RESTARTING THE NON-SYMMETRIC LANCZOS ALGORITHM VIA THE IMPLICITLY SHIFTED LR ALGORITHM*

P. S. NEGI† AND C. ARRATIA‡

Abstract. The shifted QR iteration is used as a restart procedure for the Arnoldi method for the calculation of a few dominant eigenvalues of a large matrix. We show that the underlying idea can be utilized in much the same manner via the shifted LR iteration to create a restart procedure for the non-symmetric Lanczos algorithm for eigenvalue calcuations. Additionally, we show that the (shifted) LR iteration can be performed implicitly in a manner similar to the Francis' algorithm, resulting in a bulge-chase type procedure which does not require the explicit construction of the full lower and upper triangular matrices.

Key words. LR algorithm, unsymmetric Lanczos, implicit restart

MSC codes. 68Q25, 68R10, 68U05

1. Introduction.

The Arnoldi iteration [1] is a popular Krylov space method for calculating a few eigenvalues of a large matrix. The method relies on the generation of a sequence of Krylov vectors which determine the subspace within which approximations of the eigenvalue-eigenvector pairs are obtained. Depending on the accuracy and number of eigenpair approximations needed, the Krylov space size can become exceedingly large so that the quality of the results may be limited by the available memory. Sorensen [9] introduced an elegant procedure for restarting the Arnoldi factorization based on polynomial filters, which are applied through the implicitly shifted QR iterations on the reduced Hessenberg matrix obtained through the Arnoldi method. In particular, the use of exact shifts was shown to be successful in the convergence process of the eigenspace[9] of the specified eigenvalues. The method has subsequently found widespread application through the ARPACK library [4]. The use of QR iterations ensures that the reduced matrix preserves its Hessenberg structure through the transforms that make up the restart process. If the underlying matrix is symmetric, the Arnoldi iteration reduces to the Lanczos algorithm and, the Hessenberg matrix reduces to a symmetric tridiagonal matrix. The QR iteration preserves the symmetric tridiagonal structure as well and, as pointed out by Sorensen [9], the implicit restart process applies equally well for the Lanczos method for symmetric matrices.

One would like to extend this procedure to the case of the non-symmetric Lanczos method. However, the reduced matrix that one obtains is a non-symmetric tridiagonal matrix, with the tridiagonal structure being the result of the recurrence relations of the Lanczos algorithm [8]. Since the QR iterations do not preserve the banded structure of non-symmetric matrices, a straightforward application of the restart procedure propounded by Sorensen will lead to a loss of this tridiagonal structure of the reduced matrix (the Hessenberg structure will still be preserved). This loss of structure can be circumvented if one looks to the predecessor of the QR algorithm namely, the LR algorithm proposed by Rutihauser [6, 7, 5], which has the attractive property of preserving the band structure of a matrix. This property was already pointed out by Rutihauser in [6] where the banded matrices were referred to as striped matrices.

*Submitted to the editors DATE.

Funding: This work was funded by the Fog Research Institute under contract no. FRI-454.

†Imagination Corp., Chicago, IL (prabal.negi@su.se, http://www.imag.com/~ddoe/).

‡Department of Applied Mathematics, Fictional University, Boise, ID (cristobal.arratia@su.se).

As we will show in the next section, shifted LR transforms is the appropriate generalization of the restart procedure to the case of non-symmetric Lanczos iteration. The process would necessarily require refining both the right as well as the left Krylov spaces simultaneously.

The rest of the paper is organized as follows. In the next section we start with the introduction of the non-symmetric Lanczos iteration and then develop the restart procedure. We also show that the LR algorithm can be implemented as a bulge-chase method, similar to the Francis' algorithm. In section 3 we apply the restart process to the Grcar matrix, and make some concluding remarks in section 4.

2. Non-symmetric Lanczos.

Lanczos first introduced his algorithm in [3] as a method for tridiagonalizing a matrix, but also realized that the method could be used iteratively to find eigenvalues. For an arbitry matrix A, the method generates a pair of Krylov subspaces $\{v_1, \dots, v_j\}$ and $\{w_1, \dots, w_j\}$, through repeated action of A and A^H respectively. We refer to these as the right and left Krylov spaces respectively and they satisfy the biorthogonality relation $w_i^H v_j = \delta_{ij}$. The two subspaces are generated through a three term recurrence relation which, for a Krylov space of size m, can be written in matrix form as

