A RANK-ONE REDUCTION FORMULA AND ITS APPLICATIONS TO MATRIX FACTORIZATIONS

MOODY T. CHU*, ROBERT. E. FUNDERLIC† AND GENE H. GOLUB‡

Dedicated to the Memory of Alston S. Householder.

Abstract. Let $A \in R^{m\times n}$ denote an arbitrary matrix. If $x \in R^n$ and $y \in R^m$ are vectors such that $\omega = y^T Ax \neq 0$, then the matrix $B := A - \omega^{-1}Axy^T A$ has rank exactly one less than the rank of A. This Wedderburn rank-one reduction formula is easy to prove, yet the idea is so powerful that perhaps all matrix factorizations can be derived from it. The formula also appears in places like the positive definite secant updates BFGS and DFP as well as the ABS methods. By repeatedly applying the formula to reduce ranks, a biconjugation process analogous to the Gram-Schmidt process with oblique projections can be developed. This process provides a mechanism for constructing factorizations such as LDM^T, QR and SVD under a common framework of a general biconjugate decomposition $V^T AU = \Omega$ that is diagonal and nonsingular. Two characterizations of biconjugation provide new insight into the Lanczos method including its breakdown. One characterization shows that the Lanczos algorithm (and the conjugate gradient method) is a special case of the rank-one process and in fact these processes can be identified with the class of biconjugate direction methods so that history is pushed back by about twenty years.

Key words. Rank Reduction, Matrix Factorization, Matrix Decomposition, QR Factorization, LU Factorization, Outer Product Expansion, Conjugate Directions, Conjugate Gradient Method, Lanczos Method, Singular Value Decomposition, SVD, Deflation.

AMS(MOS) subject classifications. 65F15, 65H15.

1. Introduction.

Matrix factorizations or decompositions reign supreme in providing practical numerical algorithms and theoretical linear algebra insights. Matrix factorizations are examples of perhaps the most important strategy of numerical analysis: replace a relatively difficult problem with a much easier one, e.g. triangular systems are easier to solve than full systems. The purpose of this paper is to unify matrix factorizations and several fundamental linear algebra processes within a common framework by exploiting a remarkably simple powerful idea of rank reduction due to Wedderburn.

The emphasis here is on conceptional unification of linear algebra algorithms, not on practical implementation, algorithmic details or stability. We do have high hopes that new insights associated with rank reduction and biconjugation will lead to new algorithms or important modifications to existing ones. We expect to carry out various associated numerical experiments and encourage other researchers to do likewise.

Wedderburn [18, p.69] observed that rank one matrices of the form $\omega^{-1}Axy^TA$ when subtracted from A result in a matrix of rank one less than A if $\omega = y^TAx$. Successive use of this idea (rank(A) times) will therefore be seen to provide quite general factorizations of A. This fruitful successive rank-reducing idea along with what we will call a biconjugation of the x and y vectors gives a paradigm that seems to yield

* Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. This research was supported in part by the National Science Foundation under grant DMS-9123448.

† Department of Computer Science, North Carolina State University, Raleigh, NC 27695-8206.

‡ Department of Computer Science, Stanford University, Stanford, CA 94305.

any matrix decomposition or factorization for appropriate study. Furthermore, many of the fundamental processes of numerical linear algebra such as the Gram-Schmidt, conjugate direction and Lanczos methods are shown to be in the class of the rank-reducing processes developed here. This development provides a special connection between these fundamental numerical processes and matrix factorizations.

Throughout this paper the discussion will be restricted to the real-valued matrices only. The generalization to complex-valued case should be quite obvious.

Let $A \in R^{m\times n}$ denote an arbitrary matrix. We start with the fundamental result of Wedderburn from his 1934 book [18, p.69].

THEOREM 1.1. *If $x \in R^n$ and $y \in R^m$ are vectors such that $\omega = y^T Ax \neq 0$, then the matrix*

$$B := A - \omega^{-1} Axy^T A \tag{1}$$

has rank exactly one less than the rank of A.

A converse of this result is also true. In fact, Householder with reference to Wedderburn included the following characterization of rank-one subtractivity as an exercise in his seminal book [14, Ex.34, p.33].

THEOREM 1.2. *Let $u \in R^m$ and $v \in R^n$. Then the rank of the matrix $B = A - \sigma^{-1}uv^T$ is less than that of A if and only if there are vectors $x \in R^n$ and $y \in R^m$ such that $u = Ax$, $v = A^T y$ and $\sigma = y^T Ax$, in which case $\operatorname{rank}(B) = \operatorname{rank}(A) - 1$.*

Pál Rózsa was responsible for the publication of his mentor's, E. Egerváry's [4], posthumous paper on rank reduction. We infer from this work that Egerváry was the first to prove the entire characterization of Theorem 1.2, though it appears that he was not aware of Wedderburn's earlier result. Having realized that (1) is a rank-diminishing operator and can be repeatedly utilized, Egerváry proposed a general finitely terminating scheme that unifies a variety of processes occurring in the solution of linear equations [4]. More precisely, let $A_1 := A$. So long as $A_k \neq 0$, one may apply (1) repeatedly with nearly limitless flexibility to generate a sequence of matrices $\{A_k\}$ by using

$$A_{k+1} := A_k - \omega_k^{-1} A_k x_k y_k^T A_k \tag{2}$$

for any vectors $x_k \in R^n$ and $y_k \in R^m$ for which $\omega_k := y_k^T A_k x_k \neq 0$. The sequence determined by (2) must terminate in $\operatorname{rank}(A) = \gamma$ steps since $\{\operatorname{rank}(A_k)\}$ decreases by exactly one at each step. This process will continually be used throughout the sequel and we will refer to such a process as *a rank-reducing process* and the A_k matrices will be called *Wedderburn matrices* for the rank-reducing process. Egerváry described several interesting examples, including Gaussian factorization and the Purcell-Motzkin method, to illustrate the application of the rank-reducing process. Some of his ideas resurface in our considerations in a more explicit form.

We may summarize the rank-reducing process in matrix outer-product factorization form:

$$A = \Phi\Omega^{-1}\Psi^T, \tag{3}$$

where $\Omega := \operatorname{diag}\{\omega_1, \ldots, \omega_\gamma\}$, $\Phi := [\phi_1, \ldots, \phi_\gamma] \in R^{m\times\gamma}$ and $\Psi := [\psi_1, \ldots, \psi_\gamma] \in R^{n\times\gamma}$ with

$$\begin{aligned} \phi_k &:= A_k x_k, \\ \psi_k &:= A_k^T y_k. \end{aligned}$$

Obviously, different choices of vectors $\{x_1, \ldots, x_\gamma\}$ and $\{y_1, \ldots, y_\gamma\}$ will result in different factorizations in (3). The factorization represented in (3), therefore, is extremely general. One of the purposes of this paper will be the study of the connection of (3) to some well-known decompositions of A.

The rank-reducing process of this paper is related to Stewart's notion [16] of matrices being A-conjugate. Motivated by conjugate direction methods, Stewart has developed a general conjugation algorithm from which it is shown that different choices of the parameters in his algorithm lead to various methods for solving linear systems and that these methods are closely related to well known matrix factorizations. We will see that a similar notion of biconjugation with respect to a matrix A arises naturally from the Wedderburn rank-one reduction formula which again leads to various matrix factorizations. All of these interesting connections will be explored in this paper.

The rank reduction formula (1) can further be generalized to the case where a matrix of rank possibly greater than one is subtracted. The following analogue of rank-one subtractivity along with several equivalent formulations and connections with generalized inverses has been established by Cline and Funderlic [3].

THEOREM 1.3. *Suppose $U \in R^{m\times k}$, $R \in R^{k\times k}$ and $V \in R^{n\times k}$. Then*

$$\operatorname{rank}(A - UR^{-1}V^T) = \operatorname{rank}(A) - \operatorname{rank}(UR^{-1}V^T)$$

if and only if there exist $X \in R^{n\times k}$ and $Y \in R^{m\times k}$ such that

$$U = AX, \;\; V = A^T Y \;\; \text{and} \;\; R = Y^T AX.$$

The discussion hereafter for the rank-one reduction formula can thus be extended in a similar way to the block version of rank reduction.

