OBLIQUE PROJECTION METHODS FOR LARGE SCALE MODEL REDUCTION*

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

Abstract. The aim of this paper is to consider approximating a linear transfer function $F(s)$ of McMillan degree N , by one of McMillan degree m in which $N \gg m$ and where N is large. Krylov subspace methods are employed to construct bases to parts of the controllability and observability subspaces associated with the state space realisation of $F(s)$. Low rank approximate grammians are computed via the solutions to low dimensional Lyapunov equations and computable expressions for the approximation errors incurred are derived. We show that the low rank approximate grammians are the exact grammians to a perturbed linear system in which the perturbation is restricted to the transition matrix, and furthermore, this perturbation has at most rank = 2. This paper demonstrates that this perturbed linear system is equivalent to a low dimensional linear system with state dimension no greater than m . Finally, exact low dimensional expressions for the $\mathcal{L}^\infty$ norm of the errors are derived. The model reduction of discrete time linear systems is considered via the use of the same Krylov schemes. Finally, the behaviour of these algorithms is illustrated on two large scale examples.

Key words. Lanczos, Arnoldi, iterative methods, model reduction, Krylov subspace methods, Lyapunov matrix equation, large scale systems

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

1. Introduction. The need for model reduction arises in many areas of engineering, where high order mathematical models are used to describe complex dynamical behaviour. These occur whenever models are described by partial differential equations that culminate in large linear finite element or finite difference models. For practical reasons, it is desirable to replace these high order models by low order approximations. For example, in control system applications, high order models may result in high order controllers and the subsequent implementation of these controllers is cumbersome and expensive. Consider a stable linear state-space model of the form

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

$$(2) \quad y(t) = cx(t) + du(t),$$

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 vectors b , c , and d are real with their dimensions fixed by those of $x(t)$, $u(t)$, and $y(t)$. The associated transfer function is given by $F(s) = c(sI - A)^{-1}b + d$. The task of any model reduction algorithm is to find an approximate stable model

$$(3) \quad \dot{x}_m(t) = A_m x_m(t) + b_m u(t),$$

$$(4) \quad y_m(t) = c_m x_m(t) + d_m u(t),$$

in which $x_m(t) \in \mathbb{R}^m$, with $m \ll N$ and the low order transfer function is given by $F_m(s) = c_m(sI - A_m)^{-1}b_m + d_m$. Well-established model reduction methods such as optimal Hankel norm [6] and balanced truncation [13] begin by solving the linear matrix equations

$$(5) \quad AP + PA' + bb' = 0,$$

$$(6) \quad A'Q + QA + c'c = 0,$$

* Received by the editors June 23, 1993; accepted for publication (in revised form) by P. Van Dooren, March 25, 1994.

[†] Interdisciplinary Research Centre for Process Systems Engineering, Imperial College, Exhibition Road, London SW7-2BY, England (jaimouka@ps.ic.ac.uk and kasenall@ps.ic.ac.uk).

which admit unique symmetric solutions if and only if $\lambda_i(A) + \bar{\lambda}_j(A) \neq 0$ for all i, j , and where λ_i denotes the i th eigenvalue and the overbar represents the complex conjugate. The requisite for P and Q stems from the easily computable $\mathcal{L}^\infty$ error bound [6]

$$(7) \quad \|F(s) - F_m(s)\|_\infty \leq 2 \sum_{m+1}^N \sigma_i(F(s)),$$

where the σ_i 's are the Hankel singular values of $F(s)$ defined as $\sigma_i = \lambda_i^{1/2}(PQ)$ and arranged in decreasing order of magnitude and where the norm $\|F(s)\|_\infty$ is defined as $\|F(s)\|_\infty = \sup_{\omega \in \mathbb{R}} \{\sigma_{\max}\{F(j\omega)\}\}$ in which $\sigma_{\max}(\cdot)$ denotes the largest singular value. Furthermore, P and Q are used by both methods to form the balancing transformation the effect of which is to yield $P = Q = \Sigma$ in the new coordinate system where Σ is a diagonal matrix of the Hankel singular values. Safonov and Chiang [18] developed a numerically robust variant of Moore's balanced truncation algorithm that does not require the formation of the balancing transformations; however, this variant does not obviate the need to compute P and Q . The motivation for using Krylov subspace methods in this paper is to enable low order approximate models to be computed while effecting all the computations in the low dimension. This technique was used successfully in [1] to perform model reduction in the light of a control system design for a fusion reactor. One of the aims of this paper is to justify this approach and give computable error expressions.

Related to this work are [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 realisation 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 [14] to obtain minimal realisations. In that paper, the rank of the Hankel matrix was used to determine the order of the minimal realisation. Furthermore, it was demonstrated that a minimal realisation could be constructed from the data generated by the Lanczos process. A similar approach was adopted in [7] in which the minimal realisation and its order were found to be related to the different types of breakdowns encountered in the Lanczos process. Here, too, the onset of breakdown was also given in terms of properties of the Hankel matrix. Recently, the presentations in [3], [19] reviewed the use of projection methods for large scale control problems. Both papers suggest 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.

A drawback associated with the methods above is that the computed minimal realisation may still have a high dimension. This deficiency is remedied by obtaining approximate reduced order models with low state dimension. A key issue in the development of our model reduction schemes is the efficient computation of low rank approximate solutions to the controllability and observability Lyapunov equations. This paper exploits the approach developed in our previous work [10]–[12] and that of Saad [17], both of which employ classical Krylov subspace techniques. In [17], Saad considers the low rank approximate solutions to (5) by imposing a Galerkin condition on the residual error, and [11] extends his work to the general case via the use of block schemes and gives a computable expression for the associated residual error norm. Furthermore, we addressed the problem of computing a low rank approximate solution to (5) which meets an optimality condition. The generalized minimal residual

(GMRES) method presented in [11] minimises the Frobenius norm of the residual error for which an exact computable expression is also derived. The focus of this paper is to consider the solution to the coupled Lyapunov equations (5) and (6). Two Lyapunov equation solvers are presented for which we derive computable residual error and a priori and a posteriori backward error expressions. Two model reduction algorithms are introduced for which computable $\mathcal{L}^\infty$ error expressions are provided. For ease of presentation, the developments are carried out for single-input, single-output systems, and the findings are then extended to multivariable problems and discrete-time dynamical systems.

The following summarises the contributions in this paper. Section 2 describes the type of approximations employed and justifies the use of Krylov subspace methods to solve large Lyapunov equations. Approximate solutions to (5) and (6) are obtained together with exact expressions for the errors incurred. Furthermore, it is demonstrated that the two approximate solutions are the exact solutions to a pair of perturbed Lyapunov equations. In §3, we show that the latter are the controllability and observability Lyapunov equations of a perturbed linear system; furthermore, this perturbed linear system is equivalent to a linear system of state dimension no greater than m . Computable $\mathcal{L}^\infty$ error expressions are provided that enable one to gauge the progress of the iterative method for increasing m . Section 4 employs the solution techniques of §§2 and 3 to obtain model reduction schemes for discrete time systems. Two illustrative examples in §5 show how the Lyapunov equation solvers and model reduction algorithms behave in practice, and, finally, the conclusions are found in §6.

2. Krylov subspace techniques. In practice, solutions to large Lyapunov equations (5) and (6) frequently admit good low rank approximations. In addition, one is generally interested in computing only the dominant eigenspace of the exact solution P^* , rather than P^* itself, since the dominant eigenspace of P^* is known to be associated with the dominant modes of the system described by (1) and (2) [1]. In what follows, what can be said of (5) can also be said of (6). Thus we will limit our discussion to (5) and invoke (6) only when necessary.

Ideally, we want to compute a rank m approximation P_m where $m \ll N$ such that $\|P^* - P_m\|_F$ is minimised. Throughout this paper we make use of the Frobenius norm defined as $\|Z\|_F = \sqrt{\text{trace}(ZZ')}$ in which Z' denotes the conjugate transpose of Z .

Consider the Schur decomposition of P^* given by $P^* = U\Sigma U'$ in which $U \in \mathbb{R}^{N \times N}$ is an orthogonal matrix and $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_N\}$ is a matrix of eigenvalues ordered such that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N \geq 0$. Then the optimal rank m Frobenius norm approximation of P^* is given by

$$(8) \quad P_m := U \begin{bmatrix} \Sigma_m & 0 \\ 0 & 0 \end{bmatrix} U' = U_m \Sigma_m U_m',$$

where $\Sigma_m = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_m\}$ (i.e., the first m diagonal elements of Σ) and $U_m \in \mathbb{R}^{N \times m}$ is a matrix of the first m columns of U . In [10], [11], [12], [17] a low rank approximation is computed by choosing an orthogonal matrix $V_m \in \mathbb{R}^{N \times m}$ and calculating the exact solution X_m to the reduced order Lyapunov equation

$$(9) \quad (V_m' A V_m) X_m + X_m (V_m' A' V_m) + V_m' b b' V_m = 0.$$

The estimate of P^* is then given by $P_m = V_m X_m V_m'$. Compared with the optimal approximation given in (8), it is apparent that to compute a good estimate of P^* , V_m

must be an accurate approximation of U_m or, in other words, the m most dominant eigenvectors of P^* . Unfortunately, minimising $\|P^* - P_m\|_F$ is intractable when P^* is unknown.

We therefore turn our attention to the problem of selecting V_m and X_m . Let $\mathcal{K}_m(A, b)$ be the m -dimensional Krylov space defined as

$$(10) \quad \mathcal{K}_m(A, b) = \text{span} \{ [b \quad Ab \quad A^2b \quad \dots \quad A^{m-1}b] \},$$

then, following [17], we select V_m to be an orthogonal basis of $\mathcal{K}_m(A, b)$. Throughout the remainder of this paper, we exploit approximations to the solution P^* which have the form

$$(11) \quad P_m = V_m X_m V_m',$$

where $X_m \in \mathbb{R}^{m \times m}$ is an arbitrary symmetric matrix. The key point here is that even though $P_m \in \mathbb{R}^{N \times N}$, it may be efficiently stored as the product of smaller matrices $V_m \in \mathbb{R}^{N \times m}$ and $X_m \in \mathbb{R}^{m \times m}$. We observe that P_m is symmetric for symmetric X_m and $\text{rank}(P_m) = \text{rank}(X_m) \leq m$. Similarly, associated with (6), one seeks an approximate solution of the form $Q_m := W_m Y_m W_m'$ where W_m is selected to be an orthogonal basis of m -dimensional Krylov space defined as

$$(12) \quad \mathcal{L}_m(A', c') = \text{span} \{ [c' \quad (A')c' \quad (A')^2c' \quad \dots \quad (A')^{m-1}c'] \}.$$

The remainder of this section is devoted to the appropriate selection of symmetric matrices X_m and Y_m . We begin by defining the residual error functions associated with a particular choice of X_m and Y_m as

$$(13) \quad R_m(X_m) := A(V_m X_m V_m') + (V_m X_m V_m')A' + bb',$$

$$(14) \quad S_m(Y_m) := A'(W_m Y_m W_m') + (W_m Y_m W_m')A + c'c.$$

The solution techniques presented in this paper are based on seeking a symmetric X_m and Y_m so as to give $R_m(X_m)$ and $S_m(Y_m)$ desirable properties. Sections 2.1 and 2.2 address the problem of constructing different X_m and Y_m such that $R_m(X_m)$ and $S_m(Y_m)$ have an orthogonality property with respect to the Krylov spaces defined in (10) and (12). Computable expressions for $\|R_m(X_m)\|_F$ and $\|S_m(Y_m)\|_F$ are provided. We demonstrate that the two approximate grammians are the exact solutions to a set of perturbed Lyapunov equations, thus providing an alternative means of gauging the progress of the iterative process for increasing m .

