

IMPLICITLY RESTARTED KRYLOV SUBSPACE METHODS FOR STABLE PARTIAL REALIZATIONS*

IMAD M. JAIMOUKHA[†] AND EBRAHIM M. KASENALLY[‡]

Abstract. This paper considers an implicitly restarted Krylov subspace method that approximates a stable, linear transfer function $f(s)$ of order n by one of order m , where $n \gg m$. It is well known that oblique projections onto a Krylov subspace may generate unstable partial realizations. To remedy this situation, the oblique projectors obtained via classical Krylov subspace methods are supplemented with further projectors which enable the formation of stable partial realizations directly from $f(s)$. A key feature of this process is that it may be incorporated into an implicit restart scheme. A second difficulty arises from the fact that Krylov subspace methods often generate partial realizations that contain nonessential modes. To this end, balanced truncation may be employed to discard the unwanted part of the reduced-order model. This paper proposes oblique projection methods for large-scale model reduction that simultaneously compute stable reduced-order models while discarding all nonessential modes. It is shown that both of these tasks may be effected by a single oblique projection process. Furthermore, the process is shown to naturally fit into an implicit restart framework. The theoretical properties of these methods are thoroughly investigated, and exact low-dimensional expressions for the \mathcal{L}^∞ -norm of the residual errors are derived. Finally, the behavior of the algorithm is illustrated on two large-scale examples.

Key words. Arnoldi, model reduction, Krylov subspace methods, large-scale systems, implicit restarts

AMS subject classifications. 65F10, 65F15, 93A15, 93B05, 93B07, 93B20

PII. S0895479895279873

1. Introduction. Model reduction has been practiced widely by engineers, and, until recently, the process was often based on intuition and a sound understanding of the physical principles associated with the modeling task. Chemical engineers assume that mixing is instantaneous and that packed distillation columns may be modeled using discrete trays. Electrical engineers represent transmission lines and eddy currents in the rotor cage of an induction motor by lumped circuits. Mechanical engineers routinely remove the high-frequency vibration modes from models of aircraft wings, turbine shafts, and flexible structures. The purpose of the present paper is to systematize the model reduction of large-scale dynamical systems without any a priori knowledge of their characteristics and to provide computable expressions for the errors incurred during the approximation process.

Consider the dynamical system described by a stable, linear, time-invariant state-space model of the form

$$(1) \quad \dot{x}(t) = Ax(t) + bu(t), \quad y(t) = cx(t),$$

in which $x(t)$ is the state vector of dimension n and $u(t)$ and $y(t)$ are scalar functions representing the input and the output of the system, respectively. The matrix A and

* Received by the editors January 13, 1995; accepted for publication (in revised form) by P. Van Dooren July 12, 1996.

<http://www.siam.org/journals/simax/18-3/27987.html>

[†] Interdisciplinary Research Centre for Process Systems Engineering and Department of Electrical and Electronic Engineering, Imperial College of Science, Technology and Medicine, London SW7-2BY, U.K. (jaimouka@ps.ic.ac.uk).

[‡] Bank of America NT & SA, Exposure Management, EMEA 5209, 26 Elmfield Road, Bromley, Kent BR1-1WA, U.K. The research contained in this paper was performed while this author was with the Interdisciplinary Research Centre for Process Systems Engineering, Imperial College of Science, Technology and Medicine, London SW7-2BY, U.K. (kasenall@ps.ic.ac.uk).

vectors b and c are real, with their dimensions fixed by those of $x(t)$, $u(t)$, and $y(t)$. It should be noted that the matrix A will be large and sparse in the following. Applying Laplace transforms to the system in (1) for zero initial conditions leads to a transfer function given by $f(s) = c(sI - A)^{-1}b$. The task of any model-reduction algorithm is then to find a stable approximate model of the form

$$(2) \quad \dot{x}_m(t) = A_m x_m(t) + b_m u(t), \quad y_m(t) = c_m x_m(t),$$

in which $x_m(t) \in \mathbb{R}^m$ and $m \ll n$ and where the associated low-order transfer function is given by $f_m(s) = c_m(sI - A_m)^{-1}b_m$. Well-established model reduction methods such as the optimal Hankel norm [8] and the balanced truncation [20] begin by solving the linear matrix equations

$$(3) \quad AP + PA^T + bb^T = 0 \quad \text{and} \quad A^T Q + QA + c^T c = 0,$$

which admit unique symmetric solutions if and only if $\lambda_i(A) + \bar{\lambda}_j(A) \neq 0$ for all i, j , where λ_i denotes the i th eigenvalue and the overbar represents the complex conjugate. Approximating large-scale systems via such methods is intractable because of the prohibitive storage requirements and a computational burden of $O(n^3)$ for each Lyapunov equation. The approach developed here has its roots in determining low-rank approximate solutions to (3) via the application of classical Krylov subspace methods [1, 18].

Recently, several schemes using Krylov subspace methods have been proposed for the task of large-scale model reduction. The objective of such algorithms has been to replace a high-order model by an m th-order low-dimensional approximation while effecting $O(m^2n)$ operations. For example, in [1], this type of model reduction formed an integral part of control system synthesis where high-order controllers had to be avoided because of certain engineering considerations. Krylov subspace model reduction schemes have recently been employed in linear circuit analysis, where they allow for the accurate and efficient simulation of large-scale circuits [6]. In our previous work [16, 17, 18], the focus was placed on calculating low-rank approximate solutions to (3). It was shown that these low-rank solutions were the exact controllability and observability Gramians of a dynamical system obtained by perturbing the system in (1). Finally, it was demonstrated that the reduced-order models could be readily computed via an oblique projection process based on the data generated in the course of the iterative process. It is well known, however, that partial realizations computed in this way may be unstable even if $f(s)$ is stable. Thus, one of the aims of this paper is to propose a method that yields a stable reduced-order model by effecting a second oblique projection process which we shall call a stable projection. An interesting feature of this method is that by combining the oblique projectors generated by the iterative process with those from the stable projection, one may compute stable reduced-order models by effecting a single oblique projection on $f(s)$. Consequently, one may cast the algorithm into an implicit restart setting which is similar in philosophy to that proposed in [14, 15].

Stable partial realizations via implicitly restarted Krylov subspace methods have been proposed in [14] in the context of control and in [15] in the setting of linear circuit analysis. The difference between the methods proposed here and those of [14, 15] is that, in the present setting, the integrity of both Arnoldi equations is preserved, whereas [14, 15] focus on preserving only one of these equations. One may not wish to approximate the state transition matrix independently of b and c , such as in a modal reduction scheme, but rather obtain a reduced-order projection process that

uses all the problem data. One advantage of the present approach is that it offers computable expressions for the \mathcal{L}^∞ -norm of the residual errors. Another advantage is that the theoretical properties expounded in [18] remain in force irrespective of the number of restarts performed; furthermore, the reduced-order model satisfies a moment-matching property. While the implicit restart strategy presented here focuses on model reduction, this approach may also be used to implicitly restart Krylov subspace schemes for the solution of large-scale Lyapunov equations [16, 17] and approximations to the exponential operator [22].

Krylov subspace methods are known to generate partial realizations which capture the outermost part of the spectrum of A . While this feature may be desirable in some settings, such as for the solution of stiff ordinary differential equations, it often results in reduced-order models that are unable to approximate the low-frequency characteristics of $f(s)$. A second objective addressed in this paper is to extract the nonessential modes (i.e., those modes associated with the outermost part of A 's spectrum) from the reduced-order model by effecting a balanced truncation of $f_m(s)$. We make use of a numerically robust variant of Moore's algorithm [20] which does not require the formation of the balancing transformations, known as the square root algorithm [24]. A key feature of the square root model-reduction scheme is that it may be cast as an oblique projection process and when combined with the oblique projectors derived from the iterative process yields projectors that compute a reduced-order approximation to $f(s)$ which retains the essential characteristics of $f_m(s)$. A second feature that will be exploited is that this process may be cast naturally into an implicit restart framework. It is interesting to observe that one may combine the oblique projectors from the iterative process with those obtained from the stable projection and balanced truncation to produce an implicitly restarted Krylov subspace method that computes stable partial realizations that accurately replicate the low-frequency characteristics of $f(s)$.

Related to this work are those of [4, 5], which use Krylov subspace methods to obtain bases for the controllability and observability spaces; furthermore, in [5] Boley and Golub presented a means of computing a minimal realization of a linear dynamical system from the coefficients generated in the course of the Lanczos process. The Lanczos process was also exploited by Parlett in [21] to obtain minimal realizations. In that paper, the rank of the Hankel matrix was used to determine the order of the minimal realization; furthermore, it was demonstrated that a minimal realization could be constructed from the data generated by the Lanczos process. A similar approach was adopted in [12], in which the minimal realization and its order were found to be related to the different types of breakdown encountered in the Lanczos process. Here too the onset of breakdown was given in terms of properties of the Hankel matrix. Recently, the presentations in [3, 26] review the use of projection methods for large-scale control problems. Each paper suggests the use of Krylov subspace methods as an effective tool for the model reduction of large-scale linear dynamical systems; however, no algorithms were provided.

