

# A Rational Lanczos Algorithm for Model Reduction \*

K. Gallivan and E. Grimme  
Coordinated Science Laboratory  
University of Illinois at Urbana-Champaign  
Urbana IL 61801, USA

P. Van Dooren  
Center for Systems Engineering and Applied Mechanics  
Université Catholique de Louvain  
B-1348 Louvain-la-Neuve, BELGIUM

## Abstract

This paper presents a model reduction method for large-scale linear systems that is based on a Lanczos-type approach. A variant of the nonsymmetric Lanczos method, rational Lanczos, is shown to yield a rational interpolant (multi-point Padé approximant) for the large-scale system. An exact expression for the error in the interpolant is derived. Examples are utilized to demonstrate that the rational Lanczos method provides opportunities for significant improvements in the rate of convergence over single-point Lanczos approaches.

**AMS classification:** Primary 65F15; Secondary 65G05.

**Key Words :** State space systems, nonsymmetric Lanczos algorithm, Padé approximation, rational interpolation, model reduction.

## 1 Introduction

This paper explores the use of Lanczos techniques for the reduced-order modeling of large-scale dynamical systems. A need for such reduced-order models arises in various areas of engineering such as the control of large flexible space structures [5] and the simulation of high speed circuits [6]. The system to be modeled is typically defined via a set of state space equations

$$E\dot{x}(t) = Ax(t) + bu(t) \quad \text{and} \quad y(t) = c^T x(t) + du(t) \quad (1)$$

where for simplicity, the direct-coupling term,  $d$ , will be assumed to be zero. As this paper will restrict itself to single-input single-output (SISO) systems, the input  $u(t)$  and output  $y(t)$  are scalar functions of time with  $b$  and  $c$  column vectors of length  $n$ . The system matrix,  $A \in IR^{n \times n}$ , and descriptor matrix,  $E \in IR^{n \times n}$ , are assumed to be sparse or structured (e.g., Toeplitz). We stress that such assumptions are met by large-scale problems arising from most applications. However, most existing model reduction techniques (e.g., balanced truncations and Hankel norm optimal approximations) [8] fail to take advantage of any sparsity or structure in the system matrix and are thus typically impractical for large-scale problems.

For the case where  $E$  is an identity matrix, the zero-state ( $x(0) = 0$ ) solution to the first expression in (1) is  $x(t) = \int_0^t e^{A(t-\tau)} bu(\tau) d\tau$ . Thus determining a good  $k \ll n$  order approximation,

$$\hat{E}\dot{\hat{x}}(t) = \hat{A}\hat{x}(t) + \hat{b}u(t) \quad \text{and} \quad \hat{y}(t) = \hat{c}^T \hat{x}(t), \quad (2)$$

---

\*This work has been supported in part by the National Science Foundation under grants CCR-9120105 and CCR-9209349. The work of the second author was also supported in part by a DOE Computational Science Graduate Fellowship.

is intimately connected with finding a good approximation to a matrix exponential. A method based on orthogonal Krylov projectors (the Arnoldi algorithm) is utilized in [12, 26] for approximating  $e^{At}b$ . But in fact, these concepts can be taken one step further by noting that one is really only interested in that information in  $e^{At}b$  which lies in the direction of  $c$  (one ultimately desires  $\|y - \hat{y}\|$  small for some desired range of inputs  $u$ ). Numerous papers [4, 17, 27, 28] are beginning to explore this last fact. In particular, these papers begin to investigate the use of an oblique Krylov projector (i.e., the Lanczos algorithm) for generating the reduced-order model.

The Lanczos-based approaches to model reduction are in fact connected to well-known approximations of (1) including partial realizations and/or Padé approximants [14, 29]. These approximations are centered on the transfer functions  $g(s) \equiv y(s)/u(s) = c^T(sE - A)^{-1}b$  and  $\hat{g}(s) \equiv \hat{y}(s)/\hat{u}(s) = \hat{c}^T(s\hat{E} - \hat{A})^{-1}\hat{b}$  which arise out of Laplace transforms of (1) and (2) respectively. The reduced-order model is computed so that its transfer function  $\hat{g}(s)$  shares (matches) certain attributes of the original transfer function  $g(s)$ . The Lanczos method is known to be a preferred numerical approach for computing such a model [7, 10]. Additionally, Lanczos-type methods only involve multiplication by  $A$  and  $E$  and/or solving linear equations involving  $A$  and  $E$ . Thus one can take advantage of the structure of these matrices. Avenues also exist in the Lanczos method for removing the spurious, unstable poles which may appear in the approximation [15].

Unfortunately, model reduction methods such as partial realization and Padé approximation are not acceptable in all applications. Such approximations tend to converge in a local fashion about a single frequency  $s = \sigma \in \mathcal{C}$ . The dimension of the reduced-order model can grow large before becoming an acceptable global approximation of the original system. To overcome this difficulty, several papers in the areas of control and circuits explore the use of a multi-point Padé approximant (denoted a rational interpolant in the systems literature) for approximating (1) (see for example [6, 16, 18, 32]). In rational interpolation [1] (multi-point Padé [2]), a reduced-order model is constructed whose transfer function  $\hat{g}(s)$  interpolates the value and subsequent derivatives of  $g(s)$  at multiple frequencies  $\{\sigma_1, \sigma_2, \dots, \sigma_r\}$ . Each interpolation point is selected to identify the dynamics of (1) in a specific frequency range. One avoids trying to acquire information from a single, distant point.

This paper lays the foundation for a practical computational approach to rational interpolation through the development of the novel rational Lanczos method. Being a Lanczos type method, rational Lanczos still possesses the desirable numerical qualities lacking in explicit moment matching approaches. But in a significant break from the standard Lanczos algorithms, rational Lanczos is no longer tied to a specific interpolation point. By intelligently selecting from multiple interpolation points, rational Lanczos provides an opportunity for efficiently and accurately determining models across a wide frequency range. An error expression between the transfer functions of the original and reduced-order systems is derived which may enhance the placement of the interpolation points. Given this set of interpolations points, a strategy for selecting among them arises quite cheaply out of rational Lanczos and is grounded in system theory. One does not simply match a fixed number of moments about each interpolation point. Such an approach may be unnecessary at certain interpolation frequencies and insufficient at others. Instead, selections are made from among the interpolation points as the model size grows with the goal of maximizing the amount of new information being placed into the model. A surprising benefit of this last fact is that the rational Lanczos method is driven to avoid the numerical instabilities present in the standard Lanczos method. Meaningful system theory in rational Lanczos can replace the nonintuitive, complex fixes of the standard Lanczos method (e.g., look-ahead [9, 21]).

This paper begins in §2 by describing moment matching, the Lanczos method and the connections between the two. An emphasis is placed on defining the terminology associated with both moment matching and the Lanczos method in a unified and unambiguous way. The techniques of §2 correspond to interpolation about a single point. §3 discusses the limitations of interpolating about a single frequency point and thus motivates the development of the rational Lanczos method in §4. The rational Lanczos method of §4 is constructed in a simplified manner so as to promote an understanding of the algorithm. The relation between rational Lanczos and rational interpolation is proven. §5 converts the rational Lanczos method into a model reduction tool. Examples are provided to suggest the power of the approach. An error expression for the reduced-order model is derived in §6.

## 2 Background

This section contains the background material necessary to proceed with the later development of the rational Lanczos algorithm as a model reduction method. We emphasize the need for a thorough coverage of moment matching methods, versions of the Lanczos method, and the interconnections between the two. The terminology and credit for these topics lies strewn over several application areas. It is our goal to at least begin to piece together these items in the following review.

### 2.1 Moment matching methods

The model-reduction methods of interest in this paper are those which reproduce in the reduced-order model a set of invariant attributes belonging to the transfer function  $g(s)$  of (1). To be more specific, we are interested in determining a reduced-order model which matches the first  $2k$  coefficients,  $m_j$ , arising in a power series expansion of  $g(s)$ . If  $g(s)$  is expanded about zero for example,

$$g(s) = m_0 + m_1 s + \frac{s^2}{2!} m_2 + \frac{s^3}{3!} m_3 + \dots, \quad (3)$$

the coefficients (referred to as moments in this case) satisfy  $m_j = -c^T (A^{-1}E)^j A^{-1}b$ . The reduced-order model, a Padé approximant, is constructed so that  $m_j = \hat{m}_j = -\hat{c}^T (\hat{A}^{-1}\hat{E})^{j-1} \hat{b}$  for  $j = 0, 1, 2, \dots, 2k-1$ . These moments are the value and subsequent derivatives of the transfer function  $g(s)$  evaluated at  $s = 0$ . If  $g(s)$  is expanded in a power series about infinity,

$$g(s) = d + m_{-1}s^{-1} + m_{-2}s^{-2} + m_{-3}s^{-3} + \dots, \quad (4)$$

the coefficients (referred to as Markov parameters in this case) satisfy  $m_{-j} = c^T (E^{-1}A)^{(j-1)} E^{-1}b$ . The resulting model, denoted a partial realization, possesses moments which satisfy  $m_{-j} = \hat{m}_{-j}$  for  $j = 1, 2, \dots, 2k$ . These Markov parameters are the value of the zero-state impulse response  $g(t)$  (the inverse Laplace transform of  $g(s)$ ) and subsequent derivatives of  $g(t)$  evaluated at  $t = 0$ . Power series expansions about 0 are generally of greater interest because one typically desires to reproduce the steady-state (versus the transient) response of the original system over some frequency range. The steady-state behavior of the output can be defined in terms of the frequency response of the system,  $g(i\omega)$ , where the variable  $\omega \in IR$  corresponds to real frequency and  $i = \sqrt{-1}$ . If the input  $u(t)$  includes a sinusoid of frequency  $\omega_0$ , the output  $y(t)$  contains this sinusoid at steady-state scaled in magnitude and shifted in phase by the value of  $g(i\omega_0)$ . By replacing  $s$  in (3) with the shifted variable  $s - \sigma$ , i.e.,  $g(s) = \sum_{j=0}^{\infty} (s - \sigma)^j m_j(\sigma) / j!$ , one can generate shifted moments,  $m_j(\sigma) = -c^T \{(A - \sigma E)^{-1} E\}^j (A - \sigma E)b$ , which match  $g(s)$  and its subsequent derivatives at a user-specified frequency  $\sigma$ . On the other hand, we will show shortly that the use of a shifted variable  $s - \sigma$  in an expansion of the form (4) does not effect the resulting partial realization. For these reasons, this paper will concentrate on matching moments which are the coefficients of positive powers of  $s$  (possibly shifted). Models of this type fall under the title of Padé approximants.

For quick reference, various sources for the moments to be matched are summarized in Table 1. The only listed type of approximation yet to be covered is the rational interpolant or so-called multi-point Padé approximant. The rational interpolant (which includes Padé approximation as a special case) matches moments arising out of multiple (say  $\bar{i}$ ) power series expansions. These expansions are about 0 but each is shifted by a different amount,  $\sigma_i$ ,  $i = 1, \dots, \bar{i}$ . The resulting reduced-order model is defined by the matrices  $\{\hat{A}, \hat{E}, \hat{b}, \hat{c}\}$  which satisfy

$$m_{j_i}(\sigma_i) = \hat{m}_{j_i}(\sigma_i), \quad j_i = 0, 1, \dots, 2\bar{j}_i - 1, \quad i = 1, 2, \dots, \bar{i} \quad (5)$$

where

$$m_{j_i}(\sigma_i) = -c^T \{(A - \sigma_i E)^{-1} E\}^{j_i} (A - \sigma_i E)^{-1} b, \quad (6)$$

$$\hat{m}_{j_i}(\sigma_i) = -\hat{c}^T \{(\hat{A} - \sigma_i \hat{E})^{-1} \hat{E}\}^{j_i} (\hat{A} - \sigma_i \hat{E})^{-1} \hat{b} \quad (7)$$

and  $\sum_{i=1}^{\bar{i}} \bar{j}_i = k$ . The value and subsequent derivatives of  $\hat{g}(s)$  are thus equivalent to those of  $g(s)$  at multiple interpolation frequencies. The number of data pieces matched about a given interpolation point  $\sigma_i$  is twice the user-selected value of  $\bar{j}_i$ .<sup>†</sup>