2.1. The Arnoldi process and Lyapunov equations. Next we use the well-established Arnoldi algorithm [20] to calculate the orthonormal bases V_m and W_m for the Krylov subspaces $\mathcal{K}_m$ and $\mathcal{L}_m$, respectively. The basic outline of the Arnoldi process is given next in terms of (A, b) .

ARNOLDI PROCESS

- Initialise: Compute $\beta = \|b\|_2$ and set $v_1 := b/\beta$.
- Iterate: Do $j = 1, \dots, m$
 - Compute j coefficients h_{ij} so that $\hat{v} := Av_j - \sum_{i=1}^j v_i h_{ij}$ is orthogonal to all the previous v_i 's.
 - Compute $h_{j+1,j} := \|\hat{v}\|_2$. If $h_{j+1,j} = 0$ stop, else $v_{j+1} := \hat{v}/h_{j+1,j}$.
 - End Do.

By construction, the Arnoldi process produces the matrix $V_m = [v_1, v_2, \dots, v_m]$, which forms an orthogonal basis for the Krylov subspace $\mathcal{K}_m(A, b)$. The process also yields an $m \times m$ upper Hessenberg matrix H_m , which satisfies the relation

$$(15) \quad AV_m = V_m H_m + v_{m+1} h_{m+1,m} e'_m,$$

in which e_m denotes the last column of the m -dimensional identity matrix. We observe that (15) is a combination of the first two steps of the iterative loop. From (15) it is easy to verify that $H_m = V'_m A V_m$ since $[V_m \ v_{m+1}]$ is part of an orthogonal matrix. Applying the Arnoldi process to (A', c') yields an orthogonal basis $W_m := [w_1, w_2, \dots, w_m]$ for the Krylov space $\mathcal{L}_m(A', c')$ in which $w_1 := c/\delta$ and $\delta := \|c'\|_2$. Furthermore, G_m is a lower Hessenberg matrix which satisfies

$$(16) \quad W'_m A = G_m W'_m + e_m g_{m,m+1} w'_{m+1}.$$

For convenience, we define

$$(17) \quad \hat{H}_m := (W'_m V_m)^{-1} W'_m A V_m = H_m + (W'_m V_m)^{-1} W'_m v_{m+1} h_{m+1,m} e'_m,$$

$$(18) \quad \hat{G}_m := W'_m A V_m (W'_m V_m)^{-1} = G_m + e_m g_{m,m+1} w'_{m+1} V_m (W'_m V_m)^{-1}$$

for nonsingular $(W'_m V_m)$. Observe that $\hat{H}_m$ and $\hat{G}_m$ are upper Hessenberg and lower Hessenberg matrices, respectively. The Arnoldi process generates an upper Hessenberg H_m and $(W'_m V_m)^{-1} W'_m v_{m+1} h_{m+1,m} e'_m$ has nonzero elements only in its last column; thus $\hat{H}_m$ has the same structure as H_m . Similarly, $\hat{G}_m$ is lower Hessenberg and $e_m g_{m,m+1} w'_{m+1} V_m (W'_m V_m)^{-1}$ has nonzero elements only in its last row. For implementation details of the Arnoldi process and its breakdown-free variants, we refer the reader to [4], [11].

Assuming that the approximate solution to (5) has the form $P_m := V_m X_m V'_m$ for some arbitrary symmetric $X_m \in \mathbb{R}^{m \times m}$, the residual error function associated with any solution is given by (13). Substituting (15) and (17) into (13) and assuming that $(W'_m V_m)$ is nonsingular allows the error function in (13) to be written as

$$(19) \quad R_m(X_m) = [V_m \ (I - V_m (W'_m V_m)^{-1} W'_m) v_{m+1}] \\ \times \begin{bmatrix} \hat{H}_m X_m + X_m \hat{H}'_m + e_1 \beta^2 e'_1 & X_m e_m h_{m+1,m} \\ h_{m+1,m} e'_m X_m & 0 \end{bmatrix} \\ \times \begin{bmatrix} V'_m \\ v'_{m+1} (I - W_m (V'_m W_m)^{-1} V'_m) \end{bmatrix}.$$

Similarly, associated with (6), the residual error function for any given approximate solution of the form $Q_m = W_m Y_m W'_m$ is defined in (14). Substituting (16) and (18) into (14) allows the error function to be written as

$$(20) \quad S_m(Y_m) = [W_m \ (I - W_m (V'_m W_m)^{-1} V'_m) w_{m+1}] \\ \times \begin{bmatrix} \hat{G}'_m Y_m + Y_m \hat{G}_m + e_1 \delta^2 e'_1 & Y_m e_m g_{m,m+1} \\ g_{m,m+1} e'_m Y_m & 0 \end{bmatrix} \\ \times \begin{bmatrix} W'_m \\ w'_{m+1} (I - V_m (W'_m V_m)^{-1} W'_m) \end{bmatrix}$$

for nonsingular $(W'_m V_m)$. The orthogonality conditions imposed in [10]–[12] sought to determine low rank approximate solutions $P_m := V_m X_m V'_m$ such that

$V'_m R_m(X_m) V_m = 0$. In contrast, the Arnoldi–Lyapunov solver considered in this section seeks symmetric matrices X_m and Y_m such that the residual errors $R_m(X_m)$ and $S_m(Y_m)$ satisfy orthogonality properties with respect to the Krylov subspaces $\mathcal{L}_m(A', c')$ and $\mathcal{K}_m(A, b)$, respectively. We now state the problem we wish to address.

PROBLEM 2.1. Find the approximate solutions $P_m := V_m X_m V'_m$ and $Q_m := W'_m Y_m W_m$ to (5) and (6), respectively, that satisfy the Galerkin type conditions $W'_m R_m(X_m) W_m = 0$ and $V'_m S_m(Y_m) V_m = 0$.

The following theorem gives the solution to Problem 2.1.

THEOREM 2.1. *Suppose that m steps of the Arnoldi process have been taken, that $(W'_m V_m)$ is nonsingular, and that the residual errors associated with (5) and (6) are defined by (19) and (20). Furthermore, suppose that $\lambda_i(\hat{H}_m) + \bar{\lambda}_j(\hat{H}_m) \neq 0$ for all i, j and $\lambda_i(\hat{G}_m) + \bar{\lambda}_j(\hat{G}_m) \neq 0$ for all i, j then,*

(a) $W'_m R_m(X_m) W_m = 0$ if and only if $X_m = X_m^A$, where X_m^A satisfies

$$(21) \quad \hat{H}_m X_m^A + X_m^A \hat{H}'_m + e_1 \beta^2 e'_1 = 0,$$

where e_1 is the first column of the $m \times m$ identity matrix. Under these conditions,

$$(22) \quad \|R_m^A\|_F := \|R_m(X_m^A)\|_F$$

$$(23) \quad = \sqrt{2h_{m+1,m} v'_{m+1} (I - W_m (V'_m W_m)^{-1} V'_m) R_m(X_m^A) V_m X_m^A e_m}$$

$$= \left\| \begin{bmatrix} H_m X_m^A + X_m^A H'_m + e_1 \beta^2 e'_1 & X_m^A e_m h_{m+1,m} \\ h_{m+1,m} e'_m X_m^A & 0 \end{bmatrix} \right\|_F.$$

(b) $V'_m S_m(Y_m) V_m = 0$ if and only if $Y_m = Y_m^A$, where Y_m^A satisfies

$$(24) \quad \hat{G}'_m Y_m^A + Y_m^A \hat{G}_m + e_1 \delta^2 e'_1 = 0.$$

Under these conditions,

$$(25) \quad \|S_m^A\|_F := \|S_m(Y_m^A)\|_F$$

$$(26) \quad = \sqrt{2g_{m,m+1} w'_{m+1} (I - V_m (W'_m V_m)^{-1} W'_m) S_m(Y_m^A) W_m Y_m^A e_m}$$

$$= \left\| \begin{bmatrix} G'_m Y_m^A + Y_m^A G_m + e_1 \delta^2 e'_1 & Y_m^A e_m g_{m,m+1} \\ g_{m,m+1} e'_m Y_m^A & 0 \end{bmatrix} \right\|_F.$$

Proof. Pre and postmultiplying (19) by W'_m and W_m , respectively, gives

$$W'_m R_m(X_m) W_m$$

$$= \begin{bmatrix} W'_m V_m & 0 \end{bmatrix} \begin{bmatrix} \hat{H}_m X_m + X_m \hat{H}'_m + e_1 \beta^2 e'_1 & X_m e_m h_{m+1,m} \\ h_{m+1,m} e'_m X_m & 0 \end{bmatrix} \begin{bmatrix} V'_m W_m \\ 0 \end{bmatrix}$$

$$= (W'_m V_m) (\hat{H}_m X_m + X_m \hat{H}'_m + e_1 \beta^2 e'_1) (V'_m W_m).$$

The result follows immediately since $(W'_m V_m)$ is assumed to be nonsingular. Substituting (21) into (19) gives, following some calculation,

$$(27) \quad \|R_m^A\|_F^2 := \|R_m(X_m^A)\|_F^2 = \text{trace}\{R_m^A R_m^{A'}\}$$

$$(28) \quad = 2h_{m+1,m} v'_{m+1} (I - W_m (V'_m W_m)^{-1} V'_m) R_m(X_m^A) V_m X_m^A e_m,$$

which establishes (22). Substituting $X_m := X_m^A$ into (13) allows us to factorise $R_m(X_m^A)$ as

$$R_m(X_m^A) = [V_m \ v_{m+1}] \begin{bmatrix} H_m X_m^A + X_m^A H'_m + e_1 \beta^2 e'_1 & X_m^A e_m h_{m+1,m} \\ h_{m+1,m} e'_m X_m^A & 0 \end{bmatrix} \begin{bmatrix} V'_m \\ v'_{m+1} \end{bmatrix},$$

from which (23) follows immediately since $[V_m \ v_{m+1}]$ is part of an orthogonal matrix and completes the proof of part (a).

The proof to part (b) is identical to that of part (a) except that it uses (14), (20), and (24). $\square$

An implication of the above result is that as m is increased, the residuals are confined to progressively smaller and smaller subspaces of $\mathbb{R}^{N \times N}$. This however, does not imply that the Arnoldi process will produce a sequence of nonincreasing residual error norms. The residual error norms in (22) and (25) provide a useful stopping criterion in a practical implementation of the algorithm as they allow one to economically evaluate the error norms and gauge the quality of the low rank approximations. The key points here are, first, that (21) and (24) are Lyapunov equations of dimension m that can be solved accurately using the Bartels–Stewart algorithm [2]; second, P_m and Q_m may be efficiently stored as the product of low order matrices; and, third, the residual error norm does not require the formation of the approximate solutions at each step. Instead (22), (23), (25), and (26) may be computed via low dimensional matrix products.

Remark 2.1. Computing the residual error norms $\|R_m^A\|_F$ and $\|S_m^A\|_F$ at the end of each Arnoldi process iteration using (22) and (25) may be limited to approximately $6Nm$ floating point operations if the computations are restricted to matrix vector products. Alternatively, approximately $O(m^3)$ floating point operations are needed to evaluate either (23) or (26). Thus as m increases, one would switch from evaluating (23) and (26) to computing (22) and (25) as it became more economical to do so.

The following result gives one perturbation, in exact arithmetic, of the data in (5) (and similarly for (6)) for which the low rank solutions given in Theorem 2.1 are the exact solutions. This result is reminiscent of the backward error analysis of [9].