The following summarizes the structure of this paper. Section 2 briefly describes the application of classical Krylov subspace techniques to the model-reduction problem. Section 3 considers a general implicit restart structure for arbitrary transformations, a theoretical analysis is presented and computable \mathcal{L}^∞ -norms of the error expressions are derived. Section 4 uses the framework of section 3 to produce several implicitly restarted Krylov subspace algorithms based on particular transformations such as stable projections or balancing transformations. Two numerical experiments

expounding the benefits of an implicitly restarted Krylov subspace method are found in section 5, and section 6 contains the conclusions.

2. Oblique projection methods for large-scale model reduction. This section considers oblique projection methods onto the m -dimensional Krylov subspaces

$$\begin{aligned}\mathcal{K}_m(A, b) &:= \text{span} \{ [b \ Ab \ \dots \ A^{m-1}b] \} \\ \mathcal{L}_m(A^T, c^T) &:= \text{span} \{ [c^T \ A^T c^T \ \dots \ (A^{m-1})^T c^T] \},\end{aligned}$$

which are parts of the controllability and observability subspaces, respectively. The class of iterative techniques focused on hinge on the efficient formation of well-conditioned bases for $\mathcal{K}_m(A, b)$ and $\mathcal{L}_m(A^T, c^T)$. To this end, one exploits a modified Gram–Schmidt process known as the Arnoldi process to calculate matrices $V_m = [v_1, v_2, \dots, v_m]$ and $W_m = [w_1, w_2, \dots, w_m]$, whose columns form an orthogonal basis for each of the Krylov subspaces as well as unit vectors v_{m+1} and w_{m+1} , which are orthogonal to V_m and W_m , respectively [2]. An outline of the Arnoldi process and its application to several large-scale numerical linear algebra problems may be found in [4, 7, 16, 17, 18, 19, 23]. By construction, the Arnoldi process associated with $\mathcal{K}_m(A, b)$ produces an $m \times m$ upper Hessenberg matrix H_m ; furthermore, one readily verifies that b and A satisfy

$$(4) \quad b = V_m l_m,$$

$$(5) \quad AV_m = V_m H_m + \tilde{V}_m \tilde{H}_m,$$

in which $l_m = e_1 \|b\|_2$, $\tilde{V}_m = v_{m+1}$, and $\tilde{H}_m = h_{m+1,m} e_m^T$ and where $h_{m+1,m}$ is a nonnegative scalar and e_1 and e_m are, respectively, the first and last columns of the m -dimensional identity matrix. From (5), it is easy to see that $H_m = V_m^T AV_m$ since $[V_m \ \tilde{V}_m]$ is part of an orthogonal matrix. In what follows, we refer to (4) and (5) as the controllability Arnoldi equations. Similarly, associated with $\mathcal{L}_m(A^T, c^T)$, the Arnoldi process generates a lower Hessenberg matrix G_m ; in this setting, c and A then satisfy

$$(6) \quad c^T = W_m k_m^T,$$

$$(7) \quad A^T W_m = W_m G_m^T + \tilde{W}_m \tilde{G}_m^T,$$

in which $k_m = \|c\|_2 e_1^T$, $\tilde{W}_m = w_{m+1}$, and $\tilde{G}_m^T = g_{m,m+1} e_m^T$ and where $g_{m,m+1}$ is a nonnegative scalar. It follows that $G_m = W_m^T A W_m$ since $[W_m \ \tilde{W}_m]$ is part of an orthogonal matrix. In the development below, we refer to (6) and (7) as the observability Arnoldi equations. A key difference between the restart process proposed in this paper and those of [14, 15] is that the present approach aims to preserve the integrity of (4)–(7). The main advantage of this approach is that it allows simple residual error expressions to be computed. In contrast, [14, 15] focus on preserving (5) and (7), and no residual error expressions are provided. Furthermore, we show that the implicitly restarted approximate model continues to satisfy a (modified) moment-matching property enjoyed by the oblique projection approximation without restarts.

2.1. Model reduction using Krylov subspace methods. The aim of this section is to consider the Krylov subspace techniques described above to provide

computationally efficient model-reduction schemes for large-scale dynamical systems. Denoting the transfer function corresponding to the model (1) by $f(s)$, then

$$(8) \quad f(s) = c(sI - A)^{-1}b \stackrel{s}{=} \left[\begin{array}{c|c} A & b \\ \hline c & 0 \end{array} \right], \quad A \in \mathbb{R}^{n \times n}, \quad b, c^T \in \mathbb{R}^n.$$

The model-reduction task determines a reduced-order model given by

$$f_m(s) = c_m(sI - A_m)^{-1}b_m \stackrel{s}{=} \left[\begin{array}{c|c} A_m & b_m \\ \hline c_m & 0 \end{array} \right], \quad A_m \in \mathbb{R}^{m \times m}, \quad b_m, c_m^T \in \mathbb{R}^m,$$

which approximates the high-dimensional model $f(s)$, where $m \ll n$.

Rewriting (8) as $f(s) = cf_b(s) = f_c(s)b$, where $f_b(s) = (sI - A)^{-1}b$ and $f_c(s) = c(sI - A)^{-1}$, permits us to consider $f_b(s)$ and $f_c(s)$ as the solutions to the coupled linear systems

$$(9) \quad (sI - A)f_b(s) = b \quad \text{and} \quad f_c(s)(sI - A) = c,$$

respectively. The focus of what follows is to approximate $f(s)$ by obtaining approximate solutions $f_{b,m}(s)$ and $f_{c,m}(s)$ to the linear systems (9). These approximate solutions are constructed to satisfy the following two conditions. (1) $f_{b,m}(s) \in \mathcal{K}_m(A, b)$, i.e., $f_{b,m}(s) = V_m h_m(s)$, such that $\mathcal{L}_m(A^T, c^T) \perp \{(sI - A)f_{b,m}(s) - b\}$. (2) $f_{c,m}^T(s) \in \mathcal{L}_m(A^T, c^T)$, i.e., $f_{c,m}(s) = g_m(s)W_m^T$, such that $\{f_{c,m}(s)(sI - A) - c\} \perp \mathcal{K}_m(A, b)$. Since $f_{b,m}(s)$ and $f_{c,m}(s)$ are approximate solutions to the linear systems in (9) and $f(s) = cf_b(s) = f_c(s)b$, we will consider $f_{m,1}(s) := cf_{b,m}(s)$ and $f_{m,2}(s) := f_{c,m}(s)b$ as approximations to $f(s)$. The problem we wish to solve may be stated as follows.

PROBLEM 2.1. *Find approximate solutions $f_{b,m}(s) = V_m h_m(s)$ and $f_{c,m}(s) = g_m(s)W_m^T$ to (9) which satisfy the Galerkin-type conditions*

$$(10) \quad W_m^T \{(sI - A)V_m h_m(s) - b\} = 0 \quad \forall s, \quad \{g_m(s)W_m^T(sI - A) - c\}V_m = 0 \quad \forall s.$$

For convenience, define the matrices

$$\begin{aligned} \widehat{H}_m &:= T_m^{-1}W_m^T A V_m = H_m + T_m^{-1}W_m^T \widetilde{V}_m \widetilde{H}_m, \\ \widehat{G}_m &:= W_m^T A V_m T_m^{-1} = G_m + \widetilde{G}_m \widetilde{W}_m^T V_m T_m^{-1} \end{aligned}$$

for nonsingular $T_m := W_m^T V_m$ and observe that \widehat{H}_m and \widehat{G}_m are upper and lower Hessenberg, respectively. The following theorem gives the solution to Problem 2.1.

THEOREM 2.1 (see [18]). *Suppose that m steps of the Arnoldi process have been taken and that T_m is nonsingular. Then*

1. *the Galerkin conditions in (10) are satisfied if and only if $h_m(s) = (sI - \widehat{H}_m)^{-1}l_m$ and $g_m(s) = k_m(sI - \widehat{G}_m)^{-1}$. Under these conditions, the residual error \mathcal{L}^∞ -norms are*

$$(11) \quad \|b - (sI - A)V_m h_m(s)\|_\infty = \left\| \left[\begin{array}{c|c} T_m^{-1}W_m^T \widetilde{V}_m & \\ \hline 1 & \end{array} \right] \widetilde{H}_m h_m(s) \right\|_\infty,$$

$$(12) \quad \|c - g_m(s)W_m^T(sI - A)\|_\infty = \left\| g_m(s)\widetilde{G}_m \left[\begin{array}{c|c} \widetilde{W}_m^T V_m T_m^{-1} & \\ \hline & 1 \end{array} \right] \right\|_\infty;$$

2. *the approximations*

$$(13) \quad f_{m,1}(s) = cV_m h_m(s) \stackrel{s}{=} \left[\begin{array}{c|c} T_m^{-1}W_m^T A V_m & T_m^{-1}W_m^T b \\ \hline cV_m & 0 \end{array} \right] = \left[\begin{array}{c|c} \widehat{H}_m & l_m \\ \hline k_m T_m & 0 \end{array} \right]$$

and

$$(14) \quad f_{m,2}(s) = g_m(s)W_m^T b \stackrel{s}{=} \left[\begin{array}{c|c} \frac{W_m^T A V_m T_m^{-1}}{c V_m T_m^{-1}} & \frac{W_m^T b}{0} \end{array} \right] = \left[\begin{array}{c|c} \widehat{G}_m & T_m l_m \\ \hline k_m & 0 \end{array} \right]$$

are different realizations of the same transfer function, namely, $f_m(s) = f_{m,1}(s) = f_{m,2}(s)$ for all s .

