4 can be found in the literature. One approach [41, 65, 76, 117] is to use the pair A, B to define a linear operator $L(A, B)$ on the space $\mathbf{C}(n, m)$ as follows: $L(A, B)(X) = AX - XB$. The operator $L(A, B)$ can then be represented by the $mn \times mn$ matrix $I \otimes A - B^T \otimes I$. The eigenvalues of this matrix are the numbers $\alpha_i - \beta_j$, where α_i is an eigenvalue of A and β_j is an eigenvalue of B, and hence $L(A, B)$ has a nontrivial null space if and only if A and B have a common eigenvalue.

We now investigate similarity transformations that preserve block triangular structure. The statements and proofs of Theorems 3.5 through 3.7 either appear in or are suggested by similar results in [9, 12, 59, 62, 69, 92, 99].

Theorems 3.5 and 3.6 are not explicitly stated in [9, 59, 69, 92, 99], but these papers certainly contain all of the ingredients for their proofs.

THEOREM 3.5. *Suppose A is $\mathbf{T}(n_1, n_2, \dots, n_t)$ with diagonal blocks A_i, where A_i and A_j have no common eigenvalues if $i \neq j$. Suppose S is nonsingular and $B = S^{-1}AS$. Then B is also $\mathbf{T}(n_1, n_2, \dots, n_t)$, with diagonal* *blocks B_i, where, for each i, the blocks A_i and B_i have the same eigenvalues, if and only if S is $\mathbf{T}(n_1, n_2, \ldots, n_t)$.*

Proof. Suppose B is $\mathbf{T}(n_1, n_2, \ldots, n_t)$, with diagonal blocks B_i, where, for each i, the blocks A_i and B_i have the same eigenvalues. The numbers n_i specify partitions of A and B into submatrices A_{ij} and B_{ij}; partition S in a similar manner. Now compare the two sides of the matrix equation $AS = SB$, using block matrix multiplication. Looking at the $t,1$ block on each side, we have $A_t S_{t1} = S_{t1} B_1$. Since A_t and B_1 have no common eigenvalue, Theorem 3.4 tells us that $S_{t1} = 0$. Now look at the $t-1,1$ block to find that $A_{t-1} S_{t-1,1} = S_{t-1,1} B_1$, so that $S_{t-1,1}$ is also zero. Continue in this fashion to show that $S_{i1} = 0$ for all $i > 1$. Either use induction, or repeat the same argument on subsequent columns of blocks to show that $S_{ij} = 0$ whenever $i > j$.

Conversely, suppose S is $\mathbf{T}(n_1, n_2, \ldots, n_t)$. Then $S^{-1}AS = B$ must also be $\mathbf{T}(n_1, n_2, \ldots, n_t)$. The block triangular form of S yields $S_{ii}^{-1} A_i S_{ii} = B_i$, so that A_i and B_i must be similar and have the same eigenvalues. ■

Our main interest will be the special case in which each A_i has a single eigenvalue α_i of multiplicity n_i and the matrix S is unitary.

Let A have t distinct eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_t$, where α_i has multiplicity n_i. Let J be the Jordan canonical form of A, and write $J = \mathbf{D}(J_1, J_2, \ldots, J_t)$, where the $n_i \times n_i$ block J_i is the direct sum of all of the Jordan blocks belonging to the eigenvalue α_i. Using Schur's theorem 2.1, we assume A is already in triangular form with α_1 in the first n_1 diagonal positions, α_2 in the next n_2 diagonal positions, and so on. Thus, we have $A = \mathbf{T}(A_1, A_2, \ldots, A_t)$, where each $n_i \times n_i$ diagonal block A_i is triangular with α_i in each diagonal entry. We now want to apply a further unitary similarity to transform the diagonal blocks A_i into a special form determined by the Jordan canonical form of A. We first note that J_i must be the Jordan canonical form of A_i.

THEOREM 3.6. *Let A be an $n \times n$ matrix with t distinct eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_t$ where α_i has multiplicity n_i. Let $J = \mathbf{D}(J_1, J_2, \ldots, J_t)$ be the Jordan canonical form of A, where the $n_i \times n_i$ block J_i is the direct sum of all of the Jordan blocks belonging to the eigenvalue α_i. Suppose that $A = \mathbf{T}(A_1, A_2, \ldots, A_t)$, where A_i is of size n_i, and α_i is the only eigenvalue of A_i. Then J_i is the Jordan canonical form of A_i, and A is similar to $\mathbf{D}(A_1, A_2, \ldots, A_t)$.*

Proof. Let S be a nonsingular matrix such that $S^{-1}AS = J$. The result then follows from Theorem 3.5 and by noting that $S_i^{-1} A_i S_i = J_i$. ■

A unitary matrix that is $\mathbf{T}(n_1, n_2, \dots, n_t)$ must actually be $\mathbf{D}(n_1, n_2, \dots, n_t)$, with unitary diagonal blocks. A similar result holds for normal matrices. Thus, if $N = \mathbf{T}(n_1, n_2, \dots, n_t)$ is normal, then comparing the traces of the diagonal blocks in the equation $N^*N = NN^*$, and using the fact that the trace of $N_{ij}^* N_{ij}$ is nonnegative and is zero only if $N_{ij} = 0$, shows that N must be $\mathbf{D}(n_1, n_2, \dots, n_t)$ with normal diagonal blocks. This is also a special case of the following theorem of Parker [80]: If N is a partitioned normal matrix and the characteristic polynomial of N is the product of the characteristic polynomials of the diagonal blocks, $N_1, N_2, \dots, N_t$, of N, then N must be block diagonal $\mathbf{D}(N_1, N_2, \dots, N_t)$.

Theorem 3.5, together with the fact that a unitary matrix that is $\mathbf{T}(n_1, n_2, \dots, n_t)$ must actually be $\mathbf{D}(n_1, n_2, \dots, n_t)$ with unitary diagonal blocks, yields out next result, which appears in [9, 59, 92, 99].

THEOREM 3.7. *Let A be an $n \times n$ matrix with t distinct eigenvalues $\alpha_1, \alpha_2, \dots, \alpha_t$, where α_i has multiplicity n_i. Suppose that $A = \mathbf{T}(A_1, A_2, \dots, A_t)$ where A_i is triangular of size n_i and has α_i in each diagonal entry. If U is unitary and U^*AU also has this form, then U must be $\mathbf{D}(n_1, n_2, \dots, n_t)$.*

REMARK. If A has n distinct eigenvalues, the U of Theorem 3.7 will be diagonal with diagonal entries of the form $e^{i\theta}$. Thus, Theorem 2.3 is a special case of Theorem 3.7.

Let U_i denote the ith diagonal block of the U in Theorem 3.7. The next step is to choose U_i so that $U_i^* A_i U_i$ has a special form determined by J_i. To avoid double subscripts, consider a $k \times k$ matrix B with a single eigenvalue β, and let J be the Jordan form of B. Then J is a direct sum of Jordan blocks of sizes $k_1, k_2, \dots, k_r$, where each block has β's in the diagonal positions, 1's along the superdiagonal, and 0's elsewhere. We order the k_i's so that $k_1 \geqslant k_2 \geqslant \cdots \geqslant k_r$. The numbers $k_1, k_2, \dots, k_r$ form a partition of k, called the *Segre characteristic* belonging to β [33, 61, 121]. The Jordan form of a matrix is completely specified by its eigenvalues and their corresponding Segre characteristics. An alternative description can be given by the Weyr characteristic [61, 121, 131]. The matrix $N = J - \beta I$ is nilpotent, and if p is the index of N (the smallest power of N that is zero) then $p = k_1$. Let $g_1, g_1 + g_2, g_1 + g_2 + g_3, \dots, g_1 + g_2 + g_3 + \cdots + g_p$ be the nullities of $N, N^2, N^3, \dots, N^p$. That is, for $i = 1, 2, \dots, p$, the dimension of the null space of N^i is $g_1 + g_2 + \cdots + g_i$. The list of numbers $g_1, g_2, \dots, g_p$ is called the *Weyr characteristic* corresponding to β. Since the nullity of N^i exceeds that of N^{i-1} by precisely the number of Jordan blocks of size at least i, the number g_i is the number of Jordan blocks of size greater than or equal to i. Thus the numbers of the Weyr characteristic form the conjugate partition of the Segre characteristic [121, pp. 79–80], and the Jordan form of a matrix is also completely specified by its eigenvalues and their corresponding Weyr characteristics. Note also that $g_1 \geqslant g_2 \geqslant \cdots \geqslant g_p$.

Viewing B as a linear transformation of $\mathbf{C}^k$, consider the null spaces, or kernels, of the maps $(B - \beta I)^i$. For convenience of notation, assume $\beta = 0$; thus B is nilpotent of index $p = k_1$. Since the kernel of B^i contains that of B^{i-1}, we can build a special orthonormal basis for $\mathbf{C}^k$ as follows. Start with an orthonormal set of g_1 vectors that are a basis for the kernel of B, then extend this to an orthonormal set of $g_1 + g_2$ vectors that form a basis for the kernel of B^2, and so on, until we have k orthonormal vectors such that the first $g_1 + g_2 + \cdots + g_i$ vectors in the set give an orthonormal basis for the null space of B^i. Notice that if $\mathbf{v}$ is one of the vectors added at stage i, then $B\mathbf{v}$ is in the kernel of B^{i-1} and hence is a linear combination of the first $g_1 + g_2 + \cdots + g_{i-1}$ vectors in the basis. Using these basis vectors as the columns of a $k \times k$ unitary matrix U, we see that $U^*BU = \mathbf{T}(Z_1, Z_2, \ldots Z_p)$, where the $g_i \times g_i$ block Z_i is zero. Furthermore, we can say something about the blocks in the superdiagonal positions.

Definition 3.3. Let B be a block triangular matrix, $\mathbf{T}(B_1, B_2, \ldots, B_p)$, where B_i is $g_i \times g_i$. Assume β is the only eigenvalue of B. Then we shall say B is in *Weyr form* if:

(1) $g_1 \geqslant g_2 \geqslant \cdots \geqslant g_p$.
(2) For each i, the diagonal block B_i is the scalar matrix βI_{g_i}.
(3) Each superdiagonal block $B_{i\,i+1}$ has rank g_{i+1}, and thus, since $B_{i,i+1}$ has size $g_i \times g_{i+1}$, each superdiagonal block has linearly independent columns.

Theorem 3.8. *Suppose B is $\mathbf{T}(g_1, g_2, \ldots, g_p)$ and has a single eigenvalue β. If B is in Weyr form, then $g_1, g_2, \ldots, g_p$ must be the Weyr characteristic of B.*

Proof. Since $N = B - \beta I$ is also in Weyr form and has the same Weyr characteristic as B, it suffices to prove the result for the nilpotent matrix N. Since each of the superdiagonal blocks $N_{12}, N_{23}, \ldots, N_{p-1,p}$ has linearly independent columns, and the first g_1 columns of N are zero, N has rank $g_2 + g_3 + \cdots + g_p$ and nullity g_1. The matrix N^2 has zero blocks in the diagonal and superdiagonal positions, while the next diagonal line contains the products $N_{12}N_{23}, N_{23}N_{34}, \ldots, N_{p-2,p-1}N_{p-1,p}$. But the product $N_{i,i+1}N_{i+1,i+2}$ also has linearly independent columns, and hence has rank g_{i+2}. So N^2 has rank $g_3 + g_4 + \cdots + g_p$ and nullity $g_1 + g_2$. In general,

N^r has blocks of zeros along the diagonal lines up to the rth superdiagonal line, which contains products of r consecutive $N_{i,i+1}$'s. The linear independence of the columns in each $N_{i,i+1}$ then guarantees that N^r has rank $g_{r+1} + g_{r+2} + \cdots + g_p$ and nullity $g_1 + g_2 + \cdots + g_r$. Hence, N and B have Weyr characteristic $g_1, g_2, \ldots, g_p$. ■

Although Littlewood, Sergeĭchuk, and Benedetti and Cragnolini do not mention the Weyr characteristic, they all describe and use some version of the special triangular matrices defined in Definition 3.3. Benedetti and Cragnolini do discuss the connection between these special triangular matrices and the Jordan form, and point out that the block sizes $g_1, g_2, \ldots, g_p$ form the conjugate partition of the Segre characteristic. Radjavi also defines these special triangular matrices, but does not mention the superdiagonal blocks. McRae describes an inductive procedure that leads to matrices in Weyr form—although he does not mention the superdiagonal blocks, one can show that his construction does lead to a matrix in Weyr form. Similarly, although none of these authors associates the block sizes with the Weyr characteristic, Littlewood, Sergeîchuk, and Benedetti and Cragnolini all have versions of Theorems 3.9 and 3.12, or their natural extensions to matrices with more than one eigenvalue.

THEOREM 3.9. *Let B be a $k \times k$ matrix with a single eigenvalue β. Then B has Weyr characteristic $g_1, g_2, \ldots, g_p$ if and only if B is unitarily similar to a matrix C that is* $\mathrm{T}(g_1, g_2, \ldots, g_p)$ *and in Weyr form.*

Proof. Suppose B has Weyr characteristic $g_1, g_2, \ldots, g_p$. Then, since $B - \beta I$ is nilpotent, the discussion preceding Definition 3.3 shows that B is unitarily similar to a matrix C that is $\mathrm{T}(g_1, g_2, \ldots, g_p)$ and in which the diagonal blocks are the scalar matrices βI_{g_i}. The nilpotent matrix $N = C - \beta I$ also has Weyr characteristic $g_1, g_2, \ldots, g_p$. Recalling that g_1 is the nullity of N, we see that N has rank $k - g_1 = g_2 + g_3 + \cdots + g_p$. Since the first g_1 columns of N are all zero, the remaining columns must be linearly independent. This tells us that N_{12} has linearly independent columns, and hence has rank g_2. But N^2 has blocks of zeros in both the diagonal and superdiagonal blocks, while the next line of diagonal blocks has the products $N_{i,i+1}N_{i+1,i+2}$. Since the rank of N^2 is $g_3 + g_4 + \cdots + g_p$, and since the first $g_1 + g_2$ columns of N^2 are zero, the remaining columns must be linearly independent. In particular, $N_{12}N_{23}$ must have linearly independent columns, and so N_{23} must have linearly independent columns and thus have rank g_3. In general, examination of N^r shows that the product $N_{12}N_{23}N_{34} \cdots N_{r,r+1}$ must have linearly independent columns, so an induction argument shows that $N_{r,r+1}$ has linearly independent columns and thus rank g_{r+1}. Therefore, C has Weyr form.

Conversely, if B is unitarily similar to a matrix C that is $\mathbf{T}(g_1, g_2, \ldots, g_p)$ and in Weyr form, the Theorem 3.8 tells us that $g_1, g_2, \ldots, g_p$ is the Weyr characteristic of B. ■

REMARK. Theorem 3.9 also holds if one replaces "unitarily similar" with "similar," and one can use this to obtain another canonical form for similarity [131, 132]. Like the Jordan form, it is a direct sum of diagonal blocks, where the blocks correspond to the distinct eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_t$ and the size of the ith block is the multiplicity of α_i. Each such block is a Weyr form block with diagonal scalar blocks of sizes given by the Weyr characteristic of α_i, superdiagonal blocks consisting of an identity matrix of size g_i followed by $g_{i+1} - g_i$ rows of zeros, and zero blocks elsewhere.

Benedetti and Cragnolini's versions of Definition 3.3 and Theorem 3.9 are the same as presented here—the fact that the block sizes must correspond to the Weyr characteristic guarantees that their special triangular type is the same as we have used here. Sergeĭchuk's version is slightly different because Sergeĭchuk works with lower triangular matrices. Thus, Sergeĭchuk states a rank condition on the blocks immediately below the main diagonal, and for this formulation the rows of these subdiagonal blocks are linearly independent. Littlewood's version is based on the ranges (or column spaces) of the powers of B, rather than the kernels. This results in having block sizes that increase (rather than decrease) in size, and thus the superdiagonal blocks have linearly independent rows rather than columns. For completeness, we outline Littlewood's approach.

Again, for convenience, consider a $k \times k$ matrix B with a single eigenvalue β and let N be the nilpotent matrix $B - \beta I$. Let p be the index of N, and consider the powers N^i, where i runs from 0 to $p-1$. Since the range of N^i contains that of N^{i+1}, we can construct an orthonormal basis for $\mathbf{C}^k$ by starting with an orthonormal basis for the range of N^{p-1}, extending it to an orthonormal basis for the range of N^{p-2}, then extending to an orthonormal basis for the range of N^{p-3}, and so on, until we have an orthonormal basis for the range of N; we then extend that to an orthonormal basis for $\mathbf{C}^k$. If m_i is the number of vectors added at stage i of this process, then the number $m_1 + m_2 + \cdots + m_i = r_i$ is the rank of N^{p-i}. The nullity of N^i is then $k - (m_1 + m_2 + \cdots + m_{p-i}) = m_{p-i+1} + m_{p-i+2} + \cdots + m_p$, since $r_p = k$. But this means that the numbers m_i are the numbers of the Weyr characteristic written in reverse order, from smallest to largest. If U is the $k \times k$ unitary matrix whose columns are the vectors of the orthonormal basis constructed above, then $C = U^*BU$ will be $\mathbf{T}(m_1, m_2, \ldots, m_p)$, with the ith diagonal block C_i being the scalar matrix βI and with the $m_i \times m_{i+1}$ superdiagonal block $C_{i\,i+1}$ having linearly independent rows and rank m_i. This approach leads to the following result, which is similar to Theorem 3.9 and is proved the same way.

THEOREM 3.10. *Let B be a $k \times k$ matrix with a single eigenvalue β. Then B has Weyr characteristic $g_1, g_2, \dots, g_p$ if and only if B is unitarily similar to a matrix C that is* $\mathbf{T}(g_p, g_{p-1}, \dots, g_3, g_2, g_1)$ *and of the following form:*

(1) *The ith diagonal block, C_i, is the scalar matrix βI.*

(2) *The $g_{p-i+1} \times g_{p-i}$ superdiagonal block $C_{i,i+1}$ has linearly independent rows and* rank g_{p-i+1}.

Theorems 2.1 and 3.9 show that if A is an $n \times n$ matrix with t distinct eigenvalues $\alpha_1, \alpha_2, \dots, \alpha_t$, where α_i has multiplicity n_i, then A is unitarily similar to a triangular matrix of form $\mathbf{T}(A_1, A_2, \dots, A_t)$, where the $n_i \times n_i$ block A_i corresponds to α_i and is in Weyr form. We now see what type of similarity transformation will preserve Weyr form. Littlewood, Sergeĭchuk, and Benedetti and Cragnolini all give the following theorem for the case of unitary similarity, but the same argument establishes the result for similarity.

THEOREM 3.11. *Let B be a $k \times k$ matrix with a single eigenvalue β and Weyr characteristic $g_1, g_2, \dots, g_p$. Suppose B is in Weyr form. Then $C = S^{-1}BS$ is also in Weyr form if and only if S is* $\mathbf{T}(g_1, g_2, \dots, g_p)$.

Proof. Assume that $C = S^{-1}BS$ is in Weyr form. Since C is similar to B, we know from Theorem 3.8 that C must have the same block sizes as B. Without loss of generality, we may assume $\beta = 0$, so that B is nilpotent with diagonal blocks of zeros. Partition S into blocks of the same size, and consider the block entries in both sides of $SC = BS$. On the left hand side, the first column of blocks is all zero. Thus, examining the $p-1,1$ block on the right hand side, we see that $B_{p-1\,p}S_{p1} = 0$. Since $B_{p-1\,p}$ has linearly independent columns, $S_{p1} = 0$. Now go to the $p-2,1$ block to see that $B_{p-2,p-1}S_{p-1,1} = 0$, so the linear independence of the columns of $B_{p-2,p-1}$ forces $S_{p-1,1} = 0$. Moving up the first column in this way shows that $S_{i1} = 0$ for $i = 2, 3, \dots, p$. The same argument applied to the other columns (or, more formally, an induction proof) then shows that S is $\mathbf{T}(g_1, g_2, \dots, g_p)$.

Conversely, suppose S is $\mathbf{T}(g_1, g_2, \dots, g_p)$. Let $C = S^{-1}BS$. Then C is also $\mathbf{T}(g_1, g_2, \dots, g_p)$ with diagonal blocks $C_i = S_i^{-1}B_iS_i$. Since B_i is a scalar matrix, $C_i = B_i$. Computing the superdiagonal blocks, we have $C_{i,i+1} = S_i^{-1}B_{i,i+1}S_{i+1}$ for $i = 1, \dots, p-1$, so $C_{i,i+1}$ and $B_{i,i+1}$ have the same rank, and thus the superdiagonal blocks of C have linearly independent columns. ■

The fact that a unitary matrix is $\mathbf{T}(g_1, g_2, \ldots, g_p)$ if and only if it is $\mathbf{D}(g_1, g_2, \ldots, g_p)$ now yields the following "unitary" version of Theorem 3.11.

THEOREM 3.12. *Let B be a $k \times k$ matrix with a single eigenvalue β and Weyr characteristic $g_1, g_2, \ldots, g_p$. Suppose B is in Weyr form. Then a unitary matrix U has the property that $C = U^{-1}BU$ is also in Weyr form if and only if U is $\mathbf{D}(g_1, g_2, \ldots, g_p)$.*

Again, although not stated in terms of the Weyr characteristic, the result of Theorem 3.12 appears in [9, 59, 99].

REMARK. When B is nonderogatory, the Segre characteristic of B is just the single number k, and the Weyr characteristic is $1, 1, \ldots, 1$. The diagonal blocks of a Weyr form for B are then 1×1 blocks, and the U of Theorem 3.12 must be a diagonal matrix. Combining this with Schur's triangularization theorem and Theorem 3.7 thus establishes that Theorem 2.3 also holds for nonderogatory matrices. Mitchell gives a more direct proof in [69], and thus shows that a triangular canonical form can be found for nonderogatory matrices by using the same procedure described for matrices with distinct eigenvalues.

EXAMPLE 3.1. For any triple (x, y, z) of positive numbers, let $M(x, y, z)$ be the matrix

$$\begin{pmatrix} 0 & x & y \\ 0 & 0 & z \\ 0 & 0 & 0 \end{pmatrix}.$$

Then $M(x, y, z)$ is nonderogatory, so only a diagonal unitary similarity can preserve this form. Hence, since x, y, and z are positive, two different matrices of this form cannot be unitarily similar. Note, however, that all matrices of this form are similar.

Once B is in Weyr form, we can apply an additional unitary similarity to further specialize the superdiagonal block, as done in [9].

DEFINITION 3.4. If $m \geqslant n$ and A is an $m \times n$ matrix of rank r, we say A is in *row form* if the first $m - r$ rows of A are zero.

