Model Reduction of State Space Systems via an Implicitly Restarted Lanczos Method

E.J. Grimme
D.C. Sorensen
P. Van Dooren

CRPC-TR94458
May 1994

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

This work was supported in part by ARPA, the Department of Energy, and the NSF.

Model reduction of state space systems via an implicitly restarted Lanczos method *

E.J. Grimme
Coordinated Science Lab.
University of Illinois at
Urbana-Champaign
Urbana, IL 61801

D.C. Sorensen
Computational and Applied
Mathematics Department
Rice University
Houston, TX 77251

P. Van Dooren
Coordinated Science Lab.
University of Illinois at
Urbana-Champaign
Urbana, IL 61801

Abstract

The nonsymmetric Lanczos method has recently received significant attention as a model reduction technique for large-scale systems. Unfortunately, the Lanczos method may produce an unstable partial realization for a given, stable system. To remedy this situation, inexpensive implicit restarts are developed which can be employed to stabilize the Lanczos generated model.

AMS classification: Primary 65F15; Secondary 65G05

Key Words : State space systems, model reduction, nonsymmetric Lanczos method, eigenvalues, implicit restarting .

1 Introduction

This paper employs a modified Lanczos method to acquire a stable reduced order model for a SISO (single input-single output) system described by the state space equations

$$\dot{x} = Ax + bu \tag{1}$$
$$y = cx + du. \tag{2}$$

It will be assumed throughout the following that the system matrix, $A \in \mathbb{R}^{n\times n}$, is large, sparse and stable. Such an A can arise, for example, out of finite-element discretizations of various plants including chemical processes and mechanical systems.

Conventional model reduction techniques are ill-suited for large, sparse problems due to the sheer size of A. For example, many of the "optimal" reduction strategies (balanced realization [25], Hankel norm optimal [16], etc.) require knowledge of the solutions to the Lyapunov equations

$$AG_c + G_cA^T + bb^T = 0 \quad \text{and} \quad A^TG_o + G_oA + c^Tc = 0. \tag{3}$$

Standard computational techniques for solving (3) entail $O(n^3)$ operations [3]. As $n \gg 100$ in the large sparse problem, these standard model reduction techniques are not practical in general.

As an alternative to such model reduction techniques, this paper employs an oblique Krylov projector, $\pi_k = V_kW_k^T$, to produce a k^{th} order model

$$\dot{\hat{x}} = (W_k^TAV_k)\hat{x} + (W_k^Tb)u = \hat{A}\hat{x} + \hat{b}u$$
$$\hat{y} = (cV_k)\hat{x} + du = \hat{c}\hat{x} + du$$

*This work was supported in part by ARPA (U.S. Army ORA4466.01), by ARPA (Grant 60NANB2D1272), by the Department of Energy (Contract DE-FG0f-91ER25103) and by the National Science Foundation (Grants CCR-9209349 and CCR-9120008.)

for the original system in (1) and (2). The matrices $V_k \in \mathbb{R}^{n\times k}$ and $W_k \in \mathbb{R}^{n\times k}$ are bi-orthogonal, i.e. $W_k^T V_k = I_k$. Moreover, V_k and W_k are related to Krylov spaces, $\mathcal{K}_k$, in that

$$\text{COLSP}(V_k) = \mathcal{K}_k(A, v_1) = \text{span}\{v_1, Av_1, \ldots, A^{k-1}v_1\} \tag{4}$$

$$\text{COLSP}(W_k) = \mathcal{K}_k(A^T, w_1) = \text{span}\{w_1, A^T w_1, \ldots, A^{k-1^T} w_1\}. \tag{5}$$

The utility of this Krylov projector comes from the fact that both V_k and W_k can be generated with only inner-products and matrix-vector multiplications. By taking advantage of the fact that the A matrix is sparse, one can compute the projector relatively cheaply.

But regardless of how quickly π_k can be computed, one is certainly also interested in the correspondence between the original system $\{A, b, c\}$ and reduced-order system $\{\hat{A}, \hat{b}, \hat{c}\}$. A major insight into this relationship comes from [15, 35].

Theorem 1 *Let the reduced order system $\{\hat{A}, \hat{b}, \hat{c}\}$ be a restriction of the system $\{A, b, c\}$ by the projector, π_k, where V_k and W_k are defined as in (4) and (5). If the starting vectors, v_1 and w_1, are parallel to b and c^T respectively, then the first $2k$ Markov parameters of the original and reduced-order systems are identical, i.e.,*

$$cA^{i-1}b = \hat{c}\hat{A}^{i-1}\hat{b}$$

for $1 \le i \le 2k$.

Restating Theorem 1, the reduced order model is a Padé approximation (partial realization) of the original system.

Model reduction via Padé approximation (explicit moment matching) has a long history in the literature [29, 30, 37]. Thus the observations of [35] are certainly of interest. But the concept of using oblique projectors to perform the Padé approximation can be taken one step further by forming V_k and W_k via a two-sided, nonsymmetric Lanczos method [24]. The Lanczos algorithm simultaneously computes the projector, π_k, and a tridiagonal $\hat{A}$ with only $O(k^2 n)$ operations. Employing the Lanczos method for model reduction is discussed in a multitude of recent papers including [1, 4, 23, 32, 33, 34].

Model reduction via a Krylov projector is certainly cheaper, $O(k^2 n)$, than with an "optimal" reduction technique, $O(n^3)$, when $n \gg k$. However, Padé approximation and more specifically Lanczos-based model reduction is known to suffer from three significant disadvantages.

1. Singularities in the Padé table (ill-conditioned leading submatrices in the system's Hankel matrix) lead to "serious" breakdowns in the nonsymmetric Lanczos algorithm [27].

2. Using Koenig's theorem [21], it can be shown that matching Markov parameters (corresponding to an expansion of $G(s)$ about $s = \infty$) leads to a reduced-order model $\{\hat{A}, \hat{b}, \hat{c}\}$ which tends to approximate the high frequency poles of A [22]. Thus one can expect the transient response of the reduced-order model to closely follow that of the original system. On the other hand, the steady state error will be large in general.

3. The fact that the original system is stable does not insure that the Padé reduced-order model is stable. The ℓ_2 norm of the response error $y(t) - \hat{y}(t)$ is then unbounded [8].

At least for the SISO case, the first problem can be solved. Look-ahead can be incorporated into the Lanczos method [19, 20, 13] to avoid singularities in the Padé table. And in many of the above references, the second problem is handled by moment matching about multiple frequencies [2]. For example, incorporating information from the Krylov spaces $\mathcal{K}(A^{-1}, b)$ and $\mathcal{K}(A^{-T}, c^T)$ into the projector leads to moment matching about $s = 0$. However, this paper will not dwell on these first two issues (although the second topic especially is in need of further work). Rather, we will concentrate on the stability of the reduced-order model. Note that this paper is not the first to do so. In [33, 35], the stability of the reduced-order model is insured by incorporating an inverted grammian, G_c^{-1} into the Krylov projector. However, solving for G_c from (3) and

factoring it both require $O(n^3)$ arithmetic operations. The cost is not acceptable for large scale problems and would overwhelm the computational advantage of the Lanczos-based model reduction for large n.

As an alternative, this paper addresses the stability issue by modifying the choice for the projector. If the results with the projector, π_k, are unstable, a related projector, $\bar{\pi}_k = \bar{V}_k \bar{W}_k^T$, is selected which corresponds to the new starting vectors

$$\bar{v}_1 = \zeta_{\bar{v}}(A - \mu_p I)(A - \mu_{p-1} I)\dots(A - \mu_1 I)v_1 = \zeta_{\bar{v}} \Psi_p(A) v_1 \quad (6)$$

$$\bar{w}_1 = \zeta_{\bar{w}}(A^T - \mu_p I)(A^T - \mu_{p-1} I)\dots(A^T - \mu_1 I)w_1 = \zeta_{\bar{w}} \Psi_p(A^T) w_1. \quad (7)$$

In Section 2, a new and inexpensive technique, implicitly restarting the Lanczos algorithm, is developed for directly generating this modified projector, $\bar{\pi}_k$, from π_k. Analogous to the implicitly restarted Arnoldi method of [31], this approach incorporates shifted HR steps [6, 7] into the nonsymmetric Lanczos method to produce $\bar{\pi}_k$. In Section 3, the relationship between the modification of the projector and the resulting reduced-order model is explored. A strategy is developed for choosing the parameters, μ_i, in (6,7) to stabilize the reduced-order model. In Section 4, the numerical behavior of the implicitly restarted Lanczos method is examined in more detail. In particular, the efficiency and accuracy of implicit restarts is shown to be superior to explicit techniques for computing $\bar{\pi}_k$.

2 An Implicitly Restarted Lanczos Method

The degree of success achieved in applying a Lanczos-type method to a problem is dependent upon the choice of starting vectors, v_1 and w_1. In some cases, such as the model reduction problem, one can make an educated initial guess for these starting vectors ($v_1 = b/\beta_1$ and $w_1 = c^T/\gamma_1$). However, as we shall demonstrate, there is considerable likelihood for an unstable model to arise from a stable system. The obvious choice of starting vectors to construct a model reduction may yield disastrous results. To overcome the consequences of a poor starting vector, one could explicitly compute a projector, $\bar{\pi}_k$, from $\bar{v}_1$ and $\bar{w}_1$ by performing k additional Lanczos steps. However in this section, an implicit approach is developed for generating the modified projector, $\bar{V}_k \bar{W}_k^T$, corresponding to (6,7). Given V_k and W_k, one can implicitly pass almost immediately to $\bar{V}_k$ and $\bar{W}_k$. Because of this fact, implicit restarts are more economical than explicit ones. Also, one can typically expect a higher precision in the results of the implicit method. These statements are offered at this point only to motivate the need for implicit restarts. A more complete discussion of these two observations is postponed until Section 4.

The approach taken for implicitly restarting the Lanczos method is completely analogous to one developed in [31] for the Arnoldi method. In [31], QR steps (see [17]) are combined with the Arnoldi method to yield an implicitly restarted approach. In this section, a process denoted as the HR step is incorporated into the nonsymmetric Lanczos method in order to yield Lanczos restarts. As opposed to the implicit Lanczos restarts of [9] for the symmetric case, it should be stressed that the techniques developed below are for the nonsymmetric Lanczos method. Wherever the Lanczos method is mentioned in the following, the nonsymmetric version should be assumed.

2.1 The standard Lanczos method

Before exploring restarts, a brief review of the standard nonsymmetric Lanczos algorithm is provided. This subsection primarily establishes notation. For a more detailed discussion of the algorithm (including breakdown free variants), the reader is referred to [12, 17, 19, 20]. A standard implementation of the Lanczos method is provided as Algorithm 1.

Algorithm 1

1. *Given b and c put $\beta_1 = \sqrt{|b^T c|}$ and $\gamma_1 = sign(b^T c)\beta_1$. Initiate the starting vectors as $v_1 = b/\beta_1$ and $w_1 = c^T/\gamma_1$. Define $v_0 = w_0 = 0$.*

2. *For* $j = 1$ *to* k,

   (a) *set* $\alpha_j = w_j^T A v_j$.