COROLLARY 2.2. *Suppose that m steps of the Arnoldi process have been taken and that $P_m := V_m X_m^A V'_m$ and $Q_m := W_m Y_m^A W'_m$ are the low rank approximate solutions to (5) and (6), respectively, and furthermore that X_m^A and Y_m^A satisfy (21) and (24), respectively. Then*

$$(29) \quad (A - \Delta_1)P_m + P_m(A - \Delta_1)' + bb' = 0,$$

$$(30) \quad (A - \Delta_2)'Q_m + Q_m(A - \Delta_2) + c'c = 0,$$

where

$$(31) \quad \Delta_1 = (I - V_m(W'_m V_m)^{-1}W'_m)v_{m+1}h_{m+1,m}v'_m,$$

$$(32) \quad \Delta_2 = w_m g_{m,m+1} w'_{m+1} (I - V_m(W'_m V_m)^{-1}W'_m),$$

and

$$\|\Delta_1\|_F^2 = h_{m+1,m}^2 \{1 + \|(W'_m V_m)^{-1}W'_m v_{m+1}\|_2^2\}$$

and

$$\|\Delta_2\|_F^2 = g_{m,m+1}^2 \{1 + \|(V'_m W_m)^{-1}V'_m w_{m+1}\|_2^2\}.$$

Proof. Substituting $X_m := X_m^A$ into (19) gives

$$(33) \quad \begin{aligned} & A(V_m X_m^A V_m') + (V_m X_m^A V_m') A' + b b' \\ & = (I - V_m (W_m' V_m)^{-1} W_m') v_{m+1} h_{m+1,m} e_m' X_m^A V_m' \\ & \quad + V_m X_m^A e_m h_{m+1,m} v_{m+1}' (I - W_m (V_m' W_m)^{-1} V_m'). \end{aligned}$$

Equation (29) follows by rearranging (33) and noting that $e_m' = v_m' V_m$. The expression for $\|\Delta_1\|_F$ follows from the fact that v_m and V_m are parts of an orthogonal matrix.

Similarly, (30) and the expression for $\|\Delta_2\|_F$ follow by substituting $Y_m := Y_m^A$ into (20) and using the facts that w_m and W_m are parts of an orthogonal matrix. Finally, observe that Δ_1 and Δ_2 are at most rank-1 perturbations and that $\|\Delta_1\|_F$ and $\|\Delta_2\|_F$ may be evaluated without the need to form X_m^A or Y_m^A . $\square$

Remark 2.2. Observe that $\Delta_3 := \Delta_1 + \Delta_2$ is also a perturbation on the data in A such that

$$(34) \quad (A - \Delta_3) P_m + P_m (A - \Delta_3)' + b b' = 0,$$

$$(35) \quad (A - \Delta_3)' Q_m + Q_m (A - \Delta_3) + c' c = 0.$$

Furthermore, Δ_3 is at most a rank-2 perturbation, which may be factorised as

$$(36) \quad \begin{aligned} \Delta_3 &= [w_m \quad (I - V_m (W_m' V_m)^{-1} W_m') v_{m+1}] \\ &\quad \times \begin{bmatrix} 0 & g_{m,m+1} \\ h_{m+1,m} & 0 \end{bmatrix} \begin{bmatrix} v_m' \\ w_{m+1}' (I - V_m (W_m' V_m)^{-1} W_m') \end{bmatrix}. \end{aligned}$$

Finally, a direct calculation will verify that $W_m' \Delta_1 V_m = W_m' \Delta_2 V_m = W_m' \Delta_3 V_m = 0$.

From (31) and (32) one observes that Δ_1 and Δ_2 depend only on the data generated in the course of the Arnoldi process and, as such, one can envisage an iterative scheme in which the evolutions of $\|\Delta_1\|_F$ and $\|\Delta_2\|_F$ were monitored for increasing m . This scheme would require $4Nm$ operations above each Arnoldi step and $2N(m+1)$ storage locations for V_{m+1} and W_{m+1} and a further $3m^2$ for H_m , G_m and $W_m' V_m$, respectively, the latter being updatable through matrix vector products. In contrast, an implementation that monitors the evolution of the residual error requires $25m^3 + 2\min\{Nm, m^3\}$ floating point operations and a storage of approximately $2N(m+1) + 7m^2$ locations making it computationally cheaper but with more storage needs than a backward error checking scheme.

The following procedures summarise the Arnoldi-Lyapunov solvers proposed in this section; the first monitors the residual error while the second checks the backward error evolution.

ARNOLDI-LYAPUNOV SOLVER (RESIDUAL ERROR)

- Start: Specify a tolerance $\epsilon > 0$, set an integer parameter m .
- Perform m steps of the Arnoldi process to compute $\hat{H}_m$, $h_{m+1,m}$, V_m , v_{m+1} , and β .
- Perform m steps of the Arnoldi process to compute $\hat{G}_m$, $g_{m,m+1}$, W_m , w_{m+1} , and δ .
- Compute the symmetric matrices X_m^A and Y_m^A , which uniquely satisfy the low dimensional Lyapunov equations $\hat{H}_m X_m + X_m \hat{H}_m' + e_1 \beta^2 e_1' = 0$ and $\hat{G}_m Y_m + Y_m \hat{G}_m + e_1 \delta^2 e_1' = 0$, respectively.

- Evaluate $\|R_m^A\|_F$ and $\|S_m^A\|_F$ using (22) and (25) or (23) and (26), respectively. If either $\|R_m^A\|_F > \epsilon$ or $\|S_m^A\|_F > \epsilon$ increase m and continue the Arnoldi process, otherwise, form the approximate solutions: $P_m := V_m X_m^A V_m'$ and $Q_m := W_m Y_m^A W_m'$.

ARNOLDI-LYAPUNOV SOLVER (BACKWARD ERROR)

- Start: Specify a tolerance $\epsilon > 0$, set an integer parameter m .
- Perform m steps of the Arnoldi process to compute H_m , $h_{m+1,m}$, V_m , v_{m+1} , and β .
- Perform m steps of the Arnoldi process to compute G_m , $g_{m,m+1}$, W_m , w_{m+1} , and δ .
- Evaluate $\|\Delta_1\|_F$ and $\|\Delta_2\|_F$ using Corollary 2.2, if either $\|\Delta_1\|_F > \epsilon$ or $\|\Delta_2\|_F > \epsilon$ increase m and continue the Arnoldi process, otherwise, compute the symmetric matrices X_m^A and Y_m^A which uniquely satisfy the low dimensional Lyapunov equations $\hat{H}_m X_m + X_m \hat{H}_m' + e_1 \beta^2 e_1' = 0$ and $\hat{G}_m' Y_m + Y_m \hat{G}_m + e_1 \delta^2 e_1' = 0$, respectively.
- Form the approximate solutions: $P_m := V_m X_m^A V_m'$ and $Q_m := W_m Y_m^A W_m'$.

As pointed out earlier, $\|\Delta_1\|_F$ and $\|\Delta_2\|_F$ depend entirely on the data generated in the course of the Arnoldi process, thus they are a priori perturbation bounds that may be conservative. A posteriori backward error bounds are discussed in §2.3.

2.2. The Lanczos process and Lyapunov equations. The nonsymmetric Lanczos process is an alternative algorithm used to simultaneously construct bases for $\mathcal{K}_m(A, b)$ and $\mathcal{L}_m(A', c')$. The process requires A , its transpose, and two starting vectors b and c' to construct bases for parts of the controllability and observability subspaces that meet a biorthogonality condition (i.e., $W_m' V_m = I$). The aim of this section is not to consider the different aspects of the Lanczos process, but rather to show how this algorithm may be used to efficiently solve large Lyapunov equations. For an analysis of the Lanczos process and some of its breakdown-free variants, we refer the reader to [5], [14]–[16] and the references therein. A simple version of the nonsymmetric Lanczos process is given here.

LANCZOS PROCESS

1. Start: Set $\beta_1 := \sqrt{|cb|}$ and $\delta_1 := \beta_1 \cdot \text{sign}[cb]$ and define $v_1 := b/\delta_1$ $w_1 := c'/\beta_1$.
2. Iterate: For $j = 1, 2, \dots$, do:
 - $\alpha_j := (Av_j, w_j)$,
 - $\hat{v}_{j+1} := Av_j - \alpha_j v_j - \beta_j v_{j-1}$ (when $j = 1$, take $\beta_1 v_0 = 0$),
 - $\hat{w}_{j+1} := A^T w_j - \alpha_j w_j - \delta_j w_{j-1}$ (when $j = 1$, take $\delta_1 w_0 = 0$),
 - $\beta_{j+1} := \sqrt{|\langle \hat{v}_{j+1}, \hat{w}_{j+1} \rangle|}$, $\delta_{j+1} := \beta_{j+1} \cdot \text{sign}[\langle \hat{v}_{j+1}, \hat{w}_{j+1} \rangle]$,
 - $v_{j+1} := \hat{v}_{j+1}/\beta_{j+1}$, $w_{j+1} := \hat{w}_{j+1}/\delta_{j+1}$.

If we denote $V_m = [v_1, v_2, \dots, v_m]$ and similarly $W_m = [w_1, w_2, \dots, w_m]$, we then have $W_m' V_m = I$ where $I \in \mathbb{R}^{m \times m}$ is the identity matrix. Let us denote by T_m the tridiagonal matrix $T_m \equiv \text{Tridiag}[\delta_{i+1}, \alpha_i, \beta_{i+1}]$. Then it is easy to verify that

$$(37) \quad AV_m = V_m T_m + \delta_{m+1} v_{m+1} e_m',$$

$$(38) \quad A' W_m = W_m T_m' + \beta_{m+1} w_{m+1} e_m',$$

and $W_m' AV_m = T_m$, where the vector e_m is the m th column of the identity matrix in $\mathbb{R}^{m \times m}$. Observe that there are infinitely different ways of selecting the scalars δ_{j+1} and β_{j+1} as long as they satisfy $(\hat{v}_{j+1}, \hat{w}_{j+1}) = \delta_{j+1} \beta_{j+1}$. In our application, we select $v_1 := b/\sqrt{|cb|}$ and $w_1 := c'/\sqrt{|cb|} \cdot \text{sign}[cb]$. Then setting $\delta_1 = \sqrt{|cb|}$, we have $b = V_m e_1 \delta_1$, where e_1 is the first column of the identity matrix in $\mathbb{R}^{m \times m}$. Similarly,

setting $\beta_1 = \sqrt{|cb|} \cdot \text{sign}[cb]$ yields $c' = W_m e_1 \beta_1$. Assuming that the approximate solution to (5) has the form $P_m := V_m X_m V_m'$, substituting (37) into (13) gives

$$(39) \quad R_m(X_m) = [V_m \ v_{m+1}] \begin{bmatrix} T_m X_m + X_m T_m' + e_1 \delta_1^2 e_1' & X_m e_m \delta_{m+1} \\ \delta_{m+1} e_m' X_m & 0 \end{bmatrix} \begin{bmatrix} V_m' \\ v_{m+1}' \end{bmatrix}.$$

Similarly, suppose that the approximate solution to (6) has the form $Q_m := W_m Y_m W_m'$. Then on substituting into (14), we get

$$(40) \quad S_m(Y_m) = [W_m \ w_{m+1}] \begin{bmatrix} T_m' Y_m + Y_m T_m + e_1 \beta_1^2 e_1' & Y_m e_m \beta_{m+1} \\ \beta_{m+1} e_m' Y_m & 0 \end{bmatrix} \begin{bmatrix} W_m' \\ w_{m+1}' \end{bmatrix}.$$

The Lanczos–Lyapunov solver proposed here seeks symmetric X_m and Y_m that solve Problem 2.1. The following theorem gives the solution to Problem 2.1 in the context of the Lanczos process.