Observe that if A is in row form, so is AQ for any nonsingular matrix Q. The following theorem is based on a similar result of Benedetti and Cragnolini for unitary matrices.

THEOREM 3.13. *Suppose A is an $m \times n$ matrix of rank r in row form. Let P and Q be nonsingular matrices of sizes $m \times m$ and $n \times n$, respectively. Then PAQ is in row form if and only if P^T is* $\mathbf{T}(m-r,r)$, *so that P is lower block triangular with diagonal blocks of sizes $m-r$ and r.*

Proof. If P^T is $\mathbf{T}(m-r,r)$, then computing PA using block multiplication shows that PA is in row form, and hence PAQ must be in row form.

Conversely, suppose that PAQ is in row form. Since Q is nonsingular, the first $m-r$ rows of PA must be zero, so PA must be in row form. Partition P conformally with A, into diagonal blocks P_1 and P_2 of sizes $m-r$ and r, respectively, and off-diagonal blocks P_{12} and P_{21} of sizes $(m-r)\times r$ and $r\times(m-r)$. Letting A_1 denote the $r \times n$ matrix in the last r rows of A, the first $m-r$ rows of PA are given by $P_{12}A_1$. Since PA is in row form, $P_{12}A_1 = 0$, so $P_{12} = 0$ because A_1 has linearly independent rows. Hence P is lower block triangular with diagonal blocks of sizes $m-r$ and r. ■

We shall want the unitary version of Theorem 3.13, given in [9].

THEOREM 3.14. *Let A be an $m \times n$ matrix of rank r with $m \geqslant n$. Then there exists a unitary matrix P such that PA is in row form. Furthermore, if A itself is already in row form, then P and Q are unitary matrices such that PAQ is also in row form if and only if P is* $\mathbf{D}(m-r,r)$.

Proof. For the first part, find a set of $m-r$ orthonormal vectors that are orthogonal to the columns of A, and use these $m-r$ vectors as the first $m-r$ rows of a unitary matrix P. For the second statement, apply Theorem 3.13 and use the fact that a block triangular unitary matrix must actually be block diagonal. ■

Similarly, we have column versions of Definition 3.4 and Theorems 3.13 and 3.14.

DEFINITION 3.5. If $n \geqslant m$ and A is an $m \times n$ matrix of rank r, we say A is in *column form* if the first $n-r$ columns of A are zero.

THEOREM 3.15. *Suppose A is an $m \times n$ matrix of rank r in column form. Let P and Q be nonsingular matrices of sizes $m \times m$ and $n \times n$, respectively. Then PAQ is in column form if and only if Q is* $\mathbf{T}(n-r,r)$.

THEOREM 3.16. *Let A be an $m \times n$ matrix of rank r with $n \geqslant m$. Then there exists a unitary matrix Q such that AQ is in column form. Furthermore, if A itself is already in column form, then P and Q are unitary matrices such that PAQ is also in row form if and only if Q is $\mathbf{D}(n-r, r)$.*

Although we use different notation and terminology, Definitions 3.6 and 3.7, as well as Theorems 3.17 and 3.18, are based on [9].

DEFINITION 3.6. We shall say a block triangular matrix $B = \mathbf{T}(B_1, B_2, \dots, B_p)$ is in *special Weyr form* if B is in Weyr form and, in addition, each superdiagonal block $B_{i,i+1}$ is in row form.

DEFINITION 3.7. Suppose $g_1 \geqslant g_2 \geqslant \cdots \geqslant g_p$ are positive integers. Then $(g_1, g_2, \dots, g_p)^*$ will denote the list of positive integers obtained from $(g_1, g_2, \dots, g_p)$ by replacing the number g_i with the pair of numbers $g_i - g_{i+1}, g_{i+1}$, whenever $g_i > g_{i+1}$.

For example, if we start with the list $(7,5,5,3,1,1)$, then 7 is replaced by 2,5, the second 5 is replaced by 2,3, and the 3 is replaced by 2,1 to yield

$$(7,5,5,3,1,1)^* = (2,5,5,2,3,2,1,1,1).$$

THEOREM 3.17. *Suppose $B = \mathbf{T}(g_1, g_2, \dots, g_p)$ is in special Weyr form and U is unitary. Then U^*BU is also in special Weyr form if and only if U is $\mathbf{D}((g_1, g_2, \dots, g_p)^*)$.*

Proof. Suppose U^*BU is in special Weyr form. Then U^*BU is in Weyr form, so Theorem 3.12 tells us U must be $\mathbf{D}(g_1, g_2, \dots, g_p)$. Let $C = U^*BU$; then $C_{i,i+1} = U_i^* B_{i,i+1} U_{i+1}$. Since $B_{i,i+1}$ and $C_{i,i+1}$ are both in row form, Theorem 3.14 tells us that whenever $g_i > g_{i+1}$, the matrix U_i^* must be $\mathbf{D}(g_i - g_{i+1}, g_{i+1})$. So U_i is $\mathbf{D}(g_i - g_{i+1}, g_{i+1})$ whenever $g_i > g_{i+1}$, and hence U has the desired form.

Conversely, if U has the desired block diagonal form, U is also $\mathbf{D}(g_1, g_2, \dots, g_p)$, so U^*BU is in Weyr form. Since U_i is $\mathbf{D}(g_i - g_{i+1}, g_{i+1})$ whenever $g_i > g_{i+1}$, Theorem 3.14 tells us the superdiagonal blocks of U^*BU will still be in row form, and hence U^*BU is in special Weyr form. ■

THEOREM 3.18. *Let B be a $k \times k$ matrix with a single eigenvalue β and Weyr characteristic $g_1, g_2, \dots, g_p$. Then B is unitarily similar to a matrix C that is $\mathbf{T}(g_1, g_2, \dots, g_p)$ and in special Weyr form. Furthermore, U is a* *unitary matrix such that U^*CU is also in special Weyr form if and only if U is* $\mathbf{D}(g_1, g_2, \ldots, g_p)^*$.

Proof. By Theorem 3.9, B is unitarily similar to a block triangular matrix $F = \mathbf{T}(g_1, g_2, \ldots, g_p)$ that is in Weyr form, and by Theorem 3.12, U^*FU is also in Weyr form whenever U is $\mathbf{D}(g_1, g_2, \ldots, g_p)$. Theorem 3.14 tells us that for each superdiagonal block $F_{i,i+1}$ of F, we can find a $g_i \times g_i$ unitary matrix U_i such that $U_i F_{i,i+1}$ is in row form. Let $U = \mathbf{D}(U_1, U_2, \ldots, U_p)$. Then the superdiagonal blocks of U^*FU are $U_i D_{i,i+1} U_{i+1}$ and hence are in row form, so $C = U^*FU$ is in special Weyr form and is unitarily similar to B.
The second part of the theorem follows from Theorem 3.17. ∎

For ease of notation, we defined Weyr form and special Weyr form for matrices with only one eigenvalue, but the definitions extend naturally to the general case.

DEFINITION 3.8. Let A be an $n \times n$ matrix with t distinct eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_t$, where α_i has multiplicity n_i. Suppose $A = \mathbf{T}(A_1, A_2, \ldots, A_t)$, where A_i has size n_i, and α_i is the only eigenvalue of A_i. We say A is in *Weyr form* if each diagonal block A_i is in Weyr form (Definition 3.3), and we say A is in *special Weyr form* if each diagonal block A_i is in special Weyr form (Definition 3.6).

Theorems 2.1 and 3.18 show that any matrix is unitarily similar to a matrix in special Weyr form.

THEOREM 3.19. *If A is an $n \times n$ matrix, then there exists a unitary matrix U such that U^*AU is in special Weyr form.*

We now have the main results needed to determine the diagonal blocks in the triangular canonical form. These diagonal blocks are scalar matrices of sizes determined by the Weyr characteristic, and hence by the similarity class of the matrix. The next step is to treat the blocks above the diagonal in order to establish a unique canonical form. The main tool used here is the singular value decomposition [14, 25], which is closely related to the polar decomposition [6, 135, 136]. These are both well known, especially for the nonsingular case, and are treated in many texts [33, 48, 61]; proofs are also included in [9, 12, 59, 62, 92, 99]. One can first establish the polar decomposition and then use it to derive the singular value decomposition, or, vice versa, derive the polar decomposition from the singular value decomposition. We start with the singular value decomposition.

Let A be in $\mathbf{C}(m,n)$, and let r be the rank of A. Then A^*A and AA^* are positive semidefinite Hermitian matrices of rank r having the same eigenvalues; furthermore the r nonzero eigenvalues have the same multiplicites. These nonzero eigenvalues must be positive; denote them as $\sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2$, where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$. The positive numbers $\sigma_1, \sigma_2, \ldots, \sigma_r$ are called the singular values of A. Thus, the singular values of A are the positive square roots of the nonzero eigenvalues of A^*A or, equivalently, of AA^*. We first deal with the case where $r = m = n$, that is, where A is nonsingular.

THEOREM 3.20 (Singular value decomposition: Browne [14]; Eckart and Young [25]). *Let A be a nonsingular $n \times n$ matrix. Then there exist unitary matrices U and V, and a diagonal matrix D with positive diagonal entries, such that $A = UDV$. Furthermore, in any such factorization, the diagonal entries of D must be the singular values of A.*

Proof. Let $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ be the eigenvalues of A^*A, where $\sigma_i > 0$; let D be the diagonal matrix with σ_i in the ith diagonal entry. Let $\mathbf{x}_i$ be an eigenvector corresponding to σ_i; since A^*A is Hermitian, we can assume $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are orthonormal. Let X be the unitary matrix with columns $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Then $A^*AX = XD^2$. Since A^*A is nonsingular, D is nonsingular and $A = (A^*)^{-1}XD^2X^* = [(A^*)^{-1}XD]DX^*$. Let $V = X^*$ and let $U = (A^*)^{-1}XD$. Then V is certainly unitary and $UU^* = [(A^*)^{-1}XD][DX^*A^{-1}] = (A^*)^{-1}[XD^2]X^*A^{-1} = (A^*)^{-1}[A^*AXX^*]A^{-1} = I$, so U is unitary. Thus, we have $A = UDV$, and D has positive diagonal entries σ_i.

If $A = UDV$ is such a factorization, then $A^*A = V^*D^2V$, so the diagonal entries of D^2 must be the eigenvalues of A^*A. Since the diagonal entries of D are positive, they must then be the singular values of A. ∎

We need a singular value decomposition for singular and rectangular matrices, and we need to know how much freedom one has in choosing the U and the V [9, 12, 54, 99].

THEOREM 3.21. *Let A be an $m \times n$ matrix of rank r with singular values $\sigma_1 > \sigma_2 > \cdots > \sigma_t$, where σ_i has multiplicity m_i. Let D be the $m \times n$ diagonal matrix that is the direct sum*

$$\sigma_1 I_{m_1} \oplus \sigma_2 I_{m_2} \oplus \cdots \oplus \sigma_t I_{m_t} \oplus 0_{(m-r)\times(n-r)}.$$

Then there exist unitary matrices U and V such that $A = UDV$.

Furthermore, U and V are unitary matrices such that $UDV = D$ if any only if there is an $r \times r$ block diagonal unitary matrix $S = \mathbf{D}(m_1, m_2, \ldots, m_t)$, *and unitary matrices U_2 and V_2, of sizes $m-r$ and $n-r$, respectively, such that $U=S\oplus U_2$ and $V=S^*\oplus V_2$.*

Proof. Let X be an $m\times m$ unitary matrix in which the last $m-r$ rows are orthogonal to the columns of A, and let Y be an $n\times n$ unitary matrix such that the last $n-r$ columns of Y are orthogonal to the rows of A. Then XAY has an $r\times r$ nonsingular matrix A_1 in the upper left hand corner, with the remaining $m-r$ rows and $n-r$ columns being zero. Now apply Theorem 3.20 to A_1 to find $r\times r$ unitary matrices U_1 and V_1 such that $A_1=U_1D_1V_1$, where D_1 is the $r\times r$ diagonal matrix $\sigma_1 I_{m_1}\oplus\sigma_2 I_{m_2}\oplus\cdots\oplus\sigma_t I_{m_t}$. Let $U=X^*(U_1\oplus I_{m-r})$ and let $V=(V_1\oplus I_{n-r})Y^*$. Then U and V are unitary and

$$A=X^*(XAY)Y^*=X^*\big(U_1D_1V_1\oplus 0_{(m-r)\times(n-r)}\big)Y^*=UDV.$$

Now suppose U and V are unitary matrices such that $UDV=D$. Partition U and V conformally with the blocks of D (i.e., according to the m_i's), and compare corresponding blocks on the two sides of the equation $UD=DV^*$. Since distinct diagonal blocks of D have different scalars, U must be $\mathbf{D}(m_1,m_2,\dots,m_t,m-r)$ and V must be $\mathbf{D}(m_1,m_2,\dots,m_t,n-r)$. Then since the first t diagonal blocks of D are nonzero scalar blocks, we must have $U_i=V_i^*$ for $i=1,2,\dots,t$. Putting $S=\mathbf{D}(U_1,U_2,\dots,U_t)$, we see that U and V must have the desired form. Conversely, it is clear that if U and V have this form, then $UDV=D$. ■

Definition 3.9. The diagonal matrix D of Theorem 3.21 is called the *singular value form* of A.

We now use the singular value decomposition to establish the polar decomposition; one can also prove the polar decomposition first and then derive the singular decomposition from it.

Theorem 3.22 (Polar decomposition: Autonne [6]; Wintner and Murnaghan [136]). *Let A be an $n\times n$ matrix. Then there exists a unitary matrix U and positive semidefinite matrices H and K such that $A=UH=KU$. In any such representation, the nonzero eigenvalues of H and K must be the singular values of A. Furthermore, if A is nonsingular, then H, K, and U are uniquely determined by A.*

Proof. From Theorem 3.21 we have $A=RDS$, where R and S are unitary and D is a nonnegative diagonal matrix. Then $A=RS(S^*DS)=$

$(RDR^*)RS$. Let $U = RS$, while $H = S^*DS$ and $K = RDR^*$. Then U is unitary, H and K are positive semidefinite, and $A = UH = KU$.

If $A = UH = KU$ is such a representation, then $AA^* = K^2$ and $A^*A = H^2$, so the nonzero eigenvalues of H and K must be the singular values of A.

Finally, if A is nonsingular, then K must be the unique positive definite square root of AA^*, while H is the unique positive definite square root of A^*A, so H and K are uniquely determined. But then U must be $AH^{-1} = K^{-1}A$. ∎

The polar decomposition is analogous to writing a complex number z as $re^{i\theta}$; the unitary factor plays the role of the $e^{i\theta}$, while the positive semidefinite Hermitian factor is analogous to the modulus r. There are also versions of the polar decomposition for nonsquare matrices; see [109, 135, 49].

4. THE REDUCTION PROCESS

Theorem 3.19 tells us that any matrix can be transformed to special Weyr form with a unitary similarity. The next step is to apply further unitary similarities to specify the off-diagonal blocks and eventually achieve a unique canonical form, a process outlined by Littlewood. The more recent papers by Sergeĭchuk and by Benedetti and Cragnolini differ in some details, but both use Weyr form and describe a step-by-step procedure for reducing a matrix to canonical form. However, the basic method of the reduction process appears in Brenner's earlier paper, which gives an inductive definition for a canonical form. Brenner's form is not triangular, but his inductive approach to the problem is essentially the same as those developed in the later papers. The idea is to partition the matrix, and then apply a sequence of unitary similarities that successively transform the submatrices into a special form; at each stage, one uses a unitary transformation that both preserves the form of the already reduced submatrices and reduces an additional submatrix to the desired form. Each step involves refining the partition of the matrix and further restricting the group of unitary transformations to be used in the next step; the refinement process guarantees that the construction stabilizes after a finite number of steps.

Thus, at each step, one has a partitioned matrix in which some blocks have already been reduced to their final form; the idea is to reduce the "next" block while preserving the already treated blocks. A crucial part of the argument is to identify the group of unitary transformations that preserve the already reduced blocks. One then applies a unitary similarity from that group to further refine additional blocks. We also need to specify an ordering of the blocks, in order to identify the "next" block to be reduced. Various versions of these ideas are developed in all of the papers [9, 12, 89, 99]; the specific presentation given here most closely follows [9]. Our notation and terminology differ somewhat from the cited papers. We begin with some ideas from group theory [93] and then develop some notation.

Let $\mathcal{G}$ be a group acting on the set $\mathbf{C}(n)$. We say $n \times n$ matrices A and B are *$\mathcal{G}$-equivalent* if $g(A) = B$ for some g in $\mathcal{G}$. The $\mathcal{G}$-equivalency classes are the *orbits of* $\mathcal{G}$; the action of $\mathcal{G}$ partitions $\mathbf{C}(n)$ into $\mathcal{G}$-orbits. For A in $\mathbf{C}(n)$, the *stabilizer of* A is the subgroup $\mathrm{st}(A) = \{g \in \mathcal{G} \mid g(A) = A\}$.

DEFINITION 4.1. Let $\mathcal{G}$ be a group acting on the set $\mathbf{C}(n)$. A *$\mathcal{G}$-canonical form* is a function $\mathcal{F}: \mathbf{C}(n) \to \mathbf{C}(n)$ such that:

(1) For any A in $\mathbf{C}(n)$, the matrix $\mathcal{F}(A)$ is $\mathcal{G}$-equivalent to A.

(2) $\mathcal{F}(A) = \mathcal{F}(B)$ if and only if A and B are in the same orbit, i.e., if and only if A and B are $\mathcal{G}$-equivalent.

Thus, a canonical form may be viewed as a function that chooses a single representative from each orbit. We shall say $\mathcal{F}(A)$ is the *$\mathcal{G}$-canonical form* of A.

If $\mathcal{G}$ is $\mathrm{GL}(n)$, the multiplicative group of $n \times n$ nonsingular complex matrices, acting on $\mathbf{C}(n)$ by conjugation, then A and B are $\mathrm{GL}(n)$-equivalent if and only if A and B are similar, the $\mathrm{GL}(n)$ orbits are the similarity classes, and the stabilizer of A is $\{S \in \mathrm{GL}(n) \mid S^{-1}AS = A\}$. Thus, for the action of conjugation, the stabilizer of A is the set of all nonsingular matrices that commute with A. The function $\mathcal{J}$ defined by "$\mathcal{J}(A)$ is the Jordan canonical form of A" satisfies Definition 4.1, as does the function $\mathcal{R}$ defined by "$\mathcal{R}(A)$ is the rational canonical form of A." Now consider $\mathbf{U}(n)$, the subgroup of $\mathrm{GL}(n)$ consisting of the unitary matrices. Then $\mathbf{U}(n)$ also acts on $\mathbf{C}(n)$ by conjugation, and A and B are $\mathbf{U}(n)$-equivalent if and only if A and B are unitarily similar. Our main goal here is to obtain a $\mathbf{U}(n)$-canonical form. However, as pointed out in [9, 12, 62, 89, 99], the method may also be used to construct canonical forms for sets of matrices and to obtain $\mathcal{G}$-canonical forms where $\mathcal{G}$ is a certain type of subgroup of $\mathbf{U}(n)$. We discuss this further in Section 5, in connection with Brenner's work.

DEFINITION 4.2. If A and B are unitarily similar, we write $A \sim B$.

We first need to order the complex numbers, so that we may order the eigenvalues of the matrices. One may do this in several ways. Following [9] and [99], we use the lexicographic ordering; thus, for $z = x - iy$ and $w = r + it$, we say z is greater than w if either $x > r$, or $x = r$ and $y > t$. We shall write $z > w$ to indicate that z is greater than w in this ordering. Littlewood uses a different ordering, based on the polar form $re^{i\theta}$ of a complex number.

DEFINITION 4.3. Let A be an $n \times n$ matrix with t distinct eigenvalues $\alpha_1 > \alpha_2 > \cdots > \alpha_t$ of multiplicites $n_1, n_2, \dots, n_t$. For each i, let $n_{i1} \geqslant n_{i2} \geqslant n_{i3} \geqslant \cdots \geqslant n_{ik_i}$ be the Weyr characteristic for the eigenvalue α_i. The ordered sequence

$$n_{11}, n_{12}, \dots, n_{1k_1}, n_{21}, n_{22}, \dots, n_{2k_2}, \dots, n_{i1}, \dots, n_{ik_i}, \dots, n_{t1}, \dots, n_{tk_t}$$

is called the *Weyr characteristic* of A and denoted $\omega(A)$. Let $\omega^*(A)$ be the sequence of positive integers obtained by replacing the number n_{ij} with the pair of numbers $n_{ij} - n_{i,j+1}, n_{i,j+1}$ whenever $n_{ij} > n_{i,j+1}$. The sequence $\omega^*(A)$ is called the *special Weyr characteristic* of A.

REMARK. Although $\omega^*(A)$ is constructed from $\omega(A)$ and thus is uniquely defined by the Weyr characteristic, one cannot necessarily recover the Weyr characteristic from the special Weyr characteristic. For example, if $\omega^*(A) = (1,1,1)$, then $\omega(A)$ could be either $(2,1)$ or $(1,1,1)$.

Henceforth, when we use the term "Weyr form" or "special Weyr form" we assume that the eigenvalues appear along the diagonal in the proper (i.e., lexicographic) order, and thus the sizes of the scalar diagonal blocks are given by the Weyr characteristic.

We need some notation to describe various partitions of matrices and certain special subgroups of $\mathrm{U}(n)$ used in the reduction process. These subgroups are identified and named in various ways in [9, 12, 62, 89, 99]. We use different notation and names, but most of the material in Definitions 4.4 to 4.14 is based on the definitions in [9].

DEFINITION 4.4. Let n be a positive integer. A finite ordered sequence $n_1, n_2, \dots, n_t$ of positive integers such that $n_1 + n_2 + \cdots + n_t = n$ is called an *n-sum* and is denoted $\phi(n) = (n_1, n_2, \dots, n_t)$.

Thus, an n-sum is simply an ordered partition of the number n. An n-sum determines a natural partition of an $n \times n$ matrix into blocks.

DEFINITION 4.5. Let A be an $n \times n$ matrix, and let $\phi(n) = (n_1, n_2, \dots, n_t)$ be an n-sum. Use the numbers n_i to partition the rows, and also the columns, of A into sets of the first n_1 rows, then the next n_2 rows, and so on. The corresponding partition of A into the submatrices A_{ij}, of sizes $n_i \times n_j$, is called a *$\phi(n)$-partition of* A or a *ϕ-partition of* A. The submatrices A_{ij} are called the *ϕ-blocks of* A and sometimes denoted as $A_{ij}(\phi)$.