Remark. 2.1. Throughout this paper, we assume that T_m is nonsingular, which is equivalent to a breakdown-free Arnoldi process. Furthermore, for simplicity, we only consider the model reduction of single-input single-output transfer functions. For the model reduction of multi-input multioutput systems, one needs to resort to block Arnoldi schemes. For more details of such processes, including breakdown, we refer the reader to [4, 18] and the references therein.

Thus, for the m th-order approximate model described by (2), we can take either $f_m(s) = f_{m,1}(s)$ or $f_m(s) = f_{m,2}(s)$, where $f_{m,1}(s)$ and $f_{m,2}(s)$ are given by (13) and (14), respectively. The following procedure summarizes an oblique projection method for model reduction of large-scale systems.

ALGORITHM 2.1 (Krylov subspace model-reduction algorithm).

- *Start: Specify tolerances $\gamma > 0$ and $\epsilon > 0$; set an integer parameter m .*
- *Perform m steps of the Arnoldi process with (A, b) to produce $H_m, \widetilde{H}_m, V_m, \widetilde{V}_m$, and l_m .*
- *Perform m steps of the Arnoldi process with (A^T, c^T) to produce $G_m, \widetilde{G}_m, W_m, \widetilde{W}_m$, and k_m .*
- *Form the reduced-order model from either (13) or (14).*
- *Test the \mathcal{L}^∞ -norm of the errors in (11) and (12); if either (11) $> \gamma$ or (12) $> \epsilon$, increase m and continue the Arnoldi processes.*

The reduced-order models given by (13) and (14) are computed readily from the data generated in the course of the Arnoldi processes. It is known that such models may be unstable even if $f(s)$ is stable; furthermore, such partial realizations often contain modes associated with the outer part of the spectrum of A . The following section presents a framework able to remove such undesirable features by the application of further oblique projectors within an implicit restart setting.

3. A general implicit restart framework. Implicit restart schemes were first proposed to compute a few desired eigenvalues of large sparse nonsymmetric matrices [25]. More recently, they have been exploited to compute stable partial realizations in the setting of control problems and linear circuit analysis [14, 15]. The aim of this section is to propose a general implicit restart framework based on the Arnoldi process. A key feature of the present approach, which differs from existing implicit restart schemes, is that the integrity of the controllability and observability Arnoldi equations (4)–(7) is preserved. In contrast, existing methods focus on preserving (5) and (7), which makes it difficult to establish system theoretic connections between $f_m(s)$ and $f(s)$. In the eigenvalue setting, preserving the integrity of either (4) or (6) is not essential, since the starting vector is an eigenvector estimate and does not form part of the problem data. An advantage offered by preserving the Arnoldi equations is that one may establish computable error expressions that are similar to those of Theorem 2.1; furthermore, one may demonstrate that the implicitly restarted reduced-order model may be obtained by effecting low-rank perturbations to the state-space representation of $f(s)$.

Remark. 3.1. Commonly, implicit restarts refer to restarting the Arnoldi process with updated starting vectors v_1 and w_1 . In this work, we take a broader inter-

pretation, so that restarts refer to removing, via oblique projections, all features of $f_m(s)$ that are deemed undesirable in a given application and restarting the (modified) Arnoldi process while preserving the Arnoldi equations (4)–(7).

Suppose that m steps of the Arnoldi process have been taken and that $f_{m,1}(s)$, given in (13), is the reduced-order model obtained upon the application of the oblique projection process in Theorem 2.1. Observe that $f_{m,1}(s)$ is not necessarily stable; furthermore, similar to a power method, the Arnoldi process engenders an \widehat{H}_m whose eigenvalues approximate those of A with large absolute value. The presence of such features makes $f_{m,1}(s)$ unsuitable for many practical applications such as circuit simulation. Furthermore, robust controller design methods based on the small gain theorem require that the actual model (in this case $f(s)$) and the nominal model (in this case $f_{m,1}(s)$) have the same number of poles in the closed right-half complex plane [9, 10]. Since $f(s)$ is stable from the outset, stability of $f_{m,1}(s)$ is required for such methods. Furthermore, if many eigenvalues of \widehat{H}_m approximate those of A with large absolute value, $f_{m,1}(s)$ is a poor reduced-order model since it is unable to replicate the low-frequency characteristics of $f(s)$. Suppose that these undesirable features may be extracted via the application of an additional oblique projection process; namely, define two full-column rank matrices $T_L, T_R \in \mathbb{R}^{m \times r}$ such that $T_L^T T_R = I_r$, where I_r is the $r \times r$ identity matrix and $r < m$. Then the desirable portion of $f_{m,1}(s)$ is given by

$$f_{r,1}(s) \stackrel{s}{=} \left[\begin{array}{c|c} T_L^T \widehat{H}_m T_R & T_L^T l_m \\ \hline k_m T_m T_R & 0 \end{array} \right].$$

The following is referred to as a basis change in the state-space realization of any rational $g(s)$:

$$g(s) \stackrel{s}{=} \left[\begin{array}{c|c} A & b \\ \hline c & d \end{array} \right] \xrightarrow{T} g(s) \stackrel{s}{=} \left[\begin{array}{c|c} TAT^{-1} & Tb \\ \hline cT^{-1} & d \end{array} \right],$$

where T is nonsingular. The next result establishes that the projectors T_L and T_R applied to $f_{m,1}(s)$ may be combined with the oblique projectors generated in the course of the Arnoldi process; furthermore, the composite projectors yield a reduced-order model whose structure is reminiscent of (13).

PROPOSITION 3.1. *Suppose that m steps of the Arnoldi process have been taken and that T_m is nonsingular. Let $T_R = Q_R R_R$ and $(T_m^T)^{-1} T_L = Q_L R_L$ be QR decompositions in which $Q_L, Q_R \in \mathbb{R}^{m \times r}$ are parts of orthogonal matrices and $R_L, R_R \in \mathbb{R}^{r \times r}$ are upper triangular. Define $V_r := V_m Q_R$, $W_r := W_m Q_L$, and $T_r := W_r^T V_r$. Then*

1. T_r is nonsingular and $T_r^{-1} = (Q_L^T T_m Q_R)^{-1} = R_R R_L^T$,
2. $f_{r,1}(s)$ may be expressed as

$$(15) \quad f_{r,1}(s) \stackrel{s}{=} \left[\begin{array}{c|c} T_r^{-1} W_r^T A V_r & T_r^{-1} W_r^T b \\ \hline c V_r & 0 \end{array} \right].$$

Proof. Since T_L and T_R have full-column rank, R_L and R_R are nonsingular. Using the QR decompositions and $T_L^T T_R = I_r$, $T_r = Q_L^T T_m Q_R = (R_L^T)^{-1} R_R^{-1}$, from which part 1 follows. By Theorem 2.1, $f_{r,1}(s)$ may be expressed as

$$(16) \quad \begin{aligned} f_{r,1}(s) &\stackrel{s}{=} \left[\begin{array}{c|c} T_L^T T_m^{-1} W_m^T A V_m T_R & T_L^T T_m^{-1} W_m^T b \\ \hline c V_m T_R & 0 \end{array} \right] \\ &\stackrel{s}{=} \left[\begin{array}{c|c} R_R T_L^T T_m^{-1} W_m^T A V_m Q_R & R_R T_L^T T_m^{-1} W_m^T b \\ \hline c V_m Q_R & 0 \end{array} \right] \end{aligned}$$

upon substituting $T_R = Q_R R_R$ and effecting a basis change using the transformation R_R . Substituting $(T_m^T)^{-1} T_L = Q_L R_L$ into (16) and using part 1 readily establishes the claim in part 2. \square

Remark. 3.2. Alternatively, one may apply the transformations T_L and T_R to the reduced-order model $f_{m,2}(s)$, given in (14), to yield

$$f_{r,2}(s) \stackrel{s}{=} \left[\begin{array}{c|c} T_L^T \widehat{G}_m T_R & T_L^T T_m l_m \\ \hline k_m T_R & 0 \end{array} \right].$$

On effecting the QR decompositions $T_L = Q_L R_L$ and $T_m^{-1} T_R = Q_R R_R$, one readily demonstrates that

$$(17) \quad f_{r,2}(s) \stackrel{s}{=} \left[\begin{array}{c|c} W_r^T A V_r T_r^{-1} & W_r^T b \\ \hline c V_r T_r^{-1} & 0 \end{array} \right],$$

where V_r and W_r are defined in Proposition 3.1.

Observe that V_r and W_r continue to remain orthogonal bases to parts of the controllability and observability subspaces since Q_R and Q_L are parts of orthogonal matrices. From Proposition 3.1, it appears that $f_{r,1}(s)$ may be computed by imposing a Galerkin condition on the residual of (9). Thus, similar to the arguments leading to the statement of Problem 2.1, one defines an approximate solution to the first linear system of (9) of the form $f_{b,r}(s) = V_r h_r(s)$; then we consider $f_{r,1}(s) = c f_{b,r}(s)$ as an approximation to $f(s)$. Similarly, $f_{c,r}(s) = g_r(s) W_r^T$ defines an approximate solution to the second linear system of (9), which leads to an approximation of $f(s)$ given by $f_{r,2}(s) = f_{c,r}(s) b$. The functions $h_r(s)$ and $g_r(s)$ are then computed by imposing Galerkin conditions as shown by the next corollary. For notational convenience, define $\widehat{H}_r := T_r^{-1} W_r^T A V_r$ and $\widehat{G}_r := W_r^T A V_r T_r^{-1}$.