This paper is organized as follows. In §2 we describe how the rank-reducing process can proceed without explicitly expressing the intermediate Wedderburn matrices. This postprocessing of the vectors used in a rank-reducing process of a matrix A will be called a *biconjugation process* as the (X, Y) matrices of the rank-reducing process will yield a *biconjugate pair* (U, V) that is defined by

$$V^T AU = \Omega \tag{4}$$

being diagonal and nonsingular, and (4) is called a *biconjugate decomposition* of A. It should be noted that when we use terms such as biconjugation process, biconjugatable matrices or biconjugate matrices, there is a given matrix A and the terms are relative to this matrix.

A central result that characterizes a pair of matrices being biconjugatable will be given which is closely related to fundamental matrix factorizations and biconjugate direction methods including the Lanczos method. In §3 we will develop a paradigm for associating matrix factorizations with the biconjugation process of §2. This makes use of a second characterization of biconjugation that recognizes the class of matrices that map into given biconjugate matrices; we specifically illustrate the LDM^T, Cholesky, QR and SVD factorizations. The singular value decomposition, i.e. SVD, will illustrate a rank-reducing process being carried out as a deflation method with optimal numerical stability. In §4 we argue that the well-known Lanczos process is in fact a special case of the biconjugation algorithm of §2. We specify the choice of (X, Y) that produces the biconjugate directions of the Lanczos algorithm. Furthermore, the biconjugation characterization of §2 provides insight into Lanczos breakdown with the rank-reducing process suggesting possible recovery. Each (X, Y) that effects a rank-reducing process gives rise to a biconjugation process and conversely; all of which follows from a second characterization of biconjugation. Finally, in §5 we point out that the Wedderburn rank-one reduction formula of Theorem 1.1 also appears in the ABS method except that the ABS method emphasizes a different part in the formula. In fact, the rank reduction formula also appears in the well-known BFGS and DFP secant update methods, of which an interesting geometric meaning will be discussed in a separate paper.

Historically, Wedderburn is given credit for *first* publishing how to transform a matrix into a matrix of rank one less, Theorem 1.1, cf. [14, Ex.34, p.33]. Early drafts of the book of Householder just cited were available in 1960 and it is likely that he, independent of Egerváry, also discovered the converse. Egerváry was the first to introduce and exploit the rank-reducing *process* that is used throughout this paper. He used it to produce, e.g. the LU factorization of a matrix. We only recently became aware of Egerváry's paper [4] after noticing reference to it in an exercise of Householder's [14, Ex.20, p.145]: "Apply the obvious identity

$$(u_1, \ldots, u_n)(v_1, \ldots, v_n)^T = u_1 v_1^T + \ldots u_n v_n^T$$

along with Wedderburn's theorem [Theorem 1.2 above] to obtain Egerváry's rank-reducing transformations and thereby derive the methods of triangularization and of orthogonal triangularization." Householder did not mention this exercise when (personal communication) he suggested in 1968 the rank-reducing process as a way to generalize the SVD by the use of norms other than Euclidean, see the Acknowledgement section. Stewart [16] saw the connection of conjugate direction methods and a one sided conjugation process. Though the biconjugation process of this paper originated in a primitive and unpublished form earlier [7], the connections to the Gram-Schmidt, conjugate directions (and hence Lanczos), and ABS processes are apparently new. Stewart's definition of a conjugate pair of matrices and his characterization of a matrix conjugated with respect to another [Theorem 3.1 [16]] suggested our definition of a biconjugated pair of matrices and the main characterization of biconjugation of this paper, Theorem 2.4.

2. The Biconjugation Process. Thus far, a factorization (3) resulting from the fundamental rank-reducing process defined by (2) depends explicitly on the successive rank-reduced Wedderburn matrices $\{A_k\}$. In this section we show how the same decomposition can be accomplished *without* the Wedderburn matrices explicitly appearing. It turns out that this procedure is analogous to the well-known Gram-Schmidt process. The purpose of the process is to transform any matrix pair (X, Y) from a rank-reducing process (2) to (U, V) such that $V^T AU$ is diagonal and nonsingular. We will also see that any (X, Y) that can be so transformed comes from a rank-reducing process. Furthermore, it will be shown by Theorem 2.4 that this process is unique in the sense that the resulting matrices are unique.

For convenience, we shall denote henceforth the bilinear form $y^T Ax$ for any $x \in R^n$ and $y \in R^m$ by

$$< x, y > := y^T Ax. \tag{5}$$

Note that in general $< \cdot, \cdot >$ is *not* an inner product since A may not be positive definite nor even a square matrix, but this seemingly ambiguous notation provides us with some interesting insights.

Let $\mathcal{R}(M)$ denote the range space of a general matrix M. By the definition (2) of the Wedderburn matrix A_2, it is clear that

$$x_1 \perp \mathcal{R}(A_2^T), \tag{6}$$

$$y_1 \perp \mathcal{R}(A_2). \tag{7}$$

Observe also that for any $z \in R^n$ we have

$$A_2 z = A_1 \left(z - \frac{< z, y_1 >}{< x_1, y_1 >} x_1 \right).$$

Define $u_1 := x_1$ and $v_1 := y_1$. One may think of the space R^n as the direct sum

$$R^n = \text{span}\{u_1\} \oplus \mathcal{S}(u_1, v_1)$$

where $\mathcal{S}(u_1, v_1)$ is the $(n-1)$ dimensional linear subspace

$$\mathcal{S}(u_1, v_1) := \left\{ z - \frac{< z, v_1 >}{< u_1, v_1 >} u_1 | z \in R^n \right\}.$$

So the vector

$$u_2 := x_2 - \frac{< x_2, v_1 >}{< u_1, v_1 >} u_1$$

is simply the projection of x_2 onto the subspace $\mathcal{S}(u_1, v_1)$ along the vector u_1. Similarly, one may define

$$v_2 := y_2 - \frac{< u_1, y_2 >}{< u_1, v_1 >} v_1$$

which is the *projection* of y_2 onto the $(m-1)$ dimensional subspace

$$\mathcal{T}(u_1, v_1) := \left\{ w - \frac{< u_1, w >}{< u_1, v_1 >} v_1 | w \in R^m \right\}$$

along the vector v_1. In doing so, we naturally have established the relationships:

$$\begin{aligned} Au_2 &= A_2x_2 \in \mathcal{R}(A_2) \subset \mathcal{R}(A_1) \\ v_2^TA &= y_2^TA_2 \in \mathcal{R}(A_2^T) \subset \mathcal{R}(A_1^T). \end{aligned}$$

Together with (6) and (7), it follows that

$$< u_2, v_1 >=< u_1, v_2 >= 0$$

and hence

$$\omega_2 = y_2^TA_2x_2 =< u_2, v_2 > .$$

This completes the first step of the biconjugation process described by the following theorem. The biconjugation process can now be continued by induction to obtain the following result.

THEOREM 2.1. *Suppose* rank$(A) = \gamma$. *Let* $\{x_1, \ldots, x_\gamma\}$ *and* $\{y_1, \ldots, y_\gamma\}$ *be any vectors associated with a rank-reducing process (so that* $y_k^TA_kx_k \neq 0$ *for each* k*). Then*

$$u_k := x_k - \sum_{i=1}^{k-1} \frac{< x_k, v_i >}{< u_i, v_i >} u_i, \tag{8}$$

$$v_k := y_k - \sum_{i=1}^{k-1} \frac{< u_i, y_k >}{< u_i, v_i >} v_i \tag{9}$$

are well defined for $k = 1, \ldots, \gamma$. *Furthermore, it is true that*

$$Au_k = A_kx_k, \tag{10}$$

$$v_k^TA = y_k^TA_k, \tag{11}$$

$$\omega_k = y_k^TA_kx_k =< u_k, v_k > \tag{12}$$

and that

$$< u_k, v_j >=< u_j, v_k >= 0 \tag{13}$$

for all $j < k$.

Proof. We have already seen the case when $k = 2$. Suppose now the theorem is true for all $j \leq k < \gamma$. Then $\omega_k =< u_k, v_k >\neq 0$. So u_{k+1} and v_{k+1} are well defined.