DEFINITION 4.6. If $\phi(n)=(n_1,n_2,\dots,n_t)$ is an n-sum and A is $\mathbf{T}(n_1,n_2,\dots,n_t)$, we say A is $\mathbf{T}(\phi(n))$. Similarly, if A is $\mathbf{D}(n_1,n_2,\dots,n_t)$, we say A is $\mathbf{D}(\phi(n))$.

For fixed $\phi(n)$, the set of all nonsingular matrices that are $\mathbf{T}(\phi(n))$ forms a subgroup of $\mathrm{GL}(n)$, as does the set of all nonsingular matrices that are $\mathbf{D}(\phi(n))$. We are interested in the subgroup of unitary matrices that are $\mathbf{D}(\phi(n))$.

DEFINITION 4.7. The set of all unitary matrices that are $\mathbf{D}(\phi(n))$ is called the *$\phi(n)$-subgroup of* $\mathrm{U}(n)$ and is denoted by $\mathrm{U}(\phi(n))$.

Note that $\mathrm{U}(\phi(n))$ is the direct sum of the unitary groups $\mathrm{U}(n_i)$. Radjavi [89] uses the term "unrestricted direct group" for $\mathrm{U}(\phi(n))$.

If A is a ϕ-partitioned matrix, if $U=\mathbf{D}(U_1,U_2,\dots,U_t)$ is in $\mathrm{U}(\phi(n))$, and if $B=U^*AU$ is also ϕ-partitioned, then $B_{ij}(\phi)=U_i^*A_{ij}(\phi)U_j$.

At each step of the reduction procedure we refine the partition of the matrix.

DEFINITION 4.8. Let $\phi(n)=(n_1,n_2,\dots,n_t)$ be an n-sum, and for each $i=1,\dots,t$ suppose $\phi_i(n_i)=(n_{i1},n_{i2},\dots,n_{ik_i})$ is an n_i-sum. Then the n-sum $(n_{11},n_{12},\dots,n_{1k_1},n_{21},n_{22},\dots,n_{2k_2},\dots,n_{t1},n_{t2},\dots,n_{tk_t})$ is denoted $\phi(n_1,n_2,\dots,n_t)$ or $(\phi_1,\phi_2,\dots,\phi_t)$ and is called a *refinement of* $\phi(n)$. We say *ϕ_i is the refinement of ϕ in the ith place.*

For example, if $\phi(9)=(4,3,2)$ and $\phi_1(4)=(3,1)$ while $\phi_2(3)=(1,2)$ and $\phi_3(2)=(2)$, then the corresponding refinement is $\phi(4,3,2)=(3,1,1,2,2)$.

If $\phi_1(n)=\phi(n_1,n_2,\dots,n_t)$ refines $\phi(n)$, then $\mathrm{U}(\phi_1(n))$ is a subgroup of $\mathrm{U}(\phi(n))$. Note that the ϕ_1-partition of a matrix A refines the ϕ-partition—i.e., each ϕ-block $A_{ij}(\phi)$ is further partitioned by ϕ_1 and each ϕ_1-block $A_{ij}(\phi_1)$ is contained in some ϕ-block.

If A is an $n\times n$ matrix with t distinct eigenvalues $\alpha_1>\alpha_2>\cdots>\alpha_t$ of multiplicities $n_1,n_2,\dots,n_t$, then the Weyr characteristic $\omega(A)$ is an n-sum and is a refinement of the n-sum $\phi(n)=(n_1,n_2,\dots,n_t)$. The special Weyr characteristic $\omega^*(A)$ is a refinement of $\omega(A)$. Thus, we may restate Theorems 3.12, 3.18, and 3.19 as follows.

THEOREM 4.1. *Let A be an $n\times n$ matrix with Weyr characteristic $\omega(A)=\omega(n)$ and special Weyr characteristic $\omega^*(A)=\omega^*(n)$. Then there is a matrix A_0 in special Weyr form such that $A\sim A_0$. Furthermore, if U is unitary, then U^*A_0U is also in Weyr form if and only if U is in $\mathrm{U}(\omega(n))$, and U^*A_0U is also in special Weyr form if and only if U is in $\mathrm{U}(\omega^*(n))$.*

DEFINITION 4.9. If A is a nonsingular matrix with singular values $\sigma_1 > \sigma_2 > \cdots > \sigma_t$ of multiplicites $n_1, n_2, \ldots, n_t$, then the n-sum $\sigma(n) = (n_1, n_2, \ldots, n_t)$ is called the *singular value characteristic* of A and is denoted by $\sigma(A)$.

We may then restate the nonsingular case of Theorem 3.21 as follows.

THEOREM 4.2. *Let A be a nonsingular matrix with singular values $\sigma_1 > \sigma_2 > \cdots > \sigma_t$ and singular value characteristic $\sigma(A) = (n_1, n_2, \ldots, n_t)$. Then there exist unitary matrices U and V such that $UAV = D = \sigma_1 I_{n_1} \oplus \sigma_2 I_{n_2} \oplus \cdots \oplus \sigma_t I_{n_t}$ is the direct sum of the t scalar matrices $\sigma_i I_{n_i}$. Furthermore, unitary matrices P and Q satisfy the equation $PDQ = D$ if and only if $P = Q^*$ and P is in $\mathbf{U}(\sigma(A))$.*

We now want to define certain subgroups of $\mathbf{U}(\phi(n))$ in which we require some of the diagonal blocks to be equal. We indicate this with a "tagging function" that associates the same nonzero "tag number" to blocks that are required to be equal, while tagging the unrestricted blocks with the number zero. Definition 4.10 comes directly from Benedetti and Cragnolini; Brenner and Radjavi [89] give alternative descriptions.

DEFINITION 4.10. Let $\phi(n) = (n_1, n_2, \ldots, n_t)$ be an n-sum. A *tagging function τ of $\phi(n)$* is a function from $\{1, 2, \ldots, t\}$ to $\{0, 1, 2, \ldots, p\}$ satisfying the following conditions:

(1) If $\tau(i) \neq 0$, then there is at least one $j \neq i$ such that $\tau(i) = \tau(j)$.

(2) If $\tau(i) \neq 0$ and $\tau(i) = \tau(j)$ then $n_i = n_j$.

(3) The restriction of τ to the set of minima $\{\min \tau^{-1}(\tau(i)) \mid \tau(i) \neq 0\}$ is an increasing map and is onto $\{1, 2, \ldots, p\}$.

The n-sum $\phi(n)$ together with the tagging function τ is called a *tagged n-sum* and is denoted by $\phi\tau(n) = (n_{1\,\tau(1)}, n_{2\,\tau(2)}, \ldots, n_{t\,\tau(t)})$. If $\tau(i) \neq 0$ and $\tau(i) = \tau(j)$, we say that i and j, or n_i and n_j, are *linked*.

DEFINITION 4.11. Let $\phi\tau(n)$ be a tagged n-sum with $\phi(n) = (n_1, n_2, \ldots, n_t)$. The *$\phi\tau(n)$-subgroup of* $\mathbf{U}(n)$ is

$$\left\{U = \mathbf{D}(U_1, U_2, \ldots, U_t) \in \mathbf{U}(\phi(n)) \,\middle|\, U_i = U_j \text{ whenever } i \text{ and } j \text{ are linked}\right\}$$

and is denoted by $\mathbf{U}(\phi\tau(n))$.

For example, if $n = 22$ and $\phi\tau(22) = (3_0, 2_1, 3_0, 4_2, 2_1, 2_0, 4_2, 2_1)$, then $\mathbf{U}(\phi\tau(24))$ is the set of all block diagonal matrices of the form

$\mathbf{D}(A, B, C, D, B, E, D, B)$, where A and C are in $\mathbf{U}(3)$, while B and E are in $\mathbf{U}(2)$ and D is in $\mathbf{U}(4)$. The numbers of the n-sum $\phi(n)$ tell us the sizes of the diagonal blocks, and the tagging function τ tells us which blocks must be the same—all blocks having the same nonzero tag must be the same. Brenner calls the groups $\mathbf{U}(\phi\tau(n))$ "generalized diagonal unitary groups," while Radjavi [89] uses the term "direct groups."

We can also use the tagging function to identify associated submatrices of partitioned matrices. Sergeĭchuk does something similar in his paper, where he uses "shaded" and "unshaded" blocks.

DEFINITION 4.12. Let $\phi\tau(n)$ be a tagged n-sum, and let A be an $\phi(n)$-partitioned matrix. If $\tau(i) \neq 0$ and $\tau(i) = \tau(j)$, then A_{ij} is called a *tagged submatrix* or *tagged block* and we say A is a $\phi\tau(n)$-partitioned matrix, or a *tagged partitioned matrix*.

Suppose A is a ϕ-partitioned matrix, $U = \mathbf{D}(U_1, U_2, \dots, U_t)$ is in $\mathbf{U}(\phi(n))$, and $B = U^*AU$ is also ϕ-partitioned. Then $B_{ij}(\phi) = U_i^*A_{ij}(\phi)U_j$. If we use a tagged partition, $\phi\tau(n)$, then $U_i = U_j$ whenever i and j are linked, so any tagged block is transformed by a unitary similarity.

DEFINITION 4.13. Let $\phi\tau(n)$ be a tagged n-sum, and let A be a ϕ-partitioned matrix. We say A_{ij} is a *$\phi\tau$-stable block* if $(U^*AU)_{ij} = A_{ij}$ for every U in $\mathbf{U}(\phi\tau(n))$.

Thus, the $\phi\tau$-stable blocks are preserved whenever a unitary similarity from $\mathbf{U}(\phi\tau(n))$ is applied to A. If i and j are linked, then A_{ij} is $\phi\tau$-stable if and only if $U^*A_{ij}U = A_{ij}$ for every U in $\mathbf{U}(n_i)$, which is possible only if A_{ij} is a scalar matrix. Hence, for linked subscripts i and j, the minor A_{ij} is $\phi\tau$-stable if and only if A_{ij} is scalar. If τ does not link i and j, then A_{ij} is $\phi\tau$-stable if and only if $UA_{ij}V = A_{ij}$ for every U in $\mathbf{U}(n_i)$ and every V in $\mathbf{U}(n_j)$. But then we have $UA_{ij} = A_{ij}V^*$ for every U in $\mathbf{U}(n_i)$ and every V in $\mathbf{U}(n_j)$, and thus $A_{ij} = 0$ by Theorem 3.4. So if τ does not link i and j, then A_{ij} is $\phi\tau$-stable if and only if $A_{ij} = 0$.

Finally, we want to construct refinements of tagged n-sums that preserve the original tagging.

DEFINITION 4.14. Let $\phi\tau(n) = (n_{1\tau(1)}, n_{2\tau(2)}, \dots, n_{t\tau(t)})$ be a tagged n-sum. Let $\phi(n_1, n_2, \dots, n_t) = (\phi_1, \phi_2, \dots, \phi_t)$ be a refinement of the n-sum $\phi(n)$ such that whenever n_i and n_j are linked by τ, then $\phi_i(n_i) = \phi_j(n_j)$. Let ρ be a tagging function on $\phi(n_1, n_2, \dots, n_t)$ that tags $\phi_i(n_i)$ and $\phi_j(n_j)$ exactly the same way whenever n_i and n_j are linked by τ and uses only nonzero tags on the numbers in such linked $\phi_i(n_i)$ and $\phi_j(n_j)$. In other words, linked n_i and n_j in the original tagged n-sum must induce the same partition and tags in the refinement, so that all of the pieces of the refinement preserve the original linkings. Then we say the tagged n-sum *$\phi\rho(n_1, n_2, \ldots, n_t)$ is a refinement of $\phi\tau(n)$* and that *ϕ_i is the refinement of $\phi\tau$ in the ith place.*

For example $(2_1, 3_2, 2_1, 3_2)$ is a refinement of $(5_1, 5_1)$, but $(2_0, 3_1, 2_0, 3_1)$ is not, and $(2_1, 3_2, 3_2, 2_1)$ is not. As a consequence of this definition, we see that whenever the tagged n-sum $\phi_2\tau_2(n)$ refines $\phi_1\tau_1(n)$, then $\mathbf{U}(\phi_2\tau_2(n))$ is a subgroup of $\mathbf{U}(\phi_1\tau_1(n))$. Hence if A is a ϕ_1-partitioned matrix, any $\phi_1\tau_1$-stable minor of A is unchanged by any unitary similarity from $\mathbf{U}(\phi_2\tau_2(n))$, and any ϕ_2-minor of A contained in a $\phi_1\tau_1$-stable minor is necessarily $\phi_2\tau_2$-stable. Also, we may identify an untagged n-sum $\phi(n) = (n_1, n_2, \ldots, n_t)$ with the tagged n-sum $\phi\tau(n)$, where τ is the trivial tagging function $\tau(i) = 0$ for all i, and hence any refinement of $\phi(n_1, n_2, \ldots, n_t)$ endowed with any tagging function ρ may be considered a tagged refinement of the untagged n-sum $\phi(n)$.

At each stage of the reduction process, we have a tagged n-sum $\phi\tau$, a $\phi\tau$-partitioned matrix, the subgroup $\mathbf{U}(\phi\tau(n))$, and a set of ϕ-blocks of the matrix that are $\phi\tau$-stable. We then apply one of the following four operations to the next ϕ-block that is not $\phi\tau$-stable; Theorems 3.14, 3.16, 3.21, 4.1, and 4.2 tell us that these operations are possible. The operation increases the number of $\phi\tau$-stable blocks and determines a refinement of $\phi\tau$ to be used in the next step.

Type 1. If B is $m \times n$ with $m \geqslant n$, and B has rank $r < m$, then we can find a unitary matrix U such that $UB = B_0$ is in row form. Furthermore, for unitary matrices P and Q, the matrix PB_0Q is still in row form if and only if P is $\mathbf{D}(m - r, r)$.

Type 2. If B is $m \times n$ with $m < n$, and B has rank r, then we can find a unitary matrix U such that $BU = B_0$ is in column form. Furthermore, for unitary matrices P and Q, the matrix PB_0Q is still in column form if and only if Q is $\mathbf{D}(n - r, r)$.

Type 3. If B is nonsingular with singular value characteristic $\sigma(B)$, then we can find a pair of unitary matrices U and V such that $UBV = D$, where D is the singular value form of B. Furthermore, for unitary matrices P and Q, we have $PDQ = D$ if and only if $P = Q^*$ and P is in $\mathbf{U}(\sigma(B))$.

Type 4. If B is $n \times n$, then we can find a unitary matrix U such that $U^*BU = B_0$ is in special Weyr form. Furthermore, for unitary V, the matrix V^*B_0V is in special Weyr form if and only if V is in $\mathbf{U}(\omega^*(B))$, where $\omega^*(B)$ is the special Weyr characteristic of B.

Benedetti and Cragnolini use these four operations, but other reduction schemes are possible. For example, Sergeĭchuk uses Theorem 3.21, the singular value decomposition for singular and rectangular matrices, to combine types 1–3.

We can now describe the reduction process. Let A be an $n\times n$ matrix with special Weyr characteristic $\omega^*(A)=\omega^*(n)$. Using induction on d, we show that at stage d one can construct a matrix $A(d)$, and a tagged n-sum $\phi_d\tau_d(n)$, such that $A(d)\sim A(d-1)\sim A$ and $\phi_d\tau_d$ is a refinement of $\phi_{d-1}\tau_{d-1}$. Furthermore, $A(d)$ is in special Weyr form, at least d of the ϕ_d-blocks of $A(d)$ are $\phi_d\tau_d$-stable, and a unitary similarity U preserves all of the $\phi_d\tau_d$-stable blocks of $A(d)$ only if U is in $\mathrm{U}(\phi_d\tau_d(n))$. Since each matrix $A(d)$ is upper triangular, and since the scalar blocks along the main diagonal are preserved at each stage, we need be concerned only with reducing the submatrices above the diagonal. The submatrices above the diagonal are called *upper blocks*, or *upper submatrices*. Let γ be an ordering of the set $\{(i,j)\,|\,i<j\}$, and for any partitioned matrix, use γ to order the upper blocks. One may use any ordering, but the canonical form obtained depends on the choice of ordering.

STEP 0. Find a unitary matrix U such that $A(0)=U^*AU$ is in special Weyr form. Let $\phi_0\tau_0$ be the tagged n-sum with $\phi_0=\omega^*(A)$, and τ_0 the trivial tagging function. Notice that all of the scalar diagonal blocks of $A(0)$ are preserved by the action of $\mathrm{U}(\phi_0\tau_0(n))$ and thus are $\phi_0\tau_0$-stable. Furthermore, for any pair of consecutive integers $n_{i,j-1},n_{ij}$ in $\omega(n)$ with $n_{i,j-1}>n_{ij}$, the corresponding superdiagonal block of size $n_{i,j-1}\times n_{ij}$ starts with $n_{i,j-1}-n_{ij}$ rows of zeros that form a zero block of size $(n_{i,j-1}-n_{ij})\times n_{ij}$, which is $\phi_0\tau_0$-stable. From Theorem 4.1, if a unitary similarity U preserves special Weyr form it must be in $\mathrm{U}(\omega^*(A))$, so any unitary similarity preserving the $\phi_0\tau_0$-stable blocks of $A(0)$ must belong to $\mathrm{U}(\phi_0\tau_0(n))$.

Assume, as the induction hypothesis, that at Step $d-1$ we have a matrix $A(d-1)\sim A(d-2)$ and that $A(d-1)$ is in special Weyr form. Also assume that we have a tagged n-sum $\phi_{d-1}\tau_{d-1}$ that refines $\phi_{d-2}\tau_{d-2}$, and that at least $d-1$ of the upper submatrices of the ϕ_{d-1}-partitioned matrix $A(d-1)$ are $\phi_{d-1}\tau_{d-1}$-stable. Finally, assume that a unitary similarity preserves the $\phi_{d-1}\tau_{d-1}$-stable blocks of $A(d-1)$ only if it is in $\mathrm{U}(\phi_{d-1}\tau_{d-1}(n))$.

If every upper submatrix of $A(d-1)$ is $\phi_{d-1}\tau_{d-1}$-stable, then the process ends. If not, continue to Step d.

STEP d. We know that $A(d-1)$ has upper blocks that are not $\phi_{d-1}\tau_{d-1}$-stable. Using the ordering γ, let $B=A_{ij}(d-1)$ be the first ϕ_{d-1}-block of $A(d-1)$ that is not stable. Then B is $n_i\times n_j$, where $\phi_{d-1}(n)=(n_1,n_2,\dots,n_t)$. Let r be the rank of B. We must now consider several cases.

CASE 1. Suppose that n_i and n_j are linked by τ_{d-1}, so that B is a tagged block. Then $n_i = n_j$, and we perform a type 4 operation on B. Thus, find an $n_i \times n_i$ unitary matrix U_b such that $U_b^* B U_b = B_0$ is in special Weyr form. Let $U = \mathbf{D}(U_1, U_2, \ldots, U_t)$ be a matrix in $\mathbf{U}(\phi_{d-1}\tau_{d-1}(n))$ such that $U_i = U_b$. Note that we must then have $U_k = U_b$ whenever k is linked to i by τ_{d-1}; in particular, $U_j = U_b$. Let $A(d) = U^* A(d-1) U$. Since U is in $\mathbf{U}(\phi_{d-1}\tau_{d-1}(n))$, the matrix $A(d)$ is still in special Weyr form and the $\phi_{d-1}\tau_{d-1}$-stable blocks of $A(d-1)$ are the same in $A(d)$. However, $A_{ij}(d) = U_i^* B U_j = U_b^* B U_b = B_0$ is in special Weyr form.

Let $\omega^*(n_i) = \omega^*(B) = (m_1, m_2, \ldots, m_s)$ be the special Weyr characteristic of B, and let $\phi_d(n)$ be the n-sum obtained by replacing n_r with $\omega^*(n_i)$ whenever n_r is linked to n_i by τ_{d-1}. So ϕ_d is the refinement of ϕ_{d-1} obtained by refining ϕ_{d-1} in the ith place, and also in each place linked to i, by the n_i-sum $\omega^*(B)$. Now define a tagged refinement of $\phi_{d-1}\tau_{d-1}$ by giving each of the s numbers in the list $m_1, m_2, \ldots, m_s$ a different nonzero tag, replacing each n_k that is τ_{d-1}-linked to n_i with the newly tagged numbers m_h, and then relabeling the tags, if necessary, so that the tagged n-sum $\phi_d\tau_d$ is a tagged refinement of $\phi_{d-1}\tau_{d-1}$. Then $\mathbf{U}(\phi_d\tau_d(n))$ is a subgroup of $\mathbf{U}(\phi_{d-1}\tau_{d-1}(n))$, so any ϕ_d-minors of $A(d)$ contained in the $\phi_{d-1}\tau_{d-1}$-stable blocks of $A(d-1)$ are also $\phi_d\tau_d$-stable. Furthermore, the scalar, diagonal blocks of the special Weyr form matrix B_0 are also stable under the action of $\mathbf{U}(\phi_d\tau_d(n))$. Hence, the new matrix $A(d)$ has at least d upper blocks that are $\phi_d\tau_d$-stable. Furthermore, we know that a unitary similarity will preserve the Weyr special form of $A_{ij}(d)$ only if it is $\mathbf{D}(\omega^*(n_i))$, so any unitary similarity preserving both the ϕ_{d-1}-stable blocks of $A(d)$ and the new stable blocks in the reduced submatrix $A_{ij}(d)$ must be in $\mathbf{U}(\phi_{d-1}\tau_{d-1}(n))$ and must have an ith diagonal block of the form $\mathbf{D}(\omega^*(n_i))$; hence it must be in $\mathbf{U}(\phi_d\tau_d(n))$ by the way we have defined the refinement $\phi_d\tau_d$.

Thus, we have constructed $A(d) \sim A(d-1)$ and a tagged n-sum $\phi_d\tau_d(n)$ such that $A(d)$ is in Weyr special form, $\phi_d\tau_d$ refines $\phi_{d-1}\tau_{d-1}$, the matrix $A(d)$ has at least d upper blocks that are $\phi_d\tau_d$-stable, and these $\phi_d\tau_d$-stable blocks of $A(d)$ are preserved only by unitary similarities from $\mathbf{U}(\phi_d\tau_d(n))$.

CASE 2. Suppose n_i and n_j are not linked by $n_i = n_j = r$. Then we can apply a type 3 operation to B. Thus, let U_b and V_b be unitary matrices in $\mathbf{U}(n_i)$ such that $U_b^* B V_b = D$, where D is the singular value form of B. Choose a matrix $U = \mathbf{D}(U_1, U_2, \ldots, U_t)$ in $\mathbf{U}(\phi_{d-1}\tau_{d-1}(n))$ such that $U_i = U_b$ and $U_j = V_b$, and let $A(d) = U^* A(d-1) U$. Then, again, the $\phi_{d-1}\tau_{d-1}$-stable blocks of $A(d-1)$ are not changed by this transformation, but $A_{ij}(d) = U_b^* A_{ij}(d-1) V_b = U_b^* B V_b = D$.