COROLLARY 3.2. *Suppose that the conditions of Proposition 3.1 are in force.*

Then

1. $h_r(s) = (sI - \widehat{H}_r)^{-1} T_r^{-1} W_r^T b$ if and only if $\{(sI - A)V_r h_r(s) - b\} \perp W_r$.
2. $g_r(s) = c V_r T_r^{-1} (sI - \widehat{G}_r)^{-1}$ if and only if $\{c - g_r(s) W_r^T (sI - A)\} \perp V_r$.

Proof. By direct calculation,

$$\begin{aligned} W_r^T \{(sI - A)V_r h_r(s) - b\} &= (sT_r - W_r^T A V_r) h_r(s) - W_r^T b \\ &= T_r \left\{ (sI - \widehat{H}_r) h_r(s) - T_r^{-1} W_r^T b \right\}. \end{aligned}$$

Part 1 of the corollary is readily established, since T_r is nonsingular. Part 2 is verified in a similar way. \square

The approximation to $f(s)$ is then given by $c V_r (sI - \widehat{H}_r)^{-1} T_r^{-1} W_r^T b$, which is $f_{r,1}(s)$ as defined in (15). Since $f_{r,1}(s)$ and $f_{r,2}(s)$ may be computed via the application of Galerkin conditions, it is natural to question whether the terms in (15) and (17) satisfy certain ‘‘Arnoldi-like’’ equations. The development that follows answers this question affirmatively and shows that this process naturally fits into an implicit restart framework.

Suppose that Q_{R_\perp} is the orthogonal completion of Q_R so that $[Q_R \ Q_{R_\perp}] \in \mathbb{R}^{m \times m}$ is an orthogonal matrix. Postmultiplying (5) by $[Q_R \ Q_{R_\perp}]$ enables one to express (4) and (5) as

$$(18) \quad \begin{aligned} b &= V_m [Q_R \ Q_{R_\perp}] [Q_R \ Q_{R_\perp}]^T l_m, \\ A V_m [Q_R \ Q_{R_\perp}] &= V_m [Q_R \ Q_{R_\perp}] [Q_R \ Q_{R_\perp}]^T H_m [Q_R \ Q_{R_\perp}] \\ &\quad + \widetilde{V}_m \widetilde{H}_m [Q_R \ Q_{R_\perp}], \end{aligned}$$

respectively, which leads to

$$(19) \quad b = V_r l_r + [V_m Q_{R\perp} \quad \tilde{V}_m] \begin{bmatrix} Q_{R\perp}^T l_m \\ 0 \end{bmatrix},$$

$$(20) \quad AV_r = V_r H_r + [V_m Q_{R\perp} \quad \tilde{V}_m] \begin{bmatrix} Q_{R\perp}^T H_m Q_R \\ \tilde{H}_m Q_R \end{bmatrix},$$

where $l_r = Q_R^T l_m$ and $H_r = Q_R^T H_m Q_R$. Observe that (20) is the (1,1) block of (18). Similarly, suppose that $Q_{L\perp}$ is the orthogonal completion of Q_L so that $[Q_L \quad Q_{L\perp}] \in \mathbb{R}^{m \times m}$ is an orthogonal matrix. Then postmultiplying (7) by $[Q_L \quad Q_{L\perp}]$ enables one to express (6) and (7) as

$$(21) \quad \begin{aligned} c^T &= W_m [Q_L \quad Q_{L\perp}] [Q_L \quad Q_{L\perp}]^T k_m^T, \\ A^T W_m [Q_L \quad Q_{L\perp}] &= W_m [Q_L \quad Q_{L\perp}] [Q_L \quad Q_{L\perp}]^T G_m^T [Q_L \quad Q_{L\perp}] \\ &\quad + \tilde{W}_m \tilde{G}_m^T [Q_L \quad Q_{L\perp}], \end{aligned}$$

respectively, which leads to

$$(22) \quad c^T = W_r k_r^T + [W_m Q_{L\perp} \quad \tilde{W}_m] \begin{bmatrix} Q_{L\perp}^T k_m^T \\ 0 \end{bmatrix},$$

$$(23) \quad A^T W_r = W_r G_r^T + [W_m Q_{L\perp} \quad \tilde{W}_m] \begin{bmatrix} Q_{L\perp}^T G_m^T Q_L \\ \tilde{G}_m^T Q_L \end{bmatrix},$$

where $k_r^T = Q_L^T k_m^T$ and $G_r = Q_L^T G_m Q_L$. Observe that (23) is the (1,1) block of (21). The key observation is that despite the application of an additional oblique projection process, (19)–(23) have an Arnoldi-like structure, except for the second term in the right-hand side of (19) and (22). This leads one to conclude that we may restart the iterative process with a view to improving the approximation.

3.1. An implicit restart scheme. The objective of this section is to propose an implicit restart scheme based on the Arnoldi-like equations given in (19)–(23). In order to effect a restart, consider (19) and (20), where

$$(24) \quad [V_m Q_{R\perp} \quad \tilde{V}_m] \in \mathbb{R}^{n \times (m-r+1)} \quad \text{and} \quad \begin{bmatrix} Q_{R\perp}^T H_m Q_R \\ \tilde{H}_m Q_R \end{bmatrix} \in \mathbb{R}^{(m-r+1) \times r}.$$

Suppose that the second term of (24) has $(m - r + 1)$ linearly independent rows; then the application of $(m - r)$ steps of restart would yield a basis for part of the controllability space spanned by the columns of $[V_r \quad V_m Q_{R\perp}]$. This basis is a rotation of V_m since $V_r = V_m Q_R$; therefore, $[V_r \quad V_m Q_{R\perp}]$ does not contribute to updating the reduced order model. Under such circumstances, the approximation errors would stagnate irrespective of the number of restarts employed. Therefore, for an effective restart scheme, one selects $2r < m$, which will be a standing assumption throughout this paper. Consider the QR decomposition

$$(25) \quad \begin{bmatrix} Q_{R\perp}^T l_m & Q_{R\perp}^T H_m Q_R \\ 0 & \tilde{H}_m Q_R \end{bmatrix} = Q [\tilde{l}_r \quad \tilde{H}_r],$$

where $[\tilde{l}_r \quad \tilde{H}_r] \in \mathbb{R}^{(r+1) \times (r+1)}$ is upper triangular and $Q \in \mathbb{R}^{(m-r+1) \times (r+1)}$ is part of an orthogonal matrix. Then (19) and (20) may be expressed as

$$(26) \quad b = V_r l_r + \tilde{V}_r \tilde{l}_r = [V_r \quad \tilde{V}_r] \begin{bmatrix} l_r \\ \tilde{l}_r \end{bmatrix},$$

$$(27) \quad AV_r = V_r H_r + \tilde{V}_r \tilde{H}_r = [V_r \ \tilde{V}_r] \begin{bmatrix} H_r \\ \tilde{H}_r \end{bmatrix},$$

where $\tilde{V}_r := [V_m Q_{R_\perp} \ \tilde{V}_m] Q \in \mathbb{R}^{n \times (r+1)}$. The following modified Gram–Schmidt process augments the Arnoldi-like equations (26) and (27) to yield a matrix $[V_m \ \tilde{V}_m]$, which is part of an orthogonal matrix, and an $(r + 1)$ upper Hessenberg matrix $[H_m^T \ \tilde{H}_m^T]^T$ (i.e., for $1 \leq j \leq m$, $h_{j+r+2,j} = 0$). Observe that the first r columns of $[H_m^T \ \tilde{H}_m^T]^T$ are already available from (27); furthermore, the first $2r + 1$ columns of $[V_m \ \tilde{V}_m]$ are also available prior to the restart. It is thus natural to consider a process which augments the existing data from dimension r to dimension m . The following is a modified Gram–Schmidt process that performs this task.

ALGORITHM 3.1 (an implicitly restarted modified Gram–Schmidt process).

- For $j = r + 1, r + 2, \dots, m$,
- $w := Av_j$,
- for $i = 1, 2, \dots, r + j$, $\begin{cases} h_{i,j} := w^T v_i, \\ w := w - v_i h_{i,j}, \end{cases}$
- $h_{j+r+1,j} := \|w\|_2$ and $v_{j+r+1} := w/h_{j+r+1,j}$.