THEOREM 2.3. *Suppose that m steps of the Lanczos process have been taken and that the residual errors are defined by (39) and (40). Then if $\lambda_i(T_m) + \bar{\lambda}_j(T_m) \neq 0$ for all i, j*

(a) $W_m' R_m(X_m) W_m = 0$ if and only if $X_m = X_m^L$ where X_m^L satisfies

$$(41) \quad T_m X_m^L + X_m^L T_m' + e_1 \delta_1^2 e_1' = 0.$$

If these conditions are met then the residual error norm is given by

$$(42) \quad \|R_m^L\|_F := \|R_m(X_m^L)\|_F = \sqrt{2\delta_{m+1} v_{m+1}' R_m(X_m^L) V_m X_m^L e_m}.$$

(b) $V_m' S_m(Y_m) V_m = 0$ if and only if $Y_m = Y_m^L$ where Y_m^L satisfies

$$(43) \quad T_m' Y_m^L + Y_m^L T_m + e_1 \beta_1^2 e_1' = 0.$$

If these conditions are met then the residual error norm is given by

$$(44) \quad \|S_m^L\|_F := \|S_m(Y_m^L)\|_F = \sqrt{2\beta_{m+1} w_{m+1}' S_m(Y_m^L) W_m Y_m^L e_m}.$$

Proof. The proof is essentially the same as that of Theorem 2.1 except that it uses (39) and (40). $\square$

Each low dimensional Lyapunov equation may be solved at a cost of $12.5m^3$ floating point operations and $2.5m^2$ storage locations. Observe that the fourth step of the Lanczos process reveals that δ_{j+1} and β_{j+1} differ at most by a sign. A consequence of this is that T_m is almost symmetric in the sense that its superdiagonal entries are equal to its subdiagonal elements up to an occasional sign. It is therefore natural to ask whether this structure may be exploited in the solution to the low dimensional Lyapunov equations (41) and (43). Thus, we seek a sign matrix J such that $T_m' = J T_m J$, pre and postmultiplying (41) by J shows that $Y_m^L := J X_m^L J$. Hence a saving of $12.5m^3$ operations may be incurred since one would only have to solve (41) and form the matrix J . A simple way to generate the sign matrix J is to employ the following scheme.

$$J(1, 1) := -1$$

For $i := 2$ to m

$$J(i, i) := 1 \cdot \text{sign}[J(i-1, i-1) T_m(i, i-1)]$$

End.

The following corollary gives perturbations on the data in (5) and (6) such that the low rank approximate solutions given in Theorem 2.3 are exact.

COROLLARY 2.4. *Suppose that m steps of the Lanczos process have been taken and that $P_m := V_m X_m^L V_m'$ and $Q_m := W_m Y_m^L W_m'$ are the low rank approximate solutions to (5) and (6), respectively, and furthermore that X_m^L and Y_m^L satisfy (41) and (43), respectively. Then*

$$(45) \quad (A - \Delta_1)P_m + P_m(A - \Delta_1)' + bb' = 0,$$

$$(46) \quad (A - \Delta_2)'Q_m + Q_m(A - \Delta_2) + c'c = 0,$$

where $\Delta_1 = v_{m+1}\delta_{m+1}w_m'$ and $\Delta_2 = v_m\beta_{m+1}w_{m+1}'$. Furthermore, it holds that $\|\Delta_1\|_F = \|\hat{v}_{m+1}\|_2\|w_m\|_2$ and $\|\Delta_2\|_F = \|v_m\|_2\|\hat{w}_{m+1}\|_2$.

Proof. The proof is essentially the same as that of Corollary 2.2. $\square$

It is interesting to note that $\|\Delta_1\|_F$ and $\|\Delta_2\|_F$ may be evaluated directly from the data generated in the course of the last two steps of the Lanczos process. Compared to a residual error checking scheme, an implementation based on monitoring the backward error evolution would require less computations and storage, i.e., $\|\Delta_1\|_F$ requires $2N$ operations while evaluating $\|R_m^L\|_F$ consumes in excess of $3Nm$ flops. Observe that $\Delta_3 := \Delta_1 + \Delta_2$ is a perturbation on the data in A such that $(A - \Delta_3)P_m + P_m(A - \Delta_3)' + bb' = 0$ and $(A - \Delta_3)'Q_m + Q_m(A - \Delta_3) + c'c = 0$. Furthermore, Δ_3 is at most a rank-2 perturbation that may be factorised as

$$(47) \quad \Delta_3 = \begin{bmatrix} v_m & v_{m+1} \end{bmatrix} \begin{bmatrix} 0 & \beta_{m+1} \\ \delta_{m+1} & 0 \end{bmatrix} \begin{bmatrix} w_m' \\ w_{m+1}' \end{bmatrix}.$$

Finally, a direct calculation will verify that $W_m'\Delta_1V_m = W_m'\Delta_2V_m = W_m'\Delta_3V_m = 0$.

In contrast to the Arnoldi method, the Lanczos process yields matrices V_m and W_m that are no longer orthogonal. Consequently, manipulations with either V_m or W_m may suffer from numerical difficulties because of possible poor conditioning. There is frequently a loss of biorthogonality that may be checked by rebi-orthogonalising the newly computed v_{j+1} and w_{j+1} against V_m and W_m [20]. Finally, we observe that the Lyapunov equation solvers stemming from the Lanczos process enjoy lower complexity and storage requirements than Arnoldi solvers presented in §2.1.

2.3. A posteriori backward error. Even though one might elect to monitor the evolutions of Δ_1 , Δ_2 or Δ_3 for increasing m , there is no guarantee that their norms are nonincreasing. In fact our computational experience has revealed that $\|\Delta_i\|_F$ for $i = 1, 2, 3$ may behave erratically for increasing m . The conservative nature of a priori expression for $\|\Delta_i\|_F$ for $i = 1, 2, 3$ leads us to seek alternatives that enable us to assess the quality of the results obtained when exploiting the techniques proposed in this paper. It is interesting to consider whether, for given P_m and Q_m , there exist other, possibly smaller Δ_1 and Δ_2 for which (29) and (30) hold. Thus the aim of this section is to derive a posteriori expressions Δ_4 , Δ_5 and Δ_6 for Δ_1 , Δ_2 and Δ_3 , respectively.

COROLLARY 2.5. *Suppose that m steps of the Arnoldi process have been taken and that X_m^A and Y_m^A are the solutions to the m -dimensional Lyapunov equations (21) and (24), respectively. Suppose that*

$$X_m^A = \begin{bmatrix} U_r & U_\perp \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r' \\ U_\perp' \end{bmatrix}$$

and

$$Y_m^A = [U_s \ \tilde{U}_\perp] \begin{bmatrix} \Sigma_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U'_s \\ \tilde{U}'_\perp \end{bmatrix}$$

are Schur decompositions in which $\Sigma_r \in \mathbb{R}^{r \times r}$ and $\Sigma_s \in \mathbb{R}^{s \times s}$ are nonsingular and $U_r \in \mathbb{R}^{m \times r}$ and $U_s \in \mathbb{R}^{m \times s}$. Then

$$(48) \quad (A - \Delta_4)P_m + P_m(A - \Delta_4)' + bb' = 0,$$

$$(49) \quad (A - \Delta_5)'Q_m + Q_m(A - \Delta_5) + c'c = 0,$$

where

$$(50) \quad \Delta_4 = (I - V_m(W'_m V_m)^{-1} W'_m) v_{m+1} h_{m+1, m} e'_m U_r U'_r V'_m,$$

$$(51) \quad \Delta_5 = W_m U_s U'_s e_m g_{m, m+1} w'_{m+1} (I - V_m(W'_m V_m)^{-1} W'_m).$$

Furthermore, the norms are given by $\|\Delta_4\|_F^2 = h_{m+1, m}^2 \{1 + \|(W'_m V_m)^{-1} W'_m v_{m+1}\|_2^2\} \cdot \|e'_m U_r\|_2^2$ and $\|\Delta_5\|_F^2 = g_{m, m+1}^2 \{1 + \|(V'_m W_m)^{-1} V'_m w_{m+1}\|_2^2\} \|e'_m U_s\|_2^2$.

Proof. This proof is essentially the same as Corollary 2.2 except that it uses $X_m := X_m^A := U_r \Sigma_r U'_r$ and $Y_m := Y_m^A := U_s \Sigma_s U'_s$ and the facts that U_r and U_s are parts of orthogonal matrices. A more detailed proof may be found in [12]. $\square$

Remark 2.3. It is clear from (31) and (50) that $\|\Delta_4\|_F \leq \|\Delta_1\|_F$ since $\|e'_m U_r\|_2 \leq 1$. Similarly, $\|\Delta_5\|_F \leq \|\Delta_2\|_F$ since $\|e'_m U_s\|_2 \leq 1$, in the event that X_m^A and Y_m^A have rank equal to m $\|\Delta_1\|_F = \|\Delta_4\|_F$ and $\|\Delta_2\|_F = \|\Delta_5\|_F$, respectively. Finally, observe that Δ_4 and Δ_5 are rank-1 perturbations and that $\Delta_4 + \Delta_5 =: \Delta_6$ is a perturbation that simultaneously satisfies (34) and (35).

A posteriori perturbation bounds may be derived for the nonsymmetric Lanczos process that we now state without proof.

COROLLARY 2.6. *Suppose that m steps of the Lanczos process have been successfully completed and that X_m^L and Y_m^L are the solutions to the m -dimensional Lyapunov equations (41) and (43), respectively. Suppose that*

$$X_m^A = [U_r \ U_\perp] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U'_r \\ U'_\perp \end{bmatrix}$$

and

$$Y_m^A = [U_s \ \tilde{U}_\perp] \begin{bmatrix} \Sigma_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U'_s \\ \tilde{U}'_\perp \end{bmatrix}$$

are Schur decompositions in which $\Sigma_r \in \mathbb{R}^{r \times r}$ and $\Sigma_s \in \mathbb{R}^{s \times s}$ are nonsingular and $U_r \in \mathbb{R}^{m \times r}$ and $U_s \in \mathbb{R}^{m \times s}$. Then

$$(52) \quad (A - \Delta_4)P_m + P_m(A - \Delta_4)' + bb' = 0,$$

$$(53) \quad (A - \Delta_5)'Q_m + Q_m(A - \Delta_5) + c'c = 0,$$

where

$$(54) \quad \Delta_4 = v_{m+1} \delta_{m+1} e'_m U_r U'_r W'_m,$$

$$(55) \quad \Delta_5 = V_m U_s U'_s e_m \beta_{m+1} w'_{m+1},$$

and the norms are given by $\|\Delta_4\|_F = \|\hat{v}_{m+1}\|_2 \cdot \|W_m U_r U_r' e_m\|_2$ and $\|\Delta_5\|_F^2 = \|\hat{w}_{m+1}\|_2 \cdot \|V_m U_s U_s' e_m\|_2$.

Observe that as with the Arnoldi process, if X_m^L and Y_m^L have rank equal to m , then Δ_4 and Δ_5 defined in (52) and (53) degenerate into $\Delta_1 = v_{m+1}\delta_{m+1}w_m'$ and $\Delta_2 = v_m\beta_{m+1}w_{m+1}'$, respectively. Furthermore, $\Delta_6 := \Delta_4 + \Delta_5$ is a rank-2 perturbation of the matrix A such that $(A - \Delta_6)P_m + P_m(A - \Delta_6)' + bb' = 0$ and $(A - \Delta_6)'Q_m + Q_m(A - \Delta_6) + c'c = 0$.