From (2), (10) and (11), we have

$$\begin{aligned} A_{k+1} &= A_k - \omega_k^{-1}Au_kv_k^TA \\ &= A - \sum_{i=1}^{k} \omega_i^{-1}Au_iv_i^TA \end{aligned} \tag{14}$$

where the second equality is obtained by recursion. It is now clear by direct substitution that (10) and (11) hold for $k + 1$.

It is readily seen that

$$\operatorname{span}\{x_1,\ldots,x_j\} = \operatorname{span}\{u_1,\ldots,u_j\}$$

$$\operatorname{span}\{y_1,\ldots,y_j\} = \operatorname{span}\{v_1,\ldots,v_j\}$$

for any $1 \leq j \leq k+1$. Observe from (2) that the column spaces of the Wedderburn matrices for a rank-reducing process satisfy

$$\mathcal{R}(A_1) \supset \mathcal{R}(A_2) \supset \ldots \ .$$

Since $y_k \perp \mathcal{R}(A_{k+1})$, it follows that $y_j \perp \mathcal{R}(A_{k+1})$ for all $j < k+1$. Therefore, $< u_{k+1}, v_j >= 0$ for all $j < k+1$. Similarly, one can prove $< u_j, v_{k+1} >= 0$ for all $j < k+1$ by using the fact that $x_k \perp \mathcal{R}(A_{k+1}^T)$. This proves (13).

The equality (12) now follows from (13) and (10) by rewriting y_{k+1} in terms of $v_1, \ldots, v_{k+1}$. □

Although it appears that the sequence of Wedderburn matrices $\{A_k\}$ is needed among the assumptions of Theorem 2.1, we note that the arrays $\{u_k\}$ and $\{v_k\}$ defined by (8) and (9) make no explicit reference to the Wedderburn matrices. Indeed, so long as $< u_k, v_k > \neq 0$, then (8) and (9) can be used to generate u_{k+1} and v_{k+1} directly.

We note also that symbolically there is a considerable resemblance between the formulas (8) and (9) in Theorem 2.1 and the formulas that arise from the classical Gram-Schmidt process. The main differences are that the bilinear form $< \cdot, \cdot >$ used in Theorem 2.1 is not necessarily an inner product and that Theorem 2.1 is dealing with two arrays of vectors $\{x_1, \ldots, x_\gamma\}$ and $\{y_1, \ldots, y_\gamma\}$ simultaneously. In other words, the notion of the orthogonal projection in the Gram-Schmidt process is being replaced by an *oblique* projection as is demonstrated in the discussion of u_2 and v_2.

An important implication of Theorem 2.1 is that it offers an alternative way of computing the ω_k's and the $A_k x_k$ and $y_k^T A_k$ vectors. In particular given any $\{x_1, \ldots, x_\gamma\}$ and $\{y_1, \ldots, y_\gamma\}$ satisfying $y_k^T A_k x_k \neq 0$ for each k, we are able to describe any factorization (3) of A arising from a rank-reducing process,

$$[A_1 x_1, \ldots, A_\gamma x_\gamma]\Omega^{-1}[y_1^T A_1, \ldots, y_\gamma^T A_\gamma],$$

as $AU\Omega^{-1}V^T A$, independent of the Wedderburn matrices, A_k. Furthermore, by way of the following corollary we obtain a remarkable decomposition, $V^T AU$, of A independent of Wedderburn matrices appearing. Actually Wedderburn matrices are involved in the sense that the x_j and y_j vectors (used to produce the u_j and v_j vectors in (8) and (9)) effect a rank producing process. Theorem 2.4 will cause the Wedderburn matrices to be further concealed.

COROLLARY 2.2. *Suppose A has has rank γ and $1 \leq k \leq \gamma$. Let $U_k := [u_1, \ldots, u_k]$ and $V_k := [v_1, \ldots, v_k]$ where u_i and v_j are defined in (8) and (9). Let $\Omega_k := \operatorname{diag}\{\omega_1, \ldots, \omega_k\}$. Then*

$$V_k^T A U_k = \Omega_k. \tag{15}$$

Furthermore,

$$A = AU_\gamma \Omega_\gamma^{-1} V_\gamma^T A, \tag{16}$$

and

$$M = U_\gamma \Omega_\gamma^{-1} V_\gamma^T. \tag{17}$$

is a semi-inverse [9] of A.

Proof. Identity (15) follows from (13) and (16) from (14). The semi-inverse conditions $AMA = A$ and $MAM = M$ can be checked directly by substitution. □

It is interesting to note from (17) that if A is nonsingular, then $M = A^{-1}$ can be obtained simply by *transposing* the two matrices U_n and V_n and by inverting the diagonal matrix Ω_n — an analogy to the SVD even though U_n and V_n are *not* necessarily orthogonal. We will follow Stewart's [17, pp. 29-30] distinction between the singular value *decomposition* and singular value *factorization* by referring to (15) as a decomposition of A and (16) as a factorization of A.

We now turn our attention to Stewart's notion of an A-conjugate pair of matrices [16] which was motivated quite differently from the Wedderburn rank-diminishing idea. Stewart defines (U, V) to be an A-conjugate pair if $V^T AU$ is lower triangular under the assumptions that $A, U, V \in R^{n\times n}$ are nonsingular. Given nonsingular matrices V, A and P, Stewart has proposed an algorithm that combines columns of P linearly in a suitable way to form a matrix U so that U and V are A-conjugate. Stewart called his process the *A-conjugation of P with respect to V*. We refer readers to [16] for more details on his A-conjugation process. The following definition provides a natural transition from *A-conjugation* to *biconjugation with respect to a matrix A*.

DEFINITION 2.1. *Let $A \in R^{m\times n}$, $U \in R^{n\times k}$ and $V \in R^{m\times k}$. Then (U, V) is a biconjugate pair (with respect to A) if*

$$\Omega := V^T AU \tag{18}$$

is nonsingular and diagonal. Such a decomposition is called a biconjugate decomposition of A. Factorizations of the form (16) satisfying the semi-inverse conditions associated with (17) are called biconjugate factorizations of A.

Definition 2.1 suggests a convenient special notational convention that will be used throughout the sequel. When we use the notation of (18), Ω will always refer to a nonsingular diagonal matrix with diagonal elements ω_j and such elements will always refer to the relevant diagonal elements of such a matrix Ω.

It is important to note that Theorem 2.1 provides a *biconjugation process* for a given pair of matrices (X, Y) provided that for the Wedderburn matrices, A_j, the relations $y_j^T A_j x_j \neq 0$ hold for all $1 \leq j \leq k$. Indeed, we may rewrite (8) and (9) in the following matrix form.

COROLLARY 2.3. *If $(X_k, Y_k) \in R^{n\times k} \times R^{m\times k}$ effects a rank-reducing process for A, then there are unique unit upper triangular matrices $R_k^{(x)}$ and $R_k^{(x)}$ of order k such that*

$$X_k = U_k R_k^{(x)}, \quad \textit{and} \quad Y_k = V_k R_k^{(y)}, \tag{19}$$

where U_k and V_k are matrices with columns resulting from the biconjugation process of Theorem 2.1.