Table 1: Possible Choices for the Moments to be Matched

Approximation Name(s)	Power Series Expansion	$j^{\text{th}}$ Moment
Partial Realization, Padé at $\infty$	$g(s) = \sum_{j=1}^{\infty} m_{-j} s^{-j}$	$m_{-j} = c(E^{-1}A)^{j-1}E^{-1}b = g^{(j)}(t) _{t=0}$
Padé	$g(s) = \sum_{j=0}^{\infty} m_j \frac{s^j}{j!}$	$m_j = -c(A^{-1}E)^j A^{-1}b = g^{(j)}(s) _{s=0}$
Padé (shifted)	$g(s) = \sum_{j=0}^{\infty} m_j \frac{(s-\sigma)^j}{j!}$	$m_j = -c\{(\sigma E - A)^{-1}\}^j (\sigma E - A)^{-1}b = g^{(j)}(s) _{s=\sigma}$
Rational Interpolant, Multi-point Padé	$g(s) = \sum_{j_i=0}^{\infty} m_{j_i}(\sigma_i) \frac{(s-\sigma_i)^{j_i}}{j_i!}$ for $i = 1, 2, \dots, \bar{i}$	$m_{j_i}(\sigma_i) = -c\{(\sigma_i E - A)^{-1}E\}^{j_i} (\sigma_i E - A)^{-1}b$ for $i = 1, 2, \dots, \bar{i}$

## 2.2 Moment matching through Lanczos methods

From a systems point of view, our interest in the nonsymmetric Lanczos method [19] (presented as Algorithm 1) centers on its ability to compute rectangular matrices  $W_k, V_k \in \mathbb{R}^{n \times k}$  which satisfy (i) the biorthogonality condition  $W_k^T V_k = I$  and (ii) the Krylov subspace conditions  $\text{colsp}(V_k) = \mathcal{K}_k(\Psi, r_0)$  and  $\text{colsp}(W_k) = \mathcal{K}_k(\Psi^T, q_0)$  where the Krylov subspaces are

$$\mathcal{K}_k(\Psi, r_0) = \text{span}\{r_0, \Psi r_0, \dots, \Psi^{k-1} r_0\} \quad \text{and} \quad \mathcal{K}_k(\Psi^T, q_0) = \text{span}\{q_0, \Psi^T q_0, \dots, \Psi^{k-1T} q_0\}.$$

It is the construction and use of these two Krylov subspaces which connects the Lanczos method to moment matching [29]. Note that the Krylov subspaces are shift-invariant; replacing  $\Psi$  with  $\Psi - \sigma I$  does not change the resulting subspaces.

Besides those features already mentioned, it can be easily shown that Lanczos method leads to the recursive identities

$$\Psi V_k = V_k T_k + \gamma_{k+1} v_{k+1} e_k^T \quad \text{and} \quad \Psi^T W_k = W_k T_k^T + \beta_{k+1} w_{k+1} e_k^T.$$

The standard unit vector  $e_k$  is the  $k^{\text{th}}$  column of an identity matrix of appropriate length. The matrix  $T_k = W_k^T \Psi V_k$  takes on the well-known tridiagonal form which is composed of the scalars  $\gamma_j$  below the diagonal,  $\alpha_j$  on the diagonal and  $\beta_j$  above the diagonal.

For the model reduction problem, it is important to point out that the matrix  $\Psi$  (or at least the action of  $\Psi$  on a vector) has historically been assumed to be known a priori in step (A1.4) of Algorithm 1. This condition was met by the first Lanczos-based model reduction papers in the control area [4, 17, 28]. The choices  $\Psi = A$ ,  $r_0 = b$  and  $q_0 = c$  were made while  $E$  was assumed to be an identity matrix. Given that  $\Psi = A$  is sparse, the matrix-vector products in (A1.4) are obtained with only  $\rho n$  operations where  $\rho$  is the average number of non-zero entries in a row of  $A$ . The resulting model,  $\hat{A} = T_k = W_k^T A V_k$ ,  $\hat{b} = W_k^T b$  and  $\hat{c} = V_k^T c$ , is a partial realization of the original system [14]. However, as noted in §2.1,  $k$  must typically grow large before a sufficiently accurate partial realization is acquired. The use of shifts does not help here. For example, if one replaces  $s$  in (4) with the shifted variable  $\tilde{s} = s - \sigma$  and assumes for simplicity that  $E = I$ , the resulting shifted Markov parameters are  $m_{-j} = c^T (A - \sigma I)^{j-1} b$  due to the Neumann expansion of  $(sI - A)^{-1} = ((s - \sigma)I - (A - \sigma I))^{-1} = (\tilde{s}I - (A - \sigma I))^{-1}$ . Thus the choice  $\Psi = A$  in Algorithm 1

<sup>†</sup>The restriction that an even number of moments be matched about each interpolation point is due to the form of rational Lanczos but need not hold in the most general definition of rational interpolation.

**Algorithm 1** *Nonsymmetric Lanczos [19]*


---

Input: starting vectors  $r_0$  and  $q_0$  of length  $n$ ;  
For  $k = 1$  to  $\bar{k}$ ,  
(A1.1)  $\gamma_k = \sqrt{|r_{k-1}^T q_{k-1}|}$  and  $\beta_k = \text{sign}(r_{k-1}^T q_{k-1})\gamma_k$ ;  
(A1.2)  $v_k = (r_{k-1}/\gamma_k)$  and  $w_k = (q_{k-1}/\beta_k)$ ;  
(A1.3)  $\alpha_k = w_k^T \Psi v_k$ ;  
(A1.4)  $r_k = \Psi v_k - \alpha_k v_k - \beta_k v_{k-1}$  and  $q_k = \Psi^T w_k - \alpha_k w_k - \gamma_k w_{k-1}$ ;  
end.

---

need only be shifted by  $\sigma I$ . But since the underlying Krylov subspaces are shift invariant, shifting  $s$  has no effect on the final partial realization. Only when shifting  $s$  does more than shift  $\Psi$  (e.g., Padé approximation where  $(A - \sigma I)^{-1} \neq A^{-1} - \sigma I$ ) will  $\sigma$  make an impact on the reduced-order model.

For improved accuracy, other papers select  $\Psi$  to be a rational function of  $A$  and/or  $E$ . For example, the earliest known papers on Lanczos-based model reduction (arising in structural dynamics [20, 27, 30]) chose  $\Psi = A^{-1}E$ . This selection corresponds to Padé approximation with  $\sigma = 0$  [29]. Although this choice of  $\Psi$  still fits the notation of Algorithm 1, it differs in a significant computational way from the commonly assumed choice of  $\Psi = A$ . On the surface, selecting  $\Psi$  to be a rational function of  $A$  and/or  $E$  still leads to a matrix-vector product in (A1.4). However, such a  $\Psi$  is not known a priori; more to the point, an inverse involving  $A$  and/or  $E$  should not be explicitly computed. Choosing  $\Psi$  to be a rational function requires that each matrix-vector product in (A1.4) involve the solution of a large-scale system of linear equations. But solving systems of linear equations can be much more computationally intensive than simply multiplying a known, sparse matrix times a vector. Using rational functions for  $\Psi$  (to improve accuracy or to simply handle  $E \neq I$ ) does not come without a cost. We comment on some possible approaches to minimize this additional cost in §7.

As noted above, the Lanczos algorithm is typically treated as involving a known, easily accessible  $\Psi$ . However, the use of rational functions of  $A$  and  $E$  in  $\Psi$  is examined in [23]. The so-called rational Krylov space was defined as  $\text{span}\{r_0, \Psi_1 r_0, \Psi_2 r_0, \dots, \Psi_{k-1} r_0\}$  where  $\Psi_j$  could be a rational function and where the restriction that  $\Psi_j = \Psi_1 \Psi_{j-1}$  was dropped. Hence those model reduction approaches which select  $\Psi$  to be a fixed rational function are special cases of a rational Krylov method since they enforce the relation  $\Psi_j = \Psi_1 \Psi_{j-1}$ . The resulting reduced-order model is a Padé approximant associated with a single shift  $\sigma$ . But the following section motivates interpolating  $g(s)$  at multiple frequency points. To achieve interpolation at multiple points, this paper drops the restriction that  $\Psi_j = \Psi_1 \Psi_{j-1}$ . The result is a rational Lanczos method.

### 3 Limitations of single-point interpolation

As discussed in §2, Lanczos-type algorithms with  $\Psi = (A - \sigma E)^{-1}E$  are a desirable numerical approach to computing Padé approximants. The resulting reduced-order model interpolates the transfer function  $g(s)$  and subsequent derivatives of  $g(s)$  at a single point,  $\sigma$ . However, even if one can accurately match attributes of  $g(s)$  at  $s = \sigma$ , the resulting reduced-order model may not be acceptable. Properties of Padé approximation and the Lanczos algorithm are combined in this section to indicate why the frequency response of a Lanczos-generated model tends to be only locally accurate about  $\sigma$  for reasonably small values of  $k$ . Specifically, we are interested in two convergence properties of single-point Padé approximations [2, 6]: (P1) Padé approximants are exact at the point of interpolation while accuracy is lost away from  $\sigma$  and (P2) the accuracy of the Padé approximant is lost away from  $\sigma$  more rapidly when pole(s) of the original system (the generalized eigenvalues of the pencil  $A - \lambda E$ ) are near  $\sigma$ . This second property implies that even non-dominant eigenvalues in the neighborhood of  $\sigma$  (eigenvalues near  $\sigma$  whose presence has negligible impact on the system's frequency response  $g(i\omega)$ ) can block the modeling of essential eigenvalues away from the interpolation frequency. Related to these properties are two important characteristics of the Lanczos method:

(P3) those eigenvalues which are on the outer-edge of the spectrum of the Krylov matrix  $\Psi = (A - \sigma E)^{-1}E$  tend to be well approximated by the Lanczos method and (P4) the Lanczos method tends to converge to well-separated eigenvalues first. Corresponding to this last property, the Lanczos method typically does a poor job of identifying the multiplicity of identical (or nearly identical) eigenvalues.

To examine the impact of these properties on single-point approximations, a simple  $22^{nd}$ -order system is considered for the remainder of this section. The  $E$  matrix in this example is the identity matrix. As for  $A$ , 18 eigenvalues are in the neighborhood of 0 while the remaining four have an imaginary component of  $\pm 500$ . Of these, only four eigenvalues close to the imaginary axis ( $-0.21 \pm i$  and  $-0.2 \pm 500i$  where  $i = \sqrt{-1}$ ) play a significant role in the frequency response of the system (the two peaks on the system's frequency response in Figure 1 corresponds to these two eigenvalue pairs). Thus one expects to be able to model the original system of this example with  $k < 10$ .

Unfortunately for  $k < 18$ , the single-point Lanczos generated model about 0 (a standard choice for  $\sigma$ ) fails to reflect the actual system's peak in magnitude at  $\omega = 500$ , see Figure 1. Even though most are unimportant to modeling the system, the eigenvalues around  $\sigma = 0$  are almost perfectly approximated before the high-frequency eigenvalues make an appearance. Such behavior is consistent with the two Padé properties P1 and P2. From a Lanczos point of view, one must consider the Krylov subspaces  $\mathcal{K}_k(A^{-1}, A^{-1}b)$  and  $\mathcal{K}_k(A^{-T}, c)$ . The eigenvalues of  $A^{-1}$  are shown in Figure 2. Note that those eigenvalues of  $A$  which are near  $\sigma = 0$  have reciprocals which are spread out in the spectrum of  $A^{-1}$ . On the other hand, the high-frequency eigenvalues of  $A$  correspond to four eigenvalues of  $A^{-1}$  which are all basically zero. More importantly, those eigenvalues of the system near the imaginary ( $i\omega$ ) axis appear on the outer edge of the spectrum of  $A^{-1}$ . By property P3, the eigenvalues of  $A$  close to zero and the imaginary axis converge quickly in the reduced-order model. The desired high-frequency eigenvalues are also on the outer edge of  $A^{-1}$ 's spectrum, but their convergence is hindered by property P4. Until  $k$  becomes large, the Lanczos method sees the four nearly identical eigenvalues of  $A^{-1}$  at 0 (the high-frequency poles of the initial system) as a single, real pole. The Lanczos properties confirm that  $k$  must be large before the high-frequency behavior can be modeled.

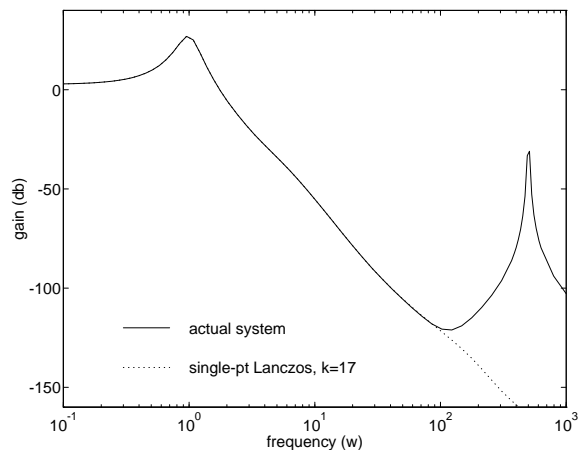


Figure 1: Frequency response of Example 1.