$$AV_m = V_m T_m + v_{m+1} e_m^T \tag{2.1a}$$

$$A^H W_m = W_m T_m^H + w_{m+1} e_m^T \tag{2.1b}$$

$$W_m^H V_m = I_m \tag{2.1c}$$

where I_m represents the Identity matrix of size m, T_m is a tri-diagonal matrix of size m and, T_m^H is the Hermitian conjugate of T_m. If either v_{m+1} or w_{m+1} vanishes it represents the convergence of the right or the left Krylov subspaces to an invariant subspace of dimension m. A more serious breakdown occurs if $w_{m+1}^H v_{m+1} = 0$ with both $v_{m+1} \neq 0$ and $w_{m+1} \neq 0$ however, we do not address that issue here. We refer the reader to [2] for a comprehensive overview on Lanczos type solvers and the related issues of breakdown.

As Sorensen points out for the Arnoldi method [9], if one is interested in an invariant subspace of dimension m, the starting vector of Krylov subspace must not contain components of the generator of a cyclic subspace of dimension greater than m. This applies equally for the right and left Krylov subspaces generated through the Lanczos recurrence relations. Hence a non-vanishing v_{m+1} (respectively w_{m+1}) implies that v_1 (respectively w_1) contains components of an invariant subspace of dimension greater than m. The idea behind restarts then is to discard the components of the starting vector v_1 (and w_1) along the unwanted dimensions, such that each restart process moves the Krylov space(s) closer to being invariant. For the Arnoldi method Sorensen [9] proposed to achieve this via polynomial filtering, *i.e.* replacing

$$v_1 \leftarrow \psi(A) v_1, \tag{2.2a}$$

$$\psi(\lambda) = (1/\tau) \Pi_{j=1}^{p} (\lambda - \mu_j). \tag{2.2b}$$

Obviously $\psi(\lambda)$ is the filtering polynomial, τ is a normalization constant and each μ_j specifies a node of the polynomial. The polynomial acts on v_1 to filter out the part of the spectrum of A that is close to each μ_j. If a particular μ_j corresponds to an exact eigenvalue of A, then components of the corresponding eigenvector are completely filtered out from v_1. The node μ_j is referred to as a shift since the application of the polynomial filtering relies on the shifted QR algorithm, where μ_j is used as the shift. As shown below for the case of a single shift, an analogous procedure can be followed using a shifted LR algorithm which achieves the same effect of applying a polynomial filter to the starting vector v_1. Starting with the Lanczos relation for the right subspace 2.1a, and adding and subtracting μV_m we obtain

$$(A-\mu I)V_m - V_m(T_m - \mu I) = v_{m+1}e_m^T \tag{2.3a}$$
$$(A-\mu I)V_m - V_m(L_1R_1) = v_{m+1}e_m^T \tag{2.3b}$$
$$(A-\mu I)V_mL_1 - V_m(L_1R_1)L_1 = v_{m+1}e_m^TL_1 \tag{2.3c}$$
$$A(V_mL_1) - (V_mL_1)(R_1L_1 + \mu I) = v_{m+1}e_m^TL_1 \tag{2.3d}$$
$$AV_m' - V_m'T_m' = v_{m+1}e_m^TL_1. \tag{2.3e}$$

Here we have set $V_m' = V_mL_1$ and $T_m' = (R_1L_1 + \mu I)$. The matrices L_1, R_1 are the lower and upper triangular matrices obtained from the LU decomposition of $(T_m - \mu I)$. The matrix L_1 can be required to be unit triangular (all entries on the main diagonal are ones), in which case the LU decomposition is unique. Furthermore, L_1 for a tridiagonal matrix only consists of one sub-diagonal (in addition to the main diagonal). One can easily recognize that the new reduced matrix T_m' is a result of one step of the shifted LR iteration. Hence T_m' retains the tridiagonal structure of the T_m [6, 5]. The relationship between generating vectors of the two spaces V_m and V_m' can be obtained by multiplying 2.3b by e_1, *i.e.*

$$\begin{aligned}(A-\mu I)V_me_1 - (V_m')R_1e_1 &= v_{m+1}e_m^Te_1\\ \implies (A-\mu I)v_1 &= v_1'\rho_{11},\end{aligned}$$

where $\rho_{11} = e_1^TR_1e_1$. This clearly shows the filtering operation done on the original vector v_1 to generate the new vector v_1'.