3. 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 systems. Denoting a transfer function by $F(s)$

$$(56) \quad F(s) = d + c(sI - A)^{-1}b \stackrel{s}{=} \left[\begin{array}{c|c} A & b \\ \hline c & d \end{array} \right]; \quad A \in \mathbb{R}^{N \times N}, \quad b, c' \in \mathbb{R}^N, \quad \text{and} \quad d \in \mathbb{R}.$$

The task is to determine a reduced order model $F_m(s)$, where

$$(57) \quad F_m(s) = d_m + c_m(sI - A_m)^{-1}b_m \stackrel{s}{=} \left[\begin{array}{c|c} A_m & b_m \\ \hline c_m & d_m \end{array} \right];$$

$$A_m \in \mathbb{R}^{m \times m}, \quad b_m, c_m' \in \mathbb{R}^m \quad \text{and} \quad d_m \in \mathbb{R},$$

that approximates the high dimensional model $F(s)$. Associated with the linear system in (56), we define the controllability and observability Lyapunov equations

$$(58) \quad AP + PA' + bb' = 0,$$

$$(59) \quad A'Q + QA + c'c = 0,$$

respectively. The low rank approximate solutions to (58) and (59) may be computed via low dimensional calculations as demonstrated in Theorem 2.1. It is natural to question whether the approximate grammians P_m and Q_m are the exact controllability and observability grammians of a perturbed linear system. From Remark 2.2, it is apparent that $P_m = V_m X_m^A V_m'$ and $Q_m = W_m Y_m^A W_m'$ are, respectively, the controllability and observability grammians of $F_\Delta(s)$, where

$$(60) \quad F_\Delta(s) = d + c(sI - A + \Delta_3)^{-1}b \stackrel{s}{=} \left[\begin{array}{c|c} A - \Delta_3 & b \\ \hline c & d \end{array} \right],$$

where X_m^A and Y_m^A satisfy (21) and (24), respectively, and Δ_3 is defined in (36). Alternatively, one may employ the relations derived for the Lanczos process in §2.2 to conclude that $P_m = V_m X_m^L V_m'$ and $Q_m = W_m Y_m^L W_m'$ are, respectively, the controllability and observability grammians of $F_\Delta(s)$ in which X_m^L and Y_m^L satisfy (41) and (43), respectively, and Δ_3 is defined in (47). The main results of this section show that (60) is equivalent to a low dimensional linear system and gives computable expression for an $\mathcal{L}^\infty$ error between the high dimensional and low order approximate model.

From (56), $F(s) = d + cf_b(s) = d + f_c(s)b$, where $f_b(s) = (sI - A)^{-1}b$ and $f_c(s) = c(sI - A)^{-1}$; for later reference, $f_b(s)$ and $f_c(s)$ may be considered as the solutions to the coupled linear systems

$$(61) \quad (sI - A)f_b(s) = b,$$

$$(62) \quad f_c(s)(sI - A) = c,$$

respectively. The focus of what follows is to approximate $F(s)$ by obtaining approximate solutions to the linear systems (61) and (62). The approximate solutions $f_{b,m}(s)$ and $f_{c,m}(s)$ to the linear systems (61) and (62) are constructed to satisfy the following two conditions: (i) $f_{b,m}(s) \in \mathcal{K}_m(A, b)$, i.e., $f_{b,m}(s) = V_m F_b(s)$, such that $\mathcal{L}_m(A', c') \perp \{(sI - A)f_{b,m}(s) - b\}$; (ii) $f'_{c,m}(s) \in \mathcal{L}_m(A', c')$, i.e., $f_{c,m}(s) = F_c(s)W'_m$, 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 (61) and (62), and $F(s) = d + cf_b(s) = d + f_c(s)b$, we consider $F_{m,1}(s) = d + cf_{b,m}(s)$ and $F_{m,2}(s) = d + f_{c,m}(s)b$ as low order approximations to $F(s)$. The problem we wish to solve may be stated as follows.

PROBLEM 3.1. Find approximate solutions $f_{b,m}(s) = V_m F_b(s)$ and $f_{c,m}(s) = F_c(s)W'_m$ to (61) and (62), respectively, which satisfy the Galerkin type conditions

$$(63) \quad W'_m \{(sI - A)V_m F_b(s) - b\} = 0 \quad \forall s,$$

$$(64) \quad \{F_c(s)W'_m(sI - A) - c\}V_m = 0 \quad \forall s.$$

The following is referred to as a basis change in the state space realisation of $F(s)$

$$(65) \quad F(s) \stackrel{s}{=} \left[\begin{array}{c|c} A & b \\ \hline c & d \end{array} \right] \xrightarrow{T} F(s) \stackrel{s}{=} \left[\begin{array}{c|c} TAT^{-1} & Tb \\ \hline cT^{-1} & d \end{array} \right],$$

where T is nonsingular. The next lemma is needed in the proof of the main result.

LEMMA 3.1. Suppose that m steps of the Arnoldi process have been completed and that Δ_3 is given by (36), then, $-W'_m(A - \Delta_3) + W'_mAV_m(W'_mV_m)^{-1}W'_m = 0$.

Proof. It holds that

$$\begin{aligned} & W'_m\Delta_3 - W'_mA(I - V_m(W'_mV_m)^{-1}W'_m) \\ &= W'_mw_m g_{m,m+1} w'_{m+1} (I - V_m(W'_mV_m)^{-1}W'_m) \\ & \quad - e_m g_{m,m+1} w'_{m+1} (I - V_m(W'_mV_m)^{-1}W'_m) \\ &= 0, \end{aligned}$$

since $W'_mw_m = e_m$. $\square$

The following theorem gives the solution to Problem 3.1.

THEOREM 3.2. Suppose that m steps of the Arnoldi process have been taken and that W'_mV_m is nonsingular, then the following are true.

(a) The Galerkin conditions in (63) and (64) are satisfied if and only if $F_b(s) = (sI - \hat{H}_m)^{-1}e_1\beta$ and $F_c(s) = \delta e'_1(sI - \hat{G}_m)^{-1}$. Under these conditions, the residual error norms are

$$(66) \quad \|b - (sI - A)V_m F_b(s)\|_\infty = h_{m+1,m} \left\| \left[\begin{array}{c} (W'_mV_m)^{-1}W'_m v_{m+1} \\ 1 \end{array} \right] \right\|_2 \|e'_m F_b(s)\|_\infty,$$

$$(67) \quad \|c - F_c(s)W'_m(sI - A)\|_\infty = g_{m,m+1} \left\| \left[\begin{array}{c} (V'_mW_m)^{-1}V'_m w_{m+1} \\ 1 \end{array} \right] \right\|_2 \|F_c(s)e_m\|_\infty.$$

(b) $F_\Delta(s)$, $F_{m,1}(s)$, and $F_{m,2}(s)$ are different realisations of the same transfer function, namely,

$$(68) \quad F_\Delta(s) \equiv F_{m,1}(s),$$

$$(69) \quad F_\Delta(s) \equiv F_{m,2}(s),$$

where

$$(70) \quad F_{\Delta}(s) \stackrel{s}{=} \left[\frac{A - \Delta_3}{c} \middle| \frac{b}{d} \right],$$

$$(71) \quad F_{m,1}(s) = d + cV_m F_b(s) \stackrel{s}{=} \left[\frac{(W'_m V_m)^{-1} W'_m A V_m}{cV_m} \middle| \frac{(W'_m V_m)^{-1} W'_m b}{d} \right] \\ = \left[\frac{\hat{H}_m}{c_m} \middle| \frac{e_1 \beta}{d} \right],$$

$$(72) \quad F_{m,2}(s) = d + F_c(s) W'_m b \stackrel{s}{=} \left[\frac{W'_m A V_m (W'_m V_m)^{-1}}{cV_m (W'_m V_m)^{-1}} \middle| \frac{W'_m b}{d} \right] = \left[\frac{\hat{G}_m}{\delta e'_1} \middle| \frac{b_m}{d} \right].$$

(c) X_m^A and $(V'_m W_m) Y_m^A (W'_m V_m)$ are the controllability and observability gram-mians of $F_{m,1}(s)$.

(d) $(W'_m V_m) X_m^A (V'_m W_m)$ and Y_m^A are the controllability and observability gram-mians of $F_{m,2}(s)$.

Proof. The residue associated with the approximate solution to (61) is $(sI - A)V_m F_b(s) - b$, premultiplying by W'_m and substituting (17) leads to

$$W'_m \{(sI - A)V_m F_b - b\} = W'_m V_m \{(sI - (W'_m V_m)^{-1} W'_m A V_m) F_b(s) \\ - (W'_m V_m)^{-1} W'_m b\} \\ = (W'_m V_m)^{-1} \{(sI - \hat{H}_m) F_b(s) - e_1 \beta\}.$$

The result follows immediately since $(W'_m V_m)$ is assumed to be nonsingular. Similarly, postmultiplying the residue associated with the approximate solution to (62) and substituting (18) leads to $F_c(s) = \delta e'_1 (sI - \hat{G}_m)^{-1}$. For the $\mathcal{L}^\infty$ error bounds, we have

$$\|b - (sI - A)V_m (sI - \hat{H}_m)^{-1} e_1 \beta\|_\infty \\ = \|V_m e_1 \beta - (V_m s - V_m H_m - v_{m+1} h_{m+1,m} e'_m)(sI - \hat{H}_m)^{-1} e_1 \beta\|_\infty \\ = h_{m+1,m} \|(I - V_m (W'_m V_m)^{-1} W'_m) v_{m+1}\|_2 \|e'_m F_b(s)\|_\infty,$$

from which follows the relation in (66); similarly, for (67), which completes the proof of part (a).

We establish that $F_{\Delta}(s) \equiv F_{m,1}(s)$ by taking the difference between the two transfer function models

$$(73) \quad F_{\Delta}(s) - F_{m,1}(s) \stackrel{s}{=} \left[\frac{A - \Delta_3}{c} \quad 0 \quad \middle| \quad \frac{b}{0} \right].$$

Consider the following basis transformation T where

$$T = \begin{bmatrix} I & 0 \\ -(W'_m V_m)^{-1} W'_m & I \end{bmatrix}; \quad \text{and} \quad T^{-1} = \begin{bmatrix} I & 0 \\ (W'_m V_m)^{-1} W'_m & I \end{bmatrix},$$

which yields

$$(74) \quad F_{\Delta}(s) - F_{m,1}(s) \\ \stackrel{s}{=} \left[\frac{A - \Delta_3}{0} \quad 0 \quad \middle| \quad \frac{b}{0} \right],$$

since $c' = W_m e_1 \delta$. By Lemma 3.1, the (2,1) block of (74) is zero, from which we conclude that $F_\Delta(s) - F_{m,1}(s) = 0$ since the realization in (74) has N unobservable and m uncontrollable modes thus establishing the first equivalence. $F_\Delta(s) \equiv F_{m,2}(s)$ follows by applying the basis transformation $T = (W'_m V_m)$ to $F_{m,2}(s)$ to give $F_{m,2}(s) \xrightarrow{T} F_{m,1}(s)$; the second equivalence is immediate, thus completing the proof of part (b).