Proof. The relations of (19) follow from (8) and (9) since the j^{th} columns of $R_k^{(x)}$ and $R_k^{(y)}$ are expressed as

$$\left[\frac{< x_j, v_1 >}{< u_1, v_1 >}, \ldots, \frac{< x_j, v_{j-1} >}{< u_{j-1}, v_{j-1} >}, 1, 0, \ldots, 0\right]^T,$$

and

$$\left[\frac{< u_1, y_j >}{< u_1, v_1 >}, \ldots, \frac{< u_{j-1}, y_j >}{< u_{j-1}, v_{j-1} >}, 1, 0, \ldots, 0\right]^T.$$

respectively. That the upper triangular matrices are unique follows because the matrices U and V in biconjugate decompositions (15) have linearly independent columns. Likewise from the characterization of rank subtractivity, Theorem 1.3, X_k and Y_k have linearly independent columns. □

Thus analogous to Stewart's A-conjugation process [16] we make the following

DEFINITION 2.2. *Given a matrix A, then a pair of matrices $(X, Y) \in R^{n\times k} \times R^{m\times k}$ is said to be biconjugatable, and biconjugated into a biconjugate pair of matrices (U, V), (see Definition 2.1), if there exist unit upper triangular matrices $R_x, R_y \in R^{k\times k}$ such that $X = UR_x$, $Y = VR_y$.*

Note from (19) that at this point the triangular matrices $R_k^{(x)}$ and $R_k^{(y)}$ depend implicitly on both V_k and U_k simultaneously. We will see, however, that this dependence is in fact immaterial (Theorem 3.6). More importantly, Corollary 2.3 will be seen to be part of a characterization of biconjugation, Theorem 3.7, which will show that if (X, Y) can be biconjugated into (U, V), then these pairs must be linked by unique unit upper triangular matrices. Furthermore, if (U, V) is a biconjugate pair and (X, Y) is biconjugatable, it does not follow that (X, Y) can be biconjugated into (U, V). Again uniqueness of the biconjugation process is the reason.

Recall that for the biconjugation process one has considerable freedom in selecting the vectors $[x_1, \ldots, x_k]$ and $[y_1, \ldots, y_k]$ to be biconjugated. Thus far, an important assumption for choosing the x_j and y_j vectors is that $y_j^T A_j x_j \neq 0$, a condition depending upon the Wedderburn matrices A_j. The following theorem analogous to Stewart's [16, Theorem 3.1] characterization of A-conjugation provides an important necessary and sufficient condition for biconjugation independent of the Wedderburn A_j matrices. This characterization provides understanding of the relation of biconjugation and some of the fundamental factorizations of linear algebra, e.g. see Corollary 2.5 and its relation to the Cholesky factorization of Theorem 3.5. In view of current wide interest in Krylov space methods, this characterization may provide useful insight into better understanding breakdown of the Lanczos algorithm as will be suggested by Theorem 4.3.

THEOREM 2.4. *Let $(X, Y) \in R^{n\times k} \times R^{m\times k}$ and $A \in R^{m\times n}$ be given. Then (X, Y) can be biconjugated if and only if $Y^T AX$ has an LDU decomposition. In that event the U, V and Ω of the biconjugate decomposition are unique and the resulting LDU factorization of $Y^T AX$ is $R_y^T \Omega R_x$, where R_x and R_y are unique upper triangular matrices from the Definition 2.2 of biconjugation.*

Proof. Suppose X and Y can be biconjugated. Then, by Definition 2.2, there exist unit upper triangular matrices R_x and R_y such that $X = UR_x$, $Y = VR_y$ and $V^TAU = \Omega$ is a nonsingular diagonal matrix. It follows that

$$Y^TAX = R_y^TV^TAUR_x = R_y^T\Omega R_x$$

is the unique unit triangular LDU decomposition of Y^TAX.

Conversely, suppose $Y^TAX = R_2^TDR_1$ is an LDU decomposition with both R_1 and R_2 unit upper triangular matrices. Then by Definition 2.2 and since R_1^{-1} and R_2^{-1} are unit upper triangular, (X, Y) biconjugates into (XR_1^{-1}, YR_2^{-1}). The uniqueness of U, V and Ω follows from the fact that any biconjugate pair resulting from (X, Y) gives rise to the same unique LDU factorization of Y^TAX. $\square$

In light of the unique nature of biconjugation as brought out by Theorem 2.4, we recapitulate the rank-reducing process, the biconjugation (algorithm) of Theorem 2.1 and the connection, Corollary 2.3, between biconjugatable matrices and the resulting biconjugate matrices: We initially provided biconjugatable matrices (X, Y) from the rank-reducing process. However, from Theorem 2.4 the biconjugation process of Theorem 2.1 can be carried out for any (X, Y) such that Y^TAX has an LDU factorization. Moreover, the process of producing a biconjugate pair (U, V) is a unique process in that U and V are unique. Corollary 2.3 used biconjugatable matrices from the rank-reducing process to infer the triangular matrices connection with the resulting biconjugate pair from the biconjugation process of Theorem 2.1. However, Theorem 2.4 shows that as long as (X, Y) is any biconjugatable pair of matrices, then the resulting biconjugate pair is unique and the associated upper triangular matrices are unique.

We conclude this section with the following observation which will be used for the LDU development of the next section.

COROLLARY 2.5. *Suppose a given pair of matrices $(X, Y) \in R^{n\times k} \times R^{m\times k}$ can be biconjugated. Then*

$$\det(Y^TAX) = \prod_{i=1}^{k} \omega_i. \tag{20}$$

Proof. By Theorem 2.4 there are unit upper triangular matrices R_x and R_y such that $Y^TAX = R_y^T\Omega R_x$. The determinant of the latter factorization gives (20). $\square$

3. Choices of X and Y. By various choices of matrices in Stewart's A-conjugation process [16], several well known matrix decompositions are obtained. In this section, we illustrate that the biconjugation process and other results from the previous section can be used to obtain similar results by varying biconjugatable X and Y.

Motivated by Corollary 2.3 which expresses the relationship between biconjugatable matrices and their resulting biconjugate pair, we begin with *upper trapezoidal* matrices where the (i, j) entry is zero whenever $i > j$.

THEOREM 3.1. *Suppose $A \in R^{m\times n}$ is of rank γ and X_γ and Y_γ are upper trapezoidal matrices in $R^{n\times\gamma}$ and $R^{m\times\gamma}$, respectively, that are biconjugated into (U_γ, V_γ). Then AU_γ*

and $A^T V_\gamma$ are lower trapezoidal matrices in $R^{m\times\gamma}$ and $R^{n\times\gamma}$, respectively. Thus the resulting biconjugate factorization (16) gives rise to a trapezoidal LDU factorization of A.

Proof. Recall that in the biconjugation process of Theorem 2.1 $[A_1x_1,\ldots,A_\gamma x_\gamma] = AU_\gamma$. We shall first examine the structure of A_2x_2 associated with the second Wedderburn matrix, A_2. Let e_i designate the i^{th} column of an identity matrix of appropriate order. Then the trapezoidal forms of X and Y imply $x_1 = \alpha e_1$ and $y_1 = \beta e_1$ for some nonzero α and β. Thus the Wedderburn matrix $A_2 = A_1 - (e_1^T A_1 e_1)^{-1} A_1 e_1 e_1^T A_1$ and therefore the first column and the first row of A_2 are identically zero. It follows that the first entry of A_2x_2 is zero.

Similarly, one can show the first k columns and the first k rows of A_{k+1} are identically zero. The proof can be completed by induction. □

It is well known that the Gaussian elimination of a square matrix can be viewed as a successive rank-reducing process. Egerváry [4] observed that this can in fact be accomplished by applying his rank-reducing process to the choices $X_n = Y_n = I_n$ with the implicit assumption that the process will not break down. Note that the matrices X and Y in Theorem 3.1 are more general. We now recast Egerváry's Gaussian elimination observation and specify a unique LDU factorization for a square matrix A. We begin with the following lemma.

LEMMA 3.2. *Let $\gamma = rank(A) \ge k$. The matrix pair (X_k, Y_k) taken as the appropriate respective "identity" matrices in $R^{n\times k}$ and $R^{m\times k}$ is biconjugatable if and only if the k^{th} leading principal minor of A is nonzero. In this case the k^{th} leading principal minor of A is given by $\prod_{i=1}^{k}\omega_i$.*

Proof. For $k \le \gamma$, $Y_k^T A X_k$ is precisely the k^{th} principal submatrix of A. The assumption that the k^{th} leading principal minor of A is nonzero implies, from Corollary 2.5, that $\omega_k \ne 0$ for all $k \le \gamma$. The biconjugation process therefore can be carried out according to Theorem 2.1. The converse holds since Corollary 2.5 implies the determinant of $Y_k^T A X_k$ is nonzero. □

THEOREM 3.3. *Let $A \in R^{n\times n}$. Then (I_n, I_n) is biconjugatable if and only if all the leading principal minors of A are nonzero. In this case the main diagonals of AU_n, $V_n^T A$ and Ω_n are identical and $A = V_n^{-T}\Omega_n U_n^{-1}$ is the unique LDM^T factorization of A.*

Proof. From Lemma 3.2 we know that biconjugation is well defined and yields well defined biconjugate U_n and V_n, and conversely. From Corollary 2.3 we also know that both U_n and V_n are nonsingular upper triangular matrices with unit diagonal entries. Since from (15) $AU_n = V_n^{-T}\Omega_n$ and $V_n^T A = \Omega_n U_n^{-1}$, the main diagonals of AU_n, $V_n^T A$ and Ω_n are all the same. □

For the symmetric case, the following theorem is very easy to prove.