   (b) *set* $r_j = Av_j - \alpha_j v_j - \gamma_j v_{j-1}$ *and* $q_j = A^T w_j - \alpha_j w_j - \beta_j w_{j-1}$.

   (c) *set* $\beta_{j+1} = \sqrt{|r_j^T q_j|}$ *and* $\gamma_{j+1} = sign(r_j^T q_j) \cdot \beta_{j+1}$

   (d) *set* $v_{j+1} = r_j/\beta_{j+1}$ *and* $w_{j+1} = q_j/\gamma_{j+1}$.

Given v_1 and w_1, the Lanczos algorithm produces the matrices $V_k = [v_1, \ldots, v_k] \in \mathbb{R}^{n \times k}$ and $W_k = [w_1, \ldots, w_k] \in \mathbb{R}^{n \times k}$ which satisfy the recursive identities

$$AV_k = V_k T_k + \beta_{k+1} v_{k+1} e_k^T \tag{8}$$

$$A^T W_k = W_k T_k^T + \gamma_{k+1} w_{k+1} e_k^T. \tag{9}$$

The vector e_k is the k^{th} standard basis vector and

$$T_k = \begin{bmatrix} \alpha_1 & \gamma_2 & & \\ \beta_2 & \ddots & \ddots & \\ & \ddots & \ddots & \gamma_k \\ & & \beta_k & \alpha_k \end{bmatrix}$$

is a truncated reduction of A. Generally and throughout this paper, the elements β_i and γ_i are chosen so that $|\beta_i| = |\gamma_i|$ and $V_k^T W_k = I$ (bi-orthogonality). One pleasing result of this bi-orthogonality condition is that multiplying (8) on the left by W_k^T yields the relationship $W_k^T A V_k = T_k$.

It will also be convenient in the following to denote the residuals $\beta_{k+1} v_{k+1}$ and $\gamma_{k+1} w_{k+1}$ as the vectors r_k and q_k, respectively. Then the relationships

$$r_k \in \mathcal{K}_{k+1}(A, v_1) \quad \text{and} \quad q_k \in \mathcal{K}_{k+1}(A^T, w_1) \tag{10}$$

arise from the Lanczos identities in (8) and (9).

Although expressions such as (8,9,10) define the Lanczos algorithm in exact terms, there are several difficulties which appear when the method is implemented numerically. Primary among these concerns is the loss of bi-orthogonality between V_k and W_k. In theory, the three-term recurrences in (8) and (9) are sufficient to guarantee $W_k^T V_k = I$. Yet in practice, it is known [26] that bi-orthogonality will in fact be lost when at least one of the eigenvalues of T_k converges to an eigenvalues of A. Bi-orthogonality is crucial if π_k is to be a valid projector of $\{A, b, c\}$ to a reduced-order system, so the extra step of full reorthogonalization will be taken in the following. During each iteration, v_{k+1} and w_{k+1} are explicitly orthogonalized against the columns of W_k and V_k respectively via one iteration of a Gram-Schmidt process. This full reorthogonalization is costly; reorthogonalized Lanczos involves $O(k^2 n)$ operation to compute V_k and W_k. However, this extra expense can be limited if the value of k can be kept sufficiently small. We refer to Section 4 for more details on these aspects.

2.2 The HR decomposition

In [9, 31], the decomposition $(T_k - \mu I) = QR$ and the corresponding QR step, $\tilde{T}_k = Q^T T_k Q$, play a key role in implicit restarts for the symmetric Lanczos method. These transformations preserve the symmetry and tridiagonality of T_k as well as the orthogonality of the updated Lanczos basis vectors. Although, symmetry is lacking in the two-sided Lanczos process defined above, the tridiagonal matrix T_k is *sign symmetric*. It turns out to be important and elegant to develop a QR-like implicit restarting scheme based on transformations

that preserve this sign-symmetry along with the tridiagonality of the T_k and the bi-orthogonality of the basis vectors.

The transformations that turn out to be useful are known as hyperbolic rotations. The corresponding implicit shift mechanism based upon these transformations has already been developed [7] and is known as the HR algorithm. A slightly modified version of the notation of [7] will be employed to describe the sign symmetry of T_k and the properties of the hyperbolic transformations.

Definition 1 *S is a signature matrix (denoted $S \in diag(\pm 1)$) if it is diagonal with 1's and/or -1's on its diagonal.*

Definition 2 *Let $S_1, S_2 \in diag(\pm 1)$. Then $H \in \mathbb{R}^{k \times k}$ is called (S_1, S_2)-unitary if $H^T S_1 H = S_2$. If $S \equiv S_1 = S_2$ then we simply say that H is S-unitary.*

Definition 3 *Let $S \in diag(\pm 1)$. Then $M \in \mathbb{R}^{k \times k}$ is S-symmetric (pseudo-symmetric) if $M^T S = SM$.*

Because the off-diagonal elements of T_k satisfy $|\beta_i| = |\gamma_i|$ for $1 < i \leq k$, T_k is S-symmetric. The exact S corresponding to T_k can be constructed as [7]

$$S_1 = \text{diag}\left(1, \text{sign}(\frac{\gamma_2}{\beta_2}), \ldots, \text{sign}(\frac{\gamma_2}{\beta_2} \cdots \frac{\gamma_k}{\beta_k})\right). \tag{11}$$

Corresponding to this more general form of symmetry, there typically exists a factorization, denoted the HR decomposition, which is more general than the QR factorization. In particular, we are interested in the HR decomposition and corresponding iteration for the specific case of real, tridiagonal, S-symmetric matrices.

Theorem 2 *Let T_k be a tridiagonal, unreduced, S_1-symmetric matrix and let $R \in \mathbb{R}^{k \times k}$ be upper-triangular. Then*

(i) There exists an (S_1, S_2)-unitary H and an upper-triangular R such that $T_k = HR$ if and only if the principal minors of $T_k^T S_1 T_k$ are non-zero and the product of the first i diagonal elements of S_2 corresponds to the sign of the i^{th} principal minor of $T_k^T S_1 T_k$ for $1 \leq i \leq k$.

(ii) If H and R satisfying (i) exist, then $\tilde{T}_k = H^{-1} T_k H = RH$ is tridiagonal and S_2-symmetric.

(iii) If H and R satisfying (i) exist and if T_k is singular, then the last row and column of $\tilde{T}_k = RH$ are zero.

Proof: For the original statement and proof of (i), see [7].

For (ii) and (iii), assume that an (S_1, S_2)-unitary H and an upper-triangular R exist such that $T_k = HR$. Then construct partitions

$$T_k = \left[\begin{array}{c|c} T_{k,k-1} & t_k \end{array}\right], \quad H = \left[\begin{array}{c|c} H_{k,k-1} & h_k \end{array}\right] \quad \text{and} \quad R = \left[\begin{array}{c|c} R_{k-1,k-1} & r \\ 0 & r_{k,k} \end{array}\right]$$

where $T_{k,k-1}$ and $H_{k,k-1}$ are the respective first $k-1$ columns of T_k and H, and $R_{k-1,k-1}$ is the leading $(k-1) \times (k-1)$ submatrix of R. The columns of $T_{k,k-1}$ are linearly independent since T_k is unreduced (all β_i's non-zero) and the columns of $H_{k,k-1}$ are linearly independent since H is nonsingular. Hence the matrix $R_{k-1,k-1}$ must be nonsingular since $T_{k,k-1} = H_{k,k-1} R_{k-1,k-1}$. It follows that $H_{k,k-1} = T_{k,k-1} R_{k-1,k-1}^{-1}$ is upper Hessenberg and that T_k is singular if and only if the scalar $r_{k,k}$ is zero.

Since $H_{k,k-1}$ is upper-Hessenberg, H is upper Hessenberg, and thus $\tilde{T}_k = RH$ is upper-Hessenberg. And as $S_1 H = H^{-T} S_2$, $\tilde{T}_k$ is S_2-symmetric because [7]

$$\tilde{T}_k^T S_2 = H^T T_k^T H^{-T} S_2 = H^T T_k^T S_1 H = H^T S_1 T_k H = S_2 H^{-1} T_k H = S_2 \tilde{T}_k.$$

Being both upper-Hessenberg and S_2-symmetric, $\tilde{T}_k$ is tridiagonal and thus (ii) is established.

If T_k is singular then $r_{k,k} = 0$ which implies that the last row of RH is zero. Because $\tilde{T}_k = RH$ is S_2-symmetric, the last column must also be zero and (iii) is proved. □

Assuming its existence, the HR decomposition and HR step (i.e., $\tilde{T}_k = H^{-1}T_kH_k$) possesses many of the desirable properties of the QR method. For the remainder of this section, it will be assumed that the HR decomposition always exists. A discussion of the existence and stability of the HR algorithm in the context of the Lanczos algorithm is provided in Section 4.

An algorithm for computing H and R for an S-symmetric tridiagonal matrix is provided in detail in [6, 7]. To briefly sketch out this technique, first define a Givens rotation [17] as

$$H_g(i,j) = \begin{bmatrix} I & & & & \\ & c & & -s & \\ & & I & & \\ & s & & c & \\ & & & & I \end{bmatrix}$$

where the $c = \cos(\theta)$ and $s = \sin(\theta)$ entries are in the $(i,i)^{th}$, $(i,j)^{th}$, $(j,i)^{th}$ and $(j,j)^{th}$ positions. Values for c and s can always be found such that

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix}\begin{bmatrix} x_i \\ x_j \end{bmatrix} = \begin{bmatrix} \sqrt{x_i^2+x_j^2} \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x_i & x_j \end{bmatrix}\begin{bmatrix} c & -s \\ s & c \end{bmatrix} = \begin{bmatrix} \sqrt{x_i^2+x_j^2} & 0 \end{bmatrix} \tag{12}$$

for an arbitrary x_i and x_j. Thus, the rotation will annihilate the entry x_j if $c = x_i/\sqrt{x_i^2+x_j^2}$. The symmetry in both equations of (12) explains why a single Givens rotation annihilates both x_j entries. When this symmetry is only present in an S-symmetric sense, one must turn to the hyperbolic rotation [17]

$$H_h(i,j) = \begin{bmatrix} I & & & & \\ & c & & s & \\ & & I & & \\ & s & & c & \\ & & & & I \end{bmatrix}.$$

If $|x_i| > |x_j|$, the values of $c = \cosh(\theta)$ and $s = \sinh(\theta)$ are selected so that

$$\begin{bmatrix} c & -s \\ -s & c \end{bmatrix}\begin{bmatrix} x_i \\ x_j \end{bmatrix} = \begin{bmatrix} \sqrt{x_i^2-x_j^2} \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x_i & -x_j \end{bmatrix}\begin{bmatrix} c & s \\ s & c \end{bmatrix} = \begin{bmatrix} \sqrt{x_i^2-x_j^2} & 0 \end{bmatrix}. \tag{13}$$

When putting $c = x_i/\sqrt{x_i^2-x_j^2}$ and $s = x_j/\sqrt{x_i^2-x_j^2}$ both x_j entries are annihilated even though there is a sign difference between the two sides of (13) (i.e., $[x_i\ x_j]^T = \mathrm{diag}(1,-1)[x_i\ -x_j]$). Alternatively if $|x_j| > |x_i|$, one can avoid complex arithmetic by constructing the rotation $H_h(i,j)$ to be

$$\begin{bmatrix} -c & s \\ s & -c \end{bmatrix}\begin{bmatrix} x_i \\ x_j \end{bmatrix} = \begin{bmatrix} \sqrt{x_j^2-x_i^2} \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x_i & -x_j \end{bmatrix}\begin{bmatrix} -c & -s \\ -s & -c \end{bmatrix} = \begin{bmatrix} \sqrt{x_j^2-x_i^2} & 0 \end{bmatrix}. \tag{14}$$

In this case, putting $c = x_i/\sqrt{x_j^2-x_i^2}$ and $s = x_j/\sqrt{x_j^2-x_i^2}$ annihilates the x_j entries. The hyperbolic rotation is now defined everywhere except for the event $|x_j| = |x_i|$. If this occurs it will cause a breakdown of the HR decomposition (see Theorem 2 (i)). However, as stated above, it will be assumed that such breakdowns do not occur until they can be more fully addressed in Section 4.