Let $\sigma(B)=\sigma(n_i)$ be the singular value characteristic of B. Let ϕ_d be the refinement of ϕ_{d-1} obtained by replacing both n_i and n_j with the n_i-sum $\sigma(n_i)$, as well as replacing n_k with the n_i-sum $\sigma(n_i)$ whenever n_k is linked to either n_i or n_j by τ_{d-1}. Now define a tagged refinement, $\phi_d\tau_d$, of $\phi_{d-1}\tau_{d-1}$ by giving each of the numbers in the n_i-sum $\sigma(n_i)$ a different nonzero tag, replacing each n_k that is τ_{d-1}-linked to either n_i or n_j with the newly tagged $\sigma(n_i)$, using the original tagging function τ_{d-1} on the remaining n_k's, and then relabeling the tags, if necessary. Note that the ϕ_d-blocks of $A(d)$ that are contained in the $\phi_{d-1}\tau_{d-1}$-stable blocks of $A(d-1)$ are also $\phi_d\tau_d$-stable, and in addition the scalar, diagonal blocks of the singular value form D are stable under the action of $\mathbf{U}(\phi_d\tau_d(n))$. Furthermore, Theorem 4.2 tells us that $UDV=D$ only if $U=V^*$ and U is $\mathbf{D}(\sigma(n_i))$, so the definition of $\phi_d\tau_d$ guarantees that a unitary similarity can preserve the old $\phi_{d-1}\tau_{d-1}$-stable blocks as well as the new stable blocks of D only if it is in $\mathbf{U}(\phi_d\tau_d(n))$.

Thus, we have $A(d)\sim A(d-1)$ with $A(d)$ in Weyr special form, the tagged n-sum $\phi_n\tau_d(n)$ is a refinement of $\phi_{d-1}\tau_{d-1}$, and $A(d)$ has at least d upper minors that are $\phi_d\tau_d$-stable; these stable blocks are preserved only by unitary similarities from $\mathbf{U}(\phi_d\tau_d(n))$.

Case 3. Suppose n_i and n_j are not linked and $r<\min\{n_i,n_j\}$.

If $n_i\geqslant n_j$, apply a type 1 operation to B. Thus, find an $n_i\times n_i$ unitary matrix P such that PB is in row form. Choose $U=\mathbf{D}(U_1,U_2,\dots,U_t)$ in $\mathbf{U}(\phi_{d-1}\tau_{d-1}(n))$ with $U_i=P^*$, and set $A(d)=U^*A(d-1)U$. As before, this transformation does not change the $\phi_{d-1}\tau_{d-1}$-stable blocks of $A(d-1)$, but the new block $A_{ij}(d)=PBU_j$ is in row form. Let ϕ_d be the refinement of ϕ_{d-1} formed by replacing n_i, and each n_k linked to n_i, with the pair of numbers n_i-r,r. If $\tau_{d-1}(i)=0$, so that n_i is not linked to any number of $\phi_{d-1}(n)$, then we extend the tagging function τ_{d-1} to a tagging function τ_d of ϕ_d by tagging n_i-r and r with zeros, and leaving the original tags alone. If $\tau_{d-1}(i)\neq 0$, then define τ_d by giving the numbers n_i-r and r different nonzero tags, replacing each n_k that is τ_{d-1}-linked to n_i with the newly tagged n_i-r and r, using the original tagging function τ_{d-1} on the remaining n_k's, and then relabeling the tags, if necessary. Then $A(d)\sim A(d-1)$ with $A(d)$ in Weyr special form, and the tagged n-sum $\phi_d\tau_d(n)$ refines $\phi_{d-1}\tau_{d-1}$. Furthermore the first n_i-r rows of PBU_j are zero and form a $\phi_d\tau_d$-stable minor, so $A(d)$ has at least d stable minors. From Theorem 3.14 and our construction of $\phi_d\tau_d$ we see that a unitary similarity preserves the stable blocks of $A(d-1)$ as well as the newly created stable block only if it is in $\mathbf{U}(\phi_d\tau_d(n))$.

If $n_i<n_j$, apply a type 2 operation to B and find an $n_j\times n_j$ unitary

matrix Q such that BQ is in column form. Choose U with $U_j = Q$, and refine ϕ_{d-1} in the jth place, as well as each place linked to j, with the pair $n_j - r, r$. The remainder of the argument is the same; note that the first $n_j - r$ columns of $U_i^*BQ = A_{ij}(d)$ form an $n_i \times (n_j - r)$ block of zeros that is $\phi_d \tau_d$-stable.

Thus, by induction, for each positive integer d, we can construct $A(d) \sim A$ such that $A(d)$ is in special Weyr form with at least d upper blocks that are $\phi_d \tau_d$-stable. However, a block upper triangular matrix has at most $n(n-1)/2$ blocks above the diagonal, so for some positive integer $e \leqslant n(n-1)/2$ the process must produce a matrix $A(e)$ in which all of the blocks are $\phi_e \tau_e$-stable. At this point the construction stops, and we claim that the function $\mathscr{F}: \mathbf{C}(n) \to \mathbf{C}(n)$ defined by $\mathscr{F}(A) = A(e)$ is a canonical form for unitary similarity. We clearly have $A \sim A(e)$ and now need to prove that $A(e) = B(e)$ whenever $A \sim B$. We again proceed by induction.

Thus, suppose $A \sim B$. Then $\omega(A) = \omega(B)$ and $\omega^*(A) = \omega^*(B)$. At step 0, both $A(0)$ and $B(0)$ are in special Weyr form and we have $A(0) \sim B(0)$. Note also that the tagged n-sum $\phi_0 \tau_0$ will be the same for both A and B, and so by Theorem 4.1 we have $U^*A(0)U = B(0)$ for some U in $\mathbf{U}(\phi_0 \tau_0(n))$.

Assume, as the induction hypothesis, that after step $d-1$ we have constructed $A(d-1)$ and $B(d-1)$, as well as tagged n-sums $\phi_{d-1}\tau_{d-1}(A)$ and $\phi_{d-1}\tau_{d-1}(B)$ satisfying the following conditions:

(1) $\phi_{d-1}\tau_{d-1}(A) = \phi_{d-1}\tau_{d-1}(B)$.

(2) There is a U in $\mathbf{U}(\phi_{d-1}\tau_{d-1}(n))$ such that $U^*A(d-1)U = B(d-1)$.

We now show that our construction guarantees that this also holds after step d.

Condition (2) implies that $A(d-1)$ and $B(d-1)$ must agree on every $\phi_{d-1}\tau_{d-1}$-stable block, so if $A_{ij}(d-1)$ is the first nonstable block of $A(d-1)$, then $B_{ij}(d-1)$ must be the first nonstable block of $B(d-1)$. Let $U = \mathbf{D}(n_1, n_2, \ldots, n_t)$ in $\mathbf{U}(\phi_{d-1}\tau_{d-1}(n))$ satisfy $U^*A(d-1)U = B(d-1)$. Then $B_{ij}(d-1) = U_i^*A_{ij}(d-1)U_j$. If n_i and n_j are linked, then $U_i = U_j$, and the Case 1 argument produces the same refinement on $\phi_{d-1}\tau_{d-1}$ for both A and B. If n_i and n_j are not linked, then we are in either Case 2 or Case 3. In Case 2, we see that $B_{ij}(d-1)$ and $A_{ij}(d-1)$ must have the same singular value decomposition; in Case 3, we see that $B_{ij}(d-1)$ and $A_{ij}(d-1)$ must have the same size and rank. Hence, in either case, $\phi_{d-1}\tau_{d-1}$ undergoes the same refinement construction. Thus, in all cases, the construction yields $\phi_d \tau_d(A) = \phi_d \tau_d(B)$. Furthermore, the construction guarantees that $A(d)$ and $B(d)$ must agree both on $\phi_{d-1}\tau_{d-1}$-stable blocks and on the new stable blocks formed in step d. So any unitary U satisfying $U^*A(d)U = B(d)$ must be in $\mathbf{U}(\phi_d \tau_d(n))$. Since $A(d) \sim B(d)$, such a U must exist.

Now, by induction, conditions (1) and (2) must hold at the completion of the process, so that $U^*A(e)U = B(e)$ for some U in $\mathrm{U}(\phi_e\tau_e(n))$. But every block of $A(e)$ is $\phi_e\tau_e$-stable, so $U^*A(e)U = A(e)$ and thus $A(e) = B(e)$.

Thus, we have the following result.

THEOREM 4.3. *Let γ be an ordering of the set of ordered pairs $\{(i,j)| i \leqslant j\}$. For each A in $\mathrm{C}(n)$, let $A(e)$ denote the matrix constructed by the process described above, using the ordering γ, and let $\phi_e\tau_e(n)$ be the corresponding tagged n-sum. Then the map $\mathscr{F}: \mathrm{C}(n) \to \mathrm{C}(n)$ defined by $\mathscr{F}(A) = A(e)$ is a canonical form for $\mathrm{C}(n)$ under the action of unitary similarity, and $\mathrm{U}(\phi_e\tau_e(n))$ is the group of unitary matrices that commute with $A(e)$.*

This completes the proof of the reduction process. We now make some remarks about the procedure and the canonical form obtained, and look at some special cases to get a better understanding of how the reduction works. First note that at step 0, the n-sum $\phi_0\tau_0$ is just the special Weyr characteristic with the trivial tagging function. Operations 1, 2, and 4 do not produce any new linkages, but merely preserve the already existing ones. Hence, new linked pairs n_i and n_j occur only when we use the singular value decomposition of operation 3. This occurs because the singular value form D is preserved only by unitary similarities of the proper block diagonal form; hence the new tagging function must link i and j in order to preserve a singular value form in the i,j block. At each step, new stable blocks are formed. For the Weyr form produced by operation 4, the new stable blocks are the scalar, diagonal blocks and the zero subblocks of the superdiagonal blocks. Operation 3 produces the scalar, diagonal blocks of the singular value form, while operations 1 and 2 produce zero blocks of rows or columns, respectively. Thus, our final matrix $A(e)$ is a ϕ_e-partitioned matrix in which each ϕ_e-block is either a block of zeros or a scalar block. When i and j are not linked by τ_e, then A_{ij} is $\phi_e\tau_e$-stable only if $A_{ij} = 0$, so any ϕ_e-block corresponding to an unlinked pair must be zero, while any tagged ϕ_e-block must be scalar. Of course, the ϕ_e-blocks may be quite small; if they are all 1×1, then we simply have an upper triangular matrix.

In general, it is not clear what information can be gleaned from the canonical form. While the Weyr special form does give the similarity class, and thus the Jordan form, of the matrix, one might hope that a canonical form would display some geometric information about the transformation, or reveal the unitary invariants of the matrix. However, it seems that the final structure of the canonical form evolves only as we apply the step-by-step construction, and thus reflects the sequence of steps performed. Sergeĭchuk and Benedetti and Cragnolini discuss this in more detail, giving examples and describing some general schemes for classifying these canonical forms. Also note that the canonical form obtained depends on the ordering γ used to order the submatrices. Some possible choices for γ might be a lexicographic ordering, or perhaps a "diagonal" ordering in which (i, j) would precede (k, r) whenever $j - i < r - k$.

In Section 2 we examined the special case where the matrix has distinct eigenvalues, and later we saw, in Section 3, that the same results apply in the nonderogatory case. If A is nonderogatory, with eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_t$ of multiplicities $n_1, n_2, \ldots, n_t$, then both $\omega(A)$ and $\omega^*(A)$ are the all 1 sequence, so $\phi_0 \tau_0 = (1, 1, 1, \ldots, 1)$, and $\mathrm{U}(\phi_0 \tau_0(n))$ is the group of diagonal, unitary matrices. The $\phi_0 \tau_0$-stable minors of $A(0)$ are the diagonal entries and any zero entries that happen to occur. For each eigenvalue α_i of A, the corresponding diagonal block of $A(0)$, of size $n_i \times n_i$, must have nonzero entries in the superdiagonal positions, because $A(0)$ is in Weyr form. At any stage in the construction, one applies only diagonal unitary similarities; thus, the zero entries are always stable, and new zero entries cannot appear. Operations of type 1 or 2 are never needed, the type 3 operation just replaces an entry with its modulus, and a type 4 operation has no effect. The reduction process thus reduces to the procedure described in Section 2, which amounts to choosing a diagonal unitary matrix that creates some positive entries above the diagonal. Which entries are converted to positive numbers, and the final structure of the group $\mathrm{U}(\phi_e \tau_e(n))$, depend on the ordering γ and the zero-nonzero structure of the matrix. In general, one expects to create $n-1$ positive entries and to end up with $\mathrm{U}(\phi_e \tau_e(n))$ being the group of unitary, scalar matrices. However, it is not always possible to have $n-1$ positive entries (see Example 2.1). Also, in some cases, $\mathrm{U}(\phi_e \tau_e(n))$ contains nonscalar diagonal matrices.

Finally, notice that while the first step of the process involves triangularizing the matrix, the general method does not really depend on this. As we shall see in Section 5, the same approach may be used to obtain other types of nontriangular canonical forms, to define $\mathrm{U}(\phi \tau(n))$ canonical forms, and to reduce sets of matrices simultaneously [9, 12, 99].

5. NONTRIANGULAR FORMS

The canonical forms developed by Littlewood [59], Mitchell [69], Sergeĭchuk [99], and Benedetti and Cragnolini [9] are all triangular. They are constructed by first applying Schur's theorem to put the matrix in triangular form, with the eigenvalues in specified order along the diagonal, and then reducing the rest of the matrix. By starting differently, but then using essentially the same reduction process, one can define other types of nontriangular canonical forms. In this section we discuss work of Brenner [12], McRae [62], and Radjavi [89]. Brenner begins by diagonalizing the positive definite Hermitian matrix A^*A, thus using the structure of the singular value form of A as the initial step. He then gives an inductive definition for a canonical form in which one assumes the form has already been defined for smaller cases. McRae describes a general approach for defining a variety of canonical forms. In [89], Radjavi develops a construction using unitary transformations to orthogonalize the rows and columns of submatrices of a partitioned matrix; Radjavi proposes another method in [90].

Brenner calls the groups $\mathrm{U}(\phi_T(n))$ *generalized diagonal groups*, and gives an inductive definition of a $\mathscr{G}$-canonical form for an $n \times n$ matrix A, under the action of any generalized diagonal group $\mathscr{G}$. The following proof is based on Brenner's presentation, but differs a bit in some details. As in Section 4, we let γ be some ordering of the set of ordered pairs (i, j) of positive integers and use γ to order the submatrices of A. Brenner's argument gives a particular order in which to consider these submatrices. We denote the $\mathscr{G}$-canonical form of A as $\mathscr{F}(A)$; associated with $\mathscr{F}(A)$ will be its stabilizer, the group of unitary matrices that commute with the matrix $\mathscr{F}(A)$.

Let A be an $n \times n$ complex matrix, and let $\mathscr{G}$ be a generalized diagonal group. If $n = 1$, then A is its own canonical form, i.e., we define $\mathscr{F}(A) = A$. We now assume, as the induction hypothesis, that for all $k \times k$ matrices A_0 and any generalized diagonal group $\mathscr{G}_0$, a $\mathscr{G}_0$-canonical form $\mathscr{F}_0(A_0)$ has been defined for A_0, and that the stabilizer of $\mathscr{F}_0(A_0)$ is a generalized diagonal group whenever either

(1) $k < n$ or
(2) $k = n$ and $\mathscr{G}_0$ is a proper subgroup of $\mathscr{G}$.

We must now show how to define a $\mathscr{G}$-canonical form for A. We consider two cases.

CASE 1. Suppose $\mathscr{G} = \mathrm{U}(n)$ is the full unitary group. Let D be the singular value form of A, as in Definition 3.9, and let the n-sum $\phi(n) = (n_1, n_2, \dots, n_t)$ be the singular value characteristic of A. Then, as in the proof of Theorem 3.20, we can find a unitary matrix U such that $U^*(A^*A)U = D^2$. Furthermore, a unitary matrix commutes with D or D^2 if and only if it is in $\mathrm{U}(\phi(n))$. Let $A_1 = U^*AU$; then $A_1^*A_1 = D^2$.

If D is a scalar matrix, then A_1, and hence A itself, is a scalar multiple of a unitary matrix, and thus A is normal. Therefore, we can find a unitary matrix V such that V^*AV is diagonal with the eigenvalues of A in lexicographic order along the diagonal. We define $\mathscr{F}(A)=V^*AV$. Obviously, $A\sim\mathscr{F}(A)$, and if $B\sim A$, then B is also normal and has the same eigenvalues and multiplicities as A, so $\mathscr{F}(B)=\mathscr{F}(A)$. Hence, this defines a canonical form for A in this case. If $m_1,m_2,\ldots,m_r$ are the multiplicities of the eigenvalues of A, and we let $\phi_1(n)=(m_1,m_2,\ldots,m_r)$, then the generalized diagonal group $\mathbf{U}(\phi_1(n))$ is the stabilizer of $\mathscr{F}(A)$.

If D is not a scalar matrix, then $\mathscr{G}_0=\mathbf{U}(\phi(n))$ is a proper subgroup of $\mathscr{G}$. Hence, by the induction hypothesis, we have defined a $\mathscr{G}_0$-canonical form, $\mathscr{F}_0(A_1)$ for A_1, and the stabilizer of $\mathscr{F}_0(A_1)$ is some generalized diagonal group $\mathbf{U}(\phi_0\tau_0(n))$. We then define $\mathscr{F}(A)$ to be $\mathscr{F}_0(A_1)$. Since $A\sim A_1$ and $A_1\sim\mathscr{F}_0(A_1)$, we have $A\sim\mathscr{F}(A)$. If $B\sim A$, then B has the same singular value form as A, so there is a unitary matrix V such that $V^*(B^*B)V=D^2$. Putting $B_1=V^*BV$, we have $B_1\sim A_1$, so there is a unitary matrix W such that $W^*B_1W=A_1$. But then, since $B_1^*B_1=A_1^*A_1=D^2$, we have $W^*D^2W=D^2$, so W must be in $\mathscr{G}_0=\mathbf{U}(\phi(n))$. Hence, B_1 and A_1 are $\mathscr{G}_0$-equivalent, and $\mathscr{F}_0(B_1)=\mathscr{F}_0(A_1)$. Hence $\mathscr{F}(B)=\mathscr{F}(A)$, and we have defined a canonical form for A, and the stabilizer of this canonical form is a generalized diagonal group.

Case 2. Now suppose that $\mathscr{G}=\mathbf{U}(\phi\tau(n))$ is a proper subgroup of $\mathbf{U}(n)$. If $U^*AU=A$ for every U in $\mathscr{G}$, we put $\mathscr{F}(A)=A$ and are done. Otherwise, partition A conformally with the n-sum $\phi\tau(n)$, and let A_{ij} be the first (according to the ordering γ) block of A that is not $\phi\tau$-stable. Let $U=\mathbf{D}(U_1,U_2,\ldots,U_t)$ be in $\mathscr{G}$, and let $B=U^*AU$. Then $B_{ij}=U_i^*A_{ij}U_j$, and we must consider two cases, depending on whether or not $\tau(i)=\tau(j)$.

If $\tau(i)=\tau(j)$, then $U_i=U_j$, and A_{ij} is an $n_i\times n_i$ matrix with $n_i<n$. Hence, by the induction hypothesis, we can find a unitary matrix U_i in $\mathbf{U}(n_i)$ such that $U_i^*A_{ij}U_i=\mathscr{F}(A_{ij})$ is in canonical form, and a generalized diagonal group $\mathbf{U}(\phi_1\tau_1(n_i))$ that is the stabilizer of $\mathscr{F}(A_{ij})$. Let $V=\mathbf{D}(V_1,V_2,\ldots,V_t)$ be a matrix in $\mathbf{U}(\phi\tau(n))$ such that $V_i=V_j=U_i$ and also $V_k=U_i$ for every k linked to i by the tagging function τ. Then $A_1=V^*AV$ will have $\mathscr{F}(A_{ij})$ as its i,j minor. Let $\phi_0\tau_0(n)$ be the tagged refinement of $\phi\tau(n)$ determined by refining $\phi\tau$ in position i, and every position linked to i by τ, with the tagged n_i-sum $\phi_1\tau_1$. Then $\mathscr{G}_0=\mathbf{U}(\phi_0\tau_0(n))$ is a proper subgroup of $\mathscr{G}=\mathbf{U}(\phi\tau(n))$; the induction hypothesis gives a $\mathscr{G}_0$-canonical form, $\mathscr{F}_0(A_1)$, for A_1, and a generalized diagonal group, $\mathscr{G}_1=\mathbf{U}(\phi_1\tau_1(n))$, which is the stabilizer of $\mathscr{F}_0(A_1)$. Define the canonical form $\mathscr{F}(A)$ of A to be $\mathscr{F}_0(A_1)$. Since A is $\mathscr{G}$-equivalent to A_1, and A_1 is $\mathscr{G}_0$-equivalent, and hence also $\mathscr{G}$-equivalent, to $\mathscr{F}_0(A_1)$, it follows that A is $\mathscr{G}$-equivalent to $\mathscr{F}(A)$. If B is $\mathscr{G}$-equivalent to A, then B must also be $\mathscr{G}$-equivalent to A_1. We then form the matrix B_1 from B in the same way we formed A_1; since the i,j minors of

both A_1 and B_1 are in canonical form, the matrices A_1 and B_1 must be $\mathscr{G}_0$-equivalent, so $\mathscr{F}_0(A_1) = \mathscr{F}_0(B_1)$ and hence $\mathscr{F}(B) = \mathscr{F}(A)$.