THEOREM 3.4. *If $A \in R^{n\times n}$ is symmetric and (X_γ, X_γ) is biconjugatable, then the resulting biconjugate pair (U_γ, V_γ) has $U_\gamma = V_\gamma$. In this case, (15) is the canonical form of A with respect to congruence and the columns of U_γ are conjugate direction vectors.*

COROLLARY 3.5. *Suppose $A \in R^{n\times n}$ is symmetric and positive definite. Then the Cholesky factorization of A can be obtained from the biconjugation process applied to*

(I_n, I_n).

Proof. Given that the leading principal minors of positive definite matrices are positive, Corollary 2.5 implies $\omega_j > 0$ for $j \leq n$, and the j^{th} leading principal minor of A is $\prod_{k=1}^{j} \omega_k = det(I_j^T A I_j)$. The pair (I_n, I_n) is biconjugatable from Theorem 3.3, and from Theorem 3.4 the biconjugate decomposition of A gives a factorization of the form $U_n^{-T} \Omega U_n^{-1}$. Therefore the Cholesky factor is $\Omega^{\frac{1}{2}} U_n^{-1}$ and it can be obtained from $AU = U^{-T}\Omega$. $\square$

From the proof of the corollary the Cholesky factor is AU and not U with the latter being related to the inverse of the Cholesky factor of A. The biconjugation process here is equivalent to Gram-Schmidt in the inner product $< x, y >:= y^T A x$ which was considered by Fox, Huskey and Wilkinson [6] and later by Hestenes and Stiefel [12] in their conjugate gradient paper.

The biconjugation process with respect to a given matrix A may be thought of as a function f acting on the space $R^{n \times k} \times R^{m \times k}$ by

$$f(X, Y) := (U, V), \tag{21}$$

where U and V are a biconjugate pair of matrices that result from the biconjugation process determined by Theorem 2.1 and therefore enter into a biconjugate decomposition (15), $V^T A U = \Omega$, of A. Corollary 2.3 tells us that if (X, Y) is biconjugated into (U, V) via (21), then

$$X = U R_x, \quad \text{and} \quad Y = V R_y, \tag{22}$$

for unique unit upper triangular matrices R_x and R_y. A converse holds, and in particular if at the outset X and Y are a biconjugate pair of matrices, then the biconjugation process returns the same X and Y. More generally the complete converse of Corollary 2.3 is

THEOREM 3.6. *If $(U, V) \in R^{n \times k} \times R^{m \times k}$ is a biconjugate pair, and R_x and R_y are arbitrary unit upper triangular matrices in $R^{k \times k}$, then*

$$(U R_x, V R_y) \tag{23}$$

can be biconjugated and the resulting biconjugate pair is exactly the initial (U, V).

Proof. That the matrices of (23) can be biconjugated follows from Theorem 2.4 since (U, V) is biconjugate pair and therefore $R_y^T V^T A U R_x$ is an LDU factorization of itself. By the Definition 2.2 of biconjugation the resulting biconjugate pair from (23) must be of the form $(U R_1, V R_2)$ where R_1 and R_2 are unit upper triangular. Therefore $R_2^T V^T A U R_1$ must be diagonal and nonsingular. It follows that $R_1 = R_2 = I$ because of the uniqueness of LDU factorizations. $\square$

So the definition of biconjugation, Corollary 2.3 and Theorem 3.6 give the next theorem, a second characterization of biconjugatability, to go along with the LDU characterization of Theorem 2.4. It should be emphasized again that these characterizations are related simply by $Y^T A X = R_y^T V^T A U R_x = R_y^T \Omega R_x$.

THEOREM 3.7. *The matrix pair (X, Y) biconjugates into (U, V) if and only if (U, V) is a biconjugate pair and there are unique unit upper triangular matrices R_x and R_y satisfying (22).*

Theorem 3.6 provides a paradigm for showing that fundamental decompositions of linear algebra are biconjugate decompositions, $V^T AU = \Omega$, of Corollary 2.2. We illustrate this idea by first considering the QR decomposition.

We shall assume that $A = QR$ has full column rank so that the columns of $Q \in R^{m\times n}$ are orthonormal and $R \in R^{n\times n}$ is nonsingular. We leave as an exercise for the reader to consider the rank deficient cases including $m < n$. The idea for the full column rank case is to have Theorem 3.6 suggest simple and biconjugatable X and Y. For example $X = Y = I$ in Theorem 3.3 leads to an LDU factorization. It is tempting to choose $U = R^{-1}$ so that the biconjugate decomposition would be $Q^T AR^{-1} = I$. Then from Theorem 3.6,

$$X = R^{-1}R_x \text{ and } Y = QR_y. \tag{24}$$

These give rise to two solvable problems. The first problem is that the equalities of (24) beg the question of producing Q and R since (24) depends on knowing Q and R *ab initio*. This problem is apparently solved by the choice $R_x = R_y = R$ so that $(X, Y) := (I, A)$ is independent of knowing Q and R. This gives rise to the second problem, a scaling one, because in Theorem 3.6 R_x and R_y must be *unit* upper triangular matrices and R in general does not have ones on its diagonal. This motivates rewriting the QR factorization of A as $Q\Psi R_1$ with Ψ a diagonal matrix such that $\Psi R_1 = R$ with R_1 being *unit* upper triangular. In place of (24) define

$$X := R_1^{-1}R_x, \text{ and } Y := Q\Psi R_y,$$

where R_x and R_y are arbitrary unit upper triangular matrices as suggested by Theorem 3.6 and noting that $(R_1^{-1}, Q\Psi)$ is indeed a biconjugate pair and thus X and Y are biconjugatable by Theorem 3.7. It is thus natural to choose $R_x = R_y = R_1$ so that the simple choice of (I, A) actually does biconjugate into $(R_1^{-1}, Q\Psi)$. Furthermore, $AU = AR_1^{-1} = Q\Psi = V$ and $V^T A = \Psi Q^T A = \Psi^2 R_1 = \Psi R$. Thus we may obtain the QR factorization from V and $V^T A$. As expected $R = \Psi R_1$ is the Cholesky factor of $A^T A = Y^T AX = R_1^T \Psi Q^T A = R_1^T \Psi^2 R_1$. Most of these QR factorization results are summarized in

THEOREM 3.8. *Let A have full column rank with a QR factorization, $A = QR$. Then (I, A) is biconjugatable and gives the biconjugate pair $(R_1^{-1}, Q\Psi) := (U, V)$ where Ψ is a diagonal matrix and $U = R_1^{-1}$ is the unit upper triangular matrix $R^{-1}\Psi$. Furthermore, $V^T A = \Psi R, AU = V$ and $V^T V = \Psi^2 := \Omega$.*

Theorem 3.8 is the culmination of our use of Theorem 3.6 to develop a paradigm for yielding a fundamental factorization and in this case the QR factorization. Though Theorem 3.3 did not make use of this paradigm for producing an LDM^T factorization, we can do so by defining

$$X := M^{-T}R_x, \;\; Y := L^{-T}R_y.$$

Then it is natural to choose $R_x = M^T$ and $R_y = L^T$ which means that if (I, I) is biconjugatable, it biconjugates into (M^{-T}, L^{-T}). Furthermore, $AU = AM^{-T} = LD$, $V^T A = L^{-1}A = DM^T$ and D takes the role of Ω in the biconjugate decomposition $L^{-1}AM^{-T} = D$.