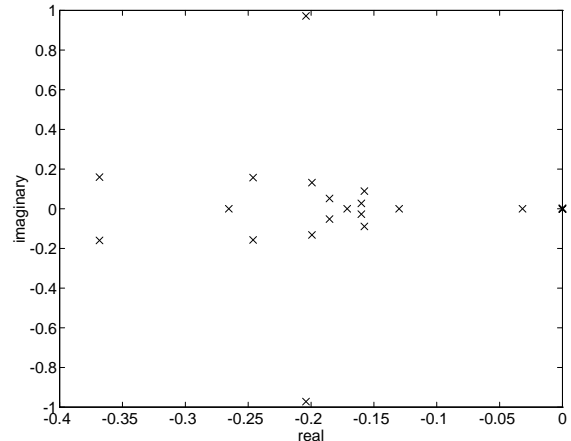


Figure 2: Eigenvalues of  $A^{-1}$ .

In general, we stress that a single-point Lanczos model will eventually model the frequency response of the system (e.g., when  $k \geq 18$ ), but the size of the reduced-order model may become large in doing so. The convergence of the single-point method is dependent on eigenvalues which are unimportant to the model. Moreover, these non-dominating eigenvalues appear in the reduced-order model. Yet if  $k$  is large and the model contains a large quantity of non-essential information, there is little value in obtaining the model. The above example is admittedly simple to clearly demonstrate these points. In §5, we obtain similar results with a real-world problem. Finally, note that one may improve the single-point results by using a different interpolation frequency. For instance, a good model arises for the above example when  $k = 12$  and  $\sigma = 20$ . However, it is not easy to locate such an interpolation point a priori. And in this example, even an optimal single-point interpolation falls short of the multi-point Padé approximation of §5.

## 4 Rational Lanczos algorithm

To avoid the difficulties inherent to single-point interpolation, one can turn to model reduction via multi-point Padé approximation [16, 32]. In multi-point approximation [2], the moments of the reduced-order model, the  $\hat{m}_{j_i}(\sigma_i)$  in (7), satisfy the moment matching condition (5). Every interpolation point,  $\sigma_i$ , is chosen to identify dynamics from a specific frequency range. One avoids trying to acquire information from a single, distant interpolation point. It is stressed once more that a Lanczos-type method is desired to avoid the numerical difficulties encountered in previous explicit moment matching methods [10].

To simplify the development of rational Lanczos, we assume in this section that a fixed number of moments ( $2\bar{j}$ ) are to be matched about each interpolation point. This restriction is not conducive to model reduction however and will be dropped in §5. We will also assume in this section that no breakdowns (divisions by zero) occur in the rational Lanczos algorithm. This second assumption is related in some ways to the first and will also be addressed in §5.

The variant of the Lanczos method employed to generate a reduced-order model  $\{\hat{E}, \hat{A}, \hat{b}, \hat{c}\}$  satisfying (5) is denoted the rational Lanczos algorithm as it was inspired by the rational Arnoldi method of [24, 25] for computing eigenvalues. In a rational Krylov method, the Krylov subspace is replaced with  $\text{span}\{r_0, \Psi_1 r_0, \dots, \Psi_{k-1} r_0\}$  where the  $\Psi_j$  are arbitrary rational functions in  $A$  and  $E$  [23]. The rational Lanczos method developed below actually computes two rational Krylov subspaces, yielding biorthogonal  $V_k$  and  $W_k$  in place of rational Arnoldi's orthogonal  $V_k$ . There are, however, numerous subtle differences between the two rational methods which are needed to insure that the oblique projector,  $\Pi_k = V_k W_k^T$ , of rational Lanczos yields a rational interpolant.

Strong similarities exist between the standard nonsymmetric Lanczos algorithm (Algorithm 1) and rational Lanczos (Algorithm 2). The key difference between the two lies in step 4 of Algorithm 2. In rational Lanczos, the matrix,  $(A - \sigma_i E)^{-1} E$ , multiplying a previous  $v$  vector varies with the interpolation point. Because this matrix is a function of  $\sigma_i$ , the union of several Krylov subspaces is computed (see Theorem 1 below). In fact, we will see that each of these Krylov subspaces corresponds to  $2\bar{j}$  moments about an interpolation frequency,  $\sigma_i$ .

We begin our analysis of Algorithm 2 by examining the case where  $k$  is a multiple of  $\bar{j}$ . This case involves the execution of step (A2.4.2) and corresponds to a change in the interpolation point from  $\sigma_i$  to  $\sigma_{i+1}$ . Note that  $\gamma_{k+1,k} v_{k+1} = r_k$  due to (A2.3). Then placing  $\tilde{r}_k$  from (A2.4.2) into the expression for  $r_k$  in (A2.6) yields

$$V_{k+1} \begin{bmatrix} \gamma_{1,k} \\ \vdots \\ \gamma_{k+1,k} \end{bmatrix} = (A - \sigma_{i+1} E)^{-1} b = \gamma_{1,0} (A - \sigma_{i+1} E)^{-1} (A - \sigma_1 E) V_k e_1 \quad (8)$$

since  $\gamma_{1,0} v_1 = (A - \sigma_1 E)^{-1} b$ . Note in (8) that the vector  $e_j$  is the  $j^{\text{th}}$  standard unit vector of appropriate length. Multiplying (8) on the right by  $(A - \sigma_{i+1} E)$  and rearranging the expression results in

$$A V_{k+1} \left( \begin{bmatrix} \gamma_{1,k} \\ \vdots \\ \gamma_{k+1,k} \end{bmatrix} - \gamma_{1,0} e_1 \right) = E V_{k+1} \left( \sigma_{i+1} \begin{bmatrix} \gamma_{1,k} \\ \vdots \\ \gamma_{k+1,k} \end{bmatrix} - \sigma_1 \gamma_{1,0} e_1 \right)$$

which can be rewritten as

$$(A - \sigma_1 E) V_{\bar{k}+1} \underbrace{\left( \begin{bmatrix} \gamma_{1,k} \\ \vdots \\ \gamma_{k+1,k} \\ 0 \end{bmatrix} - \gamma_{1,0} e_1 \right)}_{h_k} = E V_{\bar{k}+1} \underbrace{\left( \begin{bmatrix} \gamma_{1,k} \\ \vdots \\ \gamma_{k+1,k} \\ 0 \end{bmatrix} (\sigma_{i+1} - \sigma_1) \right)}_{k_k} \quad (9)$$

where  $\bar{k} = \bar{i}\bar{j}$ . When  $k$  is not a multiple of  $\bar{j}$ , step (A2.4.1) is executed and the next  $v$  vector computed is still associated with the interpolation point  $\sigma_i$ . For this case, placing the  $\tilde{r}_k$  of (A2.4.1) into the expression

**Algorithm 2** *Rational Lanczos [11]*

---

Input:  $r_0 = (A - \sigma_1 E)^{-1} b$  and  $q_0 = c$ ;  
For  $i = 1$  to  $\bar{i}$ ,  
  For  $j = 1$  to  $\bar{j}$ ,  
    (A2.1)  $k = (i - 1)\bar{j} + j$ ;  
    (A2.2)  $\gamma_{k,k-1} = \sqrt{|r_{k-1}^T q_{k-1}|}$ ;  
    (A2.3)  $v_k = (r_{k-1}/\gamma_{k,k-1})$  and  $w_k = \text{sign}(r_{k-1}^T q_{k-1}) \cdot (q_{k-1}/\gamma_{k,k-1})$ ;  
    (A2.4) if  $j < \bar{j}$  and  $i < \bar{i}$ ,  
      (A2.4.1)  $\tilde{r}_k = (A - \sigma_i E)^{-1} E v_k$  and  $\tilde{q}_k = E^T (A - \sigma_i E)^{-T} w_k$ ;  
      else if  $j = \bar{j}$  and  $i < \bar{i}$ ,  
      (A2.4.2)  $\tilde{r}_k = (A - \sigma_{i+1} E)^{-1} b$  and  $\tilde{q}_k = E^T (A - \sigma_{i+1} E)^{-T} c$ ;  
      else  
      (A2.4.3)  $\tilde{r}_k = (A - \sigma_1 E)^{-1} E v_{\bar{j}}$  and  $\tilde{q}_k = E^T (A - \sigma_1 E)^{-T} w_{\bar{j}}$ ;  
      end  
    (A2.5) if  $j \geq 2$  and  $k \neq \bar{i}\bar{j}$ ,  
      (A2.5.1)  $[\gamma_{1,k} \ \dots \ \gamma_{k,k}]^T = [0 \ \dots \ 0 \ w_{k-1}^T \tilde{r}_k \ w_k^T \tilde{r}_k]^T$  and  $[\beta_{1,k} \ \dots \ \beta_{k,k}]^T = [0 \ \dots \ 0 \ v_{k-1}^T \tilde{q}_k \ v_k^T \tilde{q}_k]^T$ ;  
      else  
      (A2.5.2)  $[\gamma_{1,k} \ \dots \ \gamma_{k,k}]^T = W_k^T \tilde{r}_k$  and  $[\beta_{1,k} \ \dots \ \beta_{k,k}]^T = V_k^T \tilde{q}_k$ ;  
      end  
    (A2.6)  $r_k = \tilde{r}_k - V_k [\gamma_{1,k} \ \dots \ \gamma_{k,k}]^T$  and  $q_k = \tilde{q}_k - W_k [\beta_{1,k} \ \dots \ \beta_{k,k}]^T$ ;  
  end  
end  
 $v_{\bar{k}+1} = (r_{\bar{k}}/\gamma_{\bar{k}+1,\bar{k}})$  where  $\gamma_{\bar{k}+1,\bar{k}} = \sqrt{|r_{\bar{k}}^T q_{\bar{k}}|}$  and  $\bar{k} = \bar{i}\bar{j}$ .

---

for  $r_k$  in (A2.6) yields

$$V_{k+1} \begin{bmatrix} \gamma_{1,k} \\ \vdots \\ \gamma_{k+1,k} \end{bmatrix} = (A - \sigma_i E)^{-1} E V_k e_k. \quad (10)$$

Multiplying (10) on the left by  $(A - \sigma_i E)$  produces

$$E V_k e_k = (A - \sigma_i E) V_{k+1} \begin{bmatrix} \gamma_{1,k} \\ \vdots \\ \gamma_{k+1,k} \end{bmatrix}$$

which can be rewritten as

$$(A - \sigma_1 E) V_{\bar{k}+1} \underbrace{\begin{bmatrix} \gamma_{1,k} \\ \vdots \\ \gamma_{k+1,k} \\ 0 \end{bmatrix}}_{h_k} = E V_{\bar{k}+1} \underbrace{\left( \begin{bmatrix} \gamma_{1,k} \\ \vdots \\ \gamma_{k+1,k} \\ 0 \end{bmatrix} (\sigma_i - \sigma_1) + e_k \right)}_{k_k}. \quad (11)$$

Combining all  $\bar{k}$  steps of Algorithm 2 yields

$$(A - \sigma_1 E) V_{\bar{k}+1} H_{\bar{k}+1,\bar{k}} = E V_{\bar{k}+1} K_{\bar{k}+1,\bar{k}} \quad (12)$$

where the columns,  $h_k$  and  $k_k$ , of  $H_{\bar{k}+1,\bar{k}}$  and  $K_{\bar{k}+1,\bar{k}}$  are defined via (9) and (11). Specifically, columns  $\bar{j}, 2\bar{j}, \dots, (\bar{i} - 1)\bar{j}$  of  $H_{\bar{k}+1,\bar{k}}$  and  $K_{\bar{k}+1,\bar{k}}$  fit the form of (9) while the remaining columns satisfy (11).