Using the different rotations defined above, the algorithm of [6] annihilates the lower off-diagonal elements of T_k by applying $(n-1)$ rotations from the left in a technique analogous to the explicit QR step for symmetric tridiagonal matrices. The product of these rotations forms H^{-1} while $H^{-1}T_k = R$. As T_k is

only S_1-symmetric, the sign symmetry of the off-diagonal terms must be taken into consideration as the transformation proceeds. Sign consistencies (inconsistencies) in the off-diagonal entries are treated with Givens (hyperbolic) rotations. With a proper ordering of Givens and hyperbolic rotations (see [6, 7] for further details), $\tilde{T}_k = H^{-1}T_kH$ can be made tridiagonal and S_2-symmetric. Additionally, it is claimed (see [7] for a proof) that this sequence of rotations which places $H^{-1}T_k$ into a unreduced, upper-triangular form with positive subdiagonal elements is unique.

As with explicit QR steps, the expense of explicit HR steps comes from the fact that both H^{-1} and H must be explicitly computed. A preferred alternative is the implicit HR step, an analogue to the Francis QR step [17]. The first implicit rotation, $H(1,2)$, is selected so that the first columns of the implicit and explicit H are equivalent. The remaining implicit rotations $H(i,i+1)$ perform a bulge-chase sweep down the tridiagonal. As the implicit HR step is completely analogous to the implicit QR step for symmetric tridiagonal matrices (including the handling of shifts and double shifts [17]), this technique will not be discussed here in detail.

Before this discussion of the HR algorithm is concluded, one additional result will be needed to aid in the development of implicit Lanczos restarts.

Lemma 1 *If H and R form an HR decomposition of $(T-\mu I)$ then $H^TR^TH^Te_1 = \rho e_1$ where $\rho = \pm r_{1,1}$.*

Proof: If H and R form an HR decomposition of $(T-\mu I)$ then by construction, $H^TS_1H = S_2$ where $S_1, S_2 \in diag(\pm 1)$ and $T^TS_1 = S_1T$. Since $e_1^TS_1 = \pm e_1^T$ it follows that

$$\begin{aligned}
\pm e_1^THRH &= e_1^TS_1(T-\mu I)H \\
&= e_1^T(T-\mu I)^TS_1H \\
&= [H^TS_1(T-\mu I)e_1]^T \\
&= [H^TS_1HRe_1]^T \\
&= [S_2Re_1]^T \\
&= \rho e_1^T
\end{aligned}$$

where $\rho = \pm r_{1,1}$ and the desired result is obtained. □

2.3 Implicit Lanczos restarts

As an intermediate step between the standard Lanczos method and the new factorization corresponding to (6,7), we will first derive a technique for arriving at the Lanczos factorization

$$A\tilde{V}_k = \tilde{V}_k\tilde{T}_k + \tilde{r}_ke_k^T \tag{15}$$
$$A^T\tilde{W}_k = \tilde{W}_k\tilde{T}_k^T + \tilde{q}_ke_k^T \tag{16}$$

which corresponds to the starting vectors

$$\tilde{v}_1 = \rho_v(A-\mu I)v_1 \quad \text{and} \quad \tilde{w}_1 = \rho_w(A^T-\mu I)w_1 \tag{17}$$

associated with the application of a real shift μ. Given that V_k and W_k are known, it would be desirable to obtain $\tilde{V}_k$ and $\tilde{W}_k$ without having to explicitly restart the Lanczos process with the vectors in (17). It would be preferable to obtain (15,16) directly from (8,9) using an implicit restart mechanism analogous to the technique introduced in [31]. This implicit restarting mechanism will now be derived.

If one obtains the decomposition $(T_k - \mu I) = HR$, (8) can be reexpressed in several different forms

$$\begin{aligned}
(A-\mu I)V_k &= V_k(T_k-\mu I) + r_ke_k^T \\
(A-\mu I)V_k &= V_k(HR) + r_ke_k^T \qquad (18) \\
(A-\mu I)V_kH &= V_kHRH + r_ke_k^TH \\
AV_kH &= V_kH(RH+\mu I) + r_ke_k^TH \\
A(V_kH) &= (V_kH)(H^{-1}T_kH) + r_ke_k^TH. \qquad (19)
\end{aligned}$$

The analogous expressions for (9) are

$$\begin{aligned}
(A^T - \mu I)W_k &= W_k(T_k^T - \mu I) + q_k e_k^T \\
(A^T - \mu I)W_k &= W_k H^{-T} H^T (R^T H^T) + q_k e_k^T \qquad (20) \\
(A^T - \mu I)W_k H^{-T} &= W_k H^{-T} H^T R^T + q_k e_k^T H^{-T} \\
A^T W_k H^{-T} &= W_k H^{-T}(RH + \mu I)^T + q_k e_k^T H^{-T} \\
A^T(W_k H^{-T}) &= (W_k H^{-T})(H^{-1} T_k H)^T + q_k e_k^T H^{-T}. \qquad (21)
\end{aligned}$$

It is convenient to define $\tilde{V}_k = V_k H$, $\tilde{W}_k = W_k H^{-T}$, and $\tilde{T}_k = H^{-1} T_k H$ so that (19) and (21) become

$$A\tilde{V}_k = \tilde{V}_k \tilde{T}_k + r_k e_k^T H \qquad (22)$$

$$A^T \tilde{W}_k = \tilde{W}_k \tilde{T}_k^T + q_k e_k^T H^{-T}. \qquad (23)$$

Multiplying (18) on the right by e_1 yields

$$(A - \mu I)v_1 = V_k H R e_1 = \tilde{V}_k R e_1 = \tilde{v}_1 r_{1,1} \qquad (24)$$

while multiplying (20) on the right by e_1 and recalling Lemma 1 yields

$$(A^T - \mu I)w_1 = W_k H^{-T} H^T R^T H^T e_1 = \tilde{W}_k H^T R^T H^T e_1 = \pm \tilde{w}_1 r_{1,1}. \qquad (25)$$

Clearly the desired result is near; new starting vectors can be obtained which fit the desired form of (17). Unfortunately, the corresponding expressions in (22,23) do not define a valid Lanczos factorization. Let $h_{i,j}$ and $h_{i,j}^{(-1)}$ be the $(i,j)^{th}$ entry in H and H^{-1} respectively. Then the residual terms in (22) and (23) are

$$r_k(h_{k,k-1}e_{k-1}^T + h_{k,k}e_k^T) \quad \text{and} \quad q_k(h_{k-1,k}^{(-1)}e_{k-1}^T + h_{k,k}^{(-1)}e_k^T)$$

rather than just vectors times e_k^T. This difficulty can be remedied however by truncating off a portion of (22,23). That is, (22) and (23) can be rewritten as

$$A\tilde{V}_k = [\tilde{V}_{k-1}, \tilde{v}_k, r_k] \left[\begin{array}{c|c} \tilde{T}_{k-1} & \tilde{\gamma}_k e_{k-1} \\ \tilde{\beta}_k e_{k-1}^T & \tilde{\alpha}_k \\ h_{k,k-1}e_{k-1}^T & h_{k,k} \end{array}\right] \qquad (26)$$

and

$$A^T\tilde{W}_k = [\tilde{W}_{k-1}, \tilde{w}_k, q_k] \left[\begin{array}{c|c} \tilde{T}_{k-1}^T & \tilde{\beta}_k e_{k-1} \\ \tilde{\gamma}_k e_{k-1}^T & \tilde{\alpha}_k \\ h_{k-1,k}^{(-1)}e_{k-1}^T & h_{k,k}^{(-1)} \end{array}\right] \qquad (27)$$

so that equating the first $k-1$ columns of (26) and (27) results in the new Lanczos identities

$$A\tilde{V}_{k-1} = \tilde{V}_{k-1}\tilde{T}_{k-1} + \tilde{r}_{k-1}e_{k-1}^T \quad \text{and} \quad A^T\tilde{W}_{k-1} = \tilde{W}_{k-1}\tilde{T}_{k-1}^T + \tilde{q}_{k-1}e_{k-1}^T. \qquad (28)$$

The new starting vectors $\tilde{v}_1$ and $\tilde{w}_1$ are still defined as in (24) and (25). The new residual vectors are derived from (26) and (27):

$$\tilde{r}_{k-1} = \tilde{\beta}_k \tilde{v}_k + h_{k,k-1} r_k \quad \text{and} \quad \tilde{q}_{k-1} = \tilde{\gamma}_k \tilde{w}_k + h_{k-1,k}^{(-1)} q_k.$$

One can also easily show that $\tilde{V}_{k-1}$, $\tilde{W}_{k-1}$, $\tilde{r}_{k-1}$, and $\tilde{q}_{k-1}$ meet the bi-orthogonality condition. Thus (28) is indeed a valid Lanczos factorization for the new starting vectors. Only one additional step of the standard Lanczos process is required to obtain (15,16) from (28).

From the above work, the extension of this technique to the multiple shift case is straightforward. One is now interested in a series of HR decompositions. The i^{th} decomposition is

$$H_i R_i = (\tilde{H}_{i-1}^{-1} T \tilde{H}_{i-1} - \mu_i I)$$

where

$$\bar{H}_{i-1} = H_1 H_2 \cdots H_{i-1}.$$

In practice, the decompositions are performed implicitly so that the H's are never explicitly formed. Pairs of complex conjugate shifts may be handled via double HR steps in real arithmetic just as in the implicitly shifted QR setting.