Observe that $H_m = V_m^T AV_m$ is satisfied. Furthermore, the associated controllability Arnoldi equations in (4) and (5) remain in force with $H_m, V_m, \tilde{H}_m, \tilde{V}_m$, and $l_m = [l_r^T \ \tilde{l}_r^T \ 0]^T$ defined by the implicitly restarted Gram–Schmidt process. These variables overwrite those computed in the previous restart step. To clarify the structure, we present an illustrative example in which $r = 2$ and $m = 6$. Writing (26) and (27) as

$$b = [v_1 \ v_2 \ | \ v_3 \ v_4 \ v_5] \begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ 0 \\ 0 \end{bmatrix}, \quad A[v_1 \ v_2] = [v_1 \ v_2 \ | \ v_3 \ v_4 \ v_5] \begin{bmatrix} \times & \times \\ \times & \times \\ \times & \times \\ 0 & \times \end{bmatrix},$$

the following structure is obtained by augmenting these equations to $m = 6$:

$$b = [v_1 \ v_2 \ v_3 \ v_4 \ v_5 \ v_6] \begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} =: V_m l_m,$$

$$A[v_1 \ v_2 \ v_3 \ v_4 \ v_5 \ v_6] = [v_1 \ v_2 \ v_3 \ v_4 \ v_5 \ v_6 \ | \ v_7 \ v_8 \ v_9] \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ \hline 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \times \end{bmatrix},$$

$$\Rightarrow \quad AV_m = [V_m \ \tilde{V}_m] \begin{bmatrix} H_m \\ \tilde{H}_m \end{bmatrix}.$$

In order to effect a restart with the observability Arnoldi equations, consider the QR decomposition

$$(28) \quad \begin{bmatrix} Q_{L\perp}^T k_m^T & Q_{L\perp}^T G_m^T Q_L \\ 0 & \tilde{G}_r^T Q_L \end{bmatrix} = U \begin{bmatrix} \tilde{k}_r^T & \tilde{G}_r^T \end{bmatrix},$$

where $\begin{bmatrix} \tilde{k}_r^T & \tilde{G}_r^T \end{bmatrix} \in \mathbb{R}^{(r+1) \times (r+1)}$ is upper triangular and $U \in \mathbb{R}^{(m-r+1) \times (r+1)}$ is part of an orthogonal matrix. Thus, the observability Arnoldi equations may be expressed as

$$(29) \quad c^T = W_r k_r^T + \tilde{W}_r \tilde{k}_r^T \quad \text{and} \quad A^T W_r = W_r G_r^T + \tilde{W}_r \tilde{G}_r^T,$$

where $\tilde{W}_r = [W_m Q_{L\perp} \tilde{W}_m] U \in \mathbb{R}^{n \times (r+1)}$ is part of an orthogonal matrix. One may now employ the implicitly restarted modified Gram–Schmidt process to augment (29) and to produce G_m, W_m, \tilde{W}_m , and k_m satisfying the Arnoldi equations (6) and (7). A reduced-order model is then formed following (13) or (14). Suppose that the undesirable features present in this model may be removed by two full-rank matrices T_L and T_R such that $T_L^T T_R = I_r$, then Proposition 3.1 may be used to extract the unwanted features by effecting a further state reduction. The restart process may then be repeated until convergence. The following is an outline of the implicitly restarted model-reduction algorithm.

ALGORITHM 3.2 (implicitly restarted model-reduction algorithm).

- **Start:** Specify m and r such that $m > 2r$.
 1. Perform m steps of the Arnoldi process with (A, b) to find $H_m, \tilde{H}_m, V_m, \tilde{V}_m$, and l_m .
 2. Perform m steps of the Arnoldi process with (A^T, c^T) to find $G_m, \tilde{G}_m, W_m, \tilde{W}_m$, and k_m .
- **Restart:** Effect the QR decompositions

$$T_R = [Q_R \ Q_{R\perp}] \begin{bmatrix} R_R \\ 0 \end{bmatrix} \quad \text{and} \quad (T_m)^{-T} T_L = [Q_L \ Q_{L\perp}] \begin{bmatrix} R_L \\ 0 \end{bmatrix}.$$

1. Effect the QR decompositions (25) and (28) which define the terms in (26), (27), and (29).
2. Evaluate the residual errors; if satisfied, form the reduced-order model using (15) or (17) and stop; otherwise, continue.
3. Effect $m - r$ implicitly restarted, modified Gram–Schmidt steps using $[V_r^T \ \tilde{V}_r^T]^T, [H_r^T \ \tilde{H}_r^T]^T$, and $[V_r \ \tilde{V}_r]$ to yield $l_m, [H_m^T \ \tilde{H}_m^T]^T$, and $[V_m \ \tilde{V}_m]$.
4. Effect $m - r$ implicitly restarted, modified Gram–Schmidt steps using $[k_r \ \tilde{k}_r], [G_r \ \tilde{G}_r]$, and $[W_r \ \tilde{W}_r]$ to yield $k_m, [G_m \ \tilde{G}_m]$, and $[W_m \ \tilde{W}_m]$.

Remark. 3.3. In practice, it is advisable to compute the QR factorizations in step 1 of the restart via the Gram–Schmidt or modified Gram–Schmidt orthogonalization processes. For full details, including the case that the left-hand sides of (25) and (28) do not have full-column rank, see [4, 13].

Two elements of Algorithm 3.2 have yet to be discussed: the first addresses the selection of appropriate T_L and T_R of the restart step; the second concerns computable error formulas for step 2 of the restart process. The second is the object of the next subsection; the first is covered in section 4.

3.2. Theoretical properties. The aim of this section is to provide a theoretical analysis of the implicit restart algorithm presented in section 3.1.

To gauge the quality of the reduced-order model as each restart is completed, the following theorem provides computable expressions for the \mathcal{L}^∞ -norm of the residual errors.

THEOREM 3.3. *Suppose that m steps of the Arnoldi process have been completed and that the controllability and observability Arnoldi equations are given by (26) and (27) and (29), respectively. The \mathcal{L}^∞ -norm of the residual errors associated with the approximate solutions to (9) are given by*

$$(30) \quad \|b - (sI - A)V_r h_r(s)\|_\infty = \left\| \begin{bmatrix} T_r^{-1}W_r^T \tilde{V}_r \\ I \end{bmatrix} \left(\tilde{l}_r + \tilde{H}_r h_r(s) \right) \right\|_\infty,$$

$$(31) \quad \|c - g_r(s)W_r^T(sI - A)\|_\infty = \left\| \left(\tilde{k}_r + g_r(s)\tilde{G}_r \right) \begin{bmatrix} \tilde{W}_r^T V_r T_r^{-1} & I \end{bmatrix} \right\|_\infty.$$

Proof. Substituting (26) and (27) into the left-hand side of (30) yields

$$\begin{aligned} \|b - (sI - A)V_r h_r(s)\|_\infty &= \left\| \begin{bmatrix} l_r - (sI - H_r)h_r(s) \\ \tilde{l}_r + \tilde{H}_r h_r(s) \end{bmatrix} \right\|_\infty \\ &= \left\| \begin{bmatrix} -T_r^{-1}W_r^T \tilde{V}_r \tilde{l}_r + (H_r - \tilde{H}_r)h_r(s) \\ \tilde{l}_r + \tilde{H}_r h_r(s) \end{bmatrix} \right\|_\infty. \end{aligned}$$

Substituting $T_r^{-1}W_r^T \times (27)$ into the (1,1) block yields the desired result. The residual error expression of (31) is derived in a similar way using (29). \square

Next, we establish properties of the reduced-order model that are reminiscent of those derived in [18]. We begin by deriving low-rank approximate solutions to (3) by imposing Galerkin-type conditions on their associated residual errors. Suppose that the low-rank Gramians have the form $P_r = V_r X_r V_r^T$ and $Q_r = W_r Y_r W_r^T$ for some symmetric matrices X_r and $Y_r \in \mathbb{R}^{r \times r}$. The residual error functions associated with a particular choice of X_r and Y_r are then defined by

$$R_r = AV_r X_r V_r^T + V_r X_r V_r^T A^T + bb^T, \quad S_r = A^T W_r Y_r W_r^T + W_r Y_r W_r^T A + c^T c.$$

The residual error functions may be factorized by using (26)–(29) to yield

$$(32) \quad \begin{aligned} R_r &= [V_r \quad (I - V_r T_r^{-1} W_r^T) \tilde{V}_r] \\ &\times \begin{bmatrix} \hat{H}_r X_r + X_r \hat{H}_r^T + T_r^{-1} W_r^T bb^T W_r (T_r^T)^{-1} & X_r \tilde{H}_r^T + T_r^{-1} W_r^T b \tilde{l}_r^T \\ \tilde{H}_r X_r + \tilde{l}_r b^T W_r (T_r^T)^{-1} & \tilde{l}_r \tilde{l}_r^T \end{bmatrix} \\ &\times \begin{bmatrix} V_r^T \\ \tilde{V}_r^T (I - W_r (T_r^T)^{-1} V_r^T) \end{bmatrix}, \end{aligned}$$

$$(33) \quad \begin{aligned} S_r &= [W_r \quad (I - W_r (T_r^T)^{-1} V_r^T) \tilde{W}_r] \\ &\times \begin{bmatrix} \hat{G}_r^T Y_r + Y_r \hat{G}_r + (T_r^T)^{-1} V_r^T c^T c V_r T_r^{-1} & Y_r \tilde{G}_r + (T_r^T)^{-1} V_r^T c^T \tilde{k}_r \\ \tilde{G}_r^T Y_r + \tilde{k}_r^T c V_r T_r^{-1} & \tilde{k}_r^T \tilde{k}_r \end{bmatrix} \\ &\times \begin{bmatrix} W_r^T \\ \tilde{W}_r^T (I - V_r T_r^{-1} W_r^T) \end{bmatrix}. \end{aligned}$$

The Arnoldi–Lyapunov solvers considered here seek symmetric X_r and Y_r such that the residual R_r and S_r satisfy orthogonality properties with respect to parts of the

controllability and observability subspaces spanned by the columns of V_r and W_r , respectively. The following theorem determines the low-rank Gramians that satisfy such orthogonality conditions.