The matrices  $H_{\bar{k}+1, \bar{k}}$  and  $K_{\bar{k}+1, \bar{k}}$  are upper-Hessenberg. The elements,  $\gamma$ , of these matrices are computed in Algorithm 2 so as to enforce a biorthogonality condition, i.e.,  $W_{\bar{k}}^T V_{\bar{k}} = I$ . However, as indicated by step (A2.5.1), a majority of the elements above the diagonals of  $H_{\bar{k}+1, \bar{k}}$  and  $K_{\bar{k}+1, \bar{k}}$  are typically zero and thus need not be computed in theory. To be precise,  $H_{\bar{k}+1, \bar{k}}$  is tridiagonal except for off-tridiagonal fill-in occurring in those columns where (using the notation of Algorithm 2)  $j = 1$  or  $j = 2$ . The structure of  $H_{\bar{k}+1, \bar{k}}$  and  $K_{\bar{k}+1, \bar{k}}$  follows from Lemma 1 (see appendix) and is a generalization of the three term recurrences present in the standard nonsymmetric Lanczos algorithm. For example if  $k = \bar{j} + 2$ , the element  $\gamma_{t, k} = w_t^T \tilde{r}_k = w_t^T (A - \sigma_2 E)^{-1} E v_{\bar{j}+2}$  is zero for  $t < k - 1$ . This last fact is due to the biorthogonality of  $V_k$  and  $W_k$  and also by (37) of Lemma 1, i.e.,  $w_t^T (A - \sigma_2 E)^{-1} \in \mathcal{K}_1 (E^T (A - \sigma_2 E)^{-T}, E^T (A - \sigma_2 E)^{-T} c) \cup \mathcal{K}_{\bar{j}} (E^T (A - \sigma_1 E)^{-T}, c)$  if  $t < k - 1$ .

Special mention should also be given to the  $\bar{k}^{th}$  columns of  $H_{\bar{k}+1, \bar{k}}$  and  $K_{\bar{k}+1, \bar{k}}$ . Due to step (A2.4.3), the  $\bar{k}^{th}$  column satisfies the general form of (11) with  $\sigma_i = \sigma_1$ . Thus  $k_{\bar{k}} = [e_{\bar{j}}^T \ 0]^T$  so that  $V_{\bar{k}+1} K_{\bar{k}+1, \bar{k}} = V_{\bar{k}} K_{\bar{k}, \bar{k}}$ . Making use of this last fact when multiplying (12) on the left by  $W_{\bar{k}}^T (A - \sigma_1 E)^{-1}$  yields

$$H_{\bar{k}, \bar{k}} = W_{\bar{k}}^T (A - \sigma_1 E)^{-1} E V_{\bar{k}} K_{\bar{k}, \bar{k}} \quad (13)$$

where  $\bar{k} = \bar{j}$ . Expressions (12) and (13) serve as the initial relations between the projector  $V_{\bar{k}} W_{\bar{k}}^T$  and  $E$  and  $A$ .

Under the assumption that  $E$  is invertible, the relations (12) and (13) were utilized in [11] to argue for

$$\hat{A} = K_{\bar{k}, \bar{k}} + \sigma_1 H_{\bar{k}, \bar{k}}, \quad \hat{E} = H_{\bar{k}, \bar{k}}, \quad \hat{b} = W_{\bar{k}}^T (A - \sigma_1 E)^{-1} b, \quad \text{and} \quad \hat{c}^T = c^T V_{\bar{k}} K_{\bar{k}, \bar{k}} \quad (14)$$

in the reduced-order model, (2). In this paper, the assumption that  $E$  is invertible is dropped. Furthermore, the remainder of this section combined with several lemmas provided in the appendix actually proves that (14) corresponds to a multi-point Padé approximation of (1) (i.e., (5) is satisfied by the model selected by (14)). To arrive at this final result, we begin by obtaining a relationship between the rational Lanczos projector  $V_k W_k^T$  and Krylov subspaces.

**Theorem 1** *If  $V_k$  and  $W_k$  are the results of the first  $k$  steps of Algorithm 2 with  $1 \leq k \leq \bar{k}$  then*

$$\text{colsp}(V_k) = \left\{ \mathcal{K}_{k - \bar{j}(i-1)} \left( (A - \sigma_i E)^{-1} E, (A - \sigma_i E)^{-1} b \right) \bigcup_{l=1}^{i-1} \mathcal{K}_{\bar{j}} \left( (A - \sigma_l E)^{-1} E, (A - \sigma_l E)^{-1} b \right) \right\} \quad (15)$$

where  $i - 1$  is the quotient of  $k/\bar{j}$ . Correspondingly,

$$\text{colsp}(W_k) = \left\{ \mathcal{K}_{k - \bar{j}(i-1)} \left( E^T (A - \sigma_i E)^{-T}, c_i \right) \bigcup_{l=1}^{i-1} \mathcal{K}_{\bar{j}} \left( E^T (A - \sigma_l E)^{-T}, c_l \right) \right\} \quad (16)$$

where the vector  $c_l$  is  $c$  if  $l = 1$  and  $(A - \sigma_l E)^{-T} c$  otherwise.

*Proof:* We prove (15) via induction. The result clearly holds for  $k = 1$  since  $\gamma_{1,0} v_1 = (A - \sigma_1 E)^{-1} b$  by choice. Assume

$$\text{colsp}(V_{k-1}) = \left\{ \mathcal{K}_{k - \bar{j}(i-1) - 1} \left( (A - \sigma_i E)^{-1} E, (A - \sigma_i E)^{-1} b \right) \bigcup_{l=1}^{i-1} \mathcal{K}_{\bar{j}} \left( (A - \sigma_l E)^{-1} E, (A - \sigma_l E)^{-1} b \right) \right\}. \quad (17)$$

For  $k > 1$ , steps (A2.4) and (A2.6) in the  $k^{th}$  iteration of Algorithm 2 yield

$$\gamma_{k, k-1} v_k = (A - \sigma_i E)^{-1} E \tilde{v}_{k-1} + \sum_{t=1}^{k-1} \gamma_{t, k-1} v_t \quad (18)$$

where  $\tilde{v}_{k-1} = b$  if (A2.4.2) is executed and  $\tilde{v}_{k-1} = v_{k-1}$  if (A2.4.1) is executed. Under the assumption (17),  $\tilde{v}_{k-1} = \tilde{v}_{k-1}^{(i)} + \tilde{v}_{k-1}^{(l)}$  where

$$\tilde{v}_{k-1}^{(i)} \in \mathcal{K}_{k - \bar{j}(i-1) - 1} \left( (A - \sigma_i)^{-1} E, (A - \sigma_i E)^{-1} b \right) \quad \text{and} \quad \tilde{v}_{k-1}^{(l)} \in \bigcup_{l=1}^{i-1} \mathcal{K}_{\bar{j}} \left( (A - \sigma_l)^{-1} E, (A - \sigma_l)^{-1} b \right). \quad (19)$$

Thus (18) can be rewritten as

$$\gamma_{k,k-1}v_k = (A - \sigma_i E)^{-1} E \tilde{v}_{k-1}^{(i)} + (A - \sigma_i E)^{-1} E \tilde{v}_{k-1}^{(l)} + \sum_{t=1}^{k-1} \gamma_{t,k-1}v_t. \quad (20)$$

The vector  $(A - \sigma_i E)^{-1} E \tilde{v}_{k-1}^{(i)}$  lies in  $\mathcal{K}_{k-\bar{j}(i-1)}((A - \sigma_i E)^{-1} E, (A - \sigma_i E)^{-1} b)$  since one is simply multiplying some power of  $(A - \sigma_i E)^{-1} E$  again by  $(A - \sigma_i E)^{-1} E$ . Elsewhere, one must use Lemma 1 of the appendix to show that  $(A - \sigma_i E)^{-1} E \tilde{v}_{k-1}^{(l)} \in \{\mathcal{K}_1((A - \sigma_i E)^{-1} E, (A - \sigma_i E)^{-1} b) \cup_{l=1}^{\bar{j}-1} \mathcal{K}_{\bar{j}}((A - \sigma_l)^{-1} E, (A - \sigma_l)^{-1} b)\}$ . Lastly,  $\sum_{t=1}^{k-1} \gamma_{t,k-1}v_t \in \text{colsp}(V_{k-1})$  where  $\text{colsp}(V_{k-1})$  is defined by (17). Combining these last three facts with (20) implies that  $V_k$  satisfies (15). The portion of the proof corresponding to (16) is the dual to that presented above.  $\square$

In [29], it is shown that an oblique Krylov projector leads to a Padé approximation about a single interpolation point. Except for certain technicalities which are handled in Lemmas 2 and 3 of the appendix, an argument in [29] can be generalized to prove that an oblique projector corresponding to a union of multiple Krylov subspaces leads to a multi-point Padé approximation. This result is given in the following theorem.

**Theorem 2** *Let the  $j^{\text{th}}$  moments of the original and reduced order systems about the interpolation point  $\sigma_i$  be  $m_j(\sigma_i) = c^T \{(A - \sigma_i E)^{-1} E\}^j (A - \sigma_i E)^{-1} b$  and  $\hat{m}_j(\sigma_i) = \hat{c}^T \{(\hat{A} - \sigma_i \hat{E})^{-1} \hat{E}\}^j (\hat{A} - \sigma_i \hat{E})^{-1} \hat{b}$  respectively. If  $\hat{A} = K_{k,k} + \sigma_1 H_{k,k}$ ,  $\hat{E} = H_{k,k}$ ,  $\hat{b} = W_k^T (A - \sigma_1 E)^{-1} b$  and  $\hat{c}^T = c^T V_k K_{k,k}$  where  $H_{k+1,k}$ ,  $K_{k+1,k}$ ,  $V_{k+1}$  and  $W_{k+1}$  are the results of Algorithm 2 with  $k = \bar{k}$  (i.e., Algorithm 2 is run to completion), then  $m_j(\sigma_i) = \hat{m}_j(\sigma_i)$  for  $i = 1, 2, \dots, \bar{i}$  and  $j = 0, 1, 2, \dots, 2\bar{j} - 1$ .*

*Proof:* Corresponding to the two-sided nature of nonsymmetric Lanczos methods, it is helpful to split up the expression for  $m_j(\sigma_i)$  as

$$m_j(\sigma_i) = \left[ c^T \{(A - \sigma_i E)^{-1} E\}^{j_1} \right] \left[ \{(A - \sigma_i E)^{-1} E\}^{j_2} (A - \sigma_i E)^{-1} b \right] \quad (21)$$

where  $j_1 = \lceil \frac{\bar{j}}{2} \rceil$  and  $j_2 = \lfloor \frac{\bar{j}}{2} \rfloor$ . If  $\Pi_k = V_k W_k^T$  is a biorthogonal projector,  $v \in \text{colsp}(V_k)$  and  $w \in \text{colsp}(W_k)$ , then  $\Pi_k v = v$  and  $w^T \Pi_k = w^T$ . Thus (21) can be rewritten as

$$m_j(\sigma_i) = \left[ c^T V_k \{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j_1} \right] \left[ \{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j_2} W_k (A - \sigma_i E)^{-1} b \right] \quad (22)$$

by the properties of the biorthogonal projector and Theorem 1. From (22),  $m_j(\sigma_i)$  is also the the  $j^{\text{th}}$  moment about  $\sigma_i$  of the restriction of (1) by  $\Pi_k$ . We must now simply show that this moment of the restricted system takes on the form specified by  $\hat{m}_j(\sigma_i)$ . Two lemmas from the appendix will be needed to relate the matrix  $W_k^T (A - \sigma_i E)^{-1} E V_k$  of (22) to the matrices  $H_{k,k}$  and  $K_{k,k}$  appearing in  $\hat{m}_j(\sigma_i)$ .