Applying p implicit restarts yields the new Lanczos factorization

$$\begin{aligned} A\bar{V}_{k-p} &= \bar{V}_{k-p}\bar{T}_{k-p} + \bar{r}_{k-p}e_{k-p}^T \\ A^T\bar{W}_{k-p} &= \bar{W}_{k-p}\bar{T}_{k-p}^T + \bar{q}_{k-p}e_{k-p}^T \end{aligned}$$

where $\bar{T}_{k-p}$, $\bar{V}_{k-p}$ and $\bar{W}_{k-p}$ are the appropriate submatrices of

$$\begin{aligned} \bar{T}_k &= \bar{H}_p^{-1} T_k \bar{H}_p \\ \bar{V}_k &= V_k \bar{H}_p \\ \bar{W}_k &= W_k \bar{H}_p^{-T}. \end{aligned}$$

The new residuals are

$$\bar{r}_{k-p} = \bar{\beta}_{k-p+1}\bar{v}_{k-p+1} + \bar{h}_{k,k-p}r_k \quad \text{and} \quad \bar{q}_{k-p} = \bar{\gamma}_{k-p+1}\bar{w}_{k-p+1} + \bar{h}_{k-p,k}^{(-1)}q_k$$

and new starting vectors are

$$\bar{v}_1 = \zeta_v(A - \mu_p I)\cdots(A - \mu_1 I)v_1 \quad \text{and} \quad \bar{w}_1 = \zeta_w(A^T - \mu_p I)\cdots(A^T - \mu_1 I)w_1.$$

In this case, p additional standard Lanczos steps are required to obtain the order-k Lanczos factorization,

$$A\bar{V}_k = \bar{V}_k\bar{T}_k + \bar{r}_k e_k^T \quad \text{and } A^T\bar{W}_k = \bar{W}_k\bar{T}_k^T + \bar{q}_k e_k^T, \tag{29}$$

corresponding to $\bar{v}_1$ and $\bar{w}_1$. However for $p \ll k$, this implicit approach represents a considerable saving in computations over the k standard Lanczos steps required for an explicit Lanczos restart.

For more details on the implicitly restarted method, the reader is referred to Section 4.

3 Model Reduction via Lanczos Techniques

Given Theorem 1, it is a simple matter to connect the standard Lanczos method to model reduction. Equation (10) indicates that one should select the starting vectors as $v_1 = b/\beta_1$ and $w_1 = c^T/\gamma_1$ so that $\pi_k = V_k W_k^T$ corresponds to the Krylov spaces $\mathcal{K}(A, b)$ and $\mathcal{K}(A^T, c^T)$ respectively. Then, $\hat{A} = W^T A V = T_k$, $\hat{b} = W^T b = e_1\beta_1$ and $\hat{c} = cV_k = e_1^T\gamma_1$ is the desired partial realization.

3.1 Performing restarts to stabilize a model

Of course, it has been repeatedly stated that the resulting Padé approximant need not be stable. In this case, we construct a modified projector which corresponds to a stable reduced-order model. Crucial for arriving at such a stabilizing projector is the proper selection of the restart shifts, μ_i. Although there is certainly an endless number of possibilities for these shifts, a practical policy arises from Theorem 2.

Lemma 2 *Let $\{\theta_1, \ldots, \theta_k\} \cup \{\mu_1, \ldots, \mu_p\}$ be a disjoint partition of the spectrum of T_{k+p} and define $\bar{T}_k$ to be the tridiagonal matrix resulting from p implicit restarts with shifts μ_1 through μ_p. Then the eigenvalues of $\bar{T}_k$ are $\{\theta_1, \ldots, \theta_k\}$.*

Proof: The proof follows from extending Theorem 2, result (iii), to the case of multiple shifts. Alternatively, a proof completely analogous to one for the Arnoldi method [31, Theorem 2.8] can be developed. □

Restarting with exactly p eigenvalues of T_{k+p} as the shifts deflates out these p eigenvalues from $\bar{T}_k$. This matrix is in fact easily seen to be the projected matrix of T_{k+p} on its stable invariant subspace since in the new coordinate system, $\bar{T}_{k+p}$ is block diagonal with leading diagonal submatrix $\bar{T}_k$. For our application, given that T_k is unstable, one needs to proceed until a T_{k+p} is determined with $q \leq p$ unstable modes (i.e., eigenvalues with real parts greater than zero). Then if q restarts are performed where each shift is an unstable mode of T_{k+p}, the resulting $\bar{T}_{k+p-q}$ is stable. Note that the condition "find T_{k+p} with at least k stable modes" is far less restrictive than finding a stable T_{k+p}.

For clarity, this approach for obtaining a stable, reduced-order model is summarized in the following algorithm.

Algorithm 2

1. *Perform k standard Lanczos steps to compute T_k, V_k, and W_k. Set q equal to the number of unstable modes in T_k and set $p = 0$.*
2. *While $q > p$,*
   (a) *Increment p.*
   (b) *Perform another standard Lanczos step to obtain T_{k+p}.*
   (c) *Set q equal to the number of unstable modes in T_{k+p}.*
3. *Obtain $\bar{T}_{k+p-q}$, $\bar{V}_{k+p-q}$, and $\bar{W}_{k+p-q}$ by implicitly restarting with q μ_i's selected to be the unstable modes of T_{k+p}.*
4. *Define the stable, reduced-order model to be $\hat{A} = \bar{T}_{k+p-q}$, $\hat{b} = \bar{W}_{k+p-q}^T b$, and $\hat{c} = c\bar{V}_{k+p-q}$.*

Remark The above algorithm is not the only way to guarantee a stable reduced system of order k. One alternative is e.g. to limit at any time the order of the model to *at most* k. One would then deflate out, say, the p unstable eigenvalues from this model and complete the system again to order k with p Lanczos steps. If this new model still has a number of unstable eigenvalues, we repeat the above procedure. The advantage of this method is that one never has to store bases larger than what one is finally interested in. This can be relevant when k and n are so large that memory (or cache) constraints become important. This method was also tested and gave similar results to Algorithm 2 in terms of total amount of work.

There is an alternative view on selecting shifts which agrees with well-known observations on the relationship between Padé approximation and stability. Note that the starting vector, v_1, can be expressed as a linear combination of the eigenvectors, y_i, of A, i.e., $v_1 = \sum_{i=1}^{n} \gamma_i y_i$. Then a shifted starting vector takes the form

$$\tilde{v}_1 = \zeta_v (A - \mu I) v_1 = -\zeta_v \sum_{i=1}^{n} \gamma_i (\mu - \lambda_i) y_i. \tag{30}$$

By assumption, the eigenvalues, λ_i, of A have negative real parts while the shifts are selected to have positive real parts. So from (30), applying such restarts to v_1 tends to emphasize those eigenvectors of A in $\tilde{v}_1$ which correspond to the high-frequency modes of the original system since the "weights" $(\mu - \lambda_i)$ are larger for these vectors. One can thus consider restarts with non-negative shifts as a way to emphasize the high-frequency modes in $\bar{\pi}_k$ and thus also in $\bar{T}_k$. Such interpretation corresponds well to the observation [29] that unstable realizations are obtained by better approximations of the original system's high-frequency modes.

Unfortunately, stabilizing a realization leads to discrepancies between the moments of the actual and reduced-order systems. The following lemma indicates that these identities are then replaced by an equal number of identities relating what could be called "modified moments".

Lemma 3 *Let $\bar{T}_k$, $\bar{V}_k$, and $\bar{W}_k$ be the result of an implicit Lanczos process with shifts characterized by the polynomial $\psi(s) = (s-\mu_p)\cdots(s-\mu_1)$. Then we have the following equations relating modified moments of the original system and the restarted Lanczos model :*

$$c\psi(A)A^{i-1}\psi(A)b = c\psi(A)\bar{V}_k\bar{T}_k^{i-1}\bar{W}_k^T\psi(A)b \tag{31}$$

for $i \leq 2k$, where $\hat{A} = \bar{T}_k = \bar{W}_k^T A\bar{V}_k$, $\hat{b} = \bar{W}_k^T b$ and $\hat{c} = c\bar{V}_k$ are the reduced order model parameters.

Proof: As $\bar{V}_k$ and $\bar{W}_k$ correspond to the starting vectors $\zeta_v(A-\mu I)b$ and $\zeta_w(A-\mu I)c$ respectively, Theorem 1 yields immediately the desired result. □

The cost of modifying the projector is thus that the resulting approximation is no longer of a Padé type. However, it should be stressed that this approximation is preferable when the true Padé approximation is unstable. We also point out that the above lemma suggests that other choices for $\hat{b}$ and $\hat{c}$ could be used rather than $\hat{b} = \bar{W}_k^T b$ and $\hat{c} = c\bar{V}_k$, but this is still under investigation.

3.2 Example: the portable Compact Disc player

The Compact Disc player is a well-known mechanism for sound reproduction. At the heart of the CD player is an optical unit (consisting of a laser diode, lenses, and photodetectors) which is mounted on the end of a radial arm [5]. The control problem consists of employing two magnets as actuators in order to (i) adjust the position of the radial arm so that the laser beam is correctly centered on one of the thousands of tracks on the disc, and (ii) adjust the position of a spring-mounted focusing lens so that the laser beam strikes an appropriate depth on the disc. The control system is thus a 2-input 2-output system. In this paper, the emphasis is on SISO systems. Therefore rather than working with the entire CD player mechanism, four smaller SISO systems are studied. In particular, this example concentrates on the relationship between the lens actuator and lens position.

Traditionally, the behavior of the lens position is represented by a third-order differential equation [5]. However, controllers designed from these simple, low-order systems experience difficulties when employed in the newer, portable Compact Disc players [5]. External physical shocks must now be taken into consideration which leads to higher order models.

A better model of the behavior of the CD player is obtained via finite element approximation of the various portions of the CD player. These models are typically large and sparse. Moreover the open loop system is stable by construction, but with very small damping. Models constructed in this way have orders varying from 100 to several thousands. The system matrix of the example used here is 120, which is already relatively small. Although this last fact is unfortunate from a model reduction point of view, this example is well suited to demonstrate both the severity of the unstable realization problem and the power of implicit restarts in overcoming this problem.

A very valid concern is the total number of Lanczos models (T_k, $W_k^T b$, cV_k, $1 \leq k \leq 120$) which are actually unstable. If there are only a few values of k for which T_k is unstable, then incorporating implicit restarts into the standard Lanczos method is unnecessary work. One could simply perform one or two more standard Lanczos steps to find a stable Krylov model. But Figure 1 demonstrates that T_k stable is the exception, not the rule, for this example. Beyond $k = 47$, an additional 49 Lanczos steps are required before another stable model is found. In general, one cannot count on stumbling upon stability at the appropriate recursion step k. The explanation of this phenomenon is that the original system has low damping and hence has poles very close to the unstable region, and the Padé approximations seem to move a few of these poles over the stability boundary.

However, employing implicit restarts with appropriate choices for the μ_i's (i.e., via Algorithm 2) quickly stabilizes the reduced-order model. Table 1 indicates the number of extra forward Lanczos steps, p, and implicit restarts, q, required to obtain a stable $\bar{T}_{k+p-q}$ given various T_k's. The main point to be drawn from this table and this example is that a model of arbitrary size can be stabilized with $q \ll k$ implicit restarts.