THEOREM 3.4. *Suppose that m steps of the Arnoldi process have been completed and that the controllability and observability Arnoldi equations are given by (26) and (27) and (29), respectively. Furthermore, suppose that $\lambda_i(\widehat{H}_r) + \bar{\lambda}_j(\widehat{H}_r) \neq 0 \quad \forall i, j$ and $\lambda_i(\widehat{G}_r) + \bar{\lambda}_j(\widehat{G}_r) \neq 0 \quad \forall i, j$. Then*

1. $W_r^T R_r W_r = 0$ if and only if

$$(34) \quad \widehat{H}_r X_r + X_r \widehat{H}_r^T + T_r^{-1} W_r^T b b^T W_r (T_r^T)^{-1} = 0.$$

Under these conditions,

$$\|R_r\|_F = \left\| \left[\begin{array}{cc} H_r X_r + X_r H_r^T + l_r l_r^T & X_r \widetilde{H}_r^T + l_r \widetilde{l}_r^T \\ \widetilde{H}_r X_r + \widetilde{l}_r l_r^T & \widetilde{l}_r \widetilde{l}_r^T \end{array} \right] \right\|_F.$$

2. $V_r^T S_r V_r = 0$ if and only if

$$(35) \quad \widehat{G}_r^T Y_r + Y_r \widehat{G}_r + (T_r^T)^{-1} V_r^T c^T c V_r T_r^{-1} = 0.$$

Under these conditions,

$$\|S_r\|_F = \left\| \left[\begin{array}{cc} G_r^T Y_r + Y_r G_r + k_r^T k_r & Y_r \widetilde{G}_r + k_r^T \widetilde{k}_r \\ \widetilde{G}_r^T Y_r + \widetilde{k}_r^T k_r & \widetilde{k}_r^T \widetilde{k}_r \end{array} \right] \right\|_F.$$

3. X_r and $T_r^T Y_r T_r$ are, respectively, the controllability and observability Gramians of $f_{r,1}(s)$.

4. $T_r X_r T_r^T$ and Y_r are, respectively, the controllability and observability Gramians of $f_{r,2}(s)$.

Proof. The proof is similar to the proofs of Theorems 2.3 and 3.2 in [18] and is omitted. \square

The following result establishes that the low-rank Gramians P_r and Q_r are the exact Gramians of a pair of perturbed Lyapunov equations. It also gives the approximations $f_{r,1}(s)$ and $f_{r,2}(s)$ defined in (15) and (17), respectively, as minimal realizations of various perturbations of $f(s)$.

THEOREM 3.5. *Suppose that m steps of the Arnoldi process have been completed and that the controllability and observability Arnoldi equations are given by (26) and (27) and (29), respectively. Suppose that $P_r := V_r X_r V_r^T$ and $Q_r := W_r Y_r W_r^T$ are the low-rank approximate solutions to (3), where X_r and Y_r satisfy (34) and (35), respectively.*

1. Define the perturbations $\Delta = (I - V_r T_r^{-1} W_r^T)$, $\Delta_1 = \Delta \widetilde{V}_r \widetilde{H}_r V_r^T$, and $\Delta_2 = W_r \widetilde{G}_r \widetilde{W}_r^T \Delta$. Then

$$(36) \quad (A - \Delta_1) P_r + P_r (A - \Delta_1)^T + (I - \Delta) b b^T (I - \Delta)^T = 0,$$

$$(37) \quad (A - \Delta_2)^T Q_r + Q_r (A - \Delta_2) + (I - \Delta)^T c^T c (I - \Delta) = 0.$$

Furthermore,

$$\|\Delta_1\|_F^2 = \|\widetilde{H}_r\|_F^2 + \|T_r^{-1} W_r^T \widetilde{V}_r \widetilde{H}_r\|_F^2,$$

$$\|\Delta_2\|_F^2 = \|\widetilde{G}_r\|_F^2 + \|\widetilde{G}_r \widetilde{W}_r^T V_r T_r^{-1}\|_F^2,$$

$$\|\Delta\|_F^2 = n - 2r + \|T_r^{-1}\|_F^2.$$

2. Define $\Delta_3 := \Delta_1 + \Delta_2$. Then $W_r^T \Delta_1 V_r = W_r^T \Delta_2 V_r = W_r^T \Delta V_r = W_r^T \Delta_3 V_r = 0$. Furthermore,

$$\begin{aligned} (A - \Delta_3)P_r + P_r(A - \Delta_3)^T + (I - \Delta)bb^T(I - \Delta)^T &= 0, \\ (A - \Delta_3)^T Q_r + Q_r(A - \Delta_3) + (I - \Delta)^T c^T c(I - \Delta) &= 0. \end{aligned}$$

3. Define $f_{\Delta_i}(s) \stackrel{s}{=} ((A - \Delta_i), (I - \Delta)b, c(I - \Delta), 0)$ for $i = 1, 2, 3$. Then

$$f_{r,1}(s) = f_{r,2}(s) = f_{\Delta_i}(s) \quad \forall s, \quad i = 1, 2, 3.$$

Proof. Substituting (34) into (32) and rearranging yields (36), while substituting (35) into (33) and rearranging yields (37). Part 2 follows by direct calculation using part 1. The proof of part 3 is similar to the proof of Corollary 3.3 in [18] and is omitted. \square

The effects of Δ_3 and Δ are to perturb A , b , and c in such a way that the nonminimal modes of the perturbed system are simultaneously uncontrollable and unobservable. The perturbation Δ_1 to the transition matrix of $f(s)$ yields $n - r$ uncontrollable modes while the perturbation Δ_2 gives rise to $n - r$ unobservable modes [18].

Observe that Δ_1 , Δ_2 , and Δ_3 are additive perturbations on the state transition matrix of $f(s)$ while Δ is a multiplicative perturbation on the input and output vectors b and c . It is interesting to observe that despite the fact that Δ_1 , Δ_2 , and Δ_3 have different Frobenius norms, each perturbed linear system is a different realization of the same transfer function.

Although the Arnoldi equations (4)–(7) continue to be satisfied after any number of restarts, the structure of the variables $H_m, \tilde{H}_m, l_m, k_m, G_m$, and \tilde{G}_m differs from that required by the Arnoldi process, namely, that

$$E_m := \begin{bmatrix} l_m & H_m \\ 0 & \tilde{H}_m \end{bmatrix}, \quad F_m := \begin{bmatrix} k_m & 0 \\ G_m & \tilde{G}_m \end{bmatrix}$$

are upper and lower triangular, respectively. Instead, E_m and F_m for the restart scheme are, respectively, r -upper and r -lower Hessenberg (i.e., for $1 \leq j \leq m$, $(E_m)_{j+r+1,j} = (F_m)_{j,j+r+1} = 0$) (see Algorithm 3.2 and Algorithm 3.1 and the subsequent discussion). This implies that the moment-matching property

$$(38) \quad cA^{i-1}b = c_m A_m^{i-1} b_m, \quad 1 \leq i \leq 2m$$

[11], which essentially follows from the triangular structure of E_m and F_m , no longer applies for the implicit restart scheme and an alternative justification of the scheme is required. In the implicitly restarted model-reduction algorithm in [14, 15], the authors give an equation similar to (38), relating modified moments of the original and restarted Lanczos model. Here, we show that (38) is still satisfied, albeit for lower values of i .

THEOREM 3.6. *Suppose that $H_m, \tilde{H}_m, l_m, k_m, G_m$, and \tilde{G}_m are output by Algorithm 3.2; define*

$$(39) \quad \begin{aligned} A_m &= H_m + T_m^{-1} W_m^T \tilde{V}_m \tilde{H}_m = T_m^{-1} W_m^T A V_m, \\ b_m &= l_m = T_m^{-1} W_m^T b, \quad c_m = k_m T_m = c V_m \end{aligned}$$

for nonsingular $T_m = W_m^T V_m$; and let k be the largest integer satisfying

$$(40) \quad k \leq \frac{m}{r+1}.$$

Then

$$(41) \quad A^{i-1}b = V_m A_m^{i-1} b_m, \quad cA^{i-1} = c_m A_m^{i-1} T_m^{-1} W_m^T, \quad 1 \leq i \leq k.$$

Hence,

$$(42) \quad cA^{i-1}b = c_m A_m^{i-1} b_m, \quad 1 \leq i \leq 2k.$$

Proof. It follows from the r -upper triangular structure of E_m that

$$(43) \quad \tilde{H}_m H_m^{i-1} l_m = 0, \quad 1 \leq i \leq k-1.$$

Hence (39) implies that

$$(44) \quad H_m^{i-1} l_m = A_m^{i-1} l_m = A_m^{i-1} b_m, \quad 1 \leq i \leq k.$$

Repeated evaluation of $b, Ab, \dots, A^i b$ using (4), (5), (43), and (44) verifies the first part of (41). A similar procedure, using G_m, \tilde{G}_m , and k_m , verifies the second part of (41). Finally, (42) follows from (41) upon noting that

$$\begin{aligned} cA^{2i-1}b &= cA^{i-1}AA^{i-1}b = c_m A_m^{i-1} T_m^{-1} W_m^T A V_m A_m^{i-1} B_m \\ &= c_m A_m^{2i-1} b_m, \quad 1 \leq i \leq k. \end{aligned}$$

Notice that when $r = 0$ (no restarts), $k = m$ and (42) reduces to (38). \square

Remark. 3.4. One difficult issue associated with any restart scheme is the choice of m . Clearly, if $m = n - 1$, then $f_{m,1}(s) = f(s)$. So the question is how small should m be to guarantee convergence of $f_{r,1}(s)$ to the balanced truncation of $f(s)$? Theorem 3.6 states that the restarted process generating $f_{m,1}(s)$ matches fewer moments than an $f_{m,1}(s)$ based on no restarts. One interpretation is that, without restarts, $f_{m,1}(s)$ tends to be a good approximation to the high-frequency component of $f(s)$, while effecting implicit restarts via the stable projection and balanced truncation of $f_{m,1}(s)$, improves the approximation at low frequencies at the expense of degraded high-frequency behavior.