If $\tau(i) \neq \tau(j)$, then we choose U_i in $\mathbf{U}(n_i)$ and U_j in $\mathbf{U}(n_j)$ so that $U_i^* A_{ij} U_j$ is in singular value form. Note that $A_{ij} \neq 0$ because A_{ij} is not $\phi\tau$-stable. Let $V = \mathbf{D}(V_1, V_2, \dots, V_t)$ be a matrix in $\mathbf{U}(\phi\tau(n))$ such that $V_i = V_k = U_i$ for every k linked to i by the tagging function τ, and such that $V_j = V_k = U_j$ for every k linked to j by the tagging function τ. Put $A_1 = V^*AV$; then the i, j minor of A_1 is the singular value form $U_i^* A_{ij} U_j$. Let q be the rank of A_{ij}, and let $m_1, m_2, \dots, m_r$ be the multiplicities of the nonzero singular values of A_{ij}. Let $\phi_1(n_i) = (m_1, m_2, \dots, m_r, n_i - q)$ and $\phi_2(n_i) = (m_1, m_2, \dots, m_r, n_j - q)$. Define a tagged refinement, $\phi_0\tau_0(n)$, of $\phi\tau(n)$ determined by refining $\phi\tau$ in position i, and every position linked to i by τ, with $\phi_1(n_i)$; refining $\phi\tau$ in position j, and every position linked to j by τ, with $\phi_2(n_i)$; and then linking the m_k's, for each $k = 1, 2, \dots, r$, with the tagging function τ_0. From Theorem 3.21, this is exactly what is needed for the group $\mathscr{G}_0 = \mathbf{U}(\phi_0\tau_0(n))$ to be a subgroup of $\mathbf{U}(\phi\tau(n))$ and for the i, j block of A_1 to be $\phi_0\tau_0$-stable. Since A_{ij} is not zero, $\mathscr{G}_0$ is a proper subgroup of $\mathscr{G}$, and so we can invoke the induction hypothesis to produce the canonical form, $\mathscr{F}_0(A_1)$, for A_1 under the action of $\mathscr{G}_0$. We then define $\mathscr{F}(A) = \mathscr{F}_0(A_1)$ to be the canonical form of A. The same argument used in the first case shows that $\mathscr{F}(A)$ is $\mathscr{G}$-equivalent to A, and if B is $\mathscr{G}$-equivalent to A then $\mathscr{F}(B) = \mathscr{F}(A)$.

This describes a very general inductive method for defining canonical forms; the use of the singular value form of A in the first step serves mainly to start the process. As with the triangular canonical form obtained in Section 4, the final result is a tagged partition $\phi_0\tau_0(n)$ and a $\phi_0\tau_0$-partitioned matrix, $\mathscr{F}(A)$, in which every $\phi_0\tau_0$-block is $\phi_0\tau_0$-stable. Thus, the tagged blocks must be scalar blocks, and the untagged blocks must be blocks of zeros. When $\mathscr{G}$ is the full unitary group and we start the process by diagonalizing A^*A, we obtain $(\mathscr{F}(A))^*(\mathscr{F}(A)) = D^2$, where D is the singular value form of A.

The Brenner inductive definition can be used to find a $\mathscr{G}$-canonical form whenever $\mathscr{G}$ is a generalized diagonal group, and thus the method may be extended to sets of matrices. Let $(A_1, A_2, \dots, A_q)$ and $(B_1, B_2, \dots, B_q)$ be two ordered q-tuples of $n \times n$ matrices. We say the A_i's are simultaneously unitarily similar to the B_i's if there is a unitary matrix U such that $U^*A_iU = B_i$ for each $i = 1, \dots, q$. One might then ask for a canonical form for such ordered q-tuples of $n \times n$ matrices. The methods developed to reduce a single matrix can be applied to such q-tuples of matrices in several ways. One approach [89] is to let A be the direct sum of the A_i's; thus $A = \mathbf{D}(A_1, A_2, \dots, A_q)$ is a block diagonal matrix of size $nq \times nq$. Let $\mathscr{G}$ be the generalized diagonal group of all unitary matrices in $\mathbf{U}(nq)$ of the form $\mathbf{D}(V,V,\dots,V)$, where V is any $n\times n$ unitary matrix. Thus, $\mathcal{G}=\mathbf{U}(\phi\tau(nq))$, with $\phi\tau(nq)=(n_1,n_1,\dots,n_1)$, where the n appears q times and the subscript 1 is the value of the tagging function τ. For any $U=\mathbf{D}(V,V,\dots,V)$ in $\mathcal{G}$, we have $U^*AU=\mathbf{D}(V^*A_1V,V^*A_2V,\dots,V^*A_qV)$, and we can use the $\mathcal{G}$-canonical form for A to obtain a canonical form for the q-tuple of A_i's. An alternative procedure [62] is to consider conformal partitions of the A_i's into submatrices. Thus, let $A_i(r,s)$ denote the r,s block of A_i under some partition. Let γ be some fixed ordering defined on the set of all triples of positive integers (i,r,s) with $1\leqslant i\leqslant q$, $1\leqslant r\leqslant n$, and $1\leqslant s\leqslant n$. Then the same inductive argument used to define a canonical form for a single matrix can be used for the set $A_1,A_2,\dots,A_q$; one simply considers the $\phi\tau$-blocks of all of the matrices. Still another method, proposed by Brenner, is to first apply a unitary matrix U in $\mathcal{G}$ such that $U^*A_1U=\mathcal{F}(A_1)$ is the $\mathcal{G}$-canonical form for A. Then let $\mathcal{G}_0$ be the stabilizer of $\mathcal{F}(A_1)$, and find the $\mathcal{G}_0$-canonical form for U^*A_2U, and so on. Sergeĭchuk shows how to deal with the more general problem of canonical forms for sets of matrices of different sizes, including nonsquare matrices, by defining a large matrix containing the given matrices as submatrices.

As another approach to the canonical form problem, McRae introduces the notion of representing a matrix as a polynomial in normal matrices. Thus, let $f(x_1,x_2,\dots,x_k)$ be a polynomial with complex coefficients in the noncommuting variables $x_1,x_2,\dots,x_k$, and suppose we can find normal matrices $N_1,N_2,\dots,N_k$ such that $A=f(N_1,N_2,\dots,N_k)$. Then one can define a canonical form for A by applying the methods above to the set $N_1,N_2,\dots,N_k$. Since normal matrices can be unitarily diagonalized, when $\mathcal{G}$ is the full unitary group, one can begin by diagonalizing N_1. However, we need to restrict the polynomial f and the allowable normal matrices N_i in order to get a unique representation for A. Thus, assume we have specified a set of conditions restricting the N_i's such that there is a unique k-tuple, $N_1,N_2,\dots,N_k$, satisfying the conditions and such that $A=f(N_1,N_2,\dots,N_k)$. We also need to assume that whenever U is unitary and $N_1,N_2,\dots,N_k$ satisfy these conditions, the k-tuple $U^*N_1U,U^*N_2U,\dots,U^*N_kU$ also satisfies the conditions, and is the unique k-tuple such that $U^*AU=f(U^*N_1U,U^*N_2U,\dots,U^*N_kU)$. As a familiar example of such a representation, consider the decomposition of A into its Hermitian and skew-Hermitian components. Here we have $A=H+iK$, where $H=(A+A^*)/2$ and $K=(A-A^*)/2i$ are Hermitian. If we define $f(x_1,x_2)=(x_1+ix_2)$ and restrict N_1 and N_2 to be Hermitian, then H and K are the unique pair such that $A=f(H,K)$. We could then define the canonical form for A by first choosing U so that $U^*HU=D$ is diagonal, letting $\mathcal{G}$ be the stabilizer of D, and then finding the $\mathcal{G}$-canonical form for U^*KU. (When U^*HU is diagonal, the matrix U^*KU has a special form, described by Taussky [115, 103].) Another possible representation is the polar decomposition for nonsingular matrices; here we have $f(x_1,x_2)=x_1x_2$ and restrict N_1 to be unitary and N_2 to be positive definite Hermitian.

We now discuss Radjavi's method [89] for constructing a canonical form. Here again, we deal with a tagged n-sum, $\phi\tau(n)$, and a $\phi\tau$-partitioned matrix. The idea is orthogonalize the rows and columns of the submatrices of A. The basic construction is based on the following two lemmas.

Lemma 5.1. *Let B be an $r\times s$ matrix with nonzero singular values $\sigma_1>\sigma_2>\cdots>\sigma_t$ of multiplicities $m_1,m_2,\dots,m_t$, and put $\phi(r)=(m_1,m_2,\dots,m_t,r-(m_1+m_2+\cdots+m_t))$. Then there exists an $r\times r$ unitary matrix U_0 such that U_0B has mutually orthogonal row vectors $\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_r$ with $\|\mathbf{x}_1\|\geqslant\|\mathbf{x}_2\|\geqslant\cdots\geqslant\|\mathbf{x}_r\|$. Furthermore, a unitary matrix V has this same property if and only if V is in the coset $\mathrm{U}(\phi(r))U_0$.*

Proof. Let U_0 be a unitary matrix such that $U_0BB^*U_0^*=D$, where D is the singular value form of B. The row vectors of U_0B are then orthogonal and satisfy the length condition. If V is any other unitary matrix with this property, then we must have $VBB^*V^*=D$, because $\sigma_1,\sigma_2,\dots,\sigma_t$ are the singular values of B. Hence $U_0^*DU_0=V^*DV$, and VU_0^* commutes with D. But then VU_0^* is in $\mathrm{U}(\phi(r))$, and so V is in the coset $\mathrm{U}(\phi(r))U_0$. Conversely, if $V=UU_0$ for some U in $\mathrm{U}(\phi(r))$, then U commutes with D, so $VBB^*V^*=D$ and VB has the required property. ■

We shall call a matrix in the coset $\mathrm{U}(\phi(r))U_0$ a *row fixer* for B.

Similarly, we can formulate and prove a column version of Lemma 5.1.

Lemma 5.2. *Let B be an $r\times s$ matrix with nonzero singular values $\sigma_1>\sigma_2>\cdots>\sigma_t$ of multiplicites $m_1,m_2,\dots,m_t$, and put $\phi(s)=(m_1,m_2,\dots,m_t,s-(m_1+m_2+\cdots+m_t))$. Then there exists an $s\times s$ unitary matrix U_0 such that BU_0 has mutually orthogonal column vectors $\mathbf{y}_1,\mathbf{y}_2,\dots,\mathbf{y}_s$ with $\|\mathbf{y}_1\|\geqslant\|\mathbf{y}_2\|\geqslant\cdots\geqslant\|\mathbf{y}_s\|$. Furthermore, a unitary matrix V has this same property if and only if V is in the coset $U_0\mathrm{U}(\phi(r))$.*

The proof of Lemma 5.2 is the same as that of Lemma 5.1; use the diagonalization of B^*B to find the matrix U_0.

Definition 5.1. An $r\times s$ matrix B is said to be *row orthogonal* if BB^* is a scalar matrix, or, equivalently, if the rows of B are mutually orthogonal and all have the same length. We say B is *column orthogonal* if B^*B is a scalar matrix, i.e., if the columns of B are mutually orthogonal and all have the same length.

Now let $\mathscr{G} = \mathrm{U}(\phi\tau(n))$ be a generalized diagonal group, and let A be an $n \times n$ matrix partitioned conformally with $\phi\tau(n)$. We use Lemmas 5.1 and 5.2 to define two operations on A. Note that if U_0 is a row fixer for B, and V is unitary, then $(U_0BV)(U_0BV)^* = U_0BB^*U_0^*$, so U_0 is a row fixer for BV. Similarly, if U_0 is a column fixer for B, then U_0 is a column fixer for VB whenever V is unitary.

OPERATION 1. Let A_{ij} be the first $\phi\tau$-block of A that is not row orthogonal. Let the nonzero singular values of A_{ij} be $\sigma_1 > \sigma_2 > \cdots > \sigma_t$, with multiplicities $m_1, m_2, \ldots, m_t$, and put $\phi_i(n_i) = (m_1, m_2, \ldots, m_t, n_i - (m_1 + m_2 + \cdots + m_t))$. Let U_i^* be a row fixer for A_{ij}. Let U be the element of $\mathscr{G}$ that has U_i in the ith diagonal block position and in every position k linked to i by τ, and has the identity matrix in the other diagonal blocks. Let $\phi_0\tau_0(n)$ be the tagged refinement of $\phi\tau(n)$ obtained by refining $\phi\tau$ in the ith place, and each place linked to i, by $\phi_i(n_i)$, and linking corresponding m_i's. Put $A_0 = U^*AU$ and $\mathscr{G}_0 = \mathrm{U}(\phi_0\tau_0(n))$. Then the i, j block of A_0 has mutually orthogonal row vectors satisfying the length condition of Lemma 5.1.

OPERATION 2. Let A_{ij} be the first $\phi\tau$-block of A that is not column orthogonal. Let the nonzero singular values of A_{ij} be $\sigma_1 > \sigma_2 > \cdots > \sigma_t$, with multiplicities $m_1, m_2, \ldots, m_t$, and put $\phi_j(n_j) = (m_1, m_2, \ldots, m_t, n_j - (m_1 + m_2 + \cdots + m_t))$. Let U_j be a column fixer for A_{ij}. Let U be the element of $\mathscr{G}$ that has U_j in the jth diagonal block position and in every position k linked to j by τ, and has the identity matrix in the other diagonal blocks. Let $\phi_0\tau_0(n)$ be the tagged refinement of $\phi\tau(n)$ obtained by refining $\phi\tau$ in the jth place, and each place linked to j, by $\phi_j(n_j)$, and linking corresponding m_i's. Put $A_0 = U^*AU$ and $\mathscr{G}_0 = \mathrm{U}(\phi_0\tau_0(n))$. The i, j block of A_0 has mutually orthogonal columns satisfying the length condition of Lemma 5.2.

Now use Operations 1 and 2 to define a third operation.

OPERATION 3. By a finite number of applications of Operation 1, we can produce a tagged refinement $\phi_1\tau_1$ of $\phi\tau$, and a matrix A_1 such that A_1 is $\mathscr{G}$-equivalent to A and every $\phi_1\tau_1$-block of A_1 is row orthogonal. We then apply Operation 2 to the matrix A_1, and repeat Operation 2 until we obtain a refinement $\phi_2\tau_2$ of $\phi_1\tau_1$ and a matrix A_2 that is $\mathscr{G}$-equivalent to A_1 and such that every $\phi_2\tau_2$-block of A_2 is column orthogonal.

Now apply Operation 3 until we obtain a refinement $\phi_e\tau_e$ of $\phi\tau$ and a matrix A_e such that A_e is $\mathscr{G}$-equivalent to A and every $\phi_e\tau_e$-block of A_e is both row and column orthogonal. Thus, the nonsquare $\phi_e\tau_e$-blocks must be zero, and any nonzero, square $\phi_e\tau_e$-block is a scalar multiple of a unitary matrix. This establishes part (1) of the following theorem.

THEOREM 5.1 (Radjavi [89]). *Let $\mathscr{G} = \mathrm{U}(\phi\tau(n))$ be a generalized diagonal group, and let A be an $n \times n$ matrix. There is an algorithm that produces a refinement $\phi_e\tau_e$ of $\phi\tau$, and a matrix A_e, such that A_e is $\mathscr{G}$-equivalent to A and the following hold*:

(1) *If A_{ij} is a nonsquare ϕ_e-block of A_e, then $A_{ij} = 0$; if A_{ij} is a square ϕ_e-block of A_e, then $A_{ij} = c_{ij}U_{ij}$ for some unitary matrix U_{ij} and some scalar c_{ij}.*

(2) *A matrix B is $\mathscr{G}$-equivalent to A if and only if B_e is $\mathrm{U}(\phi_e\tau_e)$-equivalent to A_e and the refinement of $\phi\tau$ produced by applying the algorithm to B is $\phi_e\tau_e$.*

See [89] for a complete proof.

We now apply further unitary similarities from $\mathrm{U}(\phi_e\tau_e)$ to the matrix A_e to reduce the nonzero blocks. If $A_{ij} = c_{ij}U_{ij}$ is an untagged block, we can choose a unitary matrix U in $\mathrm{U}(\phi_e\tau_e)$ with U_{ij} in the ith position and an identity matrix in the jth position. The i, j block of U^*AU is then the scalar matrix $c_{ij}I$. However, if $A_{ij} = c_{ij}U_{ij}$ is a tagged block, we choose an $n_i \times n_i$ unitary matrix U_i such that $U_i^* U_{ij} U_i$ is diagonal, with diagonal entries in the lexicographic order, and let U be a matrix in $\mathrm{U}(\phi_e\tau_e)$ with U_i in positions i and j. One must then further refine the n-sum $\phi_e\tau_e$ in the ith place in accordance with the multiplicities of the eigenvalues of U_{ij}. By repeating these operations, we eventually obtain an n-sum $\phi_f\tau_f$ and a matrix A_f such that every minor of A_f is $\phi_f\tau_f$-stable. As noted before, untagged stable blocks must be blocks of zeros, and tagged stable blocks must be scalar blocks. The matrix A_f can then be used as the canonical form for A. For a complete proof and a more detailed description of the algorithms involved, see [89]. Radjavi also points out that when $\mathscr{G}$ is the full unitary group, one can first put the matrix in triangular form and then apply the algorithm to obtain a triangular canonical form.

In [90] Radjavi proposes a different approach for reducing a set of matrices to a canonical form.

The general reduction method has been used to define a variety of canonical forms under unitary similarity. In each case it produces a canonical form and an associated tagged n-sum. The resulting canonical form is stable under the action of the direct unitary group determined by the tagged n-sum, and if one partitions the canonical form using the tagged n-sum, each of the resulting tagged blocks is scalar, while the untagged blocks must be zero. However, it is difficult to visualize these forms without considering the actual steps of the reduction process. Unitary similarity preserves both the algebraic and geometric structure of the transformation, and one would like a canonical form that reveals this structure, or perhaps exhibits unitary invariants. The triangular canonical forms of [9, 59, 99] exhibit the Weyr characteristic, and Brenner's form exhibits the singular value form, but none of these forms seems to exhibit the geometric nature of the transformation the way the Jordan and rational canonical forms display the important similarity invariants.

In [52], Kaluznin and Havidi approach the problem of unitary similarity for m-tuples, $(A_1, A_2, \dots, A_m)$ and $(B_1, B_2, \dots, B_m)$, of square matrices from a more geometric point of view. First note that when we decompose each matrix into its Hermitian and skew Hermitian components, $A_j = H_j + iK_j$ and $B_j = L_j + iM_j$, where H_j, K_j, L_j, and M_j are all Hermitian, a unitary matrix U satisfies $U^*A_jU = B_j$ if and only if $U^*H_jU = L_j$ and $U^*K_jU = M_j$. Thus, we may replace each A_j and B_j with a pair of Hermitian matrices and study the $2m$-tuples in which every entry is Hermitian. Hence, it suffices to consider the case where each of the matrices A_i and B_i is normal. The eigenspaces of an $n \times n$ normal matrix give a decomposition of $\mathbf{C}^n$ into an orthogonal direct sum of subspaces. For each A_i and B_i we have such a decomposition. Hence, we have a set of m decompositions corresponding to the A_i's, and another set of m decompositions corresponding to the B_i's; these are called *configurations*. The m-tuples $(A_1, A_2, \dots, A_m)$ and $(B_1, B_2, \dots, B_m)$ are unitarily equivalent if and only if, for each i, the normal matrices A_i and B_i have the same eigenvalues and multiplicities, and there is a unitary transformation mapping one configuration onto the other. Kaluznin and Havidi analyze the geometry of these two configurations. They consider the "angles" between pairs of corresponding subspaces in the configurations, where the angle between two subspaces is a certain operator defined in terms of the projections onto those subspaces. Using the angles between pairs of subspaces, they further decompose the eigenspaces into orthogonal direct sums of lower dimensional subspaces, repeating the process until the angles between any two components are scalar. They then analyze these reduced configurations. Thus, this geometric theory also depends on a reduction process that successively refines the configurations, just as the reduction to canonical form involves refining the partitions of the matrices.

6. UNITARY INVARIANTS

Constructing canonical forms for matrices under unitary similarity provides one way to determine when two matrices are unitarily similar—A and B are unitarily similar if and only if they have the same canonical form. However, the canonical forms are complicated, especially when the matrix is not nonderogatory, and it is not clear what information they display. Another approach is to find a set of invariants that completely characterize a matrix up to unitary similarity. In this section we survey some results about quantities or properties of a matrix that are invariant under unitary similarity. Perhaps the main result in this area is a theorem of Specht [110] showing that a certain set of traces gives a complete set of unitary invariants. This theorem has been refined by Pearcy [83], and has been generalized to certain types of operators on Hilbert spaces [19, 21, 22, 83, 85, 86]. The problem of unitary invariants is also closely related to the more general problem of finding similarity invariants for sets of matrices.

We use $\mathrm{tr}(A)$ to denote the trace of A.

Unitary similarity is a special form of similarity, so anything invariant under similarity, such as eigenvalues, characteristic and minimal polynomials, elementary divisors, and the Jordan canonical form, is also preserved by unitary similarity. However, there are other properties and quantities that are not always preserved by a similarity, but are preserved by any unitary similarity. For example, if U is unitary and $U^*AU = B$, then $U^*(A^*A)U = B^*B$, so a unitary similarity preserves the singular values of A as well as the quantity $\mathrm{tr}(A^*A) = \sum_{j=1}^{n}\sum_{i=1}^{n}|a_{ij}|^2$, known as the Frobenius norm of A. If $\sigma_1, \sigma_2, \ldots, \sigma_n$ are the singular values of A, then $\mathrm{tr}(A^*A) = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$. Similarity transformations do not generally preserve the singular values or Frobenius norm. However, although the singular values are invariant under unitary similarity, similar matrices with the same singular values need not be unitarily similar.

Example 6.1. For any triple (x, y, z) of positive numbers let

$$M(x,y,z) = \begin{pmatrix} 0 & x & y \\ 0 & 0 & z \\ 0 & 0 & 0 \end{pmatrix},$$

as in Example 3.1. Recall that all such matrices are similar, but two such matrices are unitarily similar if and only if they are identical. To find the singular values of $M(x,y,z)$, compute

$$M^*(x,y,z)M(x,y,z) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & x^2 & xy \\ 0 & xy & y^2+z^2 \end{pmatrix}.$$

This matrix has eigenvalues 0 and the pair $R \pm \sqrt{R^2 - 4(xz)^2}$, where $R = x^2 + y^2 + z^2$. Thus, whenever (x_1, y_1, z_1) and (x_2, y_2, z_2) are distinct triples of positive numbers satisfying the equations $x_1^2 + y_1^2 + z_1^2 = x_2^2 + y_2^2 + z_2^2$ and $x_1 z_1 = x_2 z_2$, the matrices $M(x_1, y_1, z_1)$ and $M(x_2, y_2, z_2)$ have the same singular values but will not be unitarily similar. For a specific example, put $x_1 = y_1 = z_1 = 1$ and $x_2 = \frac{3}{2}$, $y_2 = \sqrt{11}/6$, and $z_2 = \frac{2}{3}$.