It is also important to recognize that the implicit restarts in this example do not have a detrimental effect on the accuracy of the final, stabilized model (and, in fact, they are extremely beneficial when the original

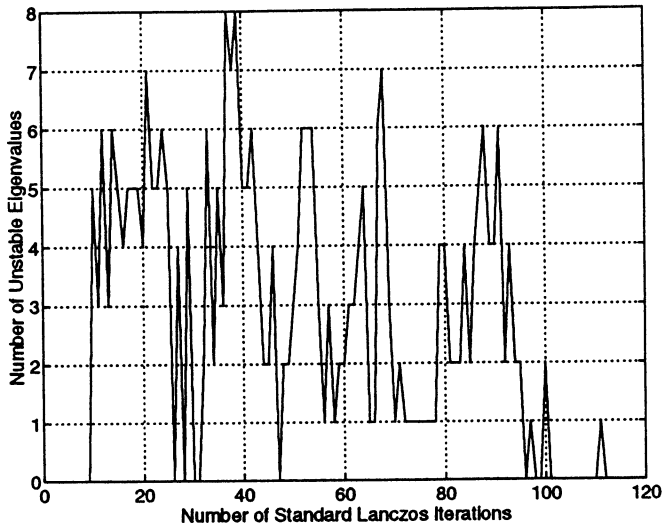

Figure 1: The number of unstable eigenvalues in T_k, where k is the number of standard Lanczos steps performed.

Table 1: Restarts Needed to Stabilize an Order-k Model

	$k = 20$	$k = 30$	$k = 40$	$k = 50$	k=60
No. Restarts, q	5	0	2	3	1
No. Forward Steps, p	5	0	4	5	5

approximation is unstable). For example, Figure 2 displays the impulse responses for both an initially stable Lanczos model ($k = 47$) and a restarted (stabilized) Lanczos model ($k = 50$). Even with a modified projector, $\bar{\pi}$, which no longer matches Markov parameters, the restarted model's response is closer to that of the actual system.

4 Numerical Properties of Implicit Lanczos Restarts

4.1 Implicit vs Explicit Restarts

In this section we analyze the differences between implicit restarts and explicit ones. As noted in Section 2, implicit restarts typically have a higher numerical accuracy than explicit restarts and moreover they are more economical to implement. These questions are further investigated below.

The amount of work involved in the implicitly restarted Lanczos method is provided in Table 2. The values of k and n are defined in Section 2. The value, α, is the average number of non-zero elements in a row of A. As an aside, it is interesting to note that full reorthogonalization is not a dominating cost in the overall process if the number of steps, k, is on the order of α.

Once a Lanczos factorization of size k is known, the amount of work, $O(pkn)$, involved in $p \ll k$ implicit restarts is split between recovering p truncated Lanczos steps and forming the new matrices $\bar{V}$ and $\bar{W}$. However, this cost is clearly preferred over the $O(k^2n)$ operations needed for an explicit Lanczos restart with full reorthogonalization. The efficiency of the implicit technique comes from avoiding the reorthogonalization required during the first $k - p$ steps of an explicit restart.

But even if one avoids full reorthogonalization (so that explicit restarts are no more expensive than

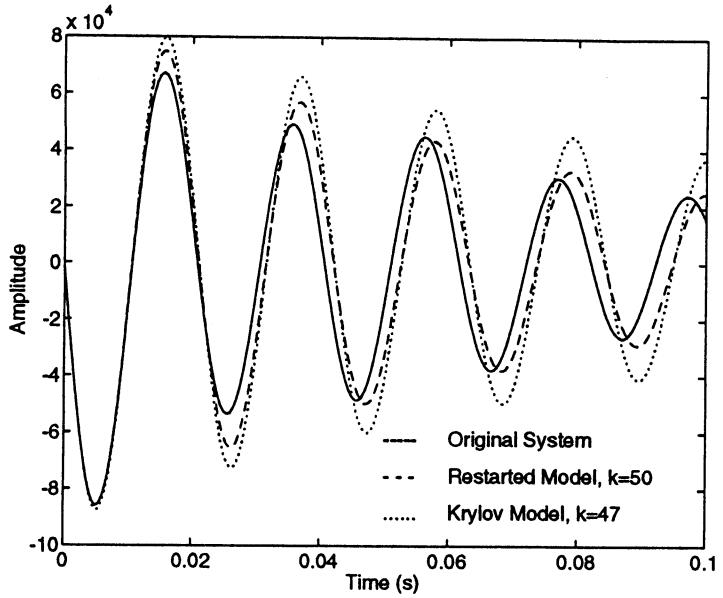

Figure 2: Impulse responses for CD player models.

Table 2: Dominating cost (in flops) of each stage of the implicit Lanczos restart

Stage of Method	Flops Required
k Lanczos steps (3-term recurrence)	$2kn(9+2\alpha)$
reorthogonalization during all k steps	$4n(k-2)^2$
p implicit restarts (obtain $\bar{T}_{k-p}$)	$12pkn$
p reorthogonalized Lanczos steps (obtain $\bar{T}_k$)	$4pn(k+\alpha)$

implicit ones), the new projector, $\bar{\pi}$, formed via implicit restarts should generally be preferred. One of the major advantages of the Lanczos method is that in order to compute the Krylov space it does not simply multiply the starting vector by powers of A. Yet in explicit restarts one is forced to directly multiply the starting vector by matrices of the form $(A-\mu_i I)$, and this is avoided by the implicit method. Even with non-zero shifts, it is quite possible that the vector $\bar{v}_1 = (A-\mu_p I)\cdots(A-\mu_1 I)v_1$ will be dominated by information from only a few of the largest (in absolute value) eigenvalues of A (recall (30)). As a result, a near *fortuitous* Lanczos breakdown [27] would soon occur at some step i and the residual vector, r_i, would consist of noise. An implicit restart, on the other hand, directly forms $\bar{V}_{k-p}$ from the wide range of information available in V_k. Even if the first columns of $\bar{V}_{k-p}$ are dominated by a few eigenvectors of A, the implicitly restarted method can call on information from V_k to accurately form the latter columns of $\bar{V}_{k-p}$.

As an example, consider a 20^{th} order system with a block diagonal A matrix :

$$A = diag\{-20e5, -19, -18, -17, -16, \ldots, -6, -5, -4, -3, \begin{bmatrix} -1 & -2 \\ 2 & -1 \end{bmatrix}\}$$

and with input and output vectors

$$b = \begin{bmatrix} -0.25837983924820 \\ 0.41004175548909 \\ -0.42821549481164 \\ -0.01782129915330 \\ 0.47742513053931 \\ 0.08416897411652 \\ -0.37205202359336 \\ -0.07836053361109 \\ -0.00548840151424 \\ -0.24356424982825 \\ 0.41565313651955 \\ -0.11773451586148 \\ 0.23599191603064 \\ 0.31613272699347 \\ 0.24274257931055 \\ -0.22546952763827 \\ -0.46635101640893 \\ 0.03846721515919 \\ -0.48151481942344 \\ 0.18042995020767 \end{bmatrix} \qquad c^T = \begin{bmatrix} 0.48617314034429 \\ 0.11196976649201 \\ -0.12413456878817 \\ -0.32969762283829 \\ -0.22794704312829 \\ -0.10595385711917 \\ 0.23352339804802 \\ -0.17224900688615 \\ 0.01094126445751 \\ -0.11016826266896 \\ 0.40200932272803 \\ -0.42931290992038 \\ -0.46207703182570 \\ -0.12867389462361 \\ 0.37785306101565 \\ -0.42360350998286 \\ 0.49580771801798 \\ 0.04031672819532 \\ -0.39674922120606 \\ -0.16416081025459 \end{bmatrix}.$$

The system is obviously stable since the eigenvalues of A are its scalar diagonal elements and $-1 \pm 2i$, $i = \sqrt{-1}$ for the 2×2 block.

After 6 standard Lanczos steps with $v_1 = b/\beta_1$ and $w_1 = c^T/\gamma_1$, the eigenvalues of the tridiagonal matrix, $T_6 = W_6^T A V_6$ are

$$\lambda(T_6) = \left\{ \begin{array}{c} -1.99999999964382 \\ 0.00000824023822 + 0.00002392349475i \\ 0.00000824023822 - 0.00002392349475i \\ 0.00000046571227 \\ -0.00002277422501 \\ -0.00006519548026 \end{array} \right\} \cdot 10^6.$$

To remove the three unstable eigenvalues from the reduced-order model, one can perform three implicit restarts as prescribed by Algorithm 2. Then the eigenvalues of the resulting tridiagonal matrix, $\bar{T}_3$, are

$$\lambda(\bar{T}_3) = \left\{ \begin{array}{c} -1.99999999964112 \\ -0.00002277422692 \\ -0.00006519510588 \end{array} \right\} \cdot 10^6.$$

Theorem 2 says that these should be the stable eigenvalues of T_6. The implicit HR steps separated the stable from the unstable eigenvalues but lost 3 to 5 figures along the way. The sensitivity of the eigenvalues of T_6 and the condition number of the transformation separating these eigenvalues, are in fact both of the order of 10^3, which explains that loss of accuracy. Alternatively, one should theoretically be able to obtain $\bar{T}_3$ by explicitly restarting the Lanczos process with the starting vectors $\bar{v}_{e1}$ and $\bar{w}_{e1}$ which (as denoted by the subscript e) are explicitly computed via (6) and (7). However, the eigenvalues of the explicitly computed $\bar{T}_{e3}$ are

$$\lambda(\bar{T}_{e3}) = \left\{ \begin{array}{c} -2.00000000000000 \\ -0.00002320939927 \\ -0.00003385741443 \end{array} \right\} \cdot 10^6. \tag{32}$$

This time we lost from 5 to 10 digits of accuracy, which is much more than expected from the conditioning of the problem. The large relative error in the smaller eigenvalues of (32) results from a near fortuitous Lanczos breakdown. Due to the stiffness of the A matrix, the explicitly restarted starting vectors, $\bar{v}_{e1}$ and $\bar{w}_{e1}$, are both close approximations to the eigenvector of A corresponding to $\lambda = -2 \cdot 10^6$. And as a result, the first residual vector, $\bar{r}_{e1}$, is very small ($\|\bar{r}_{e1}\| = 1.3458 \cdot 10^{-8}$). This severe loss of precision in the residual vectors leads to a certain degree of randomness in the vectors $\bar{v}_{e2}$ and $\bar{w}_{e2}$. Thus the eigenvalues of the reduced matrix, $\bar{T}_{e3}$, are not the eigenvalues of the implicitly generated $\bar{T}_3$ (the stable eigenvalues of T_6). In fact,

it is conceivable that an explicitly restarted $\bar{T}_e$ could be unstable even though $\bar{T}$ is stabilized via implicit restarts.

In summary, explicit restarts will lead to numerical difficulties when either the stiffness of the A matrix or the number of restarts becomes large. Thus implicit restarts, which typically avoid these problems, should be preferred for stabilizing the reduced-order model.