Since the Lanczos method creates a biorthogonal basis, one must simultaneously transform the left basis W_m to maintain the biorthogonality property. It is easy to see that the necessary transform to maintain biorthogonality is $W_m' = W_mL_1^{-H}$, since,

$$W'^HV_m' = (W_mL_1^{-H})^H(V_mL_1) = L_1^{-1}(W_m^HV_m)L_1 = I$$

We can obtain the modified Lanczos relation for the left Krylov space as

$$A^HW_m - W_m(R_1^HL_1^H + \bar{\mu}I) = w_{m+1}e_m^T \tag{2.4}$$
$$A^H(W_mL_1^{-H}) - W_m(L_1^{-H}L_1)(R_1^HL_1^H + \bar{\mu}I)L_1^{-H} = w_{m+1}e_m^TL_1^{-H} \tag{2.5}$$
$$A^H(W_mL_1^{-H}) - (W_mL_1^{-H})(L_1^HR_1^H + \bar{\mu}I) = w_{m+1}e_m^TL_1^{-H} \tag{2.6}$$
$$A^HW_m' - W_m'(T_m')^H = w_{m+1}e_m^TL_1^{-H}. \tag{2.7}$$

Conveniently the structure of the Lanczos iteration for the left Krylov space is also preserved. Noting that L_1^{-H} is upper triangular, one can again expose the relationship between the generating vectors of the two left Krylov spaces as

$$\begin{aligned}W'e_1 &= (W_mL_1^{-H})e_1\\ \implies w_1' &= w_1(e_1^TL_1^{-H}e_1).\end{aligned}$$

Clearly w_1' is simply a scalar multiple of the old vector w_1 and no filtering of the generating vector has occurred. In order to ensure that we filter the left Krylov space as well, we perform one step of the shifted LR iteration with the conjugated shift $\bar{\mu}$ on the reduced matrix $(T_m')^H$ obtained in equation 2.7. The right Krylov space is modified again to maintain orthogonality.

$$(A - \mu I)^H W'_m - W'_m (T'_m - \mu I)^H = w_{m+1} e_m^T L_1^{-H} \tag{2.8a}$$
$$(A - \mu I)^H W'_m - W'_m (L_2 R_2) = w_{m+1} e_m^T L_1^{-H} \tag{2.8b}$$
$$(A - \mu I)^H W'_m L_2 - W'_m (L_2 R_2) L_2 = v_{m+1} e_m^T L_1^{-H} L_2 \tag{2.8c}$$
$$A(W'_m L_2) - (W'_m L_2)(R_2 L_2 + \bar{\mu} I) = v_{m+1} e_m^T L_1^{-H} L_2 \tag{2.8d}$$
$$A W''_m - W''_m (T''_m)^H = w_{m+1} e_m^T L_1^{-H} L_2. \tag{2.8e}$$

One may again obtain the relation between the starting vectors as

$$(A^H - \bar{\mu} I) w'_1 = w''_1 (e_1^T R_2 e_1). \tag{2.9}$$

Obviously the appropriate transform of the right subspace to maintain orthogonality is $V''_m = V'_m L_2^{-H} = V_m L_1 L_2^{-H}$. Again, noting that L_2^{-H} is upper triangular implies that the new v''_1 is simply the scalar multiple of v'_1 and no additional filtering occurs for v_1 in this step. One can write the modified Lanczos relation as

$$A V''_m - V''_m T''_m = v_{m+1} e_m^T L_1 L_2^{-H}. \tag{2.10}$$

The above process can be repeated for p unwanted shifts. We Denote by $L_1^p = L_{11} L_{12} \ldots L_{1p}$ as the product of the lower triangular matrices generated due to p shifted-LR steps for the right Krylov space V_m, and by $L_2^p = L_{21} L_{22} \ldots L_{2p}$ as the product of the lower triangular matrices due to the p shifted-LR iterations for the left Krylov space. Then for a Krylov space size of $m = k + p$ we have two modified Lanczos relations

$$A V''_{k+p} - V''_{k+p} T''_{k+p} = v_{k+p+1} e_{k+p}^T L_1^p (L_2^p)^{-H} \tag{2.11a}$$
$$A^H W''_{k+p} - W''_{k+p} (T''_{k+p})^H = w_{k+p+1} e_{k+p}^T (L_1^p)^{-H} L_2^p. \tag{2.11b}$$

We may take a closer look at the structure of the residual matrices on the right hand side of equation (2.11a). L_1^p is a product of p matrices that are lower triangular with just one subdiagonal. L_1^p then is lower triangular with p non-zero subdiagonals. $(L_2^p)^{-H}$ is upper triangular and the product $L_1^p (L_2^p)^{-H}$ therefore has p non-zero subdiagonals. Left multiplication by e_{k+p}^T therefore has the form

$$e_{k+p}^T L_1^p (L_2^p)^{-H} = (\underbrace{0, 0 \ldots, \theta_{k+p}}_{k}, \underbrace{b_1^T}_{p})$$

where, $\theta_{k+p} = e_{k+p}^T (L_1^p (L_2^p)^{-H}) e_k$. Therefore matrix on the right hand side of (2.11b) has zeros in the first $k-1$ columns and the k^{th} column is simply $\theta_{k+p} v_{k+p+1}$. The remaining columns are non-zero in general.