When $B = U^*AU$, the matrices AA^* and BB^* have the same trace. More generally, suppose $\omega(A, A^*)$ is any word in A and A^*—that is, $\omega(A, A^*)$ is the result of taking any monomial $\omega(x, y)$ in noncommuting variables x and y and replacing x with A and y with A^*. If A and B are unitarily similar, $U^*\omega(A, A^*)U = \omega(B, B^*)$, so $\omega(A, A^*)$ and $\omega(B, B^*)$ have the same trace. Specht proved that the converse also holds, so that the set $\{\operatorname{tr}(\omega(A, A^*)) \mid \omega(x, y) \text{ is any word in } x \text{ and } y\}$ completely determines A up to unitary similarity, and thus is a complete set of unitary invariants.

THEOREM 6.1 (Specht [110]). *Let A and B be $n \times n$ complex matrices. Then A and B are unitarily similar if and only if* $\operatorname{tr}(\omega(A, A^*)) = \operatorname{tr}(\omega(B, B^*))$ *holds for every word ω.*

The idea of Specht's proof is to look at the free semigroup $\mathscr{S}(x, y)$ generated by x and y and view the maps $\omega(x, y) \to \omega(A, A^*)$ and $\omega(x, y) \to \omega(B, B^*)$ as two representations of $\mathscr{S}$. Let $\mathscr{S}(A, A^*)$ be the set of all matrices of the form $\omega(A, A^*)$. Since $\mathscr{S}(A, A^*)$ is self-adjoint [i.e., $(\mathscr{S}(A, A^*))^* = \mathscr{S}(A, A^*)$], it is fully reducible, which means that a subspace is invariant under $\mathscr{S}(A, A^*)$ if and only if its orthogonal complement is also invariant under $\mathscr{S}(A, A^*)$. Since $\operatorname{tr}(\omega(A, A^*)) = \operatorname{tr}(\omega(B, B^*))$ holds for all words ω, these two fully reducible representations have the same character and thus must be equivalent. Hence, there is a nonsingular matrix P such that $P^{-1}\omega(A, A^*)P = P^{-1}\omega(B, B^*)P$ for any word ω. In particular, $P^{-1}AP = B$ and $P^{-1}A^*P = B^*$. But then $P^{-1}AP = P^*A(P^*)^{-1}$, so PP^* commutes with A. Now let $P = HU$ be the polar decomposition of P, where U is unitary and the Hermitian matrix H is the unique positive definite square root of PP^*. Since H can be expressed as a polynomial in PP^*, the matrix H commutes with A. Thus, $P^{-1}AP = U^*(H^{-1}AH)U = U^*AU = B$.

The final step of this argument shows that if there is a nonsingular matrix S such that $S^{-1}AS = B$ and $S^{-1}A^*S = B^*$, then A and B are unitarily similar. Also, A commutes with SS^*. Conversely, if A commutes with SS^*, then $S^{-1}A^*S = (S^{-1}AS)^*$, so A and $S^{-1}AS$ must be unitarily similar.

Also of interest is the algebra $\mathscr{A}(A, A^*)$ generated by A and A^* over the complex numbers. This self-adjoint algebra is the set of all polynomial expressions $p(A, A^*)$, where $p(x, y)$ is any polynomial in the noncommuting variables x and y. Thus, the words $\omega(A, A^*)$ span $\mathscr{A}(A, A^*)$. If $\operatorname{tr}(\omega(A, A^*)) = \operatorname{tr}(\omega(B, B^*))$ holds for all words ω, then $\operatorname{tr}(p(A, A^*) = \operatorname{tr}(p(B, B^*))$ for any polynomial $p(x, y)$. We can now try to define a map $\psi : \mathscr{A}(A, A^*) \to \mathscr{A}(B, B^*)$ by the rule $\psi(p(A, A^*)) = p(B, B^*)$ for any polynomial $p(x, y)$. One can show that ψ is well defined and one-to-one by using the fact that $\operatorname{tr}(CC^*) = 0$ if and only if $C = 0$, together with the assumption $\operatorname{tr}(p(A, A^*) = \operatorname{tr}(p(B, B^*))$ for any polynomial $p(x, y)$. The map ψ thus is a $*$-isomorphism of $\mathscr{A}(A, A^*)$ onto $\mathscr{A}(B, B^*)$ such that $\psi(A) = B$. The algebra $\mathscr{A}(A, A^*)$ is an example of an H^*-algebra [32]; Bhattacharya [10] proves Specht's theorem by using the fact that representations of an H^*-algebra are equivalent if and only if their characters are equal.

These proofs of Specht's theorem also apply to two finite sets $A_1, A_2, \dots, A_t$ and $B_1, B_2, \dots, B_t$ of $n \times n$ matrices, as noted by Wiegmann [134]. One can either consider the free semigroup $\mathscr{S}(x_1, y_1, x_2, y_2, \dots, x_t, y_t)$ and the representations $\omega(x_1, y_1, x_2, y_2, \dots, x_t, y_t) \to \omega(A_1, A_1^*, A_2, A_2^*, \dots, A_t, A_t^*)$ and $\omega(x_1, y_1, x_2, y_2, \dots, x_t, y_t) \to \omega(B_1, B_1^*, B_2, B_2^*, \dots, B_t, B_t^*)$, or regard the algebras $\mathscr{A}(A_1, A_1^*, A_2, A_2^*, \dots, A_t, A_t^*)$ and $\mathscr{A}(B_1, B_1^*, B_2, B_2^*, \dots, B_t, B_t^*)$ as equivalent representations of the same H^*-algebra. In either case, the same argument proves the following generalization of Specht's theorem to finite sets of matrices.

THEOREM 6.2. *Let $A_1, A_2, \dots, A_t$ and $B_1, B_2, \dots, B_t$ be sets of $n \times n$ matrices. There is a unitary matrix U such that $U^*A_iU = B_i$ for $i = 1, 2, \dots, t$ if and only if for every word $\omega(x_1, y_1, x_2, y_2, \dots, x_t, y_t)$ in the noncommuting variables x_i and y_i we have*

$$\operatorname{tr}(\omega(A_1, A_1^*, A_2, A_2^*, \dots, A_t, A_t^*)) = \operatorname{tr}(\omega(B_1, B_1^*, B_2, B_2^*, \dots, B_t, B_t^*).$$

Specht's theorem provides a set of invariants that uniquely determine a matrix up to unitary similarity; since one can form infinitely many words in A and A^*, it gives an infinite set of invariants. Pearcy [83] has shown that a finite set of words suffice. Let $\mathscr{W}(k)$ denote the set of words in the noncommuting variables x and y in which the sum of the exponents does not exceed k.

THEOREM 6.3 (Pearcy [83]). *If A and B are $n \times n$ complex matrices, and if $\operatorname{tr}(\omega(A, A^*)) = \operatorname{tr}(\omega(B, B^*))$ for every word $\omega(x,y)$ in $W(2n^2)$, then A and B are unitarily similar.*

This is proved by first showing that $\{\omega(A, A^*) \mid \omega \text{ is in } W(n^2)\}$ must span the algebra $\mathscr{A}(A, A^*)$. The fact that CC^* has trace zero if and only if $C = 0$ is then used to show that if $\operatorname{tr}(\omega(A, A^*)) = \operatorname{tr}(\omega(B, B^*))$ holds for every word in $W(2n^2)$, then $\operatorname{tr}(\omega(A, A^*)) = \operatorname{tr}(\omega(B, B^*))$ must hold for every word ω. Thus, A and B are unitarily similar by the original Specht theorem.

The number of words in $W(2n^2)$ is less than 4^{n^2}, giving the following result.

THEOREM 6.4 (Pearcy [83]). *There is a complete set of unitary invariants for $n \times n$ complex matrices containing fewer than 4^{n^2} elements.*

The analogues of Theorems 6.1, 6.2, and 6.4 for orthogonal invariants of real matrices also hold, for any collection of traces forming a complete set of unitary invariants for $n \times n$ complex matrices is also a complete set of orthogonal invariants for $n \times n$ real matrices [83].

One might ask for a least upper bound on the number of traces needed for a complete set of unitary invariants for $n \times n$ matrices; the number 4^{n^2} is much too large. For 2×2 matrices, only three traces are needed [74], for the traces of A and A^2 determine the eigenvalues of A, and the quantity $\operatorname{tr}(A^*A)$ then determines the superdiagonal entry of the triangular canonical form described in Section 2. Murnaghan [74] also studied the case $n = 3$; but his claim that six traces suffice is not correct, as shown in [84]. Pearcy [84] shows that nine traces suffice for $n = 3$; Sibirskiĭ [106] improves this by finding a set of seven traces that suffice and form a minimal set. Using Paz's [82] results on the size of a set of words needed to span an algebra generated by a finite set of square matrices, Laffey [57] and Bhattacharya [10] point out that one needs to compute approximately $2^{n^2/3}$ traces. For matrices for which the nonzero singular values have multiplicity one, Bhattacharya gives a family of about $(2n)^n$ traces that suffice. She also gives a topological argument for the existence of $n^2 + 1$ continuous functions on $C(n)$ that give a complete set of unitary invariants. Thus, one would like to find a complete set of specific unitary invariants with the size of the set being a polynomial in n. The problem of finding sets of unitary invariants for sets of matrices is also of interest [106, 111].

The algebra $\mathscr{A}(A, A^*)$ is a closed, self-adjoint algebra of operators acting on the finite dimensional Hilbert space $\mathbf{C}^n$ and thus is a W^*-algebra. A W^*-algebra, or von Neumann algebra, is a weakly closed, self-adjoint algebra of operators on a Hilbert space $\mathscr{H}$ [98]. The structure and classification of W^*-algebras has been much studied, along with the problem of classifying Hilbert space operators up to unitary equivalence and finding unitary invariants for operators [26, 98]. Pearcy [83, 85] has shown that the Specht trace invariants are a complete set of unitary invariants for operators that generate finite W^*-algebras of Type I. Such operators can be decomposed as direct sums of homogeneous n-normal operators; Kaplansky [53] and Brown [13] gave a structure theory for homogeneous n-normal algebras. A homogeneous n-normal operator may be viewed as a continuous, complex valued function from a totally disconnected compact Hausdorff space $\mathscr{X}$ to the set of $n \times n$ complex matrices. Thus, one may regard a homogeneous n-normal operator as an $n \times n$ matrix with entries from $\mathscr{C}(\mathscr{X})$, the ring of continuous, complex valued functions over $\mathscr{X}$. Dixmier extended the notion of trace to finite W^*-algebras by showing the existence of a unique, center valued function D with tracelike properties [83].

THEOREM 6.5 (Pearcy [83]). *Let A be a homogeneous n-normal operator generating the W^*-algebra $\mathscr{R}$, and suppose B is in $\mathscr{R}$. Let $\mathscr{S}$ be any set of words $\omega(x, y)$ such that the associated traces form a complete set of unitary invariants for the set of $n \times n$ complex matrices, and let D be the unique Dixmier central trace on $\mathscr{R}$. Then if $D(\omega(A, A^*)) = D(\omega(B, B^*))$ for each word $\omega(x, y)$ in $\mathscr{S}$, there is a unitary operator U in $\mathscr{R}$ such that $U^*AU = B$.*

THEOREM 6.6 (Pearcy [83, 85]). *Let $\mathscr{R}$ be a finite W^*-algebra of type* I, *and let A and B be in $\mathscr{R}$. Let D denote the unique Dixmier central trace on $\mathscr{R}$. Then there is a unitary operator U in $\mathscr{R}$ such that $U^*AU = B$ if and only if $D(\omega(A, A^*)) = D(\omega(B, B^*))$ for each word $\omega(x, y)$.*

In [86], Pearcy and Ringrose investigate the extent to which similar results hold in a type II_1 von Neumann algebra. Deckard and Pearcy [20] and Deckard [19] also establish generalizations of the Specht theorem for compact and trace class operators [19] and Hilbert-Schmidt operators [22].

Totally disconnected compact Hausdorff spaces are known as *Stonian spaces*; [85, 20, 21] contain additional results on the algebra $M_n(\mathscr{X})$ of $n \times n$ matrices with entries from the ring $\mathscr{C}(\mathscr{X})$ of continuous complex valued functions on the Stonian space $\mathscr{X}$. For example, in [85], Pearcy shows that if $\mathscr{A}$ is an Abelian *-subalgebra of $M_n(\mathscr{X})$, then there is a unitary element U in $M_n(\mathscr{X})$ such that $U^*\mathscr{A}U$ is diagonal. This generalizes the fact that a set of commuting normal matrices over $\mathbf{C}$ can be simultaneously diagonalized. In

[20] Deckard and Pearcy prove that any matrix with entries from $\mathscr{C}(\mathscr{X})$ is unitarily equivalent to another such matrix in upper triangular form, thus extending Schur's theorem to $M_n(\mathscr{X})$. Similarity in $M_n(\mathscr{X})$ is studied in [21].

We have already noted that A and B are unitarily similar if and only if the pair (A, A^*) is simultaneously similar to the pair (B, B^*)—i.e., if and only if there is a nonsingular matrix S such that $S^{-1}AS = B$ and $S^{-1}A^*S = B^*$. Thus, unitary invariants may be viewed as a special case of invariants for pairs of matrices under the action of simultaneous similarity, or conjugation in the group GL(n). More generally, let $\mathscr{M}_m(n)$ be the space of ordered m-tuples of $n \times n$ matrices over a field of characteristic zero, and consider the action of the group GL(n) on this space by simultaneous similarity, or conjugation. Thus, define the action of a nonsingular matrix S on $\mathscr{M}_m(n)$ by the rule $S[(A_1, A_2, \ldots, A_m)] = (S^{-1}A_1S, S^{-1}A_2S, \ldots, S^{-1}A_mS)$. For any word $\omega(x_1, x_2, \ldots, x_n)$ in the noncommuting variables x_i, the quantity $\operatorname{tr}(\omega(A_1, A_2, \ldots, A_m))$ is invariant under this action. Procesi [88] has shown that these trace invariants generate the ring of polynomial invariants for the action of GL(n) on $\mathscr{M}_m(n)$ by simultaneous similarity. A polynomial invariant is a polynomial p in the mn^2 entries of the m matrices $A_1, A_2, \ldots, A_m$ such that $p(A_1, A_2, \ldots, A_m) = p(S^{-1}A_1S, S^{-1}A_2S, \ldots, S^{-1}A_mS)$ for every S in GL(n).

THEOREM 6.7 (Procesi [88]). *Any polynomial invariant of an m-tuple* $(A_1, A_2, \ldots, A_m)$ *of* $n \times n$ *matrices is a polynomial in the invariants* $\operatorname{tr}(\omega(A_1, A_2, \ldots, A_m))$, *where* $\omega(x_1, x_2, \ldots, x_n)$ *runs over the set of all monomials in the noncommuting variables* $x_1, x_2, \ldots, x_n$.

The set of polynomial invariants does not, in general, completely characterize the orbit of $(A_1, A_2, \ldots, A_m)$, and thus these trace invariants do not classify m-tuples of $n \times n$ matrices under simultaneous similarity. For example, suppose $A_1, A_2, \ldots, A_m$ and $B_1, B_2, \ldots, B_m$ are two m-tuples of triangular matrices such that A_i and B_i have the same diagonal entries for each i. Then $\operatorname{tr}(\omega(A_1, A_2, \ldots, A_m)) = \operatorname{tr}(\omega(B_1, B_2, \ldots, B_m))$ will hold for every word ω, but the m-tuples $(A_1, A_2, \ldots, A_m)$ and $(B_1, B_2, \ldots, B_m)$ need not be simultaneously similar. In [30], Friedland solves the problem of classifying pairs (A_1, A_2) of complex matrices under simultaneous similarity. He also shows that when the characteristic polynomial, $p(\lambda, x) = \det(\lambda I - (A_1 + xA_2))$, of the pencil $A_1 + xA_2$ is irreducible, the trace invariants of Theorem 6.7 do completely determine the orbit of the pair (A_1, A_2) under simultaneous similarity. [In the previous example of two sets of triangular matrices, the characteristic polynomial $p(\lambda, x_2, x_3, \ldots, x_m) = \det(\lambda I - (A_1 + x_2A_2 + x_3A_3 + \cdots + x_mA_m))$ factors into m linear factors, where the ith factor corresponds to the ith diagonal position of the matrices A_i.] Motzkin and Taussky [70, 71] study the characteristic polynomial $p(\lambda, x)$ in their work on matrices having the L-property; Friedland generalizes one of their main results in [29].

Results similar to Theorem 6.7 also hold for the action of the orthogonal group on $\mathscr{M}_m(n)$, and for the action of the unitary group $\mathrm{U}(n)$ on m-tuples of $n \times n$ complex matrices [88].

7. THE NUMERICAL RANGE

Another unitary invariant of a matrix is its numerical range, defined by Toeplitz in [119].

DEFINITION 7.1. Let A be an $n \times n$ complex matrix. The *numerical range of* A, denoted $\mathscr{W}(A)$, is the set of complex numbers $\{\mathbf{x}^*A\mathbf{x} \mid \mathbf{x} \text{ is in } \mathbf{C}^n \text{ and } \|\mathbf{x}\| = 1\}$.

The numerical range is also called the *field of values*. The definition also applies to operators of a Hilbert space into itself. More generally, numerical ranges have been defined for operators on normed spaces and for elements of unital normed algebras [11].

Since $\mathscr{W}(A)$ is the image of the unit ball in $\mathbf{C}^n$ under the continuous map $\mathbf{x} \to \mathbf{x}^*A\mathbf{x}$, it is a closed compact subset of the complex plane; a well-known theorem of Toeplitz [119] and Hausdorff [42] says that $\mathscr{W}(A)$ is also convex. Choosing $\mathbf{x}$ to be a unit eigenvector of A shows that $\mathscr{W}(A)$ contains all of the eigenvalues of A. Hence, $\mathscr{W}(A)$ contains the closed convex hull of the eigenvalues of A, a polygonal region with eigenvalues as vertices. We denote the closed, convex hull of a set $\{\zeta_1, \zeta_2, \dots, \zeta_t\}$ of complex numbers as $\mathscr{H}(\zeta_1, \zeta_2, \dots, \zeta_t)$.

If U is unitary, $\mathscr{W}(A) = \mathscr{W}(U^*AU)$, so unitarily similar matrices have the same numerical range. However, similar matrices generally do not have the same numerical range. A direct computation shows that if D is diagonal with diagonal entries $d_1, d_2, \dots, d_n$, then $\mathscr{W}(D) = \mathscr{H}(d_1, d_2, \dots, d_n)$. Hence, if A is normal, with eigenvalues $\alpha_1, \alpha_2, \dots, \alpha_n$, then $\mathscr{W}(A) = \mathscr{H}(\alpha_1, \alpha_2, \dots, \alpha_n)$. When $n \leqslant 4$, the converse holds, but $\mathscr{W}(A) = \mathscr{H}(\alpha_1, \alpha_2, \dots, \alpha_n)$ can hold for a nonnormal matrix when $n \geqslant 5$ [72, 104].

If A has eigenvalues $\alpha_1, \alpha_2, \dots, \alpha_n$, then $\mathscr{W}(A)$ must contain the polygonal region $\mathscr{H}(\alpha_1, \alpha_2, \dots, \alpha_n)$; in general, $\mathscr{H}(\alpha_1, \alpha_2, \dots, \alpha_n)$ will be a proper subset of $\mathscr{W}(A)$. Givens [34] has shown that $\mathscr{H}(\alpha_1, \alpha_2, \dots, \alpha_n)$ is the intersection of the sets $\mathscr{W}(S^{-1}AS)$, where S ranges over all nonsingular matrices. From the polar decomposition of S and the fact that the numerical range is invariant under any unitary similarity, it follows that $\mathscr{H}(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is also the intersection of all of the sets $\mathscr{W}(H^{-1}AH)$ where H ranges over all positive definite Hermitian matrices.

A matrix H is Hermitian if and only if its numerical range is a subset of the real numbers; in this case $\mathscr{W}(H)$ is the closed interval $[\lambda_{\min}(H), \lambda_{\max}(H)]$, where $\lambda_{\min}(H)$ is the smallest eigenvalue of H and $\lambda_{\max}(H)$ is the largest eigenvalue. Writing $A = H + iK$, where H and K are Hermitian, we see $\mathscr{W}(A)$ is contained in the rectangle bounded by the vertical lines $x = \lambda_{\min}(H)$ and $x = \lambda_{\max}(H)$, and the horizontal lines $y = \lambda_{\min}(K)$ and $y = \lambda_{\max}(K)$.

For a 2×2 matrix A, the set $\mathscr{W}(A)$ is an ellipse and the eigenvalues of A are the foci of the ellipse [119]. Furthermore, if we apply a unitary similarity to put A in triangular form:

$$U^*AU = \begin{pmatrix} \alpha_1 & \rho \\ 0 & \alpha_2 \end{pmatrix},$$

then $|\rho|$ is the length of the minor axis of the ellipse [73]. Thus, for 2×2 matrices, A and B are unitarily similar if and only if $\mathscr{W}(A) = \mathscr{W}(B)$. This does not hold for $n \geqslant 3$; for example, whenever H and K are Hermitian matrices with the same minimum and maximum eigenvalues, they will have the same numerical range, but need not be similar if $n \geqslant 3$.

Let $\mathbf{e}_i$ denote the ith unit coordinate vector. Then $\mathbf{e}_i^* A \mathbf{e}_1 = a_{ii}$, so $\mathscr{W}(A)$ contains the diagonal entries of A. More generally, if B is any principal submatrix of A, then $\mathscr{W}(B) \subseteq \mathscr{W}(A)$, as can be seen by restricting the vectors $\mathbf{x}$ in $\mathbf{x}^* A\mathbf{x}$ to have nonzero entries only in components corresponding to the rows and columns that specify the submatrix B. Now let α be an eigenvalue of A and apply a unitary similarity U so that $U^*AU = A_1$ is upper triangular with α in the 1,1 entry. If B_{1j} is the 2×2 principal submatrix of A_1 formed from rows and columns 1 and j, then $\mathscr{W}(B_{1j})$ is an ellipse with α as one of its foci and a minor axis of length equal to the modulus of the $1, j$ entry of A_1. This ellipse must be contained in $\mathscr{W}(A)$. Hence, if α is a boundary point of $\mathscr{W}(A)$, entries 2 through n of the first row of A_1 must be zero. Thus, if an eigenvalue α of A lies on the boundary of $\mathscr{W}(A)$, there is a unitary matrix U such that U^*AU is $\mathbf{D}(1, n-1)$, where the 1×1 block is the number α. This fact has been noted by Röseler [92], Kippenhahn [54], and Donoghue [23], and more recently in [27, 51, 66].

Suppose β is in $\mathscr{W}(A)$ and $\mathbf{x}^* A\mathbf{x} = \beta$. If U is a unitary matrix with the vector $\mathbf{x}$ in the first column, then β is the 1,1 entry of U^*AU. Since $\mathscr{W}(A)$ is a convex set and contains the diagonal entries of A, the number $\mathrm{tr}(A)/n$ is in $\mathscr{W}(A)$. Hence, A is unitarily similar to a matrix with the number $\mathrm{tr}(A)/n$

in the 1,1 position. One can now use an induction argument to show that A is unitarily similar to a matrix with every diagonal entry equal to $\mathrm{tr}(A)/n$, a result of Parker [79].