4.2 Breakdowns in the HR decomposition

To this point in the discussion, it has been assumed that the HR decomposition always exists. Yet it is easy to see that this assumption does not always hold. If there are two starting vectors, $(A-\mu I)v_1$ and $(A-\mu I)^T w_1$, for which an explicit Lanczos restart breaks down, it is impossible to tridiagonalize A with any projector corresponding to these starting vectors [27]. The HR decomposition of $(T_k-\mu I)$ cannot exist in this situation; otherwise a tridiagonal $\tilde{T}_k$ could be constructed.

This turns out to be the only way that breakdowns in the HR decomposition can occur. A *serious* breakdown [27] of the Lanczos process beginning with starting vectors $(A-\mu I)v_1$ and $(A-\mu I)^T w_1$ will occur at the j-th step *if and only if* the j-th (hyperbolic) rotation of the implicit restart process fails to exist. This is stated and proved formally in the following

Theorem 3 *Suppose the tridiagonal matrix T_k in (8) and (9) is unreduced and let $\mu \in \mathbb{R}$. Let $H(j,j+1)$ be the j^{th} rotation required in the HR decomposition of $(T_k-\mu I)$. If the first $j-1$ rotations of the decomposition exist, then a finite $H(j,j+1)$ fails to exists if and only if $\tilde{q}_{j-1}^T \tilde{r}_{j-1} = 0$, where $\tilde{q}_{j-1}, \tilde{r}_{j-1}$ are the non-zero updated Lanczos residuals.*

Proof: Assume that the first $j-1$ rotations $H(i,i+1), 1 \le i \le j-1$ exist and let

$$\begin{bmatrix} H_j & 0 \\ 0 & I \end{bmatrix} \equiv \prod_{i=1}^{j-1} H(i,i+1). \tag{33}$$

Then

$$A\tilde{V}_j = \tilde{V}_j \tilde{T}_j + \beta_{j+1} v_{j+1} e_j^T H_j \tag{34}$$

$$A^T \tilde{W}_j = \tilde{W}_j \tilde{T}_j^T + \gamma_{j+1} w_{j+1} e_j^T H_j^{-T}, \tag{35}$$

where $\tilde{T}_j = H_j^{-1} T_j H_j$, $\tilde{V}_j = V_j H_j$, and $\tilde{W}_j = W_j H_j^{-T}$. The orthogonality relations together with (34) and (35) imply

$$(A - \tilde{\alpha}_{j-1} I)\tilde{v}_{j-1} = \tilde{v}_{j-2}\tilde{\gamma}_{j-1} + \hat{v}_j \hat{\beta}_j + v_{j+1} w_{j+1}^T A \tilde{v}_{j-1} \tag{36}$$

$$(A^T - \tilde{\alpha}_{j-1} I)\tilde{w}_{j-1} = \tilde{w}_{j-2}\tilde{\beta}_{j-1} + \hat{w}_j \hat{\gamma}_j + w_{j+1} v_{j+1}^T A^T \tilde{w}_{j-1}. \tag{37}$$

The leading $(j+1)\times(j+1)$ principal submatrix of

$$\left(\prod_{i=1}^{j-1} H(i,i+1)\right)^{-1} T_k \prod_{i=1}^{j-1} H(i,i+1))$$

is

$$(\tilde{W}_j, w_{j+1})^T A (\tilde{V}_j, v_{j+1}) = \begin{bmatrix} \tilde{\alpha}_1 & \tilde{\gamma}_2 & & & & \\ \tilde{\beta}_2 & \ddots & \ddots & & & \\ & \ddots & \tilde{\alpha}_{j-2} & \tilde{\gamma}_{j-1} & & \\ & & \tilde{\beta}_{j-1} & \tilde{\alpha}_{j-1} & \hat{\gamma}_j & \tilde{w}_{j-1}^T A v_{j+1} \\ & & & \hat{\beta}_j & \hat{\alpha}_j & \hat{w}_j^T A v_{j+1} \\ & & & w_{j+1}^T A \tilde{v}_{j-1} & w_{j+1}^T A \hat{v}_j & \alpha_{j+1} \end{bmatrix}, \tag{38}$$

where the *hatted* quantities denote entries and vectors that would change if the HR update were continued beyond step $(j-1)$. It is readily verified that this matrix is sign-symmetric and that the rotation $H(j, j+1)$ is required to be a hyperbolic rotation if and only if

$$sign(\hat{\gamma}_j \tilde{w}_{j-1}^T A v_{j+1}) = -sign(\hat{\beta}_j w_{j+1}^T A \tilde{v}_{j-1}) \tag{39}$$

and this hyperbolic rotation will fail to exist if and only if

$$|\hat{\beta}_j| = |w_{j+1}^T A \tilde{v}_{j-1}| \neq 0.$$

If the rotation fails to exist, there is no loss in generality to assume that

$$\hat{\beta}_j = w_{j+1}^T A \tilde{v}_{j-1}, \tag{40}$$

since this may be arranged with a diagonal similarity scaling involving only sign changes. From the recurrence relations, it is easily shown that

$$\begin{aligned}
\tilde{\gamma}_j \tilde{\beta}_j \tilde{w}_j^T \tilde{v}_j &= \tilde{q}_{j-1}^T \tilde{r}_{j-1} && (41)\\
&= (\tilde{w}_{j-1}^T (A - \tilde{\alpha}_{j-1} I) - \tilde{\beta}_{j-1} \tilde{w}_{j-2}^T)((A - \tilde{\alpha}_{j-1} I)\tilde{v}_{j-1} - \tilde{v}_{j-2}\tilde{\gamma}_{j-1}) && (42)\\
&= \tilde{w}_{j-1}^T (A - \tilde{\alpha}_{j-1} I)^2 \tilde{v}_{j-1} - \tilde{\beta}_{j-1}\tilde{\gamma}_{j-1}, && (43)
\end{aligned}$$

However, if the two conditions (39), (40) hold, then the sign-symmetry implies that

$$\hat{\gamma}_j = -\tilde{w}_{j-1}^T A v_{j+1}$$

and this together with the relations (36)(37) and the identity $\tilde{W}_j^T \tilde{V}_j = I$, implies

$$\begin{aligned}
\tilde{w}_{j-1}^T (A - \tilde{\alpha}_{j-1} I)^2 \tilde{v}_{j-1} &= \tilde{w}_{j-1}^T (A - \tilde{\alpha}_{j-1} I)(\tilde{v}_{j-2}\tilde{\gamma}_{j-1} + \hat{v}_j \hat{\beta}_j + v_{j+1} w_{j+1}^T A \tilde{v}_{j-1}) && (44)\\
&= \tilde{w}_{j-1}^T A \tilde{v}_{j-2}\tilde{\gamma}_{j-1} + (\hat{\gamma}_j + \tilde{w}_{j-1}^T A v_{j+1})\tilde{\beta}_j && (45)\\
&= \tilde{\beta}_{j-1}\tilde{\gamma}_{j-1}. && (46)
\end{aligned}$$

Hence $\tilde{w}_j^T \tilde{v}_j = 0$ as claimed since the assumption that T_k is unreduced implies $\tilde{\gamma}_j \tilde{\beta}_j \neq 0$.

This argument has shown that an HR breakdown implies a serious Lanczos breakdown. The opposite implication follows from the uniqueness of the $(j+1)$-step Lanczos factorization. This uniqueness implies that the Lanczos process with starting vectors $(A-\mu I)v_1$ and $(A-\mu I)^T w_1$ must produce precisely the same j-step factorization as the implicitly restarted Lanczos process applied to the $(j+1)$-step factorization if $T_{j+1} - \mu I$ has an HR factorization. The existence of this factorization implies that (39), (40) cannot hold and (44) therefore could not be obtained. Hence, $\tilde{q}_{j-1}^T \tilde{r}_{j-1} \neq 0$ concluding the proof. □

Although not typically treated as such, serious Lanczos breakdowns can be considered to be the result of an ill-conditioned problem. The breakdown is the result of a poor choice for starting vectors and is not due to instabilities of the Lanczos method. For similar reasons, it is not disappointing that the implicitly restarted approach breaks down for a select set of ill-conditioned problems. One can only expect as much.

In [7], it is shown that there are at most $k(k+1)$ shifts, μ, for which the HR decomposition of $(T_k - \mu I)$ fails to exist. Hence avoiding ill-conditioned problems may demand that one take better advantage of the degrees of freedom afforded by the shifts μ_i. Even the order in which the shifts, μ_i, are applied to T_k can be important. Alternatively, one could relax the requirement that $\tilde{T}_k$ be tridiagonal and use a *lookahead* technique [19, 20]. However, the loss of tridiagonality in $\tilde{T}_k$ introduces extra complexity into the execution of additional implicit restarts. We leave the development of lookahead and/or improved shift selection in implicit restarts as an area for future work.

4.3 Stability of the implicit restart

There are some important questions regarding the numerical stability of the implicit restart mechanism. The first of these is related to the general two sided Lanczos process: What is the numerical quality of the bi-orthogonality in the vectors computed through the recursions in finite precision? Secondly: How is this numerical orthogonality and the accuracy of the eigenvalues of T_k effected during the implicit restarting due to the possible ill-conditioning of the S-orthogonal transformations?

Without some form of reorthogonalization, the Lanczos algorithm is numerically unstable (i.e., $W_k^T V_k$ will eventually drift far from the identity matrix). Yet by implementing reorthogonalization, one loses the efficiency of three-term recurrences. The technique of [10] may be generalized to maintain this bi-orthogonality. Our limited experience with this indicates that it is possible to maintain numerical bi-orthogonality at the level of round-off in working precision. The fact that the number of columns in W_k and V_k remain below a fixed modest size assures the cost of maintaining the bi-orthogonality can remain reasonable. However, further numerical experience with this technique is needed in this setting.

The next question is the numerical stability of hyperbolic rotations in this context. The ill-conditioning of these transformations has been intimately tied to serious Lanczos breakdown in the previous section. However, even though this ill-conditioning is possible, there are good reasons to use hyperbolic rotations anyway. The fact that these transformations preserve tridiagonality, sign-symmetry and bi-orthogonality is very attractive. It is natural to ask if there are any other transformations which retain these properties and yet have better conditioning.

In fact, there is a much wider class of decompositions that could be utilized to construct an implicit restart mechanism. Any decomposition of the form $\hat{H}\hat{R}$ where $\hat{H} = HD$, $\hat{R} = D^{-1}R$, and D is a diagonal scaling could be used in an implicit restart. For example, in [18] a prototype of the implicitly restarted Lanczos method was developed with $\hat{H}$ computed via LR decompositions. However, this more general $\hat{H}$ is generally not (S_1, S_2)-unitary. For this reason alone, one might argue that an $\hat{H}$ based on Givens and hyperbolic rotations is the best selection. However there is a more compelling justification. Of all possible transformations which preserve the tridiagonal structure when applied to an S-symmetric tridiagonal matrix the S-orthogonal transformations are optimally conditioned.

This result will be developed through a sequence of lemmas. The first result establishes the optimal conditioning for the two by two case. In the two by two case it is only necessary to consider a hyperbolic rotation since the condition number of a Givens rotation is 1.