A very similar structure is obtained for the residual matrix in the right hand side of equation (2.11b) with zeros in the first $k-1$ columns and the k^{th} column being equal to $\phi_{k+p} w_{k+p+1}$, with $\phi_{k+p} = e_{k+p}^T ((L_1^p)^{-H} L_2^p) e_k$

Partitioning the matrices such that

$$V''_{k+p} = (V''_k, V''_p), \qquad T''_{k+p} = \begin{pmatrix} T''_k & \delta_{k+1} e_k e_1^T \\ \beta_{k+1} e_1 e_k^T & T''_p \end{pmatrix}, \tag{2.12a}$$
$$W''_{k+p} = (W''_k, W''_p), \quad (T''_{k+p})^H = \begin{pmatrix} (T''_k)^H & \bar{\beta}_{k+1} e_k e_1^T \\ \bar{\delta}_{k+1} e_1 e_k^T & (T''_p)^H \end{pmatrix}, \tag{2.12b}$$

with the length of the e_i vectors understood to be such that the resulting matrices are consistent. We can write the modified Lanczos relations of (2.11b) and (2.11a) as

$$A(V_k'', V_p'') = (V_k'', V_p'') \begin{pmatrix} T_k'' & \delta_{k+1} e_k e_1^T \\ \beta_{k+1} e_1 e_k^T & T_p'' \end{pmatrix} + \left(\theta_{k+p} v_{k+p+1} e_k^T, M_v\right),$$

$$A^H(W_k'', W_p'') = (W_k'', W_p'') \begin{pmatrix} (T_k'')^H & \bar{\beta}_{k+1} e_k e_1^T \\ \bar{\delta}_{k+1} e_1 e_k^T & (T_p'')^H \end{pmatrix} + \left(\phi_{k+p} w_{k+p+1} e_k^T, M_w\right),$$

Finally, equating the individual columns on both sides and discarding columns $k+1, \ldots, k+p$ we are left with the new Krylov spaces of order k and the Lanczos relations

$$AV_k'' - V_k'' T_k'' = v_{k+1}'' e_k^T, \tag{2.13a}$$

$$A^H W_k'' - W_k'' (T_k'')^H = w_{k+1}'' e_k^T, \tag{2.13b}$$

$$(W_k'')^H V_k = I. \tag{2.13c}$$

The new residual vectors are defined as

$$v_{k+1}'' = \beta_{k+1} V_p'' e_1 + \theta_{k+p} v_{k+p+1} \tag{2.14a}$$

$$w_{k+1}'' = \bar{\delta}_{k+1} W_p'' e_1 + \phi_{k+p} w_{k+p+1} \tag{2.14b}$$

Acknowledgments. We would like to acknowledge the assistance of volunteers in putting together this example manuscript and supplement.

REFERENCES

[1] W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quarterly of applied mathematics, 9 (1951), pp. 17–29.
[2] M. H. Gutknecht, *Lanczos-type solvers for nonsymmetric linear systems of equations*, Acta numerica, 6 (1997), pp. 271–397.
[3] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, (1950).
[4] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*, SIAM, 1998.
[5] H. Rutihauser, *Lectures on Numerical Mathematics*, Birkhauser, 1990.
[6] H. Rutishauser, *Solution of eigenvalue problems with the lr-transformation*, National Bureau of Standards, Applied Mathematics Series, 49 (1958), pp. 47–81.
[7] H. Rutishauser and H. Schwarz, *Thelr transformation method for symmetric matrices*, Numerische Mathematik, 5 (1963), pp. 273–289.
[8] Y. Saad, *The lanczos biorthogonalization algorithm and other oblique projection methods for solving large unsymmetric systems*, SIAM Journal on Numerical Analysis, 19 (1982), pp. 485–506.
[9] D. C. Sorensen, *Implicit application of polynomial filters in ak-step arnoldi method*, Siam journal on matrix analysis and applications, 13 (1992), pp. 357–385.