Using $(W'_m V_m)^{-1} W'_m b = e_1 \beta$, it follows that X_m^A is the controllability grammian of (71). For the observability grammian, we observe that

$$(75) \quad (W'_m V_m)^{-1} W'_m A V_m = (W'_m V_m)^{-1} (G_m W'_m + g_{m,m+1} e_m w'_{m+1}) V_m,$$

$$(76) \quad c V_m = \delta e'_1 W'_m V_m.$$

Substituting (75) and (76) into (71) gives the observability Lyapunov equation

$$(77) \quad Z(W'_m V_m)^{-1} (G_m W'_m + g_{m,m+1} e_m w'_{m+1}) V_m + V'_m (W_m G'_m + w_{m+1} e'_m g_{m,m+1}) (V'_m W_m)^{-1} Z + V'_m W_m e_1 \delta^2 e'_1 W'_m V_m = 0,$$

for the realization in (71); the unknown grammian is Z . Pre and postmultiplying (77) by $(V'_m W_m)^{-1}$ and $(W'_m V_m)^{-1}$ yields

$$(78) \quad (V'_m W_m)^{-1} Z (W'_m V_m)^{-1} (G_m + g_{m,m+1} e_m w'_{m+1} (W'_m V_m)^{-1}) + (G'_m + (V'_m W_m)^{-1} w_{m+1} e'_m g_{m,m+1}) (V'_m W_m)^{-1} Z (W'_m V_m)^{-1} + e_1 \delta^2 e'_1 = 0,$$

from which we deduce that $Z = (V'_m W_m) Y_m^A (W'_m V_m)$ is the observability grammian of (71) and completes the proof of part (c).

The proof to part (d) is identical to part (c) except that it uses (72). $\square$

Remark 3.1. We note that the Galerkin type conditions of (63) and (64) are analogous to parts (a) and (b) of Theorem 2.1, respectively; similarly, the residual error norm expressions in (66) and (67) are analogous to (22) and (25), respectively. In a practical implementation, the equalities in (66) and (67) enable us to economically monitor the progress of the iterative process at each step. An open issue is that of obtaining computable expressions for the $\mathcal{L}^\infty$ error $\|F(s) - F_{m,i}(s)\|_\infty$.

Remark 3.2. An implication of the orthogonality property reported in Remark 2.2 is that the perturbation is confined to a progressively smaller subspace of $\mathbb{R}^{N \times N}$ for increasing m . Thus Δ_3 may be employed as a stopping condition for an iterative model reduction algorithm in which $\|\Delta_3\|_F$ reports on the size of the perturbation that is effected on A to obtain a reduced order model of state dimension no greater than m .

Remark 3.3. It is interesting to observe that the realisations of $F_{m,1}(s)$ and $F_{m,2}(s)$ given in Theorem 3.2 depend only on the data generated in the course of the two Arnoldi processes, and that despite the relations between the low dimensional realisations and the low order Lyapunov equations in (21) and (24), one can construct $F_{m,1}(s)$ and $F_{m,2}(s)$ without having to form X_m^A and Y_m^A .

It is clear from Corollary 2.2 that Δ_3 defined in (36) is not the only perturbation that leads to the reduced order models of Theorem 3.2. The effect of Δ_3 is to perturb A in such a way that the nonminimal modes in the perturbed system are simultaneously uncontrollable and unobservable, while, to obtain a reduced order model, it is sufficient to perturb A so as to obtain either $N - m$ uncontrollable or unobservable modes.

COROLLARY 3.3. *Suppose that m steps of the Arnoldi processes have been completed and that Δ_1 and Δ_2 are defined by (31) and (32) respectively, then,*

$$F_{m,3}(s) \stackrel{s}{=} \left[\begin{array}{c|c} A - \Delta_1 & b \\ \hline c & d \end{array} \right] \equiv \left[\begin{array}{c|c} (W'_m V_m)^{-1} W'_m A V_m & (W'_m V_m)^{-1} W'_m b \\ \hline c V_m & d \end{array} \right]$$

$$\begin{aligned}
&= \left[\frac{\hat{H}_m}{c_m} \middle| \frac{e_1 \beta}{d} \right] \stackrel{s}{=} F_{m,1}(s), \\
F_{m,4}(s) &\stackrel{s}{=} \left[\frac{A - \Delta_2}{c} \middle| \frac{b}{d} \right] \equiv \left[\frac{W'_m A V_m (W'_m V_m)^{-1}}{c V_m (W'_m V_m)^{-1}} \middle| \frac{W'_m b}{d} \right] \\
&= \left[\frac{\hat{G}_m}{\delta e'_1} \middle| \frac{b_m}{d} \right] \stackrel{s}{=} F_{m,2}(s).
\end{aligned}$$

Proof. We establish that $F_{m,3}(s) \equiv F_{m,1}(s)$ by taking the difference between the two transfer function models

$$(79) \quad F_{m,3}(s) - F_{m,1}(s) \stackrel{s}{=} \left[\begin{array}{c|c} A - \Delta_1 & 0 \\ 0 & (W'_m V_m)^{-1} W'_m A V_m \end{array} \middle| \begin{array}{c} b \\ (W'_m V_m)^{-1} W'_m b \end{array} \right].$$

Consider the following basis transformation T where

$$T = \left[\begin{array}{cc} I & -V_m \\ 0 & I \end{array} \right], \quad \text{and} \quad T^{-1} = \left[\begin{array}{cc} I & V_m \\ 0 & I \end{array} \right],$$

which yields

$$\begin{aligned}
(80) \quad &F_{m,3}(s) - F_{m,1}(s) \\
&\stackrel{s}{=} \left[\begin{array}{cc|c} A - \Delta_1 & (A - \Delta_1)V_m - V_m(W'_m V_m)^{-1} W'_m A V_m & 0 \\ 0 & (W'_m V_m)^{-1} W'_m A V_m & (W'_m V_m)^{-1} W'_m b \\ \hline c & 0 & 0 \end{array} \right].
\end{aligned}$$

Since $b = V_m e_1 \beta$, a routine calculation shows that the (1,2) block of (81) is zero, from which we conclude that $F_{m,3}(s) - F_{m,1}(s) = 0$ since the realisation in (81) has m unobservable and N uncontrollable modes thus establishing the first equivalence. $F_{m,4}(s) \equiv F_{m,2}(s)$ follows in an analogous way, but by applying the transformation

$$T = \left[\begin{array}{cc} I & 0 \\ -W'_m & I \end{array} \right]$$

to the difference $F_{m,4}(s) - F_{m,2}(s)$, thus completing the proof. $\square$

It is apparent from Theorem 3.2 and Corollary 3.3 that

$$(81) \quad \left[\frac{A - \Delta_3}{c} \middle| \frac{b}{d} \right] \equiv \left[\frac{A - \Delta_1}{c} \middle| \frac{b}{d} \right] \quad \text{and} \quad \left[\frac{A - \Delta_3}{c} \middle| \frac{b}{d} \right] \equiv \left[\frac{A - \Delta_2}{c} \middle| \frac{b}{d} \right],$$

which may be demonstrated by, respectively, applying the basis changes

$$T = \left[\begin{array}{cc} I & 0 \\ V_m(W'_m V_m)^{-1} W'_m & I \end{array} \right] \quad \text{and} \quad T = \left[\begin{array}{cc} I & V_m(W'_m V_m)^{-1} W'_m \\ 0 & I \end{array} \right]$$

to the differences between each transfer function in (81). The perturbation Δ_1 to the transition matrix of $F(s)$ yields $N - m$ uncontrollable modes, while the perturbation Δ_2 gives rise to $N - m$ unobservable modes. 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 realisation of the same transfer function.

The following theorem states the findings above in the context of the Lanczos process.

THEOREM 3.4. *Suppose that m steps of the Lanczos process have been taken and that Δ_1 , Δ_2 and Δ_3 are defined in Corollary 2.4 and (47), then, the following are true.*

(a) *The Galerkin conditions in (63) and (64) are satisfied if and only if $F_b(s) = (sI - T_m)^{-1}e_1\delta_1$ and $F_c(s) = \beta_1 e'_1(sI - T_m)^{-1}$; furthermore, the residual error norms are*

$$(82) \quad \|b - (sI - A)V_m F_b(s)\|_\infty = \|\hat{v}_{m+1}\|_2 \cdot \|e'_m F_b(s)\|_\infty,$$

$$(83) \quad \|c - F_c(s)W'_m(sI - A)\|_\infty = \|\hat{w}_{m+1}\|_2 \cdot \|F_c(s)e_m\|_\infty.$$

(b) $F_\Delta(s)$ and $F_m(s)$ are different realisations of the same transfer function, namely, $F_\Delta(s) \equiv F_m(s)$ where

(84)

$$F_\Delta(s) \stackrel{s}{=} \left[\frac{A - \Delta_3}{c} \middle| \frac{b}{d} \right] \quad \text{and} \quad F_m(s) \stackrel{s}{=} \left[\frac{W'_m A V_m}{\beta_1 e'_1} \middle| \frac{e_1 \delta_1}{d} \right] = \left[\frac{T_m}{\beta_1 e'_1} \middle| \frac{e_1 \delta_1}{d} \right].$$

(c) X_m^L and Y_m^L are the controllability and observability grammians of $F_m(s)$.

$$(d) \quad \left[\frac{A - \Delta_1}{c} \middle| \frac{b}{d} \right] \equiv \left[\frac{A - \Delta_2}{c} \middle| \frac{b}{d} \right] \equiv \left[\frac{A - \Delta_3}{c} \middle| \frac{b}{d} \right] \equiv \left[\frac{T_m}{\beta_1 e'_1} \middle| \frac{e_1 \delta_1}{d} \right].$$

Proof. The proof is essentially the same as that of Theorem 3.2 and Corollary 3.3 except that it uses the fact that $W'_m(A - \Delta_3) - W'_m A V_m W'_m = 0$. $\square$

The appeal of the Lanczos process in model reduction stems from the simplicity of the formulas derived in Theorem 3.4. The reduced order model and the backward error perturbations are expressed in terms of the data generated in the course of the Lanczos process. Thus despite the relations with Lyapunov equations, the reduced order model may be constructed without the need to form X_m^L and Y_m^L . The Arnoldi–Lanczos schemes suggested in this paper are not always guaranteed to yield stable reduced order models. A possible remedy to this problem is to employ an implicitly restarted Lanczos process to compute stable reduced order models. This method was proposed in [8].

The following procedure summarises the findings of this section.

KRYLOV SUBSPACE MODEL REDUCTION ALGORITHM

- Start: Specify tolerances $\gamma > 0$ and $\epsilon > 0$, set an integer parameter m .
Arnoldi
- Perform m steps of the Arnoldi process with (A, b) to produce $\hat{H}_m$, $h_{m+1,m}$, V_m , v_{m+1} , and β .
- Perform m steps of the Arnoldi process with (A', c') to produce $\hat{G}_m$, $g_{m,m+1}$, W_m , w_{m+1} , and δ .
- Form the reduced order model from either (71) or (72).
- Test the $\mathcal{L}^\infty$ errors in (66) and (67), if either (66) $> \epsilon$ or (67) $> \epsilon$ increase m and continue the Arnoldi process.
- Lanczos*
- Perform m steps of the Lanczos process with A , b , c' to produce T_m , V_m , v_{m+1} , W_m , w_{m+1} , δ_1 , δ_{m+1} , β_1 , and β_{m+1} .
- Form the reduced order model from (84).

- Test the $\mathcal{L}^\infty$ errors in (82) and (83), if either (82) $> \epsilon$ or (83) $> \epsilon$ increase m and continue the Lanczos process.

Remark 3.4. The Arnoldi- and Lanczos-based model reduction algorithms presented above are mathematically equivalent in the sense that they produce the same low order transfer function. This equivalence can be established by using (15), (16), (37), and (38) and the fact that V_m (similarly, W_m) computed from the Arnoldi and the Lanczos methods are different bases for the same space.