Suppose that r is given and that $f_{m,1}(s)$ must match at least a given number of moments of $f(s)$ at $s = \infty$. Then (40) and (42) suggest a minimum value of m to guarantee the moment-matching condition.

4. Stable projection and balanced truncation. The objective of this section is to suggest transformations T_L and T_R which enable Algorithm 3.2 to form stable partial realizations that retain the low-frequency characteristics of $f(s)$.

Suppose that m steps of the Arnoldi process have been completed and that $f_m(s)$ is an unstable partial realization of $f(s)$. The approach proposed here then determines T_L and T_R , which effects a stable projection of $f_m(s)$. In other words, the application of T_L and T_R to $f_m(s)$ yields $f_{m+}(s)$, where $f_m(s) = f_{m+}(s) + f_{m-}(s)$ in which $f_{m+}(s)$ is stable and $f_{m-}(s)$ is antistable. For the purposes of the present discussion, suppose that $f_m(s) = c_m(sI - A_m)^{-1}b_m$, then transform A_m to a block-ordered real Schur form

$$T_1 A_m T_1^T = A_s = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

in which T_1 is orthogonal, $A_{11} \in \mathbb{R}^{p \times p}$ is stable, and $A_{22} \in \mathbb{R}^{(m-p) \times (m-p)}$ is anti-stable. Following this change of basis, the system of first-order differential equations is described by

$$(45) \quad \dot{x} = A_s x + T_1 b_m u, \quad y = c_m T_1^T x.$$

The next step of the stable projection process is to eliminate the (1,2) block of A_s by solving the Sylvester equation $A_{11}X - XA_{22} + A_{12} = 0$, which has a solution due to the inertia properties of A_{11} and A_{22} ; see [8]. Applying the basis change $T_2 = \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix}$ to the linear dynamical system in (45) yields

$$T_2 A_s T_2^{-1} =: \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, \quad T_2 T_1 b_m =: \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad c_m T_1^T T_2^{-1} =: [c_1 \quad c_2].$$

With this decomposition complete, we conclude that

$$T_{L_1} := T_1^T T_2^T \begin{bmatrix} I \\ 0 \end{bmatrix} = T_1^T \begin{bmatrix} I \\ -X^T \end{bmatrix} \quad \text{and} \quad T_{R_1} := T_1^T T_2^{-1} \begin{bmatrix} I \\ 0 \end{bmatrix} = T_1^T \begin{bmatrix} I \\ 0 \end{bmatrix}.$$

Finally, the stable part of $f_m(s)$ is

$$f_{m+}(s) := \left[\begin{array}{c|c} T_{L_1}^T A_m T_{R_1} & T_{L_1}^T b_m \\ \hline c_m T_{R_1} & 0 \end{array} \right] = c_1 (sI - A_{11})^{-1} b_1.$$

This selection of T_L and T_R may be used in Algorithm 3.2 to yield an implicitly restarted model-reduction algorithm that is reminiscent of [14, 15].

Similar to a power method, Krylov subspace methods generate partial realizations in which the spectrum of A_m is known to approximate the outer part of the spectrum of A . The presence of such eigenvalues contributes little to the low-frequency characteristics of a dynamical model and may be removed without altering the model's behavior. It is therefore natural to consider the application of a model-reduction step to $f_{m+}(s)$ whose purpose is to extract any redundant modes that might be present. To this end, one resorts to either the square root or the Schur-based algorithms expounded in [24]. Suppose that the state dimension of $f_{m+}(s)$ is p ; then the following procedure determines the additional transformations, T_{L_2} and T_{R_2} , which extract the undesirable modes.

ALGORITHM 4.1 (square root algorithm).

- Calculate the solutions P_s and Q_s to the Lyapunov equations

$$A_{11} P_s + P_s A_{11}^T + b_1 b_1^T = 0 \quad \text{and} \quad A_{11}^T Q_s + Q_s A_{11} + c_1^T c_1 = 0.$$

- Effect the factorizations $P_s = L_r L_r^T$ and $Q_s = L_o L_o^T$, and compute the singular value decomposition $\widehat{U} \Sigma_p \widehat{V}^T = L_o^T L_r$, where $\Sigma_p = \text{diag}(\sigma_1, \dots, \sigma_p)$ and $\sigma_1 \geq \dots \geq \sigma_p$.
- Suppose that the first r modes of $f_{m+}(s)$ are to be retained. Then define the transformations $T_{L_2} = L_o \widehat{U}_r \Sigma_r^{-1/2} \in \mathbb{R}^{r \times p}$ and $T_{R_2} = L_r \widehat{V}_r \Sigma_r^{-1/2} \in \mathbb{R}^{r \times p}$, where $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$ and \widehat{U}_r and \widehat{V}_r are the first r columns of \widehat{U} and \widehat{V} , respectively.

We could also use the Schur-based algorithm in [24], which would yield different T_{L_2} and T_{R_2} but would result in a different realization of the same reduced-order model. In [24], Safonov and Chiang show that the square root and Schur-based

model-reduction algorithms are equivalent to Moore’s balanced truncation method [20]. In the present setting, the reduced-order model is given by

$$(46) \quad f_r(s) = \left[\begin{array}{c|c} \frac{T_{L_2}^T A_{11} T_{R_2}}{c_1 T_{R_2}} & \frac{T_{L_2}^T b_1}{0} \\ \hline & \end{array} \right] =: \left[\begin{array}{c|c} A_r & b_r \\ \hline c_r & 0 \end{array} \right].$$

In a practical implementation, suppose that $f_m(s)$ is an unstable partial realization that contains several redundant modes. Then, it is natural to combine the stable projection and balanced truncation processes to form composite transformations that yield an $f_r(s)$ which is both stable and free of redundant modes. Such composite projectors are given by

$$T_L = T_{L_1} T_{L_2} = T_1^T \left[\begin{array}{c} I \\ -X^T \end{array} \right] L_o \hat{U}_r \Sigma_r^{-1/2} \quad \text{and} \quad T_R = T_{R_1} T_{R_2} = T_1^T \left[\begin{array}{c} I \\ 0 \end{array} \right] L_r \hat{V}_r \Sigma_r^{-1/2},$$

which may be applied to $f_m(s)$ to yield $f_r(s)$ in a single step. It is interesting to observe that an implicit restart scheme may be obtained for any T_R and T_L provided that $T_L^T T_R = I_r$.

The following corollary establishes that the oblique projection methods of this paper are closely related to balanced truncation, namely, that both methods satisfy orthogonality conditions with respect to oblique projectors.

COROLLARY 4.1. *Suppose that A_r , b_r , and c_r are defined in (46) and define $h_{bal}(s) = (sI - A_r)^{-1} b_r$ and $g_{bal}(s) = c_r (sI - A_r)^{-1}$. Then*

$$(sI - A_{11}) T_{R_2} h_{bal}(s) - b_1 \perp T_{L_2} \quad \text{and} \quad g_{bal}(s) T_{L_2}^T (sI - A_{11}) - c_1 \perp T_{R_2}.$$

Proof. The proof is similar to the proof of Corollary 3.2. □

5. Numerical experiments. The purpose of this section is to illustrate, with the help of two examples, the behavior of the implicitly restarted model-reduction algorithm presented in sections 3 and 4. The tests reported here were performed on a Sparc-10 Sun workstation using Pro-MATLAB 4.2 which carries out operations to a unit round off of 2.22×10^{-16} .

The first problem is set up with $A \in \mathbb{R}^{n \times n}$, where $n = 100$ and the top left-hand 4×4 block of A is set to

$$\left[\begin{array}{cccc} -0.01 & 0.1 & 0 & 0 \\ -0.1 & -0.01 & 0 & 0 \\ 0 & 0 & -0.1 & 0.5 \\ 0 & 0 & -0.5 & -0.1 \end{array} \right],$$

while the remaining nonzero elements of A are uniformly distributed in $(0, -1)$ and are all located on the leading diagonal. Consequently, all the system poles are real except for four, which are $-0.01 \pm 0.1j$ and $-0.1 \pm 0.5j$. The first 10 elements of b and c are uniformly distributed in $[0, 1]$, while the 90 remaining elements are uniformly distributed in $[0, 1/25]$. We take $m = 10$ and $r = 4$. The infinity norm of $f(s)$ may be computed from $\max |f(j\omega)| \forall \omega \in \mathbb{R}$. Table 1 shows the evolution of the \mathcal{L}^∞ -norm of the error expressions of (30) and (11), denoted here by Err2 and Err3, respectively, against the number of restarts. The first column of Table 1 shows Err1, which denotes the \mathcal{L}^∞ -norm of the error $f_{bal}(s) - f_{r,1}(s)$, where $f_{bal}(s)$ is the r th-order balanced truncation of $f(s)$. The table indicates that Err2 and Err3 fall in magnitude as the number of restarts increases; however, as is well known, Galerkin

TABLE 1

 \mathcal{L}^∞ -error norms associated with the implicitly restarted Arnoldi model-reduction scheme.