We proceed by concentrating on the right hand side of (22), i.e.,

$$\begin{aligned} & \{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j_2} W_k^T (A - \sigma_i E)^{-1} b \\ &= \{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j_2} W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) (A - \sigma_1 E)^{-1} b \\ &= \{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j_2} W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) V_k W_k^T (A - \sigma_1 E)^{-1} b. \end{aligned} \quad (23)$$

By using (40) of Lemma 2 and recalling that  $(A - \sigma_1 E)^{-1} b = \gamma_{1,0}v_1$ , (23) can be rewritten as

$$\{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j_2} [K_{k,k} - W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) r_k e_k^T (\sigma_1 - \sigma_i)] J_{k,k} e_1 \gamma_{1,0}$$

where  $J_{k,k} = [K_{k,k} + H_{k,k}(\sigma_1 - \sigma_i)]^{-1}$ . This most recent expression can be further simplified to

$$\{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j_2} K_{k,k} J_{k,k} e_1 \gamma_{1,0} \quad (24)$$

due to (45) in Lemma 3. Applying (41) of Lemma 2 once to (24) yields that (23) is equivalent to

$$\begin{aligned} & \{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j_2-1} \\ & [K_{k,k} J_{k,k} H_{k,k} + W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) r_k e_k^T \{I + (\sigma_1 - \sigma_i) J_{k,k} H_{k,k}\}] J_{k,k} e_1 \gamma_{1,0} \end{aligned}$$

which can again be further simplified via (45) to

$$\{W_k^T(A - \sigma_i E)^{-1} E V_k\}^{j_2-1} K_{k,k} J_{k,k} H_{k,k} J_{k,k} e_1 \gamma_{1,0}.$$

After the repeated use of (41) and (45) in a similar fashion, one finally obtains

$$\{W_k^T(A - \sigma_i E)^{-1} E V_k\}^{j_2} W_k^T(A - \sigma_i E)^{-1} b = K_{k,k} \{J_{k,k} H_{k,k}\}^{j_2} J_{k,k} W_k^T(A - \sigma_1 E)^{-1} b$$

so that

$$m_j(\sigma_i) = c^T V_k \{W_k^T(A - \sigma_i E)^{-1} E V_k\}^{j_1} K_{k,k} \{J_{k,k} H_{k,k}\}^{j_2} J_{k,k} W_k^T(A - \sigma_1 E)^{-1} b. \quad (25)$$

We now concentrate on the left side of (22). Applying (41) of Lemma 2 to (25) yields

$$m_j(\sigma_i) = c^T V_k \{W_k^T(A - \sigma_i E)^{-1} E V_k\}^{j_1-1} \left[ K_{k,k} J_{k,k} H_{k,k} + \right. \\ \left. W_k^T(A - \sigma_i E)^{-1} (A - \sigma_1 E) r_k e_k^T \{I + (\sigma_1 - \sigma_i) J_{k,k} H_{k,k}\} \right] \{J_{k,k} H_{k,k}\}^{j_2} J_{k,k} W_k^T(A - \sigma_1 E)^{-1} b.$$

which becomes

$$c^T V_k \{W_k^T(A - \sigma_i E)^{-1} E V_k\}^{j_1-1} K_{k,k} \{J_{k,k} H_{k,k}\}^{j_2+1} J_{k,k} W_k^T(A - \sigma_1 E)^{-1} b$$

since the residual vector,  $r_k$ , drops out due to (46). The repeated application of (41) and (46) in an analogous manner yields

$$m_j(\sigma_i) = c^T V_k K_{k,k} \{J_{k,k} H_{k,k}\}^j J_{k,k} W_k^T(A - \sigma_1 E)^{-1} b. \quad (26)$$

The right side of (26) is in fact  $\hat{m}_j(\sigma_i)$  given the definitions of  $\hat{A}$ ,  $\hat{E}$ ,  $\hat{b}$ ,  $\hat{c}$  and  $J_{k,k} = (\hat{A} - \sigma_i \hat{E})^{-1}$ .  $\square$

## 5 Model reduction with the rational Lanczos algorithm

Using the rational Lanczos method of §4, one can model the  $22^{nd}$  order problem of §3 with much smaller values of  $k$  than required by the single-point interpolation about 0. For instance, consider interpolating this system about the points  $\{.1, 1, 10, 100, 1000\}$ . Matching four moments about  $\sigma_4 = 100$  and two moments about each of the other four points generates a  $6^{th}$  order rational interpolant. Figure 3 demonstrates that the frequency response of the  $6^{th}$  order model is nearly identical to the frequency response of the original,  $22^{nd}$  order system. Recall that the single point interpolation of §3 did not yield a response with such accuracy until  $k$  grew to be 18. By utilizing multiple interpolation points, the size of  $k$  was reduced from approximately  $n$  to a value consistent with the amount of important dynamics in the original system. Of course, selecting a proper combination of interpolation frequencies,  $\sigma_i$ , and the number of moments,  $2\bar{j}$ , to be matched about each  $\sigma_i$  is by no means a trivial matter. In this section, a technique is developed for implementing the rational Lanczos as a model reduction tool.

For ease of computation and for lack of better application specific information, the  $\bar{i}$  interpolation points are fixed in this paper with a log-linear spacings over a frequency range,  $\omega_{min}$  to  $\omega_{max}$ , specified by the user. We refer the reader to [18] for a discussion of point selection in the context of rational interpolation of the frequency response. The interpolation points are spread over the positive real axis with  $\sigma_1 = \omega_{min}$  and  $\sigma_{\bar{i}} = \omega_{max}$ . The moments generated about each  $\sigma_i$  tend to yield information pertaining to the original system's response in the neighborhood of the frequency  $\sigma_i$ . One way of justifying this last statement is to examine the eigenvalues of the reduced-order model in an approach analogous to §3. For a  $\sigma_i$  between  $\omega_{min}$  and  $\omega_{max}$ , those eigenvalues of (1) with imaginary components  $\gg \sigma_i$  appear in the spectrum of  $(A - \sigma_i E)^{-1} E$  as a cluster at 0. Those eigenvalues with an imaginary component  $\ll \sigma_i$  appear in the spectrum of  $(A - \sigma_i E)^{-1} E$  as a cluster at  $1/\sigma_i$ . The remaining eigenvalues of the original system tend to be well spread in the spectrum of  $(A - \sigma_i E)^{-1} E$  with those eigenvalues near the imaginary axis on the edge of the spectrum. By generalizing the Lanczos properties discussed in §3, one can expect that the inclusion of moments about  $\sigma_i$  leads to a reduced-order model that has some eigenvalues which approximate those of

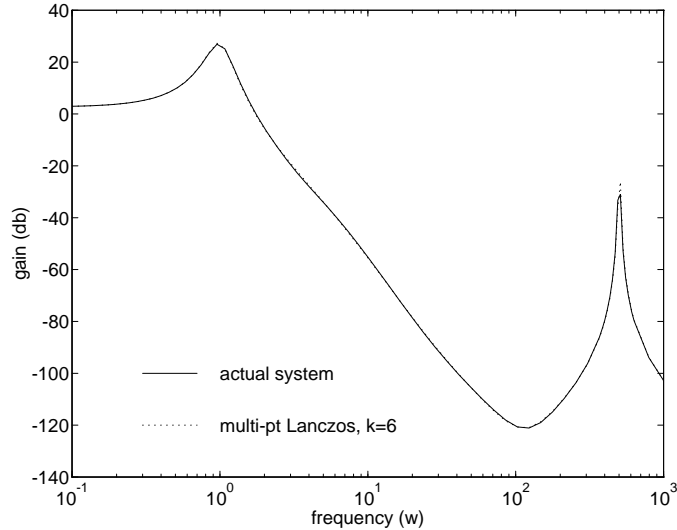


Figure 3: Frequency response for Example 1.

$(A - \lambda E)$  in the neighborhood of  $\pm i\sigma_i$ . Modeling the eigenvalues in the proximity of  $\pm i\sigma_i$  tends to in turn lead to a reduced-order model whose transfer function approximates  $g(s)$  for frequencies near  $\sigma_i$ .

Choosing real interpolation points leads to a rational Lanczos algorithm which avoids complex computations (assuming the original system (1) is real). However, complex (imaginary) points may be preferred since one is in fact interested in interpolating  $g(s)$  along the imaginary axis. The rational Lanczos algorithm is not restricted to real  $\sigma_i$ . For example, one might combine Algorithm 2 and the methods of [25] to arrive at a complex interpolation point, rational Lanczos method. The selection and implementation of complex interpolation points will be discussed in a forthcoming paper.

Besides placing the interpolation points, one must also be concerned with how many moments are to be matched about each of these points. In §4, the number of moments about each interpolation point was fixed a priori at  $(2\bar{j})$ . The first  $\bar{j}$  rational Lanczos iterations corresponded to the interpolation point  $\sigma_1$ , the next  $\bar{j}$  iterations were associated with  $\sigma_2$ , etc. Although this approach allowed for a simpler development of Theorems 1 and 2, it is not preferred for acquiring an acceptable reduced-order model. Rather, we would hope to choose an interpolation point in the  $(k+1)^{st}$  rational Lanczos iteration which yields in some sense the greatest improvement between the  $k^{th}$  and  $(k+1)^{st}$  order models<sup>‡</sup>. One should no longer simply perform all of the rational Lanczos iterations corresponding a given interpolation point consecutively. To begin to formalize these statements, consider a somewhat more generalized pair of residuals

$$r_k^{(i)} = (A - \sigma_i E)^{-1} E v^{(i)} - V_k W_k^T (A - \sigma_i E)^{-1} E v^{(i)}, \quad v^{(i)} \in \text{colsp}(V_k), \quad (27)$$

$$q_k^{(i)} = E^T (A - \sigma_i E)^{-T} w^{(i)} - W_k V_k^T E^T (A - \sigma_i E)^{-T} w^{(i)}, \quad w^{(i)} \in \text{colsp}(W_k), \quad (28)$$

where the superscripts  $(i)$  are added to explicitly denote the dependence of  $r_k^{(i)}$  and  $q_k^{(i)}$  on the choice of  $\sigma_i$ . The  $(k+1)^{st}$  iteration (and also the vectors  $v_{k+1}$  and  $w_{k+1}$ ) will be said to correspond to the specific interpolation point  $\sigma_{i_{k+1}^*} \in \{\sigma_1, \dots, \sigma_{\bar{i}}\}$  if  $v_{k+1}$  and  $w_{k+1}$  lie parallel to  $r_k^{(i_{k+1}^*)}$  and  $q_k^{(i_{k+1}^*)}$  respectively. In performing the  $(k+1)^{st}$  iterations (computing  $v_{k+1}$  and  $w_{k+1}$ ), one can thus choose from among  $\bar{i}$  residual pairs. This ability to select from among  $\bar{i}$  residual pairs per iteration reflects the freedom provided by multiple interpolation points. The goal is to choose the residual pair in the  $(k+1)^{st}$  iteration so that the  $(k+1)^{st}$  order model is in some sense the best possible improvement over the  $k^{th}$  order model.

<sup>‡</sup>The  $k^{th}$  order reduced-order model is defined in this paper to be the restriction of (1) by the projector  $V_k W_k^T$ . For the special case where  $k = \bar{k}$ , (13) holds and thus the reduced-order model can equivalently be defined via (14). It is stressed that one cannot in general define the reduced-order model in the form of (14) for  $k < \bar{k}$ .

Motivated by the discussion of the previous paragraph, Algorithm 3 is proposed as a version of the rational Lanczos algorithm suited for the model reduction problem. Algorithm 3 is a twist on Algorithm 2 which does not require that all of the moments corresponding to a given interpolation point be computed consecutively. Nor does Algorithm 3 demand that the same number of moments be computed about each interpolation point. Rather, Algorithm 3 attempts to select from among the  $\bar{i}$  interpolation points to acquire an acceptable reduced-order model. Its ability to do so is demonstrated with an example at the end of this section. However, we first consider several features of Algorithm 3. The values of  $v^{(i)}$  and  $w^{(i)}$  in (27) and (28) must be specified. The criterion used to select from among the  $\bar{i}$  interpolation points in each iteration (A3.1) must be developed. And the structure of the  $\bar{k}^{th}$  order model generated by Algorithm 3 must be presented.