The rank-reducing process and Theorem 3.6 can be related to the SVD. When the SVD of a matrix A is written with the number of columns in U and V being equal to the rank of A, the SVD of a matrix A, $V^T AU = \Omega$, displays naturally a biconjugate pair (U, V) and a related nonsingular matrix Ω. From Theorem 3.6 the biconjugation process returns (U, V) from $(X, Y) := (U, V)$. Thus we have

THEOREM 3.9. *Suppose vectors $[x_1, \ldots, x_k]$ and $[y_1, \ldots, y_k]$ in the process of biconjugation are, respectively, the right singular vectors and the left singular vectors of A. Then $u_j = x_j$ and $v_j = y_j$ for all $j \leq k$.*

Furthermore, it is well known and easy to show that

$$A = AU\Omega^{-1}V^T A, \tag{25}$$

so that the singular value factorization (25) is a biconjugate factorization (16) of A.

Because of Galois theory it is not possible to solve directly general polynomial equations and thus there is no simple biconjugatable (X, Y) that produces the SVD. However, there is a natural and numerically optimal rank-reducing process that leads to the SVD. Moreover this SVD producing rank-reducing process does not require an additional biconjugation process. The SVD of a matrix A is probably the most numerically appealing of the biconjugate decompositions (15) because as we shall now see, the singular values provide maximal denominators, ω_i's, in the rank-reducing process. From the viewpoint of numerical stability it is ideal relative to the Euclidean norm at each stage of the rank-reducing process to choose x_k and y_k so as to

$$\text{Maximize} \quad \omega_k = y_k^T A_k x_k = < u_k, v_k > \tag{26}$$

$$\text{Subject to} \quad x_k^T x_k = 1, \; y_k^T y_k = 1. \tag{27}$$

We now demonstrate that the decomposition (15) produced by meeting this particular requirement in the rank-reducing process is precisely the singular value decomposition of A.

It can easily be verified by Lagrange multipliers that a necessary condition of a stationary point for (26) is

$$A_k^T \tilde{y}_k = \tilde{\sigma}_k \tilde{x}_k \tag{28}$$

$$A_k \tilde{x}_k = \tilde{\sigma}_k \tilde{y}_k. \tag{29}$$

Some more general discussion can be found in [2]. Indeed, the maximal value occurs at $\tilde{\sigma}_k = \|A_k\|_2$ which is the largest singular value of the Wedderburn matrix A_k and it will therefore be shown to be the kth singular value ω_k of A. The corresponding solution $\tilde{y}_k$ and $\tilde{x}_k$, respectively, become the kth left singular vector y_k and the kth right singular vector x_k of A. This fact has been used to prove the existence of the singular value decomposition of A (See, for example, [10, Theorem 2.5.1]). That the singular values of the Wedderburn matrices of this particular rank-reducing process are singular values of A follows from the following theorem. Thus this rank-reducing process is a singular value deflation process.

THEOREM 3.10. *Let the nonzero singular values of A be ordered so that $\omega_1 \geq \omega_2 \geq \ldots \geq \omega_\gamma > 0$. If in the rank-reducing process $\omega_{k+1} = \|A_{k+1}\|_2$, then the corresponding x_{k+1} and y_{k+1} vectors may be chosen to be the corresponding singular vectors of A, and the largest singular value of A_{k+1} is the $(k+1)^{th}$ largest singular value of A.*

Proof. We shall first determine the singular values and vectors of A_2. Substituting (28) and (29) into the definition of the second Wedderburn matrix (2) gives

$$A_2 = A_1 - \omega_1 v_1 u_1^T.$$

But it is well known, e.g. [10, Theorem 2.5.2], that the singular value factorization for $A_1 := A$ can be expressed as an outer product expansion

$$A = \sum_{i=1}^{\gamma} \omega_i v_i u_i^T,$$

so that

$$A_2 = \sum_{i=2}^{\gamma} \omega_i v_i u_i^T.$$

Thus $\overline{\sigma}_2$, $\overline{x}_2$ and $\overline{y}_2$ obtained from (26) and (27) may be taken to be ω_2, u_2 and v_2, respectively. The proof can be completed by induction. $\square$

4. Relation to the Lanczos Process. The purpose of this section is to relate the Lanczos process to the biconjugation process of §2. The two biconjugation characterizations of Theorem 2.4 and Theorem 3.7 will be used to discern and portray a biconjugatable pair, $(R, \overline{R})$ that biconjugates into $(P, \overline{P})$. By consequence of this process $R^{-1}AR$ becomes the classical tridiagonal matrix. This development should provide additional insight to and understanding of the Lanczos process.

For a general square matrix A, the Lanczos process [15] can be described as follows: Given two nonorthogonal vectors r_1 and $\overline{r}_1$, define $p_1 := r_1$ and $\overline{p}_1 := \overline{r}_1$. Then for $k = 1, 2, \ldots$, define the recurrence relations

$$\begin{aligned} r_{k+1} &:= r_k - \alpha_k A p_k \\ \overline{r}_{k+1} &:= \overline{r}_k - \alpha_k A^T \overline{p}_k \end{aligned} \tag{30}$$

and

$$p_{k+1} := r_{k+1} + \beta_k p_k \tag{31}$$

$$\overline{p}_{k+1} := \overline{r}_{k+1} + \beta_k \overline{p}_k \tag{32}$$

where

$$\alpha_k := \frac{\overline{r}_k^T r_k}{\overline{p}_k^T A p_k} \tag{33}$$

$$\beta_k := \frac{\overline{p}_k^T A r_{k+1}}{\overline{p}_k^T A p_k}. \tag{34}$$

If we define

$$P_k := [p_1, \ldots, p_k] \text{ and } \overline{P}_k := [\overline{p}_1, \ldots, \overline{p}_k], \tag{35}$$

and

$$R_{k+1} := [r_1, \ldots, r_{k+1}], \quad \overline{R}_{k+1} := [\overline{r}_1, \ldots, \overline{r}_{k+1}],$$

it can be proved (See, for example, [5]) that

THEOREM 4.1. *Assuming that the Lanczos process does not break down, then*

$$\overline{P}_k^T A P_k := \Omega_k \tag{36}$$

and $\overline{R}_k^T R_k := D_r$ *are nonsingular diagonal matrices with (36) (along with (33) and (34) well defined) defining the biconjugate pair* $(P_k, \overline{P}_k)$ *of (35).*

The Lanczos process is perhaps better recognized in matrix form. After the k^{th} step, we have from (30)

$$A P_k D_k = R_{k+1} C_k$$

where

$$\begin{aligned} D_k &:= \operatorname{diag}\{\alpha_1, \ldots, \alpha_k\}, \\ C_k &:= \begin{bmatrix} 1 & 0 & & \ldots & 0 \\ -1 & 1 & 0 & & 0 \\ 0 & -1 & 1 & & \\ \vdots & & & \ddots & \\ & & & \ddots & 1 \\ 0 & & & \ldots & -1 \end{bmatrix} \in R^{(k+1)\times k}. \end{aligned}$$

We also obtain from (31) and (32) the conditions of the second characterization of biconjugation, Theorem 3.7, i.e.

$$R_{k+1} = P_{k+1} B_k \text{ and } \overline{R}_{k+1} = \overline{P}_{k+1} B_k, \tag{37}$$

where

$$B_k := \begin{bmatrix} 1 & -\beta_1 & 0 & \ldots & 0 \\ 0 & 1 & -\beta_2 & & \\ \vdots & & 1 & \ddots & \\ 0 & & & \ddots & -\beta_k \\ 0 & & & & 1 \end{bmatrix} \in R^{(k+1)\times(k+1)}. \tag{38}$$

It follows that

$$\begin{aligned} A R_{k+1} &= A P_{k+1} B_k \\ &= \left[R_{k+1} C_k D_k^{-1} B_{k-1}, A r_{k+1}\right] && (39) \\ &= [R_{k+1} T_k, \bigcirc] + [\bigcirc, A r_{k+1}] && (40) \end{aligned}$$

where $T_k := C_k D_k^{-1} B_{k-1}$ is a "tridiagonal" matrix in $R^{(k+1)\times k}$.