When $n = 2$, the boundary of the numerical range is an ellipse and thus is an algebraic curve of degree 2. For general n, the numerical range is also determined by an algebraic curve, as follows. Let $A = H + iK$ be an $n \times n$ matrix, where H and K are Hermitian; recall that H and K are uniquely determined. The set of all matrices of the form $xH + yK$, where x and y are variable, is called the *pencil generated by H and K*. The polynomial $f(x, y, z) = \det(zI - xH - yK)$ is the characteristic polynomial of the pencil $xH + yK$. This polynomial is homogeneous of degree n in the three variables x, y, z, and determines two different curves in the complex projective plane. If we view the triple (x, y, z) as representing a point in the projective plane, then the homogeneous equation $f(x, y, z) = 0$ represents a curve of degree n. However, if we regard the triple (x, y, z) as coordinates for a line in the projective plane, i.e., as line coordinates, then the solutions to $f(x, y, z) = 0$ represent a set of lines forming the envelope of an algebraic curve. The degree n of the line coordinate equation $f(x, y, z) = 0$ is called the *class* of this curve and is the number of tangent lines to the curve one can draw from any fixed point. Murnaghan [73] and Kippenhahn [54] showed that the lines of support of the numerical range of A satisfy the equation $f(-x, -y, z) = 0$, and the numerical range of A is the closed, convex hull of this curve of class n. Furthermore, the n eigenvalues of A correspond to the n real foci of this curve [54, 73, 94, 102]. Kippenhahn also showed that any singular point $\alpha = a + ib$ on this curve must be an eigenvalue of A; furthermore, there is a unitary matrix U such that U^*AU is $\mathbf{D}(1, n-1)$, where the 1×1 block is the number α. Donoghue [23] showed that if T is an operator of a Hilbert space into itself, and if $\mathscr{W}(T)$ is closed, then any nondifferentiable boundary point is an eigenvalue of T.

Degree two curves also have class two, so the ellipse obtained in the case $n = 2$ does have class two. Since $\mathscr{W}(A)$ determines A up to unitary similarity when $n = 2$, the polynomial $f(x, y, z)$ must also determine A up to unitary similarity when $n = 2$. In fact, for $n = 2$, the polynomial $\det(xH + yK)$ suffices to determine $A = H + iK$ up to unitary similarity [1, 116]. These results are not true for $n \geqslant 3$; see [101, 102] for examples of matrices that are not unitarily similar, but do have the same polynomial $f(x, y, z)$.

The numerical range has been generalized in many ways [8, 11, 34, 35, 36, 40, 49, 63, 64, 67, 77]. For example, there is the Bauer field of values [8, 77], the k-numerical range [40], the c-numerical range [36], the C-numerical range [36], the G-bilinear range [67]. We shall not discuss these here; see [35, 49] for a survey and references. However, we will discuss the matrix range introduced by Arveson [4, 2], who shows that an irreducible compact operator

T on a Hilbert space can be classified up to unitary similarity by a sequence of invariants, $\mathscr{W}_n(T)$, which are generalizations of the numerical range. The set $\mathscr{W}_n(T)$ is the set of $n \times n$ complex matrices obtained by applying certain completely positive linear maps to T.

Let $\mathscr{A}$ and $\mathscr{B}$ be C*-algebras. A linear map $\psi : \mathscr{A} \to \mathscr{B}$ is said to be *positive* if $\psi(A) \geqslant 0$ whenever $A \geqslant 0$. For example, if $\mathscr{A} = \mathbf{C}(m)$ and $\mathscr{B} = \mathbf{C}(n)$, and P is any $m \times n$ matrix, the map $\psi(A) = P^*AP$ is positive, for whenever A is a positive semidefinite Hermitian matrix, so is P^*AP. Stinespring [112] introduced the notion of a completely positive map. For any positive integer q, let $\mathscr{A}(q)$ be the C*-algebra of $q \times q$ matrices with entries from $\mathscr{A}$; one may regard $\mathscr{A}(q)$ as the tensor product $\mathbf{C}(q) \otimes \mathscr{A}$. We can extend the map ψ to a linear map $\psi_q : \mathscr{A}(q) \to \mathscr{B}(q)$ by applying ψ to each entry of a matrix in $\mathscr{A}(q)$.

DEFINITION 7.2 (Stinespring [112]). The positive map $\psi : \mathscr{A} \to \mathscr{B}$ is said to be *completely positive* if $\psi_q : \mathscr{A}(a) \to \mathscr{B}(q)$ is positive for each positive integer q.

When $\mathscr{B}$ is an algebra of operators on a Hilbert space, the following theorem characterizes completely positive maps.

THEOREM 7.1 (Stinespring [112]). *Let $\mathscr{A}$ be a* C*-*algebra with identity, let $\mathscr{H}$ be a Hilbert space, and let ψ be a linear map from $\mathscr{A}$ to operators on $\mathscr{H}$. Then ψ is completely positive if and only if ψ has the form $\psi(A) = V^*\rho(A)V$, where V is a bounded operator from $\mathscr{H}$ to a Hilbert space $\mathscr{K}$, and ρ is a *-representation of $\mathscr{A}$ into operators on $\mathscr{K}$.*

If $\mathscr{A} = \mathbf{C}(m)$, the algebra $\mathscr{A}(q)$ is the set of $qm \times qm$ complex matrices, viewed as the tensor product $\mathbf{C}(q) \otimes \mathbf{C}(m)$. Given a matrix M in $\mathbf{C}(qm)$, we partition M into q^2 blocks M_{jk}, each of size $m \times m$, and write $M = (M_{jk})$. If $\psi : \mathbf{C}(m) \to \mathbf{C}(n)$ is a linear map, then $\psi_q : \mathbf{C}(qm) \to \mathbf{C}(qn)$ is defined by $\psi_q(M) = (\psi(M_{jk}))$; note that $\psi(M_{jk})$ is in $\mathbf{C}(n)$, so $(\psi(M_{jk}))$ is a matrix of size $qn \times qn$. If $\psi(A) = P^*AP$ for some $m \times n$ matrix P, then $\psi(M_{jk}) = P^*M_{jk}P$, and we have $\psi_q(M) = (I \otimes P)^*M(I \otimes P)$. Thus, for any matrix P, the map $\psi(A) = P^*AP$ is completely positive. However, there are positive maps that are not completely positive [3, 112].

For any finite set of $m \times n$ matrices $P_1, P_2, \ldots, P_t$, the map $\psi : \mathbf{C}(m) \to \mathbf{C}(n)$ defined by $\psi(A) = \sum_{i=1}^{t} P_i^*AP_i$ is completely positive; Choi [16] shows that every completely positive map from $\mathbf{C}(m)$ to $\mathbf{C}(n)$ must have this form. The set of completely positive maps from $\mathbf{C}(m)$ to $\mathbf{C}(n)$ is thus the positive cone generated by the maps $\psi(A) = P^*AP$. We can rewrite the map $\psi(A) =$

$\Sigma_{i=1}^{t} P_i^* A P_i$ in the form given by Theorem 7.1 as follows. Define $\rho: \mathbf{C}(m) \to \mathbf{C}(mt)$ by putting $\rho(A) = I_t \otimes A$; the matrix $\rho(A)$ is then the block diagonal matrix formed from the direct sum of t copies of A. Let $P = (P_1 \ P_2 \ \cdots \ P_t)^T$ be the $tm \times n$ matrix formed by stacking the t matrices P_i into a column. Then $P^*(I_t \otimes A)P = P^* \rho(A) P = \psi(A)$.

We can now define the set $\mathscr{W}_n(T)$, where T is an operator on a Hilbert space. Let $\mathbf{C}^*(T)$ denote the $\mathbf{C}^*$-algebra generated by T and the identity.

Definition 7.3 (Arveson, [2, 4]). Let n be a positive integer, and let T be an operator on a Hilbert space $\mathscr{H}$. Then $\mathscr{W}_n(T)$ is the set of all $n \times n$ matrices of the form $\psi(T)$, where ψ ranges over all completely positive maps of $\mathbf{C}^*(T)$ into $\mathbf{C}(n)$ that preserve the identity.

If $n = 1$, the set $\mathscr{W}_1(T)$ is the closure of the ordinary numerical range $\mathscr{W}(T)$ [4]. The sequence $\{\mathscr{W}_1(T), \mathscr{W}_2(T), \dots\}$ is called the *matrix range of* T. When $\mathscr{H} = \mathbf{C}(m)$, Choi's characterization of completely positive maps from $\mathbf{C}(m)$ to $\mathbf{C}(n)$ gives

$$\mathscr{W}_n(T) = \left\{ \sum_{i=1}^{t} P_i^* T P_i \,\middle|\, P_1, P_2, \dots, P_t \text{ are } m \times n \text{ matrices such that } \sum_{i=1}^{t} P_i^* P_i = I_n \right\}.$$

The matrix range gives a complete set of unitary invariants for irreducible compact operators.

Theorem 7.4 (Arveson [2, 4]). *Let S and T be irreducible compact operators on the Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$, respectively. Then S and T are unitarily equivalent if and only if $\mathscr{W}_n(S) = \mathscr{W}_n(T)$ for every positive integer n. If $\mathscr{H}$ and $\mathscr{K}$ are finite dimensional, with dimension n, then S and T are unitarily equivalent if and only if $\mathscr{W}_n(S) = \mathscr{W}_n(T)$.*

Thus, $n \times n$ matrices A and B are unitarily similar if and only if $\mathscr{W}_n(A) = \mathscr{W}_n(B)$.

8. UNITARY REDUCIBILITY

Since smaller matrices are easier to work with, it is often useful to reduce a matrix to block diagonal form. In this section we discuss some results about reducing a matrix to block diagonal form with a unitary similarity.

DEFINITION 8.1. We say an $n \times n$ matrix A is *unitarily reducible* if there exists an unitary matrix U such that U^*AU is block diagonal. A matrix that is not unitarily reducible is said to be *unitarily irreducible*.

The term *unitarily decomposable* is sometimes used for the concept we have called *unitarily reducible*.

Definition 8.1 can be extended to sets of matrices.

DEFINITION 8.2. We say a set Ω of $n \times n$ matrices is $\mathbf{T}(n_1, n_2, \dots, n_t)$ if every matrix in Ω is $\mathbf{T}(n_1, n_2, \dots, n_t)$. We say Ω is $\mathbf{D}(n_1, n_2, \dots, n_t)$ if every matrix in Ω is $\mathbf{D}(n_1, n_2, \dots, n_t)$. If there is a unitary matrix U and an integer $t > 1$ such that $U^*\Omega U$ is $\mathbf{D}(n_1, n_2, \dots, n_t)$, then we say Ω is *unitarily reducible*.

Let $\mathscr{A}(A, A^*)$ be the algebra generated by A and A^* over $\mathbf{C}$. Since $(U^*AU)^* = U^*A^*U$, the algebra $\mathscr{A}(A, A^*)$ is unitarily reducible if and only if A is unitarily reducible. Put $H = (A + A^*)/2$ and $K = (A - A^*)/2i$. Then H and K are the unique Hermitian matrices such that $A = H + iK$, and H and K generate the same algebra as A and A^*.

THEOREM 8.1. *Let A be an $n \times n$ matrix. The following are equivalent:*

(1) *A is unitarily reducible to a matrix that is $\mathbf{D}(k, n-k)$, where $0 < k < n$.*

(2) *The algebra $\mathscr{A}(A, A^*)$ is unitarily reducible to an algebra of matrices that is $\mathbf{D}(k, n-k)$.*

(3) *There is a k-dimensional subspace $\mathscr{U}$ of $\mathbf{C}^n$ such that both $\mathscr{U}$ and $\mathscr{U}^\perp$ are invariant under A, where $\mathscr{U}^\perp$ is the orthogonal complement of $\mathscr{U}$.*

(4) *There is a k-dimensional subspace $\mathscr{U}$ of $\mathbf{C}^n$ such that $\mathscr{U}$ is invariant under both A and A^*.*

(5) *There is a k-dimensional subspace $\mathscr{U}$ of $\mathbf{C}^n$ such that $\mathscr{U}$ is invariant under every matrix in $\mathscr{A}(A, A^*)$.*

(6) *There is a k-dimensional subspace $\mathscr{U}$ of $\mathbf{C}^n$ such that $\mathscr{U}$ is invariant under both H and K, where $H = (A + A^*)/2$ and $K = (A - A^*)/2i$.*

(7) *There is a k-dimensional subspace $\mathscr{U}$ of $\mathbf{C}^n$ such that $\mathscr{U}$ is invariant under every matrix in $\mathscr{A}(H, K)$.*

The equivalence of (1) and (3) follows by letting U be a unitary matrix in which the first k columns form an orthonormal basis for $\mathscr{U}$ and the remaining $n-k$ columns form an orthonormal basis for $\mathscr{U}^{\perp}$, and then forming U^*AU. The equivalence of (3) and (4) comes from using the fact that $\langle \mathbf{x}, A\mathbf{y}\rangle = \langle A^*\mathbf{x}, \mathbf{y}\rangle$ to show that $\mathscr{U}$ is invariant under A if and only if $\mathscr{U}^{\perp}$ is invariant under A^*. The remaining conditions follow from the fact that $\mathscr{U}$ is invariant under both A and A^* if and only if $\mathscr{U}$ is invariant under $\mathscr{A}(A, A^*)$, and $\mathscr{U}$ is invariant under both A and A^* if and only if $\mathscr{U}$ is invariant under both H and K. Theorem 8.2 essentially restates results of Specht [109]; see also [100, 57].

The algebra $\mathscr{A}(A, A^*)$ is equal to all of $\mathbf{C}(n)$ if and only if A is unitarily irreducible; this follows from condition (5) of Theorem 8.1 and the following theorem of Burnside [15, 50].

THEOREM 8.2 (Burnside [15]). *Let $\mathscr{A}$ be an algebra of $n \times n$ complex matrices, and suppose that no nontrivial subspace of $\mathbf{C}^n$ is invariant under $\mathscr{A}$. Then $\mathscr{A} = \mathbf{C}(n)$.*

Suppose $U^*AU = \mathbf{D}(A_1, A_2, \ldots, A_t)$, where each A_i is $n_i \times n_i$ and is unitarily irreducible. Using Theorem 8.1, we have an n_i-dimensional subspace $\mathscr{U}_i$ corresponding to the ith block; the subspaces $\mathscr{U}_i$ are mutually orthogonal and $\mathbf{C}^n$ is the direct sum of the $\mathscr{U}_i$'s. If $n_i = 1$, then A and A^* have a common eigenvector that generates $\mathscr{U}_i$. Each n_i is one if and only if A is normal; in this case $\mathscr{A}(A, A^*)$ is a commutative algebra. The number t and the numbers n_i are uniquely determined by the matrix A, but may not be easy to determine. A theorem of Schur [96, 108] relates the number t to the number of distinct eigenvalues in the matrices that commute with $\mathscr{A}(A, A^*)$.

THEOREM 8.3 (Schur [96]). *Let A be an $n \times n$ complex matrix, and suppose $U^*AU = \mathbf{D}(A_1, A_2, \ldots, A_t)$, where each A_i is $n_i \times n_i$ and is unitarily irreducible. Let $\mathscr{C}(A, A^*)$ be the set of matrices that commute with every matrix in the algebra $\mathscr{A}(A, A^*)$, and let r be the largest number of distinct eigenvalues of any matrix in $\mathscr{C}(A, A^*)$. Then $r = t$.*

Proof. Since $(U^*AU)^* = U^*A^*U$, we have $U^*\mathscr{A}(A, A^*)U = \mathscr{A}(U^*AU, U^*A^*U)$ and $U^*\mathscr{C}(A, A^*)U = \mathscr{C}(U^*AU, U^*A^*U)$, so we may assume that A itself is $\mathbf{D}(A_1, A_2, \ldots, A_t)$. Clearly, any matrix that is $\mathbf{D}(n_1, n_2, \ldots, n_t)$ and has scalar matrices in each diagonal block will commute with $\mathscr{A}(A, A^*)$, so $t \leqslant r$.

Now suppose that B is in $\mathscr{C}(A, A^*)$ and that B has r distinct eigenvalues $\beta_1, \beta_2, \ldots, \beta_r$ of multiplicities $m_1, m_2, \ldots, m_r$. We can find a unitary matrix U

such that $U^*BU = \mathbf{T}(B_1, B_2, \ldots, B_r)$, where B_i is $m_i \times m_i$ and β_i is the only eigenvalue of B_i. Transforming A and A^* by this same unitary similarity, U, we then deal with U^*BU and $\mathscr{A}(U^*AU, U^*A^*U)$. For notational convenience, assume B is already in the form $\mathbf{T}(B_1, B_2, \ldots, B_r)$. Let M be any matrix in $\mathscr{A}(A, A^*)$; then $BM = MB$. Partition M into blocks conformal with the block structure of B. The argument used to prove Theorem 3.5 then shows that M is $\mathbf{T}(m_1, m_2, \ldots, m_r)$. Hence $\mathscr{A}(A, A^*)$ is $\mathbf{T}(m_1, m_2, \ldots, m_r)$. But then $\mathscr{A}(A, A^*)$ must be $\mathbf{D}(m_1, m_2, \ldots, m_r)$, for whenever M is in $\mathscr{A}(A, A^*)$, then M^* is also in $\mathscr{A}(A, A^*)$. Therefore, $r \leqslant t$. Since we have already seen that $t \leqslant r$, we have $t = r$. ■

Theorem 8.3 establishes a connection between the decomposition of the algebra $\mathscr{A}(A, A^*)$ into a direct sum of irreducible components and the commutant $\mathscr{C}(A, A^*)$. An analogous result holds for strongly closed, self-adjoint algebras of Hilbert space operators–such an algebra is decomposable if and only if its center contains an operator that is not a scalar multiple of the identity. A W^*-algebra whose center consists of scalar multiples of the identity is called a factor; a general W^*-algebra can be regarded as a direct integral of such factors [98].

Theorem 8.3 concerns the number of blocks in the decomposition $U^*AU = \mathbf{D}(A_1, A_2, \ldots, A_t)$. Results linking the sizes of the blocks to polynomial identities for matrix algebras may be found in [7, 55, 100, 127]. Some special results apply when the blocks are of size at most two or three. Two $n \times n$ matrices that satisfy quadratic polynomials generate an algebra of dimension at most $2n$ [31]. Applying this to the pair of matrices A and A^* shows that if the minimal polynomial of A has degree two, then there is a unitary matrix U such that U^*AU is block diagonal with blocks of size one or two. Each of the blocks must satisfy that same polynomial of degree two, so one can say more about the structure of the blocks.

Let $A = H + iK$, where H and K are Hermitian, and let $f(x, y, z)$ be the characteristic polynomial of the pencil $xH + yK$. Motzkin and Taussky [70, 71] studied the characteristic polynomial $f(x, y, z)$ in their work on matrices with property L.

DEFINITION 8.3. Let A and B be $n \times n$ matrices. If there are orderings $\alpha_1, \alpha_2, \ldots, \alpha_n$ of the eigenvalues of A and $\beta_1, \beta_2, \ldots, \beta_n$ of the eigenvalues of B such that $xA + yB$ has eigenvalues $x\alpha_i + y\beta_i$ for all values of x and y, then A and B are said to have *property L*.

Let $h_1, h_2, \ldots, h_n$ be the eigenvalues of H, and let $k_1, k_2, \ldots, k_n$ be the eigenvalues of K. The polynomial $f(x, y, z)$ factors into the n linear factors $(z - h_i y - k_i z)$ if and only if the eigenvalues of $aH + bK$ are the n numbers $ah_i + bk_i$ for every choice of the complex coefficients a and b, that is, if and only if H and K have property L. Motzkin and Taussky [70] proved that a pair of Hermitian matrices have property L if and only if they commute. This result also holds for pairs of normal matrices [133], but does not hold for general pairs of matrices. Observe that we can then simultaneously diagonalize H and K, so the matrix A is normal and can be unitarily diagonalized. More generally, if A is unitarily reducible to a matrix that is $\mathbf{D}(n_1, n_2, \dots, n_t)$, then $f(x, y, z)$ factors into t factors $f_j(x, y, z)$, where $f_j(x, y, z)$ has degree n_j and corresponds to the jth diagonal block. More precisely, if $A_j = H_j + iK_j$, then $f_j(x, y, z) = \det(zI - xH_j - yK_j)$. However, the converse is not true, for A can be unitarily irreducible even if $f(x, y, z)$ factors [101]. Kippenhahn [54] showed that if the minimal polynomial of $xH + yK$ has degree two, then A is unitarily reducible to a block diagonal matrix with blocks of size at most two; this also follows from the previous paragraph. He conjectured that whenever the minimal polynomial has degree less than n, then A will be unitarily reducible. This conjecture is correct for $n \leqslant 5$, but is false in general, as shown by Laffey [56] and Waterhouse [126]. However, if the minimal polynomial of $xH + yK$ has degree three, then A is unitarily reducible to a block diagonal matrix with blocks of size at most three [103].

Other results on the unitary equivalence of sets of Hermitian matrices, or orthogonal equivalence of sets of real symmetric matrices, may be found in [1, 38, 44, 116]; see [44, 45] for work on simultaneous reduction of sets of matrices by unitary congruence.

Radjavi and Rosenthal [91] showed that every nonscalar operator on a separable complex Hilbert space has a matrix representation with no zero entries. Thus, any nonscalar $n \times n$ complex matrix is unitarily similar to a matrix with no zero entries. This result is used to show that if A is any nonscalar operator, then there is an operator B such that A and B have no common invariant subspace [91].

There are several sets of inequalities comparing the eigenvalues, singular values, and diagonal elements of a matrix; there are also inequalities linking the eigenvalues of the Hermitian and skew Hermitian parts of a matrix to the real and imaginary parts of its eigenvalues. The Weyl [130] inequalities apply to the eigenvalues and singular values, R. C. Thompson [118] has established inequalities for singular values and diagonal entries, and S. Sherman and C. J. Thompson [105] have inequalities comparing the eigenvalues of the skew Hermitian part of a matrix with the imaginary parts of the eigenvalues. In many cases, these inequalities give necessary and sufficient conditions for the existence of a matrix with prescribed singular values, eigenvalues, diagonal entries, etc. [47, 118]. C.-K. Li [58] has studied matrices for which these inequalities become equalities and shown that in many cases such matrices must be unitarily reducible.