---

**Algorithm 3** *Rational Lanczos for model reduction*

---

For  $i = 1$  to  $\bar{i}$ ,

Input:  $r_0^{(i)} = \tilde{r}_0^{(i)} = (A - \sigma_i E)^{-1} b$  and  $q_0^{(i)} = \tilde{q}_0^{(i)} = c$  and  $\tau_i = 1$ ;

end

For  $k = 1$  to  $\bar{k}$ ,

(A3.1) set  $i_k^*$  to be the value of  $i = 1, 2, \dots, \bar{i}$  which maximizes  $\sigma_i^2 |r_{k-1}^{(i)T} q_{k-1}^{(i)}|$ ;

(A3.2)  $\gamma_{k,k-1}^{(i_k^*)} = \sqrt{|r_{k-1}^{(i_k^*)T} q_{k-1}^{(i_k^*)}|}$ ;

(A3.3)  $v_k = (r_{k-1}^{(i_k^*)} / \gamma_{k,k-1}^{(i_k^*)})$  and  $w_k = \text{sign}(r_{k-1}^{(i_k^*)T} q_{k-1}^{(i_k^*)}) \cdot (q_{k-1}^{(i_k^*)} / \gamma_{k,k-1}^{(i_k^*)})$ ;

(A3.4)  $\tilde{r}_k^{(i_k^*)} = (A - \sigma_{i_k^*} E)^{-1} E v_k$  and  $\tilde{q}_k^{(i_k^*)} = E^T (A - \sigma_{i_k^*} E)^{-T} w_k$ ;

(A3.5) for  $t = 1$  to  $k$ ,

if  $t \geq \tau_{i_k^*}$ ,

$\gamma_{t,k}^{(i_k^*)} = w_t^T \tilde{r}_k^{(i_k^*)}$  and  $\beta_{t,k}^{(i_k^*)} = v_t^T \tilde{q}_k^{(i_k^*)}$ ;

else

$\gamma_{t,k}^{(i_k^*)} = 0$  and  $\beta_{t,k}^{(i_k^*)} = 0$ ;

end

end

(A3.6)  $r_k^{(i_k^*)} = \tilde{r}_k^{(i_k^*)} - \sum_{t=\tau_{i_k^*}}^k v_t \gamma_{t,k}^{(i_k^*)}$  and  $q_k^{(i_k^*)} = \tilde{q}_k^{(i_k^*)} - \sum_{t=\tau_{i_k^*}}^k w_t \beta_{t,k}^{(i_k^*)}$ ;

(A3.7)  $\tau_{i_k^*} = k$ ;

(A3.8) for  $i \neq i_k^*$ ,

if  $k = 1$ ,  $q_0^{(i)} = \tilde{q}_0^{(i)} = E^T (A - \sigma_i E)^{-T} w_1$ ; end

$r_k^{(i)} = r_{k-1}^{(i)} - v_k w_k^T r_{k-1}^{(i)}$  and  $q_k^{(i)} = q_{k-1}^{(i)} - w_k v_k^T q_{k-1}^{(i)}$ ;

$\tilde{r}_k^{(i)} = \tilde{r}_{k-1}^{(i)}$  and  $\tilde{q}_k^{(i)} = \tilde{q}_{k-1}^{(i)}$ ;

$\begin{bmatrix} \gamma_{1,k}^{(i)} & \dots & \gamma_{k,k}^{(i)} \end{bmatrix} = \begin{bmatrix} \gamma_{1,k-1}^{(i)} & \dots & \gamma_{k-1,k-1}^{(i)} & w_k^T \tilde{r}_{k-1}^{(i)} \end{bmatrix}$ ;

end

end

$v_{\bar{k}+1} = (r_{\bar{k}}^{(i_1^*)} / \gamma_{\bar{k}+1,\bar{k}}^{(i_1^*)})$  where  $\gamma_{\bar{k}+1,\bar{k}}^{(i_1^*)} = \sqrt{|r_{\bar{k}}^{(i_1^*)T} q_{\bar{k}}^{(i_1^*)}|}$ .

---

Algorithm 3 specifies that the vectors  $v^{(i)}$  and  $w^{(i)}$  in (27) and (28) be the rightmost columns of  $V_k$  and  $W_k$  respectively which also correspond to the interpolation point  $\sigma_i$ . This choice insures that the order- $\bar{k}$  model generated by Algorithm 3 is still a multi-point Padé approximation of the original system. Specifically,  $2\bar{j}_i$  moments are matched about  $\sigma_i$  where  $\bar{j}_i$  is the total number of times  $i_k^* = i$  in step 1 of Algorithm 3 for  $k = 1, \dots, \bar{k}$ . A proof of this statement will not be provided as it is simply a more tedious version of that which is already in §4. All of the results developed in §4 can be adapted for Algorithm 3; only the quantity and ordering of the moments computed about each  $\sigma_i$  vary from before.

The prescribed choice of  $v^{(i)}$  and  $w^{(i)}$  also leads to one other interesting point. Note that  $\bar{\nu}$  pairs of residuals,  $r_k^{(i)}$  and  $q_k^{(i)}$ , are carried from one iteration to the next. But only one pair of residuals is actually incorporated into the projector in each iteration (the pair which hopefully leads to the greatest improvement between the  $k^{\text{th}}$  and  $(k+1)^{\text{st}}$  order models). Only two new residuals are computed in steps (A3.4) through (A3.6) to replace the pair selected in (A3.1). The other residual pairs can be carried over into subsequent iterations after cheap updates in step (A3.8). Hence only one pair of residuals (only the solutions to two linear equations) need be computed per iteration. The other residuals pairs can be carried over from one iteration to the next because the values of  $v^{(i)}$  and  $w^{(i)}$  used in computing the  $i^{\text{th}}$  residual pair depend on  $\sigma_i$  but not  $k$ . The vector  $v^{(i)}$  used in  $r_k^{(i)}$  is chosen to be the most recent column of  $V_k$  which was also formed via multiplication by  $(A - \sigma_i E)^{-1} E$ .

One of the most important components of Algorithm 3 is (A3.1). Based on the  $\bar{\nu}$  values of  $\sigma_i^2 |r_{k-1}^{(i)T} q_{k-1}^{(i)}|$ , an interpolation point is selected for use in the  $k^{\text{th}}$  iteration. A justification for this selection criterion arises out of the following result.

**Theorem 3** *Let  $V_k W_k^T$  be the projector formed via the first  $k < \bar{k}$  iterations of Algorithm 3 and assume  $\sigma_i$  was chosen to be the desired interpolation point of (A3.1) in  $j_i$  of these previous  $k$  iterations. If  $i \neq i_1^*$ , then  $q_k^{(i)T} r_k^{(i)}$  is proportional to the difference between  $m_{2j_i+1}(\sigma_i) = c^T \{(A - \sigma_i E)^{-1} E\}^{2j_i+1} (A - \sigma_i E)^{-1} b$  and  $\hat{m}_{2j_i+1}(\sigma_i)$ . Otherwise,  $q_k^{(i)T} r_k^{(i)}$  is proportional to  $m_{2j_i^*}(\sigma_{i_1^*}) - \hat{m}_{2j_i^*}(\sigma_{i_1^*})$ .*

*Proof:* The distinction between  $i = i_1^*$  and  $i \neq i_1^*$  is a minor technicality arising out of the fact that the first vector in  $W_k$  needs to be  $c$  rather than  $E^T (A - \sigma_{i_1^*})^{-T} c$ . A proof is only provided for  $i \neq i_1^*$ .

Given the definitions of  $r_k^{(i)}$  and  $q_k^{(i)}$  in (27) and (28),

$$q_k^{(i)T} r_k^{(i)} = w^{(i)T} \{(A - \sigma_i E)^{-1} E\}^2 v^{(i)} - w^{(i)T} (A - \sigma_i E)^{-1} E V_k W_k^T (A - \sigma_i E)^{-1} E v^{(i)}. \quad (29)$$

From Theorem 1, it is known that  $v^{(i)}$  is a linear combination of vectors which serve as a basis for the union of  $\bar{\nu}$  Krylov subspaces. In particular, because  $\sigma_i$  was chosen to be the interpolation point in  $j_i$  of the previous  $k$  iterations,  $v^{(i)}$  can be written as  $\bar{v}^{(i)} + \tilde{v}^{(i)}$  where

$$\bar{v}^{(i)} = \gamma_{\Pi}^{(i)-1} \{(A - \sigma_i)^{-1} E\}^{j_i-1} (A - \sigma_i E)^{-1} b \in \text{colsp}(V_k)$$

and

$$\tilde{v}^{(i)} \in \mathcal{K}_{j_i-1}((A - \sigma_i E)^{-1} E, (A - \sigma_i E)^{-1} b) \bigcup_{l \neq i} \mathcal{K}_{j_l}((A - \sigma_l E)^{-1} E, (A - \sigma_l E)^{-1} b) \subset \text{colsp}(V_k).$$

Combining the expression for  $\tilde{v}^{(i)}$  with Lemma 1 yields  $(A - \sigma_i E)^{-1} E \tilde{v}^{(i)} \in \text{colsp}(V_k)$  so that  $V_k W_k^T (A - \sigma_i E)^{-1} E v^{(i)} = V_k W_k^T \gamma_{\Pi}^{(i)-1} (A - \sigma_i)^{-1} E \bar{v}^{(i)} + (A - \sigma_i E)^{-1} E \tilde{v}^{(i)}$ . Although not essential to the proof, an inspection of Algorithm 3 shows that  $\gamma_{\Pi}^{(i)} = \prod_{t=1}^k \tilde{\gamma}_{t,t-1}$  where  $\tilde{\gamma}_{t,t-1} = \gamma_{t,t-1}$  if  $i_t^* = i$  and  $\tilde{\gamma}_{t,t-1} = 0$  otherwise. Similarly, one can write  $w^{(i)}$  as  $\bar{w}^{(i)} + \tilde{w}^{(i)}$  where  $\bar{w}^{(i)} = \beta_{\Pi}^{(i)-1} \{E^T (A - \sigma_i E)^{-T}\}^{j_i} c$ ,  $\tilde{w}^{(i)T} (A - \sigma_i)^{-1} E V_k W_k^T = \tilde{w}^{(i)T} (A - \sigma_i)^{-1} E$ , and  $\beta_{\Pi}^{(i)} = \pm \gamma_{\Pi}^{(i)}$ . Using all of these facts in (29) yields

$$\begin{aligned} q_k^{(i)T} r_k^{(i)} &= \bar{w}^{(i)T} \{(A - \sigma_i E)^{-1} E\}^2 \bar{v}^{(i)} - \bar{w}^{(i)T} (A - \sigma_i E)^{-1} E V_k W_k^T (A - \sigma_i E)^{-1} E \bar{v}^{(i)} \\ &= \pm \gamma_{\Pi}^{(i)-2} (c^T \{(A - \sigma_i E)^{-1} E\}^{j_i+1} (I - V_k W_k^T) \{(A - \sigma_i E)^{-1} E\}^{j_i} (A - \sigma_i E)^{-1} b). \end{aligned} \quad (30)$$

As  $m_{2j_i+1}(\sigma_i) = c^T \{(A - \sigma_i E)^{-1} E\}^{2j_i+1} (A - \sigma_i E)^{-1} b$ , the proof is complete if the term  $c^T \{(A - \sigma_i E)^{-1} E\}^{j_i+1} V_k W_k^T \{(A - \sigma_i E)^{-1} E\}^{j_i} (A - \sigma_i E)^{-1} b$  in (30) is  $\hat{m}_{2j_i+1}(\sigma_i)$ . To quickly demonstrate this last fact, we employ a small trick from [7]. Note that the original system (1) can be rewritten as

$$(A - \sigma_i E)^{-1} E \dot{x}(t) = (A - \sigma_i E)^{-1} A x(t) + (A - \sigma_i E)^{-1} b u(t) \quad \text{and} \quad y(t) = c^T x(t)$$

so that the restriction of (1) by  $V_k W_k^T$  becomes

$$W_k^T (A - \sigma_i E)^{-1} E V_k \dot{\hat{x}}(t) = W_k^T (A - \sigma_i E)^{-1} A V_k \hat{x}(t) + W_k^T (A - \sigma_i E)^{-1} b u(t); \quad \hat{y}(t) = c^T V_k \hat{x}(t). \quad (31)$$

Taking the Laplace transform of (31) yields that the transfer function of the order- $k$  model is

$$\begin{aligned}\hat{g}(s) &= c^T V_k [W_k^T (A - \sigma_i E)^{-1} A V_k - s W_k^T (A - \sigma_i E)^{-1} E V_k]^{-1} W_k (A - \sigma_i E)^{-1} b \\ &= c^T V_k [I + (\sigma_i - s) W^T (A - \sigma_i E)^{-1} E V_k]^{-1} W_k^T (A - \sigma_i E)^{-1} b.\end{aligned}\quad (32)$$

The  $(2j_i + 2)^{nd}$  coefficient of the power series expansion of (32) in terms of  $(s - \sigma_i)$  is  $c^T V_k \{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{2j_i+1} W_k (A - \sigma_i E)^{-1} b$  which is equal to  $c^T \{(A - \sigma_i E)^{-1} E\}^{j_i+1} V_k W_k^T \{(A - \sigma_i E)^{-1} E\}^{j_i} (A - \sigma_i E)^{-1} b$ .  $\square$

The quantity  $r_k^{(i)T} q_k^{(i)}$  is proportional to the absolute error between the first unmatched moment about  $\sigma_i$  of the original system and the reduced-order model of dimension  $k$ . In fact,  $\gamma_n^{(i)^2} q_k^{(i)T} r_k^{(i)}$  is an exact expression for the absolute error. The use of the absolute moment error appears to be best suited for imaginary interpolation points and will be reported on elsewhere in the future. In this paper, real shifts are employed (recall the discussion at the beginning of this section), and an approximation of the relative error in the moments,  $\sigma_i^2 q_k^{(i)T} r_k^{(i)}$ , seems to be most useful. The scalar  $\sigma_i^2$  normalizes the residuals against the distance from the interpolation point to the  $i\omega$  axis.