In practice, it is desirable to perform model reduction of multivariable linear systems in which $F(s) = D + C(sI - A)^{-1}B$ where $B \in \mathbb{R}^{N \times p}$ and $C \in \mathbb{R}^{q \times N}$. To this end, one may employ block Krylov schemes to compute the bases V_m and W_m of $\mathcal{K}_m(A, B)$ and $\mathcal{L}_m(A', C')$, respectively. The reduced order models obtained using such schemes are given by

$$F_{m,1}(s) \stackrel{s}{=} \left[\frac{(W'_m V_m)^{-1} W'_m A V_m}{C V_m} \mid \frac{(W'_m V_m)^{-1} W'_m B}{D} \right] \quad \text{and}$$

$$F_{m,2}(s) \stackrel{s}{=} \left[\frac{W'_m A V_m (W'_m V_m)^{-1}}{C V_m (W'_m V_m)^{-1}} \mid \frac{W'_m B}{D} \right]$$

for the block Arnoldi process, assuming that $(W'_m V_m)$ is nonsingular, and by

$$F_m \stackrel{s}{=} \left[\frac{W'_m A V_m}{C V_m} \mid \frac{W'_m B}{D} \right]$$

for the block Lanczos process. We refer the reader to [5], [11], [15] for the implementation details of block Krylov methods. In the absence of breakdown, one may derive a nonsymmetric block Lanczos algorithm based on a generalisation of the vector scheme suggested in [5]. The behaviour of such an algorithm remains an open area of research in the presence of breakdowns.

Suppose that m steps of the Arnoldi process have been taken and that either $h_{m+1,m} = 0$ or $g_{m,m+1} = 0$. Three possible scenarios may then arise. The first, $h_{m+1,m} = 0$, implies that v_{m+1} may not be computed and the process yields the exact solution; $P^* = P_m = V_m X_m^A V'_m$; $Q_m = W_m Y_m^A W'_m$ is a low rank approximate solution to (6). The matrix V_m forms an orthogonal basis for the controllable space, which implies that the reduced order models given by Theorem 3.2 are equivalent to the high order model $F(s)$ in (56). The second, $g_{m,m+1} = 0$, implies that w_{m+1} may not be computed and the process yields the exact solution; $Q^* = Q_m = W_m Y_m^A W'_m$; $P_m = V_m X_m^A V'_m$ is a low rank approximate solution to (5). In this case, W_m forms an orthogonal basis for the observability space and the reduced order models of Theorem 3.2 are equivalent to $F(s)$. The third, $h_{m+1,m} = 0$ and $g_{m,m+1} = 0$, implies that v_{m+1} and w_{m+1} may not be formed and the process yields exact grammians; here, $F_{m,i}(s) \equiv F(s)$ for $i = 1, 2$. A similar type of breakdown is experienced in the Lanczos process if either $\hat{v}_{m+1} = 0$ or $\hat{w}_{m+1} = 0$. The key observation associated with these types of breakdowns is that the reduced order models are minimal realisations of the high order $F(s)$; this connection was first established in [14]. The Lanczos process might suffer from a breakdown in which $\hat{v}_{m+1} \neq 0$ and $\hat{w}_{m+1} \neq 0$ and yet $\hat{v}'_{m+1} w_{m+1} = 0$; similarly, for the Arnoldi scheme, $(W'_m V_m)$ might be singular. The implications of such breakdown on the model reduction algorithms is not well understood and is the focus of ongoing research.

Observe that if the solutions to the m -dimensional Lyapunov equations have rank $< m$, $F_{m,i}(s)$ for $i = 1, 2$ are not minimal realisations. By exploiting X_m and Y_m one

readily computes the balancing transformations and the balanced $F_{m,i}(s)$'s may then be truncated to yield minimal realisations [13]. An alternative, and numerically sound approach, is to utilise the hybrid scheme proposed in [1]. This procedure combines the Krylov subspace methods to the balancing procedure via orthogonal transformations without computing the potentially ill-conditioned balancing transformation. For the implementation details, we refer the reader to [1].

4. Model reduction of large scale discrete time systems. The focus of this section is to extend the findings above to obtain model reduction schemes suited to large scale discrete time systems. The extensions are fairly straightforward and are stated without proofs.

The need for discrete time model reduction schemes arises when high order models described by difference equations are approximated by those of lower dimension. High order models arise in such areas as the modeling of digital circuits, communication networks, and model identification. Consider the discrete time linear dynamical system

$$(85) \quad x_{k+1} = Ax_k + bu_k \quad A \in \mathbb{R}^{N \times N}, \quad b \in \mathbb{R}^N,$$

$$(86) \quad y_k = cx_k + du_k \quad c' \in \mathbb{R}^N, \quad d \in \mathbb{R}.$$

Associated with this linear system are the controllability and observability Lyapunov equations defined as

$$(87) \quad APA' - P + bb' = 0,$$

$$(88) \quad A'QA - Q + c'c = 0,$$

respectively. Low rank solutions $P_m = V_m X_m V_m'$ and $Q_m = W_m Y_m W_m'$ are sought such that the residuals $AP_m A' - P_m + bb'$ and $A'Q_m A - Q_m + c'c$ satisfy Galerkin-type conditions on the Krylov spaces $\mathcal{L}_m(A', c')$ and $\mathcal{K}_m(A, b)$, respectively. Assuming that $W_m' V_m$ is nonsingular allows the residual error to be expressed as

$$(89) \quad R_m(X_m) = [V_m \quad (I - V_m(W_m' V_m)^{-1} W_m') v_{m+1}] \\ \times \begin{bmatrix} \hat{H}_m X_m \hat{H}_m' - X_m + e_1 \beta^2 e_1' & X_m e_m h_{m+1,m} \\ h_{m+1,m} e_m' X_m & h_{m+1,m}^2 e_m' X_m e_m \end{bmatrix} \\ \times \begin{bmatrix} V_m' \\ v_{m+1}' (I - W_m (V_m' W_m)^{-1} V_m') \end{bmatrix}.$$

Similarly, associated with (88), the residual error function for any given approximate solution of the form $Q_m = W_m Y_m W_m'$ may be written as

$$(90) \quad S_m(Y_m) = [W_m \quad (I - W_m (V_m' W_m)^{-1} V_m') w_{m+1}] \\ \times \begin{bmatrix} \hat{G}_m' Y_m \hat{G}_m - Y_m + e_1 \delta^2 e_1' & Y_m e_m g_{m,m+1} \\ g_{m,m+1} e_m' Y_m & g_{m,m+1}^2 e_m' Y_m e_m \end{bmatrix} \\ \times \begin{bmatrix} W_m' \\ w_{m+1}' (I - V_m (W_m' V_m)^{-1} W_m') \end{bmatrix},$$

where V_m , v_{m+1} , $h_{m+1,m}$ and W_m , w_{m+1} , $g_{m,m+1}$ are defined from the data generated in the course of the two Arnoldi processes associated with $\mathcal{K}_m(A, b)$ and $\mathcal{L}_m(A', c')$ respectively. Furthermore, $\hat{H}_m$ and $\hat{G}_m$ are defined in (17) and (18), respectively. The following theorem solves Problem 2.1 in the context of discrete time systems.

THEOREM 4.1. *Suppose that m steps of the Arnoldi process have been taken and that the residual errors associated with (87) and (88) are defined by (89) and (90). Furthermore, suppose that $|\lambda_i(\hat{H}_m)(\bar{\lambda}_j(\hat{H}_m))^{-1}| \neq 1$ for all i, j and $|\lambda_i(\hat{G}_m)(\bar{\lambda}_j(\hat{G}_m))^{-1}| \neq 1$ for all i, j then,*

(a) $W'_m R_m(X_m) W_m = 0$ if and only if $X_m = X_m^A$, where X_m^A satisfies

$$(91) \quad \hat{H}_m X_m^A \hat{H}'_m - X_m^A + e_1 \beta^2 e'_1 = 0.$$

Under these conditions,

$$\|R_m^A\|_F := \|R_m(X_m^A)\|_F = \left\| \begin{bmatrix} H_m X_m^A H'_m - X_m^A + e_1 \beta^2 e'_1 & X_m^A e_m h_{m+1,m} \\ h_{m+1,m} e'_m X_m^A & h_{m+1,m} e'_m X_m^A e_m \end{bmatrix} \right\|_F.$$

(b) $V'_m S_m(Y_m) V_m = 0$ if and only if $Y_m = Y_m^A$, where Y_m^A satisfies

$$(92) \quad \hat{G}'_m Y_m^A \hat{G}_m - Y_m^A + e_1 \delta^2 e'_1 = 0.$$

Under these conditions,

$$\|S_m^A\|_F := \|S_m(Y_m^A)\|_F = \left\| \begin{bmatrix} G'_m Y_m^A G_m - Y_m^A + e_1 \delta^2 e'_1 & Y_m^A e_m g_{m,m+1} \\ g_{m,m+1} e'_m Y_m^A & g_{m,m+1} e'_m Y_m^A e_m \end{bmatrix} \right\|_F.$$

(c) There exist rank-1 perturbations $\Delta_1 = (I - V_m(W'_m V_m)^{-1} W'_m) v_{m+1} h_{m+1,m} v'_m$ and $\Delta_2 = w_m g_{m,m+1} w'_{m+1} (I - V_m(W'_m V_m)^{-1} W'_m)$ such that

$$\begin{aligned} (A - \Delta_1) P_m (A - \Delta_1)' - P_m + b b' &= 0, \\ (A - \Delta_2)' Q_m (A - \Delta_2) - Q_m + c' c &= 0, \end{aligned}$$

and

$$\|\Delta_1\|_F^2 = h_{m+1,m}^2 \{1 + \|(W'_m V_m)^{-1} W'_m v_{m+1}\|_2^2\}$$

and

$$\|\Delta_2\|_F^2 = g_{m,m+1}^2 \{1 + \|(V'_m W_m)^{-1} V'_m w_{m+1}\|_2^2\}.$$

Remark 4.1. Observe that $\Delta_3 := \Delta_1 + \Delta_2$ is a perturbation on the data in A such that

$$\begin{aligned} (A - \Delta_3) P_m (A - \Delta_3)' - P_m + b b' &= 0, \\ (A - \Delta_3)' Q_m (A - \Delta_3) - Q_m + c' c &= 0. \end{aligned}$$

Furthermore, Δ_3 is at most a rank-2 perturbation which may be factorised as

$$\begin{aligned} \Delta_3 &= [w_m \quad (I - V_m(W'_m V_m)^{-1} W'_m) v_{m+1}] \begin{bmatrix} 0 & g_{m,m+1} \\ h_{m+1,m} & 0 \end{bmatrix} \\ &\quad \times \begin{bmatrix} v'_m \\ w'_{m+1} (I - V_m(W'_m V_m)^{-1} W'_m) \end{bmatrix}. \end{aligned}$$

Finally, a direct calculation will verify that $W'_m \Delta_1 V_m = W'_m \Delta_2 V_m = W'_m \Delta_3 V_m = 0$.

Note that despite the form of the Lyapunov equations in (87) and (88), the results of Theorem 4.1 are the same as those obtained in §2.1. Theorem 3.2 may be restated in its exact form as presented in §3. The main differences are that V_m and W_m are derived from the two Arnoldi processes associated with $\mathcal{K}_m(A, b)$ and $\mathcal{L}_m(A', c')$, respectively, and where A , b , and c' originate from the discrete time model defined in (85) and (86); furthermore, X_m^A and Y_m^A are the solutions to the low dimensional discrete time Lyapunov equations (91) and (92), respectively. One may apply the Lanczos process to the discrete time model reduction problem in a similar fashion as presented in §3 to obtain a variant of Theorem 3.4 the derivation of which we omit.