Lemma 4 *Suppose* $|a| > |b|$ *and consider any* 2×2 *matrix* P *which satisfies*

$$P^{-1} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix} \quad \textit{and} \quad \begin{bmatrix} a & -b \end{bmatrix} P = \begin{bmatrix} \beta & 0 \end{bmatrix} \tag{47}$$

for the vector $[a\ b]$. *The hyperbolic rotation*

$$H = \begin{bmatrix} c & s \\ s & c \end{bmatrix} \quad \textit{with} \quad c = a/\sqrt{a^2 - b^2}, s = b/\sqrt{a^2 - b^2}$$

satisfies (47) and possesses the smallest condition number of any matrix P *satisfying (47).*

Proof: Conditions (47) imply $\alpha\beta = a^2 - b^2$ and that any such P must have the form

$$P = \begin{bmatrix} a/\alpha & b/\beta \\ b/\alpha & a/\beta \end{bmatrix}.$$

Thus P satisfies

$$P = \theta HD \ \text{ where } \ D = \begin{bmatrix} 1 & 0 \\ 0 & \delta \end{bmatrix},$$

with $\delta = \frac{\alpha}{\beta}$ and $\theta = \frac{\sqrt{a^2-b^2}}{\alpha}$. Now, $cond(P) = cond(HD)$ so consider the singular values $\sigma_1 \geq \sigma_2$ of H and the singular values $\hat{\sigma}_1 \geq \hat{\sigma}_2$ of HD. Observe that

$$(\hat{\sigma}_1\hat{\sigma}_2)^2 = det(DH^T HD) = \delta^2 det(H^T H) = \delta^2(\sigma_1\sigma_2)^2$$

and that

$$\hat{\sigma}_1^2 + \hat{\sigma}_2^2 = trace(DH^T HD) = (1+\delta^2)(c^2+s^2) = (1+\delta^2)trace(H^T H)/2 = (1+\delta^2)({\sigma_1}^2 + {\sigma_2}^2)/2.$$

Thus (assuming *wlog* that $\delta > 0$)

$$\frac{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}{\hat{\sigma}_1\hat{\sigma}_2} = \frac{{\sigma_1}^2 + {\sigma_2}^2}{\sigma_1\sigma_2}\frac{1}{2}(\delta + 1/\delta).$$

Since the function $\phi(\tau) \equiv \tau + 1/\tau$ satisfies $\phi(\tau) \geq 2$ for $\tau > 0$ it follows that

$$\hat{\kappa} + 1/\hat{\kappa} = (\kappa + 1/\kappa)\frac{1}{2}(\delta + 1/\delta) \geq (\kappa + 1/\kappa)$$

where $\hat{\kappa} = cond(P)$ and $\kappa = cond(H)$. Finally, the fact that $\phi(\tau)$ is strictly increasing for $\tau > 1$ implies that $\hat{\kappa} \geq \kappa$ with equality holding if and only if $\delta = 1$ and hence $P = \theta H$. □

Of all possible P's satisfying (47), the hyperbolic rotation has the smallest possible condition number. The following corollary indicates when this value will be one.

Corollary 1 *Assume a hyperbolic rotation is used in (47). Then as $b/a \to 0$ or $a/b \to 0$, the condition number of the hyperbolic rotation approaches one and as $|b/a| \to 1$ the condition number of the hyperbolic rotation approaches ∞.*

Proof: Assume $|b| < |a|$ Then as above,

$$cond(H) = cond\left(\begin{bmatrix} 1 & b/a \\ b/a & 1 \end{bmatrix}\right) = \frac{1+\rho}{1-\rho}.$$

where $\rho = |b/a|$ and the result follows immediately. The case $|a| < |b|$ has a similar proof. □
When the difference between a and b is large, the hyperbolic rotation is ideally conditioned. Of course as shown in Theorem 3, the problems arise when a and b are approximately equal.

Let us now turn to the general case.

Lemma 5 *Suppose H is (S_1, S_2)-unitary. Then there are permutations P_1 and P_2 such that $\hat{H} \equiv P_1 H P_2^T$ is S-unitary, i.e.*

$$\hat{H}^T S \hat{H} = S \quad with \quad S = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}$$

and with $p \leq q$.

Proof: If $H^T S_1 H = S_2$ then H must be nonsingular and Sylvester's inertia theorem indicates that S_1 and S_2 have the same inertia. Both sides of the relation $H^T S_1 H = S_2$ may be multiplied by -1 if necessary to arrange that S_1 and S_2 each have p ones and q negative ones on their respective diagonals where $p \leq q$. Hence there are permutations P_1 and P_2 such that $S = P_1 S_1 P_1^T = P_2 S_2 P_2^T$ with

$$S = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}.$$

Therefore,

$$P_2 H^T P_1^T P_1 S_1 P_1^T P_1 H P_2^T = P_2 S_2 P_2^T$$

so that $\hat{H}^T S \hat{H} = S$ as claimed. □
If D is diagonal then $P_1 H D P_2^T = \hat{H}\hat{D}$ with $\hat{D} = P_2 D P_2^T$. Therefore the desired result will be established in general if it is established for right-diagonal scalings of S-orthogonal matrices H. In order to obtain this result it will be useful to develop an analogy to the C-S decomposition of unitary matrices [11].

Lemma 6 *Suppose H is S-orthogonal, i.e. $H^T S H = S$ where*

$$S = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}.$$

with $p \leq q$. Then there are orthogonal matrices U_1 and V_1 of order p , orthogonal matrices U_2 and V_2 of order q and non-negative diagonal matrices Γ , Σ of order p such that

$$H = \left[\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right] \left[\begin{array}{c|cc} \Gamma & \Sigma & 0 \\ \hline \Sigma & \Gamma & 0 \\ 0 & 0 & I_{q-p} \end{array}\right] \left[\begin{array}{c|c} V_1^T & 0 \\ \hline 0 & V_2^T \end{array}\right]$$

with $\Gamma^2 = \Sigma^2 + I_p$.

Proof: Partition H conforming to S so that

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$$

with H_{11} of order p and H_{22} of order q. Since $SH^TS = H^{-1}$ we have $SH^TSH = HSH^TS$ and it follows from the partitioning that

$$H_{11}^T H_{11} = H_{21}^T H_{21} + I_p \quad H_{11} H_{11}^T = H_{12} H_{12}^T + I_p \tag{48}$$

and that

$$H_{22}^T H_{22} = H_{12}^T H_{12} + I_q \quad H_{22} H_{22}^T = H_{21} H_{21}^T + I_q. \tag{49}$$

Let $H_{11} = U_1 \Gamma V_1^T$ be the Singular Value Decomposition (SVD) of H_{11} and let $U_2 \hat{\Sigma} V_2^T$ be the SVD of H_{22}. From (48) we see that

$$\Gamma^2 = V_1^T (H_{11}^T H_{11}) V_1 = V_1^T (H_{21}^T H_{21}) V_1 + I_p$$

and

$$\Gamma^2 = U_1^T (H_{11} H_{11}^T) U_1 = U_1^T (H_{12} H_{12}^T) U_1 + I_p.$$

Hence,

$$\Sigma^2 \equiv \Gamma^2 - I_p = U_1^T (H_{12} H_{12}^T) = V_1^T (H_{21}^T H_{21}) V_1,$$

with Σ^2 a non-negative diagonal matrix. The diagonal elements matrix Σ therefore must be the singular values of both H_{12} and H_{21} . Also, it follows that the left singular vectors of H_{12} can be chosen to be the columns of U_1 while the right singular vectors of H_{21} can be chosen to be the columns of V_1. Equations (49) imply that the left singular vectors of H_{21} can be chosen to be the columns of U_2 and the right singular vectors of H_{12} can be chosen to be the columns of V_2. Thus the SVD's of these matrices are of the form

$$H_{21} = U_2 \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V_1^T \quad \text{and} \quad H_{12} = U_1 \begin{bmatrix} \Sigma & 0 \end{bmatrix} V_2^T$$

and it also follows from (49) that

$$\hat{\Sigma} = \begin{bmatrix} \Gamma & 0 \\ 0 & I_{q-p} \end{bmatrix}$$

and this concludes the proof. □

This result is interesting in its own right because it establishes a Hyperbolic C-S decomposition for S-orthogonal matrices. The following corollary is an immediate consequence:

Corollary 2 *If H is S-orthogonal with S as in Lemma (6) . Then the singular values of H are the diagonal elements of the matrices*

$$\Gamma + \Sigma, \quad \Gamma - \Sigma, \quad \text{and} \quad I_{q-p},$$

where Γ, Σ are as in Lemma (7) . Moreover, if σ is a singular value of H then its reciprocal $\frac{1}{\sigma}$ is also a singular value. Hence

$$cond(H) = (\gamma_1 + \sigma_1)^2$$

where γ_1 is the largest diagonal element of Γ and and $\gamma_1^2 = \sigma_1^2 + 1$.

Proof: The result follows immediately from the relation

$$\begin{bmatrix} \Gamma & \Sigma \\ \Sigma & \Gamma \end{bmatrix} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} = \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \begin{bmatrix} \Gamma + \Sigma & 0 \\ 0 & \Gamma - \Sigma \end{bmatrix}$$

and the fact that

$$(\Gamma + \Sigma)(\Gamma - \Sigma) = I,$$

with the largest diagonal element of Γ satisfying $\gamma_1^2 = \sigma_1^2 + 1 \geq 1$. □

Finally, we are in a position to establish the optimal conditioning of these S-orthogonal transformations.