We conclude this section with a theorem of McRae [62]. Let A have t distinct eigenvalues $\alpha_1, \alpha_2, \dots, \alpha_t$ of multiplicities $n_1, n_2, \dots, n_t$. Then A^* has

t distinct eigenvalues $\bar{\alpha}_i$ of multiplicities n_i for $i=1,\dots,t$. Let $\mathscr{U}_i$ be the eigenspace of A corresponding to α_i, so $\mathscr{U}_i$ is the null space of $A-\alpha_i I$. Let $\mathscr{V}_i$ be the eigenspace of A^* corresponding to $\bar{\alpha}_i$. Then $\mathscr{U}_i=\mathscr{V}_i$ for every i if and only if A is normal. In general, while $\mathscr{U}_i$ and $\mathscr{V}_i$ must have the same dimension, they are different subspaces. Let $\mathscr{W}_i=\mathscr{U}_i\cap\mathscr{V}_i$, and let k_i be the dimension of $\mathscr{W}_i$. Then $\mathscr{W}_i$ is invariant under both A and A^*, and if we use $\mathscr{W}_i$ as the subspace $\mathscr{U}$ of Theorem 8.1, the corresponding $k_i\times k_i$ blocks of A and A^* will be the scalar blocks $\alpha_i I_{k_i}$ and $\bar{\alpha}_i I_{k_i}$, respectively. This leads to the following decomposition.

THEOREM 8.5 (McRae [62]). *Let A be an $n\times n$ complex matrix with t distinct eigenvalues $\alpha_1,\alpha_2,\dots,\alpha_t$ of multiplicites $n_1,n_2,\dots,n_t$. Let $\mathscr{U}_i$ be the eigenspace of A corresponding to α_i, let $\mathscr{V}_i$ be the eigenspace of A^* corresponding to $\bar{\alpha}_i$, and let k_i be the dimension of $\mathscr{U}_i\cap\mathscr{V}_i$. Then there is a unitary matrix U such that $U^*AU=\mathbf{D}(D,C)$, where D is a diagonal matrix of size $k=k_1+k_2+\cdots+k_t$ in which α_i appears k_i times on the diagonal. Furthermore, if $V^*AV=\mathbf{D}(D_1,C_1)$ is any other decomposition of this type, with D_1 diagonal, then D_1 has size at most k, and if D_1 is $k\times k$, then D_1 can differ from D only in the order of the diagonal entries, and $C\sim C_1$.*

Note that A is normal if and only if $n=k_1+k_2+\cdots+k_t$; one may think of Theorem 8.5 as telling us how to split off the "normal part" of A.

I would like to thank Roger Horn for encouraging me to do this survey and for his many helpful suggestions. Thanks also to Shmuel Friedland, who drew my attention to Reference [9], and Bryan Cain, who informed me of References [2–4]. I first became interested in this subject as a graduate student, and am grateful to Olga Taussky Todd, my thesis advisor, for her invaluable guidance. Many thanks to the interlibrary loan office of McCabe Library, Swarthmore College. Most of all, thanks to Emi Horikawa and Meg Spencer of Cornell Science Library, Swarthmore College, for their generous and expert assistance with the literature search.

REFERENCES

1 A. A. Albert, On the orthogonal equivalence of sets of real symmetric matrices, *J. Math. and Mech.* 7:219–235 (1958).
2 W. B. Arveson, Unitary invariants for compact operators, *Bull. Amer. Math. Soc.* 76:88–91 (1970).
3 W. B. Arveson, Subalgebras of C^*-algebras, *Acta Math.* 123:141–224 (1969).
4 W. B. Arveson, Subalgebras of C^*-algebras II, *Acta Math.* 128:271–308 (1972).
5 W. B. Arveson, *An Invitation to C^*-Algebras*, Springer-Verlag, New York, 1976.

6 L. Autonne Sur les groupes linéaires, réels et orthogonaux, *Bull. Soc. Math. France* 30:121–134 (1902).
7 G. P. Barker, L. Q. Eifler, and T. P. Kezlan, A non-commutative spectral theorem, *Linear Algebra Appl.* 20:95–100 (1978).
8 F. L. Bauer, On the field of values subordinate to a norm, *Numer. Math.* 4:103–113 (1962).
9 R. Benedetti and P. Cragnolini, Versal families of matrices with respect to unitary conjugation, *Adv. in Math.* 54:314–335 (1984).
10 R. Bhattacharya, On the Unitary Invariants of an $n \times n$ matrix, Ph.D. Thesis, Indian Statistical Inst., New Delhi, 1987.
11 F. F. Bonsall and J. Duncan, *Numerical Ranges of Operators on Normed Spaces and of Elements of Normed Algebras*, London Math. Soc. Lecture Note Ser. 2, Cambridge U.P., London, 1971.
12 J. Brenner, The problem of unitary equivalence, *Acta. Math.* 86:297–308 (1951).
13 A. Brown, The unitary equivalence of binormal operators, *Amer. J. Math.* 76:144–434 (1954).
14 E. T. Browne, The characteristic roots of a matrix, *Bull. Amer. Math. Soc.* 36:705–710 (1930).
15 W. Burnside, On the condition of reducibility of any group of linear substitutions, *Proc. Lond. Math. Soc.* (2) 3:430–434 (1905).
16 M.-D. Choi, Completely positive linear maps on complex matrices, *Linear Algebra Appl.* 10:285–290 (1975).
17 J. L. Coolidge, *A Treatise on Algebraic Plane Curves*, Oxford U.P., Oxford, 1931.
18 J. C. Currie, Unitary-canonical matrices, Abstract 264, *Bull. Amer. Math. Soc.* 56:321 (1950).
19 D. Deckard, Complete sets of unitary invariants for compact and trace-class operators, *Acta Sci. Math. (Szeged)* 28:9–20 (1967).
20 D. Deckard and C. Pearcy, On matrices over the ring of continuous complex valued functions on a Stonian space, *Proc. Amer. Math. Soc.* 14:322–328 (1963).
21 D. Deckard and C. Pearcy, On continuous matrix-valued functions on a Stonian space, *Pacific J. Math.* 14:857–869 (1964).
22 D. Deckard and C. Pearcy, On unitary equivalence of Hilbert-Schmidt operators, *Proc. Amer. Math. Soc.* 16:671–675 (1965).
23 W. F. Donoghue, Jr., On the numerical range of a bounded operator, *Mich. Math. J.* 4:261–263 (1957).
24 M. P. Drazin, J. W. Dungey, and K. W. Gruenberg, Some theorems on commutative matrices, *J. London Math. Soc.* 26:221–228 (1951).
25 C. Eckart and G. Young, A principal axis transformation for non-Hermitian matrices, *Bull. Amer. Math. Soc.* 45:118–121 (1939).
26 J. Ernest, Charting the operator terrain, *Mem. Amer. Math. Soc.* 6, No. 171 (1976).
27 K. Fan, A remark on orthogonality of eigenvectors, *Linear and Multilinear Algebra* 23:283–284 (1988).

28 M. Fiedler, Geometry of the numerical range of matrices, *Linear Algebra Appl.* 37:81–96 (1981).

29 S. Friedland, A generalization of the Motzkin-Taussky theorem, *Linear Algebra Appl.* 36:103–109 (1981).

30 S. Friedland, Simultaneous similarity of matrices, *Adv. in Math.* 50:189–265 (1983).

31 F. J. Gaines, T. J. Laffey, and H. M. Shapiro, Pairs of matrices with quadratic minimal polynomials, *Linear Algebra Appl.* 52/53:289–292 (1983).

32 P. X. Gallagher and R. J. Proulx, Orthogonal and unitary invariants of families of subspaces, in *Contributions to Algebra, a Collection of Papers Dedicated to Ellis Kolchin* (Hyman Bass, Phyllis J. Cassidy, and Jerald Kovacic, Eds.) Academic, New York, 1977, pp. 157–164.

33 F. R. Gantmacher, *The Theory of Matrices*, Vols. 1, 2, Chelsea, New York, 1959.

34 W. Givens, Fields of values of a matrix, *Proc. Amer. Math. Soc.* 3:206–209 (1953).

35 M. Goldberg, On certain finite dimensional numerical ranges and numerical radii, *Linear and Multilinear Algebra* 7:329–342 (1979).

36 M. Goldberg and E. G. Straus, Elementary inclusion relations for generalized numerical ranges, *Linear Algebra Appl.* 18:1–24 (1977).

37 J. P. Gram, Ueber die Entwickelung reeler Functionen in Reihen mittelst der Methode der kleinsten Quadrate, *Crelles J.* 94:41–73 (1883).

38 F. A. Grunbaum, Simultaneous unitary equivalence, *Linear and Multilinear Algebra* 4:79–84 (1976).

39 P. Halmos, *Finite-Dimensional Vector Spaces*, Van Nostrand, Princeton, 1958.

40 P. Halmos, *A Hilbert Space Problem Book*, Van Nostrand, Princeton, 1967.

41 Robert E. Hartwig, The resultant and the matrix equation $AX = SB$, *SIAM J. Appl. Math.* 22:538–544 (1972).

42 F. Hausdorff, Der Wertvorrat einer Bilinearform, *Math. Z.* 3:314–316 (1919).

43 F. Herbut, P. Loncke, and M. Vujicic, Canonical form for matrices under unitary congruence transformations II: Congruence-normal matrices, *SIAM J. Appl. Math.* 26:794–805 (1974).

44 Y. Hong and R. Horn, On simultaneous reduction of families of matrices to triangular or diagonal form by unitary congruences, *Linear and Multilinear Algebra* 17:271–288 (1985).

45 Y. Hong, R. Horn, and C. R. Johnson, On the reduction of pairs of Hermitian or symmetric matrices to diagonal form by congruence, *Linear Algebra Appl.* 73:213–226 (1986).

46 Y. Hong and R. Horn, A characterization of unitary congruence, *Linear and Multilinear Algebra* 25:105–119 (1989).

47 A. Horn, On the eigenvalues of a matrix with prescribed singular values, *Proc. Amer. Math. Soc.* 5:4–7 (1954).

48 R. Horn and C. Johnson, *Matrix Analysis*, Cambridge U.P., Cambridge, 1985.

49 R. Horn and C. Johnson, *Topics in Matrix Analysis*, Cambridge U.P., Cambridge, 1990.

50 N. Jacobson, *Lectures in Abstract Algebra, Vol. II, Linear Algebra*, Van Nostrand, New York, 1953.

51 C. R. Johnson, Normality and the numerical range, *Linear Algebra Appl.* 15:89–94 (1976).
52 L. A. Kaluznin and H. Havidi, Geometric theory of the unitary equivalence of matrices, *Soviet Math. Dokl.* 7:1044–1047 (1966) [English transl. of *Dokl. Akad. Nauk SSSR* 169:1009–1012 (1966)].
53 I. Kaplansky, Algebras of type I, *Ann. of Math.* 56:460–472 (1952).
54 R. Kippenhahn, Über den Wertevorrat einer Matrix, *Math. Nachr.* 6:193–228 (1951–52).
55 T. J. Laffey, Simultaneous quasidiagonalization of complex matrices, *Linear Algebra Appl.* 16:189–201 (1977).
56 T. J. Laffey, A counterexample to Kippenhahn's conjecture on Hermitian pencils, *Linear Algebra Appl.* 51:179–182 (1983).
57 T. J. Laffey, Simultaneous reduction of sets of matrices under similarity, *Linear Algebra Appl.* 84:123–138 (1986).
58 C.-K. Li, Matrices with some extremal properties, *Linear Algebra Appl.* 101:255–267 (1988).
59 D. E. Littlewood, On unitary equivalence, *J. London Math. Soc.* 28:314–322 (1953).
60 N. H. McCoy, On the characteristic roots of matric polynomials, *Bull. Amer. Math. Soc.* 42:592–600 (1936).
61 C. C. MacDuffee, *The Theory of Matrices*, Springer-Verlag, Berlin, 1933.
62 V. V. McRae, On the Unitary Similarity of Matrices, Ph.D. Dissertation, Catholic Univ. of America Press, Washington, 1955.
63 M. Marcus and P. Andresen, Constrained extrema of bilinear functionals, *Monatsh. Math.* 84:219–235 (1977).
64 M. Marcus and I. Filippenko, Nondifferentiable boundary points of the higher numerical range, *Linear Algebra Appl.* 21:217–232 (1978).
65 M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*, Prindle, Weber & Schmidt, Boston, 1964.
66 M. Marcus and C. Pesce, Computer generated numerical ranges and some resulting theorems, *Linear and Multilinear Algebra* 20:121–157 (1987).
67 M. Marcus and M. Sandy, The *G*-bilinear range, *Linear and Multilinear Algebra* 11:317–322 (1982).
68 B. E. Mitchell, Normal and diagonalizable matrices, *Amer. Math Monthly*, 60:94–96 (1953).
69 B. E. Mitchell, Unitary transformations, *Canad. J. Math* 6:69–72 (1954).
70 T. S. Motzkin and O. Taussky, Pairs of matrices with property L, *Trans. Amer. Math. Soc.* 73:108–114 (1952).
71 T. S. Motzkin and O. Taussky, Pairs of matrices with property L, II, *Trans. Amer. Math. Soc.* 80:387–401 (1955).
72 B. N. Moyls and M. D. Marcus, Field convexity of a square matrix, *Proc. Amer. Math. Soc.* 6:981–983 (1955).
73 F. D. Murnaghan, On the field of values of a square matrix, *Proc. Nat. Acad. Sci. U.S.A.* 18:246–248 (1932).
74 F. D. Murnaghan, On the unitary invariants of a square matrix, *Proc. Nat. Acad. Sci. U.S.A.* 18:185–189 (1932).

75 F. D. Murnaghan and A. Wintner, A canonical form for real matrices under orthogonal transformations, *Proc. Nat. Acad. Sci. U.S.A.* 17:417–420 (1931).
76 H. Neudecker, A note on Kronecker matrix products and matrix equation systems, *SIAM J. Appl. Math.* 17:603–606 (1969).
77 N. Nirschl and H. Schneider, The Bauer fields of values of a matrix, *Numer. Math.* 6:355–365 (1964).
78 W. V. Parker, Characteristic roots and the field of values of a matrix, *Duke Math. J.* 15:439–442 (1948).
79 W. V. Parker, Sets of complex numbers associated with a matrix, *Duke Math. J.* 15:711–715 (1948).
80 W. V. Parker, A note on normal matrices, *Amer. Math Monthly*, 61:330–331 (1954).
81 V. Paulsen, Continuous canonical forms for matrices under unitary equivalence, *Pacific J. Math.* 76:129–142 (1978).
82 A. Paz, An application of the Cayley-Hamilton theorem to matrix polynomials in several variables, *Linear and Multilinear Algebra* 15:161–170 (1984).
83 C. Pearcy, A complete set of unitary invariants for operators generating finite W^*-algebras of type I, *Pacific J. Math.* 12:1405–1416 (1962).
84 C. Pearcy, A complete set of unitary invariants for 3×3 complex matrices, *Trans. Amer. Math. Soc.* 104:425–429 (1962).
85 C. Pearcy, On unitary equivalence of matrices over the ring of continuous complex-valued functions on a Stonian space, *Canad. J. Math.* 15:323–331 (1963).
86 C. Pearcy and J. R. Ringrose, Trace-preserving isomorphisms in finite operator algebras, *American J. Math.* 90:444–455 (1968).
87 E. Primrose, *Plane Algebraic Curves*, Macmillan, London, 1955.
88 C. Procesi, The invariant theory of $n\times n$ matrices, *Adv. in Math.* 19:306–381 (1976).
89 H. Radjavi, On unitary equivalence of arbitrary matrices, *Trans. Amer. Math. Soc.* 104:363–373 (1962).
90 H. Radjavi, Simultaneous unitary invariants for sets of matrices, *Canad. J. Math.* 20:1012–1019 (1968).
91 H. Radjavi and P. Rosenthal, Matrices for operators and generators of $B(H)$, *J. London Math. Soc.* (2) 2:557–560 (1970).
92 H. Röseler, Normalformen von Matrizen gegenüber unitären Transformationen, Dissertation, Universitätsverlag von Robert Noske in Borna-Leipzig, Leipzig, 1933.
93 J. Rotman, *The Theory of Groups*, Allyn and Bacon, Boston, 1973.
94 G. Salmon, *A Treatise on the Higher Plane Curves*, Hodges, Foster, and Figgis, Dublin, 1879.
95 E. Schmidt, Zur Theorie der linearen und nichtlinearen Intergralgleichungen, *Math. Ann.* 63:433–476 (1907).
96 I. Schur, Beiträge zur Theorie der Gruppen linearer homogener Substitutionen, *Trans. Amer. Math. Soc.* 10:159–175 (1909).
97 I. Schur, Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen, *Math. Ann.* 66:488–510 (1909).

98 J. T. Schwartz, *W-* Algebras*, Gordon and Breach, New York, 1967.

99 V. V. Sergeĭchuk, Classification of linear operators in a finite-dimensional unitary space, *Functional Anal. Appl.* 18(3):224–230 (1984).

100 H. Shapiro, Simultaneous block triangularization and block diagonalization of sets of matrices, *Linear Algebra Appl.* 25:129–137 (1979).

101 H. Shapiro, On a conjecture of Kippenhahn about the characteristic polynomial of a pencil generated by two Hermitian matrices. I, *Linear Algebra Appl.* 43:201–221 (1982).

102 H. Shapiro, A conjecture of Kippenhahn about the characteristic polynomial of a pencil generated by two Hermitian matrices. II, *Linear Algebra Appl.* 45:97–108 (1982).

103 H. Shapiro, Hermitian pencils with a cubic minimal polynomial, *Linear Algebra Appl.* 48:81–103 (1982).

104 H. Shapiro and O. Taussky, Alternative proofs of a theorem of Moyls and Marcus on the numerical range of a square matrix, *Linear and Multilinear Algebra* 8:337–340 (1980).

105 S. Sherman and C. J. Thompson, Equivalence on eigenvalues, *Indiana Univ. Math. J.* 21:807–814 (1972).

106 K. S. Sibirskiĭ, Unitary and orthogonal invariants of matrices, *Soviet Math Dokl.* 8:36–40 (1967) [English transl. of *Dokl. Akad. Nauk SSSR* 172:40–43 (1967)].

107 R. R. Smith and J. D. Ward, Matrix ranges for Hilbert space operators, *Amer. J. Math.* 102:1031–1081 (1980).

108 W. Specht, Zur Theorie der Gruppen linearer Substitutionen, *Jahresber. Deutsch. Math.-Verein.* 47:43–55 (1937).

109 W. Specht, Zur Theorie der Gruppen linearer Substitutionen, II, *Jahresber. Deutsch. Math.-Verein.* 49:207–215 (1939).

110 W. Specht, Zur Theorie der Matrizen, II, *Jahresber. Deutsch. Math.-Verein.* 50:19–23 (1940).

111 A. J. M. Spencer, Theory of invariants, in *Continuum Physics* (A. Cemal Eringen, Ed.), Academic, New York, 1971, pp. 239–353.

112 W. F. Stinespring, Positive functions on C^*-algebras, *Proc. Amer. Math. Soc.*, 6:211–216 (1955).

113 J. J. Sylvester, Sur l'equation en matrices $px = xq$, *C. R. Acad. Sci. Paris* 99:67–71, 115–116 (1884).

114 O. Taussky, Commutativity in finite matrices, *Amer. Math. Monthly* 64:229–235 (1957).

115 O. Taussky, A weak property L for pairs of matrices, *Math. Z.* 71:463–465 (1959).

116 O. Taussky, On the congruence transformation of a pencil of real symmetric matrices to a pencil with identical characteristic polynomial, *Linear Algebra Appl.* 52/53:687–691 (1983).

117 O. Taussky and H. Wielandt, Linear relations between higher additive commutators, *Proc. Amer. Math. Soc.* 13:732–735 (1962).

118 R. C. Thompson, Singular values, diagonal elements and convexity, *SIAM J. Appl. Math.* 32:39–63 (1977).

119 O. Toeplitz, Das algebraische Analogon zu einem Satze von Fejer, *Math. Z.* 2:187–197 (1918).
120 N.-K. Tsing, The constrained bilinear form and the C-numerical range, *Linear Algebra Appl.* 56:195–206 (1984).
121 H. W. Turnbull and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices*, Blackie, London, 1932.
122 P. M. Van Dooren, On the use of unitary state-space transformations, In *Linear Algebra and Its Role in Systems Theory*, Proceedings of the AMS-IMS-SIAM Conference, 29 July–4 Aug. 1984, (R. A. Brualdi et al., Eds.), Amer. Math. Soc., Providence, 1985, pp. 447–463.
123 K. Veselic, On the Jordan form in a unitary space, *Linear Algebra Appl.* 8:507–514 (1974).
124 M. Vujičić, F. Herbut, and G. Vujičić, Canonical form for matrices under unitary congruence transformations I: Conjugate-normal matrices, *SIAM J. Appl. Math.* 23:225–238 (1972).
125 G. E. Wall, On the conjugacy classes in the unitary, symplectic, and orthogonal groups, *J. Austral. Math. Soc.* 3:1–62 (1963).
126 William C. Waterhouse, A conjectured property of Hermitian pencils, *Linear Algebra Appl.* 51:173–177 (1983).
127 J. F. Watters, Simultaneous quasi-diagonalization of normal matrices, *Linear Algebra Appl.* 9:103–117 (1974).
128 U. Wegner, Über die Frobeniusschen Kovarianten, *Monatsh. Math. u. Phys.* 40:201–208 (1933).
129 R. Westwick, A theorem on numerical range, *Linear and Multilinear Algebra* 2:311–315 (1975).
130 H. Weyl, Inequalities between the two kinds of eigenvalues of a linear transformation, *Proc. Nat. Acad. Sci. U.S.A.* 35:408–411 (1949).
131 E. Weyr, Répartition des matrices en espèces et formation de toutes les espèces, *C. R. Acad. Sci. Paris* 100:966–969 (1885).
132 E. Weyr, Zur Theorie der bilinearen Formen, *Monatsh. Math. u. Phys.* 1:163–236 (1890).
133 N. Wiegmann, A note on pairs of normal matrices with property L, *Proc. Amer. Math. Soc.* 4:35–36 (1953).
134 N. Wiegmann, Necessary and sufficient conditions for unitary similarity, *J. Austral. Math. Soc.* 2:122–126 (1961/62).
135 J. Williamson, A polar representation of singular matrices, *Bull. Amer. Math Soc.* 41:118–123 (1935).
136 A. Wintner and F. D. Murnaghan, On a polar representation of non-singular square matrices, *Proc. Nat. Acad. Sci. U.S.A.* 17:676–678 (1931).

Received 15 September 1989; final manuscript accepted 21 April 1990