Given no other information, it makes little sense to match a moment in the  $(k + 1)^{st}$  iteration if the error between that moment and the corresponding moment of the order- $k$  model is already small. Rather, one should logically direct their effort towards a value of  $\sigma_i$  where the error in the first unmatched moment (e.g.,  $m_{2j_i+1}(\sigma_i)$  versus  $\hat{m}_{2j_i+1}(\sigma_i)$ ) is large. After choosing  $\sigma_i$  as the interpolation point for the  $(k + 1)^{st}$  iteration, and performing this iteration, Theorem 2 tells us that this error,  $m_{2j_i+1}(\sigma_i) - \hat{m}_{2j_i+1}(\sigma_i)$ , becomes 0. This concludes our justification of step 1 of Algorithm 3. By choosing a  $\sigma_i$  among the  $\bar{\tau}$  possibilities which maximizes  $\sigma_i^2 r_k^{(i)T} q_k^{(i)}$ , one hopes to add as much beneficial information as possible to the reduced-order model in the  $(k + 1)^{st}$  iteration. The selection at (A3.1) is in some sense locally optimal and is perhaps the best one can hope for given the limited quantity of information available at the  $(k + 1)^{st}$  iteration.

Finally, one should note that the dot-product of the residuals is an infamous quantity in the standard (single interpolation point) non-symmetric Lanczos algorithm. The occurrence of a zero or near zero dot-product with  $r_k \neq 0$ ,  $q_k \neq 0$  is termed a serious breakdown [22] as it leads to division by zero in the algorithm. A large amount of effort has been placed towards working around this breakdown in the standard non-symmetric Lanczos algorithm, e.g., the look-ahead Lanczos method [9, 21]. The serious breakdown itself is known to be connected with system theory [22]. For example, if  $r_{k+1}^T q_{k+1} = 0$ , the order- $k$  and order- $k + 1$  models of the single-point Lanczos method would share the same minimal realization. This fact is entirely consistent with Theorem 3. Rational Lanczos tends to avoid such breakdowns since one works to maximize a scaled version of  $r_k^{(i)T} q_k^{(i)}$ . Selecting an interpolation point with the goal of maximizing the amount of new information included in the order- $(k + 1)$  over the order- $k$  model leads to a fortuitous by-product: a tendency to naturally avoid breakdowns. Of course given that one can only reasonably access a finite number of interpolation points  $\ll n$ , pathological cases exist where none of the  $\bar{\tau}$  points yield new information at the  $k^{th}$  iteration. Look-ahead or additional, new interpolation points would be required in this case. Such a situation has not yet been encountered in practice and is seemingly unlikely unless the reduced-order model has actually converged over the specified frequency range.

The last topic to be covered with respect to Algorithm 3 is the structure of the matrices making up the reduced-order model of dimension  $\bar{k}$ . This reduced-order model again takes the form of (14). Related to (9) and (11), the  $k^{th}$  columns of  $H_{\bar{k}+1, \bar{k}}$  and  $K_{\bar{k}+1, \bar{k}}$  are defined for Algorithm 3 through the expression

$$(A - \sigma_{i_1}^* E) V_{\bar{k}+1} \underbrace{\left( \begin{bmatrix} \gamma_{1,k}^{(i_k^*)} \\ \vdots \\ \gamma_{k+1,k}^{(i_k^*)} \\ 0 \end{bmatrix} - \gamma_{1,0}^{(i_1^*)} e_1 \right)}_{h_k} = E V_{\bar{k}+1} \underbrace{\left( \begin{bmatrix} \gamma_{1,k}^{(i_{k+1}^*)} \\ \vdots \\ \gamma_{k+1,k}^{(i_{k+1}^*)} \\ 0 \end{bmatrix} (\sigma_{i_{k+1}^*} - \sigma_{i_1^*}) \right)}_{k_k}$$

if  $i_{k+1}^* \notin \{i_1^*, \dots, i_k^*\}$ . Otherwise,

$$(A - \sigma_{i_1^*} E)V_{\bar{k}+1} \underbrace{\begin{bmatrix} \gamma_{1,k}^{(i_{k+1}^*)} \\ \vdots \\ \gamma_{k+1,k}^{(i_{k+1}^*)} \\ 0 \end{bmatrix}}_{h_k} = EV_{\bar{k}+1} \underbrace{\left( \begin{bmatrix} \gamma_{1,k}^{(i_{k+1}^*)} \\ \vdots \\ \gamma_{k+1,k}^{(i_{k+1}^*)} \\ 0 \end{bmatrix} (\sigma_{i_{k+1}^*} - \sigma_{i_1^*}) + e_{\tau_{i_k^*}} \right)}_{k_k}.$$

As in §4, a number of the  $\gamma$  elements making up  $H_{\bar{k},\bar{k}}$  and  $K_{\bar{k},\bar{k}}$  are zero due to step (A3.5). Figure 4 provides as an example the structure of a simple  $H_{12,11}$  matrix constructed by switching back and forth between interpolation points  $\sigma_1$  and  $\sigma_2$ .  $H_{\bar{k},\bar{k}}$  is primarily tridiagonal with nonzero elements only appearing above the tridiagonal when a change in the interpolation point occurs. These changes in the interpolating point are indicated by the dashed partitioning of  $H_{12,11}$  in Figure 4. Using arguments similar to those provided in §4, one can show that the first nonzero element in the  $k^{\text{th}}$  column of  $H_{\bar{k},\bar{k}}$  is in the  $\tilde{k}^{\text{th}}$  row. The value of  $\tilde{k}$  is the index of the next to last column vector of  $V_k$  which was constructed using the same interpolation point as  $v_{k+1}$ . This next to last behavior is a generalization of the three-term recurrences of single-point, nonsymmetric Lanczos. In single-point Lanczos, the next to last column vector of  $V_k$  is always  $v_{k-1}$ .

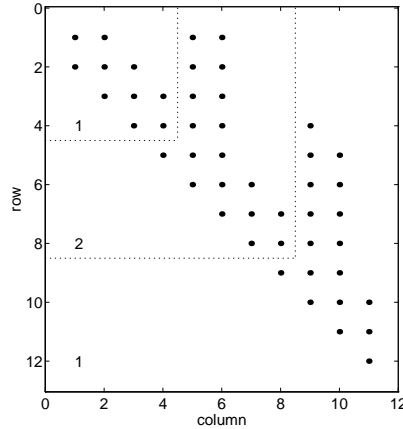


Figure 4: Structure of a sample matrix  $H_{12,11}$ . The first four columns of  $H_{12,11}$  correspond to the vectors  $v_2$  through  $v_5$  with interpolation point  $\sigma_1$ . The last three columns of  $H_{12,11}$  also correspond to  $\sigma_1$  while the middle four columns correspond to  $\sigma_2$ .

From the definition of  $\tilde{k}$ , the upper bandwidth of  $H_{k,k}$  can be restricted to  $\bar{\tau}$  if the interpolation points are perfectly interspersed, i.e., if the interpolation points are chosen so that the difference  $k - \tilde{k}$  is always equal to  $\bar{\tau}$ . Algorithm 2, on the other hand, computes all of its iterations corresponding to a given interpolation point in order. Hence the strategy of Algorithm 2 leads to off-tridiagonal spikes which always rise to the first row. In between these two extremes, the interpolation points are chosen with respect to improving the reduced-order model and structures similar to Figure 4 typically result. It may be advantageous in future work to achieve shorter recurrences by compromising between the proposed interpolation selection strategy (based on Theorem 3) and perfectly interspersed interpolation points.

It must be stressed that the structure of the  $H_{\bar{k},\bar{k}}$  and  $K_{\bar{k},\bar{k}}$  matrices holds in theory. In practice, the use of short recurrences in any Lanczos type algorithm will lead to a gradual loss of biorthogonality between  $V_k$  and  $W_k$  as  $k$  increases. Super-tridiagonal elements of  $H_{\bar{k},\bar{k}}$  will in turn be non-zero. The loss of biorthogonality between  $V_k$  and  $W_k$  is well-known [13] and complete fixes (e.g., complete reorthogonalization) are expensive. In the following example, ignoring the loss of biorthogonality versus complete reorthogonalization had a



negligible impact on the results. A more detailed study of the effects of loss of orthogonalization in the context of model reduction remains an area of current and future work.

As a second example, we consider a  $120^{th}$  order SISO system which describes the dynamics between the lens actuator and radial arm position of a portable compact disc player discussed in [31]. The transfer function of this system is shown as a solid line in Figure 5. Due to physical constraints on the size of the system's controller, a model with  $k \leq 15$  is desired. Rational Lanczos (Algorithm 3), using in order the 15 interpolation points  $\{10^5, 10^3, 100, 10^4, 10^4, 10, 10^5, 10^5, 100, 100, 100, 100, 100, 100, 100\}$  yields a frequency response which is nearly identical to that of the actual system. On the other hand, the difference between the frequency responses of an order-15 single-point Lanczos model about  $\sigma = 0$  and the original system is significant for  $\omega > 100$ . The transfer function for a single point model about  $\omega_{max} = 10^5$  is also displayed in Figure 5. The error for this single-point model is large for  $\omega < 1000$ . The convergence of the single-point models is delayed by the difficulties discussed in §3.

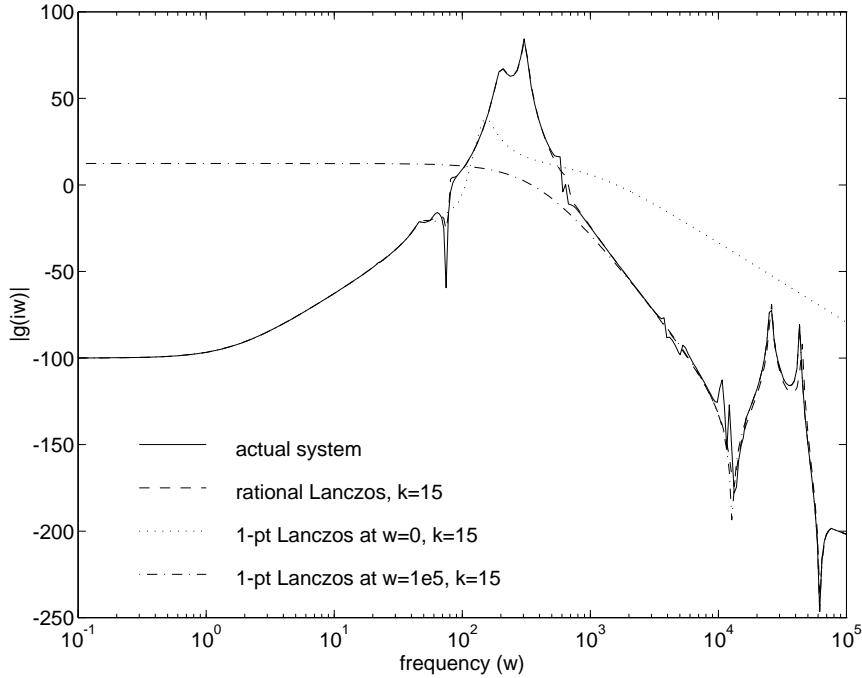


Figure 5: Frequency response for Example 2.

## 6 Modeling error

A knowledge of the error between the original system and reduced-order model is important for several reasons. In simulation, one needs to know that the response of the reduced-order model is sufficiently close to that of the original system. In control, one hopes to construct a controller from the reduced-order model which is robust enough to yield acceptable performance with the actual system. Even in performing model reduction with rational Lanczos, we would like to make improved choices for future interpolation points based on the error up to the  $k^{th}$  iteration. In all of these cases, it is desirable to quantify the error in terms of the differences between the frequency responses of the original system,  $g(s)$ , and of the reduced-order model,  $\hat{g}(s)$ .

**Theorem 4** *If  $\hat{g}(s)$  is the transfer function of the reduced-order model defined via (14), then*

$$\varepsilon(s) = g(s) - \hat{g}(s) = \left\{ c^T (sE - A)^{-1} \tilde{b} \right\} f \left( \frac{1}{s - \mu} \right) \quad (33)$$

where  $\tilde{b} = (A - \mu E)r_{\bar{k}}$ ,  $f(s) = e_k^T (H_{\bar{k},\bar{k}} - sK_{\bar{k},\bar{k}})^{-1} e_1 \gamma_{1,0}$  and  $\mu$  is the first interpolation frequency, e.g.,  $\sigma_{i_1}$  in Algorithm 3.