The vector r_{k+1} defined in (30) may be thought of as the residual

$$r_{k+1} = b - Az_{k+1} \tag{41}$$

where z_{k+1} is defined by the recurrence relation

$$z_{k+1} := z_k + \alpha_k p_k \tag{42}$$

with a suitable initial value z_1 that gives rise to the residual r_1. If at a certain stage we have $r_{k+1} = 0$, then it follows that z_{k+1} solves the linear system $Az = b$. In this case, (40) may be written as

$$AR_k = R_k T \tag{43}$$

where T is the $k \times k$ tridiagonal matrix by deleting the last row of T_k. Likewise, we may consider $\overline{r}_{k+1} = b - A^T \overline{z}_{k+1}$ with $\overline{z}_{k+1} := \overline{z}_k + \alpha_k \overline{p}_k$.

We claim that the Lanczos process is a special case of the biconjugation process of Theorem 2.1. More precisely, we have

THEOREM 4.2. *If R_k and $\overline{R}_k$ from (37) are well defined in a Lanczos process that does not break down, then $(X, Y) := (R_k, \overline{R}_k)$ biconjugates into $(P_k, \overline{P}_k)$ of (35)*

Proof. Theorem 4.1 points out that $(P_k, \overline{P}_k)$ is a biconjugate pair and the necessary unit upper triangular relations for biconjugation are immediate from (37). Thus the conclusion follows from Theorem 3.7. □

From (33) and (34) we conclude that the Lanczos process will not break down in exact arithmetic as long as $\omega_k \neq 0$. Suppose $\mathrm{rank}(A) = \gamma$. Then by Theorem 2.1 one should be able to continue the Lanczos process for (at most) γ steps. Nevertheless, we note that the subsequent behavior of the Lanczos process is completely predestined by the first choice of r_1 and $\overline{r}_1$. It is possible, therefore, that the Lanczos process will break down prematurely because r_k and $\overline{r}_k$ are fixed vectors. In contrast, the vectors x_k and y_k in the general rank-reducing and biconjugation process can be more flexible and are independent of x_1 and y_1. In other words, the biconjugation process may be carried forward to the next stage by a suitable choice of x_k and y_k so long as $\mathrm{rank}(A_k) \neq 0$. In principle the Wedderburn matrix A_k can be computed from $(P_k, \overline{P}_k)$ by using Theorem 1.3. More precisely, we have

$$A_k = A - AP_{k-1}\Omega_{k-1}^{-1}\overline{P}_{k-1}^T A.$$

To take advantages of the three-term recursion relationship in the Lanczos process and the extra flexibility of the biconjugation process, a hybrid method naturally suggests itself with the choice of $X_n = [r_1, \ldots, r_k, x_{k+1}, \ldots, x_n]$ (and similarly of Y_n) where we use the Lanczos process until it nearly breaks down and then we switch to the more general biconjugation process. In particular, if A is nonsingular, then the solution z^* to the equation $Az = b$ may be expressed, by (17), as

$$z^* = \sum_{i=1}^{n} \frac{v_i^T b}{\omega_i} u_i. \tag{44}$$

At the cost of computing (8) and (9), the biconjugation process therefore defines an iterative method that converges to z^* in n steps.

Another result that may prove to be important in understanding *breakdowns* and restart strategies of the Lanczos process is

THEOREM 4.3. *If the Lanczos process can be carried through to completion, then*

$$\overline{R}_i^T A R_i \tag{45}$$

has LDU factorizations for $i = 1, \ldots, n$.

Proof. Theorem 2.4 characterizes when the biconjugation of $(R_n, \overline{R}_n)$ can be carried out in terms of (45) having an LDU factorization. We should note that the values of α_k and β_k are well defined from (33) and (34) if $(P_n, \overline{P}_n)$ is a biconjugate pair. □

Additionally Theorem 4.3 and the biconjugation characterization of Theorem 2.4 provide us with the LDU factorization, $B^T \Omega B$, of $\overline{R}^T AR$ from (37), i.e.

$$\overline{R}^T AR = B^T \Omega B = \begin{bmatrix} \omega_1 & -\omega_1\beta_1 & 0 & \ldots & 0 & 0 \\ -\omega_1\beta_1 & \omega_2 + \omega_1\beta_1^2 & -\omega_2\beta_2 & & & \\ 0 & -\omega_2\beta_2 & \omega_2 + \omega_1\beta_1^2 & & & 0 \\ \vdots & & \ddots & \ddots & \vdots & \vdots \\ 0 & & & & & -\omega_{n-1}\beta_{n-1} \\ 0 & \ldots & 0 & \ldots & -\omega_{n-1}\beta_{n-1} & \omega_n + \omega_{n-1}\beta_{n-1}^2 \end{bmatrix},$$

a real symmetric tridiagonal matrix, where the β's are defined by (34) and the ω's are the $\overline{p}^T Ap$'s of Theorem 4.1. In the classical Lanczos development, e.g. [14, pp. 17-18] or [11], there is a matrix $Y^T AX$ that is expressed as a tridiagonal matrix, $T = X^{-1}AX$, times a diagonal matrix, $Y^T X$. In this development the columns of X and Y come from Krylov sequences associated with A and A^T respectively. In the development here, $R^{-1}AR$ is tridiagonal as $X^{-1}AX$ is in the classical treatment, but $\overline{R}^T AR$ is automatically real symmetric tridiagonal. The following corollary is a known consequence of the Lanczos process and it follows neatly from the development here. It is known that there are always starting vectors for which the Lanczos process goes through to completion. Therefore from Theorem 4.3 and the comments that follow we have

COROLLARY 4.4. *Given an arbitrary real matrix of order* n, *there are matrices* $\overline{R}$ *and* R *such that* $\overline{R}^T AR$ *is real symmetric tridiagonal of order* n.

It is well known that the Lanczos method is a biconjugate direction method, i.e. $\overline{P}_k^T AP_k$ is nonsingular and diagonal. If A is symmetric positive definite and $r_1 = \overline{r}_1$, then $\overline{P}_k = P_k$, and it is well known that the Lanczos process is equivalent to the conjugate gradient method. More generally let U at the outset determine a set of conjugate directions, i.e. $U^T AU = \Omega$ is nonsingular and diagonal. Then (33), $z_{k+1} = z_k + \alpha u_k$, and an equivalent form of (30), i.e. $r_{k+1} = b - Az_{k+1}$, taken together define the class of conjugate direction methods of which the conjugate gradient method is a member.

For the Lanczos method the imposition of positive definiteness on A prevents the Lanczos method from *breakdown*. It is evident that with appropriate scaling, (43) suggests an iterative procedure for tridiagonalizing A by orthogonal transformations.

Theorem 4.2 shows the Lanczos process is a special case of biconjugation and Theorem 3.6 suggests this is not surprising. Theorem 3.6 implies that the biconjugate directions associated with a biconjugate direction method, e.g. Lanczos method, can be obtained from the biconjugate process applied to an infinite number of biconjugatable matrices. In addition if a biconjugate pair (U, V) is associated with a biconjugate direction method, then the associated vectors would effect a rank-reducing process. Conversely, an (X, Y) associated with a rank-reducing process is biconjugatable. Therefore (X, Y) determines a biconjugate direction method. Thus the class of biconjugate direction methods can be identified with the class of matrix pairs (X, Y) that can be obtained from the fundamental rank-reducing processes (2).