Restarts	Err1	Err2	Err3	Restarts	Err1	Err2	Err3
0	.3245	.7737	.3414	8	.0032	.7284	.0563
1	.1782	.7609	.3158	9	.0026	.7284	.0491
2	.0790	.7294	.2317	10	.0021	.7284	.0493
3	.0360	.7294	.1344	11	.0018	.7284	.0463
4	.0201	.7291	.1013	12	.0015	.7285	.0474
5	.0117	.7287	.0656	13	.0012	.7285	.0457
6	.0069	.7288	.0712	14	.0009	.7285	.0469
7	.0044	.7285	.0607	15	.0007	.7285	.0456

conditions of the type in (10) do not guarantee a nonincreasing evolution of Err2 or Err3. Note that Err2 stagnates after an initial drop indicating convergence. Our experience with similar examples indicates that Err1 always tends to zero (for large enough m), which implies that $f_{r,1}(s)$ converges to $f_{bal}(s)$, although proving this remains an open question. Observe that Err2 and Err3 do not converge to zero. This follows from the fact that when approximating $f(s)$ by a stable k th-order model $f_k(s)$, $\|f(s) - f_k(s)\|_\infty$ is greater than or equal to the $(k+1)$ st Hankel singular value of $f(s)$ [8].

It is known that for many restart algorithms there exists a value of m below which convergence of the solution is very slow or not possible [25]. The second example illustrates that this is also the case for the restart algorithm presented here. The problem is set up with $A \in \mathbb{R}^{n \times n}$, where $n = 300$. The eigenvalues of A are all in the open left half plane; the real parts are uniformly distributed in the interval $[-1, 0)$, while the imaginary parts are randomly distributed in the interval $[-5, 5]$. $b, c^T \in \mathbb{R}^{n \times 1}$ are random.

TABLE 2

Evolution of relative \mathcal{L}^∞ -error norms for $m = 60$.

Restarts	1	2	3	4	5	6	7	8	9
E (%)	64	79	64	68	66	123	67	87	82

TABLE 3

Evolution of relative \mathcal{L}^∞ -error norms for $m = 70$.

Restarts	1	2	3
E (%)	333	23	4

Tables 2 to 4 illustrate the evolution of the percentage error $E = \|f_{r,1}(s) - f_{bal}(s)\|_\infty / \|f_{bal}(s)\|_\infty$ as a function of m and the number of restarts. Here $r = 5$ and $f_{bal}(s)$ is the r th-order balanced truncation of $f(s) = c(sI - A)^{-1}b$. Note that for $m = 60$, no convergence occurs; for $m = 70$, $f_{r,1}(s)$ converges to within 4% of $f_{bal}(s)$ after three restarts; while for $m = 75$, convergence is within 0.1% after only two restarts. This example is typical of our numerical experience. Determining a least value of m to guarantee convergence is still an unresolved problem and is under investigation (see Remark 3.4).

6. Conclusions. This paper presents and tests a model-reduction algorithm for large-scale, stable, linear, and time-invariant dynamic systems. We have developed a

TABLE 4
Evolution of relative \mathcal{L}^∞ -error norms for $m = 75$.

Restarts	1	2
E (%)	4.7	0.1

technique which combines the oblique Krylov subspace projectors with further projectors in order to obtain stable reduced-order approximate models that are able to approximate the low-frequency behavior of the dynamic system. We also established that this technique fits naturally within an implicit restart framework that defines an iterative procedure able to refine approximations. Exact low-dimensional expressions for the \mathcal{L}^∞ -norm of the residual errors are also derived. Our numerical experiments on several large-scale examples indicate that this process converges to the balanced truncation of the dynamic system; however, formally establishing this claim remains an open research problem. In place of the Arnoldi process, one may equally employ the Lanczos algorithm to derive an implicitly restarted scheme for stable partial realization. The derivation of such a scheme follows similar lines to those presented in this paper, except that in the Lanczos setting the biorthogonality of $[V_m \ \tilde{V}_m]$ and $[W_m \ \tilde{W}_m]$ is enforced [5, 21].

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*, 30th Conference on Decision and Control, England, 1991.

[2] W. ARNOLDI, *The principle of minimised iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.

[3] D. L. BOLEY, *Krylov space methods on state-space control models*, CS Technical report, TR92-18, Department of Computer Science, University of Minnesota, Minnesota, 1992.

[4] D. L. BOLEY AND G. H. GOLUB, *The Lanczos-Arnoldi algorithm and controllability*, Systems Control Lett., 4 (1984), pp. 317–327.

[5] D. L. BOLEY AND G. H. GOLUB, *The nonsymmetric Lanczos algorithm and controllability*, Systems Control Lett., 16 (1991), pp. 97–105.

[6] P. FELDMANN AND R. W. FREUND, *Efficient Linear Circuit Analysis by Padé Approximation via the Lanczos Process*, Technical report No. 11274-940217-02TM, AT&T Bell Laboratories, Murray Hill, NJ, 1994.

[7] C. W. GEAR AND Y. SAAD, *Iterative solution of linear equations in ODE codes*, SIAM. J. Sci. Statist. Comput., 4 (1983), pp. 583–601.

[8] K. GLOVER, *All optimal Hankel norm approximations of linear multivariable systems and their \mathcal{L}^∞ error bounds*, Inter. J. Control, 39 (1984), pp. 1115–1193.

[9] K. GLOVER, *Multiplicative approximation of linear multivariable systems with \mathcal{L}_∞ error bounds*, American Control Conference, Seattle, WA, pp. 1705–1709, 1986.

[10] M. GREEN AND D. J. N. LIMEBEER, *Linear Robust Control*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[11] W. B. GRAGG AND A. LINDQUIST, *On the partial realization problem*, Linear Algebra Appl., 50 (1983), pp. 277–319.

[12] G. H. GOLUB, B. KÄGSTRÖM, AND P. VAN DOOREN, *Direct block tridiagonalisation of single-input single-output systems*, System Control Lett., 18 (1992), pp. 109–120.

[13] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 2nd ed., John Hopkins University Press, Baltimore, MD, 1990.

[14] E. J. GRIMME, D. C. SORENSEN, AND P. VAN DOOREN, *Stable partial realizations via an implicitly restarted Lanczos method*, American Control Conference, Baltimore, MD, 1994.

[15] K. GALLIVAN, E. J. GRIMME, AND P. VAN DOOREN, *Asymptotic waveform evaluation via a restarted Lanczos method*, Appl. Math. Lett., 7 (1994), pp. 75–80.

[16] I. M. JAIMOUKHA, E. M. KASENALLY, AND D. J. N. LIMEBEER, *Numerical solution of large scale Lyapunov equations using Krylov subspace methods*, 31st Conference on Decision

- and Control, Tucson, AZ, 1992.
- [17] I. M. JAIMOUKHA AND E. M. KASENALLY, *Krylov subspace methods for solving large Lyapunov equations*, SIAM J. Numer. Anal., 31 (1994), pp. 227–251.
 - [18] I. M. JAIMOUKHA AND E. M. KASENALLY, *Oblique projection methods for large scale model reduction*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 602–627.
 - [19] E. M. KASENALLY, *Analysis of some Krylov subspace methods for large matrix equations*, IRC-PSE report, No C93-14, Imperial College, University of London, 1992.
 - [20] B. C. MOORE, *Principal component analysis in linear systems: controllability, observability and model reduction*, IEEE Trans. Automat. Control, AC-26 (1981), pp. 17–31.
 - [21] B. PARLETT, *Reduction to tridiagonal form and minimal realizations*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 567–593.
 - [22] Y. SAAD, *Analysis of some krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 29 (1991), pp. 209–228.
 - [23] Y. SAAD, *Numerical solution of large Lyapunov equations*, in Signal Processing, Scattering, Operator Theory and Numerical Methods, M. A. Kaashoek, J. H. Van Schuppen, and A. C. M. Ran, eds., Birkhäuser, Boston, Cambridge, MA, pp. 503–511, 1990.
 - [24] M. G. SAFONOV AND R. Y. CHIANG, *A Schur method for balanced-truncation model reduction*, IEEE Trans. Automat. Control, AC-34 (1989), pp. 729–733.
 - [25] D.C. SORENSEN, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.
 - [26] P. VAN DOOREN, *Numerical linear algebra techniques for large scale matrix problems in systems and control*, 31st Conference on Decision and Control, Tucson, AZ, 1992.

Copyright of SIAM Journal on Matrix Analysis & Applications is the property of Society for Industrial and Applied Mathematics and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.