5. Numerical experiments. The purpose of this section is to illustrate with the help of two examples the behaviour of the error formulas presented in §§2 and 3. The tests reported here were performed on a Sparc-10 Sun workstation using Pro-MATLAB 4.0 which carries out operations to a unit round-off of 2.22×10^{-16} .

Example 1. The aim of the first example is to illustrate the behaviour of the backward error formulas associated with the Lyapunov equation solvers of §2. This example has been derived from the discretisation of a partial differential equation of the form

$$(93) \quad \begin{aligned} -\Delta u(x, y, t) + \alpha \frac{\partial^2 u(x, y, t)}{\partial x \partial y} + 2\beta e^{-xy} \left(\frac{\partial u(x, y, t)}{\partial x} + \frac{\partial u(x, y, t)}{\partial y} \right) \\ - 2\gamma u(x, y, t) = \frac{\partial u(x, y, t)}{\partial t}, \\ x, y \in \Omega, \\ u(x, y, 0) = u_0 \quad x, y \in \Omega, \\ u(x, y, t) = \sigma(x, y), \quad t \geq 0, \quad x, y \in \partial\Omega, \end{aligned}$$

where Ω denotes the set of points in the unit square $(0, 1) \times (0, 1)$ and $\partial\Omega$ denotes the boundary. Setting $\alpha := 3$, $\beta := 0.5$, and $\gamma := 20$, the discretisation is carried out using 102 points in both the x and y directions leading to a linear system of the form

$$(94) \quad F(s) = \begin{cases} \dot{w}(t) = Aw(t) + bg(t), \\ f(t) = cw(t), \end{cases}$$

where A is nonsymmetric sparse matrix of dimension $N = 100^2$ with $nz = 88804$ nonzero elements (i.e., with a density of approximately 0.09%). The vector e is the vector of ones and $b := e * 0.1$, finally, c is a random vector in $\mathbb{R}^{1 \times N}$. Tests with other choices of b and c showed similar results.

Figures 1 and 2 report on the evolutions of $\|\Delta_i\|_F$, $i = 1, 2, 4, 5$ for the Arnoldi and Lanczos–Lyapunov equation solvers. The tests show that $\|\Delta_{4,5}\|_F$ (the lower traces) are significantly smaller than $\|\Delta_{1,2}\|_F$ (the upper traces) when either scheme is employed. Figures 1 and 2 confirm the fact that the a priori backward error bounds are conservative and do not provide an accurate means of gauging the progress of the iterative process. However, the a posteriori bounds indicate that accurate approximate solutions may be obtained for small m . The a posteriori perturbations Δ_4 and Δ_5 may not be the smallest perturbations for which one can compute an m th order approximate model; the problem of determining such perturbations remains open.

Example 2. The aim of this experiment is to test the effectiveness of the model reduction schemes proposed in §3. The problem is set up with $A \in \mathbb{R}^{N \times N}$ where $N = 100$ and the top left-hand 2×2 block of A is set to

$$\begin{bmatrix} -1 & 100 \\ -100 & -1 \end{bmatrix},$$

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 two that are $-1 \pm 100j$. The first five elements of b and c are uniformly distributed in $[0, 1]$ while the 95 remaining elements are uniformly distributed in $[0, 1/25]$. The infinity norm of $F(s)$ is given by the radius of the smallest circle centered at the origin to enclose the Nyquist plot of Fig. 3; alternatively, it may be computed from $\max |F(j\omega)|$

FIG. 1. Evolutions of $\|\Delta_{1,2}\|_F$ (upper traces) and $\|\Delta_{4,5}\|_F$ (lower traces) using the Arnoldi-based method in Example 1.

for all $\omega \in \mathbb{R}$. In this case, $\|F(j\omega)\|_\infty = 3.0716$. Table 1 shows the evolution of the $\mathcal{L}^\infty$ error expression of (66) and (67) denoted here by Err1 and Err2, respectively, for the Arnoldi model reduction algorithm. The table indicates that Err1 and Err2 fall in

TABLE 1
Residual error norms associated with the Arnoldi model reduction scheme in Example 2.

m	Err1	Err2	m	Err1	Err2	m	Err1	Err2
1	1.2998e+1	1.3398e+1	2	1.2912e+0	1.2052e+1	3	5.2256e−1	1.7498e−1
4	8.2001e−1	3.4535e−1	5	4.5508e−1	3.2630e−1	6	2.0198e−1	1.9200e−1
7	1.2076e−1	1.0679e−1	8	7.4857e−2	6.8587e−2	9	5.6592e−2	6.2173e−2
10	4.8653e−2	5.6193e−2	11	5.7017e−2	5.9666e−2	12	5.9939e−2	5.5473e−2
13	7.2936e−2	6.9952e−2	14	7.1812e−2	7.5504e−2	15	9.5773e−2	8.8814e−2

magnitude as m increases. However, as is well known, Galerkin conditions of the type in (63) and (64) do not guarantee a nonincreasing evolution of Err1 and Err2. To maintain orthogonal and biorthogonal bases for $\mathcal{K}_m(A, b)$ and $\mathcal{L}_m(A', c')$, we resorted to reorthogonalisation and rebiorthogonalisation for the Arnoldi and Lanczos processes, respectively. As predicted by Proposition 3.4, the sequence of Lanczos errors, (82) and (83), were the same as Table 1. Figure 3 compares the frequency responses of $F(s)$ and three low order approximate realisations obtained from the Arnoldi model reduction scheme. The low order approximate models have four, six, and eight states, respectively. Each frequency response is performed over $[10^{-3}\text{rads/s}, 10^3\text{rads/s}]$ with the lowest frequency point close to the positive real axis and the highest frequency point almost at the origin. Observe that the four frequency responses are indistinguishable over high frequencies with a progressively smaller error over low frequencies

FIG. 2. Evolutions of $\|\Delta_{1,2}\|_F$ (upper traces) and $\|\Delta_{4,5}\|_F$ (lower traces) using the Lanczos-based method in Example 1.

as the state dimension is increased. This behaviour is reflected in the $\mathcal{L}^\infty$ error being significantly smaller than $\|F(s)\|_\infty$ for increasing state dimension as shown in Table 2. Table 2 also lists the $\mathcal{L}^\infty$ forward error associated with the balanced truncation algorithm [13].

The tabulated results indicate that the balanced truncation algorithm enjoys a rapid forward error decay for increasing state dimension while the convergence rate of the Krylov based method is much slower. This superior convergence behaviour is due in part to the large volume of computation associated with solving (5) and (6).

TABLE 2

Forward error $\|F(s) - F_m(s)\|_\infty$ using the Arnoldi-Lanczos and balanced truncation model reduction schemes.

m	Krylov	bal/trun	m	Krylov	bal/trun	m	Krylov	bal/trun
1	2.8417e+0	7.4438e-1	5	6.6195e-1	1.4232e-2	9	1.2821e-1	3.9387e-6
2	3.0724e+0	1.4023e+0	6	2.3708e-1	1.5040e-3	10	1.1953e-1	6.6290e-7
3	2.2053e+0	7.4556e-1	7	1.6710e-1	2.1184e-4	11	1.1108e-1	7.8445e-8
4	1.5677e+0	1.1623e-1	8	1.4313e-1	2.2361e-5	12	1.0036e-1	5.7788e-9

6. Conclusions. The aim of this paper has been to present and test several model reduction algorithms suitable for computing low dimensional approximate models of large scale continuous and discrete time systems. By exploiting Krylov subspace methods, approximate solutions to the controllability and observability Lyapunov equations have been found for which the residual errors satisfy Galerkin type conditions. Furthermore, low dimensional expressions have been derived for the residual error norms and a priori and a posteriori backward perturbation norms.

FIG. 3. Frequency responses for the four models of Arnoldi based approximations, Original —; four-state -.-.-; six-state - - -; and eight-state

We show that the low rank approximate grammians are the exact grammians to a perturbed linear system in which the perturbation has at most rank = 2. We demonstrate that this perturbed linear system is equivalent to a low dimensional linear system with state dimension no greater than m . Furthermore, exact low dimensional expressions for an associated $\mathcal{L}^\infty$ error are presented.

Acknowledgment. The authors wish to thank Professor Paul Van Dooren for helpful discussions and advice on the subject matter of this paper.

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 Conf. Decision and Control, Brighton, England, 1991, pp. 1165–1170.
- [2] R. H. BARTELS AND W. STEWART, *Solution of the matrix equation $AX + XB = C$* , Comm. ACM, 15 (1972), pp. 820–826.
- [3] D. L. BOLEY, *Krylov space methods on state-space control models*, CS tech. report, TR92-18, University of Minnesota, Minneapolis, 1992.
- [4] D. L. BOLEY AND G. H. GOLUB, *The Lanczos-Arnoldi algorithm and controllability*, Systems Control Lett., 4 (1984), pp. 317–327.
- [5] ———, *The nonsymmetric Lanczos algorithm and controllability*, Systems Control Lett., 16 (1991), pp. 97–105.
- [6] K. GLOVER, *All optimal Hankel norm approximations of linear multivariable systems and their $\mathcal{L}^\infty$ error bounds*, Internat. J. Control, 39 (1984), pp. 1115–1193.
- [7] G. H. GOLUB, B. KÄGSTRÖM AND P. VAN DOOREN, *Direct block tridiagonalisation of single-input single-output systems*, Systems Control Lett., 18 (1992), pp. 109–120.
- [8] E. J. GRIMME, D. C. SORESENSEN, AND P. VAN DOOREN, *An implicitly restarted non-symmetric Lanczos algorithm for sparse or unstructured matrices*, Appl. Math. Lett., 1994, pp. 75–80.

- [9] N. J. HIGHAM, *Perturbation theory and backward error for $AX - XB = C$* , BIT, 33 (1993), pp. 124–136.
- [10] I. M. JAIMOUKHA, E. M. KASENALLY, AND D. J. N. LIMEBEER, *Numerical solution of large scale Lyapunov equations using Krylov subspace methods*, 31st Conf. Decision and Control, Tucson, AZ, 1992, pp. 1927–1932.
- [11] I. M. JAIMOUKHA AND E. M. KASENALLY, *Krylov subspace methods for solving large Lyapunov equations*, SIAM J. Numer. Anal., 31 (1994), pp. 227–251.
- [12] 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.
- [13] B. C. MOORE, *Principal component analysis in linear systems: controllability, observability and model reduction*, IEEE Trans. Auto. Contr., AC-26 (1981), pp. 17–31.
- [14] B. PARLETT, *Reduction to tridiagonal form and minimal realisations*, SIAM J. Math. Anal. Appl., 13 (1992), pp. 567–593.
- [15] Y. SAAD, *On the rates of convergence of the Lanczos and block Lanczos methods*, SIAM J. Numer. Anal., 17 (1980), pp. 687–706.
- [16] ———, *The Lanczos bi-orthogonalisation algorithm and other oblique projection methods for solving large unsymmetric systems*, SIAM J. Numer. Anal., 19 (1982), pp. 485–506.
- [17] ———, *Numerical solution of large Lyapunov equations*, Signal Processing, Scattering, Operator Theory and Numerical Methods, M. A. Kaashoek, J. H. Van Schuppen and A. C. M. Ran, eds., Birkhäuser, Boston, 1990, pp. 503–511.
- [18] M. G. SAFONOV AND R. Y. CHIANG, *A Schur method for balanced model reduction*, Proc. American Control Conf., Atlanta, GA, 1988.
- [19] P. VAN DOOREN, *Numerical linear algebra techniques for large scale matrix problems in systems and control*, 31st Conf. Decision and Control, Tucson, Arizona, 1992, pp. 1933–1938.
- [20] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.