5. Relation to the ABS Method. Given $X \in R^{m \times n}$, the ABS method [1] is a class of algorithms aimed at solving the linear system $Xz = b$. The purpose of this section is to show the relation of the ABS process to the rank-reducing process. In particular we point out a linear algebra process for which X is fixed at the outset, Y is arbitrary and the matrix being rank-reduced is not the focus of attention and in fact is arbitrary. The ABS process has been derived by analogy of the quasi-Newton methods under the condition that the approximation $z^{(i+1)}$ at the i^{th} iteration solves the first i equations. Let x_i^T be the i^{th} row of X and b_i the i^{th} component of b. A basic ABS method can be described as follows:

ALGORITHM 5.1. (Basic ABS Algorithm)
Choose $z^{(1)} \in R^n$ *and* $A_1 \in R^{n \times n}$ *arbitrarily.*
for $i = 1 : m$,
$\tau_i := x_i^T z^{(i)} - b_i$
if $A_i x_i \neq 0$,

$$p_i := A_i^T g_i \quad \text{(Search Direction)} \tag{46}$$

$$z^{(i+1)} := z^{(i)} - \frac{\tau_i}{x_i^T p_i} p_i \quad \text{(Line Search)} \tag{47}$$

$$A_{i+1} := A_i - A_i x_i y_i^T A_i \quad \text{(Matrix Update)} \tag{48}$$

else,
if $\tau_i = 0$, (The system is redundant)

$$z^{(i+1)} := z^{(i)}$$
$$A_{i+1} := A_i$$

else,
stop (The system is inconsistent)
end
end
end

In (46) and (48) the vectors g_i and y_i are chosen arbitrarily in R^n except subject to the conditions that $g_i^T A_i x_i \neq 0$ and

$$y_i^T A_i x_i = 1. \tag{49}$$

From (49), it follows that (48) determines a special rank-reducing process defined by (2). It is worth mentioning that (48) was introduced to mandate that $z^{(i+1)}$ solves the first i equations subject to the lowest possible rank change of A_i.

An interesting comparison is that the ABS method emphasizes the three free parameters A_1, the vectors g_i and y_i at the i^{th} stage, while the general rank-reducing process emphasizes two Wedderburn vectors, free parameters x_i and y_i. The vector parameter g_i in (46) is independent of the rank-reducing process. On the other hand, it is interesting to note that the initial matrix A_1, arbitrary in the ABS method, is normally fixed and the main focus of attention in a rank-reducing process, whereas the parameter x_i, free in the general rank-reducing process, is fixed to be the i^{th} row of the system to be solved in the ABS method.

An algorithm formally associated with the ABS class corresponds to the choices $A_1 = I$, $g_i = x_i$ and $y_i = x_i / x_i^T A_i x_i$ [1, 13]. But with $A_1 = I$, the Wedderburn process (See Theorem 2.1) becomes precisely the Gram-Schmidt process since the bilinear form (5) is exactly the conventional inner product in R^n. From the definition of x_i, the decomposition (19) becomes $X^T = UR$ which, after normalizing the columns of U, is the QR decomposition of X^T, see Theorem 3.8.

6. Summary. We have demonstrated a biconjugation process, based on the Wedderburn rank-one reduction formula, that provides a common framework under which well-known factorizations of matrices can be derived by selecting different values for the parameters X_γ and Y_γ. The following table provides a compendium of the different cases we have discussed. For simplicity, we have assumed that the pairs (X_γ, Y_γ) in the table are biconjugatable. Theorem 2.4 and Theorem 3.7 provide characterizations of biconjugation.

A	X_γ	Y_γ	Factor./Alg.	Ref.
I_n	from system	X_γ	Gram-Schmidt	§2&5
$m \times n$	upper trapez., arbitrary	upper trapez., arbitrary	trapez. LDU	Thm. 3.1
$n \times n$	I_n	I_n	LDM^T	Thm. 3.3
$n \times n$, $A = A^T$	arbitrary	X_r	Congruence	Thm. 3.4
$n \times n$, SPD	I_n	I_n	Cholesky	Cor. 3.5
$n \times n$	I_n	A	QR	Thm. 3.8
$m \times n$	r. sing. vec.	l. sing. vec.	SVD	Thm. 3.9
$n \times n$	residual of A	residual of A^T	Lanczos	Thm. 4.2
arbitrary	from system	arbitrary	ABS	§5

7. Acknowledgements. We would like to specially acknowledge the late Alston Householder's strong influence on this research. When Bob Funderlic went to him in the late 1960's for a possible dissertation topic, Professor Householder suggested putting various norm conditions on the Wedderburn rank-reducing formula to obtain analogous decompositions to the SVD. That is if $\| Ax \| = \| A \|$, the latter being the operator norm corresponding to the former vector norm, with $\| x \| = 1$, then $\omega_1 := \| A \| = y^T Ax$ for some vector y such that the dual norm of y is one. The pursuit of Householder's idea resulted in an ellipsoidal norm generalization of the SVD [7], numerous characterizations of rank subtractivity and their relation to generalized inverses [3], characterizations of generalized von Neumann symmetric (gauge) norms [7] and orthant monotonic norms [8], and four of the unpublished results [7] of this paper, i.e. Theorems 2.1 and 3.9, and Corollaries 2.5 and 2.2. The latter corollary which provides biconjugate decompositions was motivated by the need to eliminate the appearance of Wedderburn matrices in order to make possible generalized SVD decompositions. It should perhaps not be considered too much a disappointment that nonlinearity introduced by norms other than ellipsoidal ones acted to preclude additional norm generalizations of the SVD.

It is an additional tribute to Professor Householder that another of his students, Professor Stewart [16], saw the connection between conjugate direction methods and an important characterization of conjugation [16, Theorem 3.1]. Stewart's result naturally suggested our central biconjugation result, Theorem 2.4, that unifies much of the research of this paper. Dr. Michele Benzi and Professor Carl Meyer noticed the connection of our work with Stewart's [16] and we thank them for calling our attention to it. Dr. Benzi also pointed out the relation of the Cholesky factorization of Corollary 3.5 to the work of Fox, Huskey and Wilkinson [6] and that of Hestenes and Stiefel [12]. Related work by Dr. Benzi and Professor Meyer (associated with preconditioners among other issues) will appear in forthcoming publications.

Finally, thanks go to the two anonymous referees for their professional and painstaking work. They suggested the summary section with its table and made other valuable suggestions that clarified the presentation.

REFERENCES

[1] J. Abaffy and E. Spedicato, ABS Projection Algorithms, Mathematical Techniques for Linear and Nonlinear Equations, Halsted Press, Chichester, 1989.

[2] M. T. Chu and J. L. Watterson, Multivariate eigenvalue problem, Part I: Algebraic Theory and a Power method, SIAM J. Sci. Comput., 14(1993), 1089-1106.

[3] R. E. Cline and R. E. Funderlic, The rank of a difference of matrices and associated generalized inverses, Linear Alg. Appl., 24(1979), 185-215.

[4] E. Egerváry, On rank-diminishing operators and their applications to the solution of linear equations, Z. Angew. Math. Phys., 11(1960), 376-386.

[5] R. Fletcher, Conjugate gradient methods for indefinite systems, Proc. Dundee Conference Numer. Anal., ed. G. A. Watson, Springer-Verlag, New York, 1976, 73-89.

[6] L. Fox, H. D. Huskey and J. H. Wilkinson, Notes on the solution of algebraic linear simultaneous equations, J. Mech. Appl. Math., 1(1948), 149-173.

[7] R. E. Funderlic, Norms and semi-inverses, Ph.D. Thesis, University of Tennessee, Knoxville, Tennessee; Union Carbide Report CTC-35, Oak Ridge, Tennessee, 1970.

[8] R. E. Funderlic, Some characterizations of orthant monotonic norms, Linear Alg. Appl., 28(1979), 77-83.
[9] J. S. Frame, Matrix operations and generalized inverses, IEEE Spectrum, 1(1964), 209-220.
[10] G. H. Golub and C. F. Van Loan, Matrix Computations, 2nd ed., Johns Hopkins University Press, Baltimore, 1989.
[11] M. H. Gutknecht, The unsymmetric Lanczos algorithms and their relations to Padé approximation, continued fractions, and the qd algorithm, Preliminary Proceedings of the Copper Mountain Conference on Iterative Methods, 1990.
[12] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Stand., 49(1952), 409-436.
[13] H. Y. Huang, A direct method for the general solution of a system of linear equations, J. Optim. Theory Appl., 16(1975), 429-445.
[14] A. S. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell, New York, 1964; and Dover, New York, 1975.
[15] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Nat. Bur. Stand., 45(1950), 255-282.
[16] G. W. Stewart, Conjugate direction methods for solving systems of linear equations, Numer. Math., 21(1973), 285-297.
[17] G. W. Stewart and J. Sun, Matrix Perturbation Theory, Academic Press, 1990.
[18] J. H. M. Wedderburn, Lectures on Matrices, Colloquium Publications, vol. XVII, American Mathematical Society, New York, 1934; and Dover, New York 1964.