Lemma 7 *Suppose H is S-orthogonal, i.e. $H^T S H = S$. Let D be any nonsingular diagonal matrix with the same order as H. Then*

$$cond(HD) \geq cond(H).$$

Proof: Let D be an arbitrary diagonal scaling and partition

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$$

with D_1, D_2 diagonal and of order p and q respectively. Then

$$HD = \left[\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right] \left[\begin{array}{c|cc} \Gamma & \Sigma & 0 \\ \hline \Sigma & \Gamma & 0 \\ 0 & 0 & I_{q-p} \end{array}\right] \left[\begin{array}{c|c} V_1^T D_1 & 0 \\ \hline 0 & V_2^T D_2 \end{array}\right].$$

Since $U_1$and U_2 are orthogonal it is sufficient to consider the matrix

$$\tilde{H} \equiv \left[\begin{array}{c|cc} \Gamma & \Sigma & 0 \\ \hline \Sigma & \Gamma & 0 \\ 0 & 0 & I_{q-p} \end{array}\right] \left[\begin{array}{c|c} V_1^T D_1 & 0 \\ \hline 0 & V_2^T D_2 \end{array}\right].$$

and to note that the submatrix consisting of the *first* and $(p+1)^{st}$ rows of this matrix is

$$M \equiv \begin{bmatrix} \gamma_1 & \sigma_1 \\ \sigma_1 & \gamma_1 \end{bmatrix} \begin{bmatrix} e_1^T V_1^T D_1 & 0 \\ 0 & e_1^T V_2^T D_2 \end{bmatrix},$$

assuming that the diagonal elements of Γ (and hence Σ) are ordered in decreasing order. Defining $\delta_i = \|D_i V_i e_1\|$ we also have that the singular values of M are the same as the singular values of the two by two matrix

$$\hat{M} = \begin{bmatrix} \gamma_1 & \sigma_1 \\ \sigma_1 & \gamma_1 \end{bmatrix} \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix}$$

since $MM^T = \hat{M}\hat{M}^T$. Now, since $\hat{M}\hat{M}^T$ is a principal submatrix of $\tilde{H}\tilde{H}^T$ it follows from the Cauchy Interlacing Theorem that

$$\tilde{\sigma}_1 \geq \hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \tilde{\sigma}_n$$

where $\tilde{\sigma}_1, \tilde{\sigma}_n$ are respectively the largest and smallest singular values of HD and $\hat{\sigma}_1, \hat{\sigma}_2$ are respectively the largest and smallest singular values of $\hat{M}$ thus

$$cond(\hat{M}) = \frac{\hat{\sigma}_1}{\hat{\sigma}_2} \leq \frac{\tilde{\sigma}_1}{\tilde{\sigma}_n} = cond(HD).$$

From Lemma (4) it follows that

$$cond\left(\begin{bmatrix} \gamma_1 & \sigma_1 \\ \sigma_1 & \gamma_1 \end{bmatrix}\right) \leq cond(\hat{M}) \text{ but } cond(H) = cond\left(\begin{bmatrix} \gamma_1 & \sigma_1 \\ \sigma_1 & \gamma_1 \end{bmatrix}\right)$$

and this chain of inequalities prove the lemma. □

This result is somewhat comforting, but it does not imply numerical stability in any sense. Further investigation on this issue is required. From the results established in Section 4.2 we do know that the condition of H must be intrinsically linked to the phenomena of serious breakdown in the two-sided Lanczos process. Learning how to avoid these breakdowns will probably be an important aspect of establishing the stability of this procedure.

5 Concluding Remarks

Applying implicit restarts to the Krylov spaces $\mathcal{K}_k(A, b)$ and $\mathcal{K}_k(A^T, c^T)$ is an efficient approach for insuring that the impulse response of the reduced-order system is bounded. But with or without restarts, it is rather obvious that the projector corresponding to these Krylov spaces does a poor job of modeling the low-frequency response of the CD player. Similar results are observed in the modeling of Tokamak plasmas in [1]. The order of k must approach that of n before sufficient low-frequency information is included in the Lanczos model. As mentioned in the introduction, a solution to this problem is to incorporate information from A^{-1} into the Krylov spaces. Work along these lines is still needed to guarantee the accurate reproduction of the low-frequency dynamics. But regardless of this fact, we stress the observation of [29]: the number of spurious, unstable modes in the reduced-order model tends to increase as emphasis is placed on the low-frequency modes. Incorporating A^{-1} into the Krylov spaces may increase the need for implicit restarts.

It should also be noted that although this paper explored the computation of stable models for stable problems, there are also applications in which unstable reduced-order models must be accurately computed for unstable plants. For stable, open loop problems (i.e., systems in which no feedback is applied), the need for stable reduced-order models in undeniable. One example of this class of problems is the simulation of large-scale circuits [14, 28]. But for problems in which feedback is present, one is typically more interested in the accurate modeling of the unstable modes. The stabilizing controller for a large-scale plant is typically developed through the analysis of a reduced-order model. Thus in this situation, is of the utmost importance that the unstable modes of the large-scale system appear in some form in the reduced-order model. The use of implicit restarts to achieve this remains an area for future research.

Lanczos methods are already being applied to model reduction problems in the area of control [4, 23, 34]. However, in many applications, sparse systems occur in implicit state space systems rather than in explicit ones. In other words, instead of working with (1) and (2), the more general state space equation

$$E\dot{x} = Ax + Bu \tag{50}$$

$$y = Cx + Du \tag{51}$$

should be treated. In this case, one must be concerned with the inversion of E for high-frequency moments or A for low frequency moments. But for the more general case (for example, when direct methods result in a high degree of "fill-in"), it would seem that one must turn to iterative strategies for approximating those matrix-vector products involving inverses. Moreover block versions of the above schemes ought to be

developed for dealing with the case where B and C are not just vectors (see e.g. the CD player example used in section 3.2).

The occurrence of ill-conditioned table entries is well-studied in the Lanczos algorithm [27], where it is termed a "serious" breakdown. By employing look-ahead into the Lanczos method, [19, 20, 13], one possesses a powerful tool for detecting and avoiding ill-conditioned table entries. Along similar lines, block versions of these algorithms should be developed in order to cope appropriately with the MIMO case.

References

[1] M. M. M. Al-Husari, B. Hendel, I. M. Jaimoukha, E. M. Kasenally, D. J. N. Limebeer and A. Portone, "Vertical stabilisation of Tokamak plasmas," *Proc. IEEE 30th Conf. on Decision and Control*, (Brighton, England), 1991.

[2] G. A. Baker Jr., *Essentials of Padé Approximants*, New York, NY: Academic Press, 1975.

[3] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation $AX + XB = C$," *Commun. ACM*, vol. 15, pp. 820–826, 1972.

[4] D. L. Boley, "Krylov space methods on state-space control models," Technical Report, Univ. of Minnesota, Minneapolis, MN 55455, 1992.

[5] O. H. Bosgra, G. Schoolstra and M. Steinbuch, "Robust control of a compact disc player," in *Proc. IEEE 31st Conf. on Decision and Control*, (Tucson, AZ), 1992.

[6] M. A. Brebner and J. Grad, "Eigenvalues of $Ax = \lambda Bx$ for real symmetric matrices A and B computed by reduction to a pseudosymmetric form and the HR process, *Linear Algeb. Its Appl.*, vol. 43, pp. 99–118, 1982.

[7] A. Bunse-Gerstner, "An analysis of the HR algorithm for computing the eigenvalues of a matrix, *Linear Algeb. Its Appl.*, vol. 35, pp. 155–173, 1981.

[8] C. I. Byrnes and A. Lindquist, "The stability and instability of partial realizations," *Sys. Control Lett.*, vol. 2, pp. 99–105, 1982.

[9] D. Calvetti, L. Reichel and D. C. Sorensen, "An implicitly restarted Lanczos method for large symmetric eigenvalue problems," *Electronic Trans. Numer. Anal.*, vol. 2, pp. 1–21, 1994.

[10] J. Daniel, W. B. Gragg, L. Kaufman and G. W. Stewart, "Reorthogonalization and stable algorithms for updating the Gram-Schmidt QR factorization," *Math. Comp.*, vol. 30, pp. 772–795, 1976.

[11] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation III" *SIAM J. Numer. Anal.*, vol. 7, pp. 1–46, 1970.

[12] R. W. Freund, G. H. Golub and N. M. Nachtigal, "Iterative solution of linear systems," *Acta Numer.*, vol. 1, pp. 57–100, 1992.

[13] R. W. Freund and N. M. Nachtigal, "QMR: a quasi-minimal residual method for non-Hermitian linear systems", *Numer. Math.*, vol. 60, pp. 315–339, 1991.

[14] K. Gallivan, E. Grimme and P. Van Dooren, "Asymptotic waveform evaluation via a Lanczos method," to appear *Appl. Math. Lett.*, 1994.

[15] W. B. Gragg and A. Lindquist, "On the partial realization problem," *Linear Algeb. Its Appl.*, vol. 50, pp. 277–319, 1983.

[16] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds," *Int. J. Control*, vol. 39, pp. 1115–1193, 1984.

[17] G. H. Golub and C. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins University Press, 1989.

[18] E. Grimme, D. Sorensen and P. Van Dooren, "Stable partial realizations via an implicitly restarted Lanczos method,", to appear *Proc. of ACC*, (Baltimore, MD),1994.

[19] M. H. Gutknecht, "A completed theory of the unsymmetric Lanczos process and related algorithms, part I" *SIAM J. Matrix Anal. Appl.*, vol. 13, pp. 594–639, 1992.

[20] M. H. Gutknecht, "A completed theory of the unsymmetric Lanczos process and related algorithms, part II" *SIAM J. Matrix Anal. Appl.*, vol. 15, pp. 15–58, 1994.

[21] A. S. Householder, *The Numerical Treatment of a Single Nonlinear Equation*, New York, NY: McGraw-Hill Book Company, 1970.

[22] X. Huang, "A Survey of Padé approximations and their applications in model reduction," Technical Report, Carnegie Mellon Univ., Pittsburgh, PA 15213, 1990.

[23] I. M. Jaimoukha and E. M. Kasenally, "Oblique projection methods for large scale model reduction," to appear *SIAM J. Matrix Anal. Appl.*, 1994.

[24] C. Lanczos "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators," *J. Res. Nat. Bur. Standards*, vol. 45, pp. 255–282, 1950.

[25] B. C. Moore, "Principal component analysis in linear systems: controllability, observability and model reduction," *IEEE Trans. Autom. Control*, vol. AC-26, pp. 17–32, 1981.

[26] C. C. Paige, The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices, Ph.D. Dissertation, University of London, UK, 1971.

[27] B. N. Parlett, "Reduction to tridiagonal form and minimal realizations," SIAM J. Matrix Anal. Appl., vol. 13, pp. 567–593, 1992.

[28] V. Raghavan, R. A. Rohrer, L. T. Pillage, J. Y. Lee, J. E. Bracken and M. M. Alaybeyi, "AWE-inspired," Proc. IEEE Custom Integrated Ciruits Conf., 1993.

[29] Y. Shamash, "Stable reduced-order models using Padé type approximations," *IEEE Trans. Autom. Control*, vol. AC-19, pp. 615–616, 1974.

[30] Y. Shamash, "Model reduction using the Routh stability criterion and the Padé approximation technique," *Int. J. Control*, vol. 21, pp. 475–484, 1975.

[31] D. C. Sorensen, "Implicit application of polynomial filters in a K-step Arnoldi method," *SIAM J. Matrix Anal. Appl.*, vol. 13, pp. 357–385, 1992.

[32] T. J. Su and R. R. Craig Jr., "Model reduction and control of flexible structures using Krylov vectors," *J. Guidance, Control, and Dynamics*, vol. 14, pp. 260–267, 1991.

[33] T. J. Su and R. R. Craig Jr., "An unsymmetric Lanczos algorithm for damped structural dynamics systems," *Proc. 33rd Conf. on Structures, Structural Dynamics and Materials*, 1992.

[34] P. Van Dooren, "Numerical linear algebra techniques for large scale matrix problems in systems and control," *Proc. IEEE 31st Conf. on Decision and Control*, (Tucson, AZ), 1992.

[35] C. D. Villemagne and R. E. Skelton, "Model reduction using a projection formulation," *Int. J. Control*, vol. 46, pp. 2141–2169, 1987.

[36] J. H. Wilkinson, *The Algebraic Eigenvalue Problem.* Clarendon Press: Oxford, England, 1965.

[37] H. Xiheng, "FF-Padé method of model reduction in frequency domain", *IEEE Trans. Autom. Control*, vol. AC-32, pp. 243–246, 1987.