*Proof:* From (14) and the definition of the transfer function,

$$\begin{aligned} g(s) - \hat{g}(s) &= c^T (sE - A)^{-1} b - cV_{\bar{k}}K_{\bar{k},\bar{k}} (H_{\bar{k},\bar{k}}(s - \mu) - K_{\bar{k},\bar{k}})^{-1} W_{\bar{k}}^T (A - \mu E)b \\ &= c^T (sE - A)^{-1} (A - \mu E) \left\{ (A - \mu E)^{-1} b + ((A - \mu E)^{-1} E(s - \mu) - I) V_{\bar{k}}K_{\bar{k},\bar{k}} \cdot \right. \\ &\quad \left. (H_{\bar{k},\bar{k}}(s - \mu) - K_{\bar{k},\bar{k}})^{-1} e_1 \gamma_{1,0} \right\}. \end{aligned} \quad (34)$$

Adapting an argument from [17, Theorem 3.4], recall (12) and rewrite (34) as

$$\begin{aligned} \varepsilon(s) &= c^T (sE - A)^{-1} (A - \mu E) \left\{ V_{\bar{k}} - (V_{\bar{k}+1} H_{\bar{k}+1,\bar{k}}(s - \mu) - V_{\bar{k}} K_{\bar{k},\bar{k}}) (H_{\bar{k},\bar{k}}(s - \mu) - K_{\bar{k},\bar{k}})^{-1} \right\} e_1 \gamma_{1,0} \\ &= c^T (sE - A)^{-1} (A - \mu E) \left\{ (s - \mu)r_{\bar{k}} e_k^T (H_{\bar{k},\bar{k}}(s - \mu) - K_{\bar{k},\bar{k}})^{-1} e_1 \gamma_{1,0} \right\} \\ &= c^T (sE - A)^{-1} (A - \mu E) r_{\bar{k}} e_k^T (H_{\bar{k},\bar{k}} - (s - \mu)^{-1} K_{\bar{k},\bar{k}})^{-1} e_1 \gamma_{1,0}. \quad \square \end{aligned}$$

The error expression (33) is in fact identical in form to the rational interpolation error already derived in [18, Section 3]. However, the rational interpolation algorithm of [18] assumes no more than one moment is matched about each interpolation frequency  $\sigma_i$  and does not make use of the Lanczos algorithm. Analogous to the comments of [18, Section 3], several points should be made concerning the modeling error,  $\varepsilon(s)$ . First,  $f(s)$  corresponds to a  $\bar{k}$ -dimensional system which can be evaluated cheaply. Thus the modeling error can be expressed as a scalar function times the frequency response of the original system except that the input vector  $b$  is replaced by  $\tilde{b}$ . As a rough estimate of the error is typically sufficient, one can approximate (33) as  $f(1/(s - \mu))$  times some low-order approximation for  $c^T (sE - A)^{-1} \tilde{b}$ . The restriction of  $\{A, E, \tilde{b}, c\}$  by  $\Pi_k$  is not a good candidate for this approximation though because the error is orthogonal to the projection, e.g.,  $W_{\bar{k}}^T (A - \mu E)^{-1} \tilde{b} = 0$ .

As suggested by the proof of Theorem 4, the modeling error (33) is also related to the residual errors derived in [17]. It is proposed in [17] that the norm of the residual error,  $b - (sI - A)V_{\bar{k}}(sI - H_{\bar{k},\bar{k}})^{-1} W_{\bar{k}}^T b$ , should be made small when performing model reduction via the nonsymmetric Lanczos method (note that [17] assumed  $E = I$  and an expansion point at infinity so that  $H_{\bar{k},\bar{k}}$  is a tridiagonal, low-order approximation for  $A$  in that paper). For rational Lanczos and thus rational interpolation, the methods of [17] can easily be adapted to express the appropriate residual error as

$$\begin{aligned} \varepsilon_r(s) &= (A - \mu E)^{-1} b - ((A - \mu E)^{-1} E(s - \mu) - I) V_{\bar{k}}K_{\bar{k},\bar{k}} (H_{\bar{k},\bar{k}}(s - \mu) - K_{\bar{k},\bar{k}}) W_{\bar{k}}^T (A - \mu E)^{-1} b \\ &= r_{\bar{k}} e_k^T (H_{\bar{k},\bar{k}} - (s - \mu)^{-1} K_{\bar{k},\bar{k}})^{-1} e_1 \gamma_{1,0} \\ &= r_{\bar{k}} f \left( \frac{1}{s - \mu} \right). \end{aligned}$$

Striving for a small residual error as suggested in [17] can therefore be thought of as approximating  $\varepsilon(s)$  with  $\varepsilon_r(s)$  where  $\|r_{\bar{k}}\|$  is used as an estimate for  $c^T (sE - A)^{-1} \tilde{b}$ .

With error expressions, one should be able to actively adapt the set of interpolation points as  $k$  increases to rapidly address those frequencies where large error still exists. With the standard Lanczos algorithm on other hand, one can do nothing to promote convergence away from the interpolation point except increase the value of  $k$ . The use of error expressions in conjunction with rational Lanczos is currently being explored.

## 7 Conclusion

This paper showed that the rational Lanczos method (Algorithm 2) leads to Padé approximants about multiple interpolation frequencies. The approach of [7] is a special case of the rational Lanczos method where

only one interpolation point is allowed. The earlier methods of [4, 20, 27] are an even more specialized case of the rational Lanczos approach where only an interpolation point at zero is permitted. Given multiple interpolation points, this paper presented an easily computable criterion based on the inner-product  $r_k^T q_k$  for choosing among the possibilities. Two examples were provided to indicate why model reduction via a rational Lanczos method has the potential for significant improvement over existing single-point Lanczos approaches. Utilizing multiple interpolation points provides the freedom to search out the dominant dynamics of the system. The convergence of a single-point interpolant, on the other hand, can be slowed by the presence of non-dominating dynamics.

Linear time-invariant, SISO systems were considered for model reduction. A block rational Lanczos algorithm has been developed and will be available in a forthcoming paper. However, other issues still require additional work. Complex interpolation points, the moment error expressions of §5 and the transfer function error expressions of §6 should be better utilized to increase the effectiveness of the rational Lanczos method. Approaches for inverting  $(A - \sigma_i E)$  are also needed. Sparse matrix factorizations or iterative techniques must be utilized to avoid large computational costs.

## Appendix

**Lemma 1** *If  $\sigma$  and  $\varsigma$  are two nonidentical interpolation points, then*

$$(A - \sigma E)^{-1} E \{ (A - \varsigma E)^{-1} E \}^{j-1} (A - \varsigma E)^{-1} b \in \quad (35)$$

$$\left\{ \text{span}\{(A - \sigma E)^{-1} b\} \cup \mathcal{K}_j((A - \varsigma E)^{-1} E, (A - \varsigma E)^{-1} b) \right\} \quad (36)$$

and

$$E^T (A - \sigma E)^{-T} \{ E^T (A - \varsigma E)^{-T} \}^{j-1} c \in \left\{ \text{span}\{E^T (A - \sigma E)^{-T} c\} \cup \mathcal{K}_j(E^T (A - \varsigma E)^{-T}, c) \right\} \quad (37)$$

for any value of  $j \geq 1$ .

*Proof:* We prove (36). The key is to note that  $(A - \sigma E)^{-1}$  can be rewritten as

$$\begin{aligned} (A - \sigma E)^{-1} &= (A - \sigma E)^{-1} (A - \varsigma E) (A - \varsigma E)^{-1} \\ &= (A - \sigma E)^{-1} (A - \sigma E + (\sigma - \varsigma) E) (A - \varsigma E)^{-1} \end{aligned}$$

which yields

$$(\sigma - \varsigma) (A - \sigma E)^{-1} E (A - \varsigma E)^{-1} = (A - \sigma E)^{-1} - (A - \varsigma E)^{-1}. \quad (38)$$

Using (38), (36) follows via induction. If  $j = 1$ , multiplying (38) on the right by  $b$  gives

$$(A - \sigma E)^{-1} E (A - \varsigma E)^{-1} b = (\sigma - \varsigma)^{-1} \{ (A - \sigma E)^{-1} - (A - \varsigma E)^{-1} \} b$$

and (36) is satisfied. Next assume that (36) holds for  $j = 1, \dots, (\bar{j} - 1)$ . Multiplying (38) on the right by  $E \{ (A - \varsigma E)^{-1} E \}^{\bar{j}-2} (A - \varsigma E)^{-1} b$  yields

$$\begin{aligned} (\sigma - \varsigma) (A - \sigma E)^{-1} E \{ (A - \varsigma E)^{-1} E \}^{\bar{j}-1} (A - \varsigma E)^{-1} b = \\ (A - \sigma E)^{-1} E \{ (A - \varsigma E)^{-1} E \}^{\bar{j}-2} (A - \varsigma E)^{-1} b - \{ (A - \varsigma E)^{-1} E \}^{\bar{j}-1} (A - \varsigma E)^{-1} b. \end{aligned} \quad (39)$$

Thus under the assumption that (36) holds for  $j = \bar{j} - 1$ , (39) shows that (36) also holds for  $j = \bar{j}$ . The induction step and thus (36) hold in general. The proof of (37) is the dual to that provided for (36).  $\square$

**Lemma 2** *Let  $W_{k+1}$ ,  $V_{k+1}$ ,  $K_{k+1,k}$  and  $H_{k+1,k}$  be the results of Algorithms 2 or 3 with  $k = \bar{k}$  (i.e., the algorithm is run to completion), then*

$$W_{\bar{k}}^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) V_{\bar{k}} = [K_{\bar{k},\bar{k}} - W_{\bar{k}}^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) r_{\bar{k}} e_{\bar{k}}^T (\sigma_1 - \sigma_i)] J_{\bar{k},\bar{k}} \quad (40)$$

and

$$W_k^T (A - \sigma_i E)^{-1} E V_k K_{k,k} = K_{k,k} J_{k,k} H_{k,k} + W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) r_k e_k^T \{I + J_{k,k} H_{k,k} (\sigma_1 - \sigma_i)\} \quad (41)$$

where  $J_{k,k} = [K_{k,k} + H_{k,k} (\sigma_1 - \sigma_i)]^{-1}$ .

*Proof:* Recalling (12) gives  $(A - \sigma_1 E)^{-1} E V_{k+1} K_{k+1,k} = V_{k+1} H_{k+1,k}$ . Multiplying both sides of this expression by  $(\sigma_1 - \sigma_i)$  and adding  $V_{k+1} K_{k+1,k}$  to both sides yields

$$[I + (A - \sigma_1 E)^{-1} E (\sigma_1 - \sigma_i)] V_{k+1} K_{k+1,k} = V_{k+1} [K_{k+1,k} + H_{k+1,k} (\sigma_1 - \sigma_i)],$$

which can be rearranged as

$$V_{k+1} K_{k+1,k} = [I + (A - \sigma_1 E)^{-1} E (\sigma_1 - \sigma_i)]^{-1} V_{k+1} [K_{k+1,k} + H_{k+1,k} (\sigma_1 - \sigma_i)]. \quad (42)$$

Multiplying both sides of (42) on the left by  $W_k^T$ , recalling that  $V_{k+1} K_{k+1,k} = V_k K_{k,k}$ , and noting that  $V_{k+1} H_{k+1,k} = V_k H_{k,k} + r_k e_k^T$  since  $H_{k+1,k}$  is upper-Hessenberg yields

$$K_{k,k} = W_k^T [I + (A - \sigma_1 E)^{-1} E (\sigma_1 - \sigma_i)]^{-1} \{V_k [K_{k,k} + H_{k,k} (\sigma_1 - \sigma_i)] + (\sigma_1 - \sigma_i) r_k e_k^T\}$$

so that

$$W_k^T [I + (A - \sigma_1 E)^{-1} E (\sigma_1 - \sigma_i)]^{-1} V_k = K_{k,k} J_{k,k} + W_k^T [I + (A - \sigma_1 E)^{-1} E (\sigma_1 - \sigma_i)]^{-1} r_k e_k^T J_{k,k} (\sigma_1 - \sigma_i). \quad (43)$$

$$K_{k,k} J_{k,k} + W_k^T [(A - \sigma_1 E)^{-1} \{(A - \sigma_1 E) + E (\sigma_1 - \sigma_i)\}]^{-1} r_k e_k^T J_{k,k} (\sigma_1 - \sigma_i) \\ K_{k,k} J_{k,k} + W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) r_k e_k^T (\sigma_1 - \sigma_i). \quad (44)$$

To acquire (41), note that multiplying (12) on the left by  $W_k^T (A - \sigma_i E)^{-1}$  yields

$$W_k^T (A - \sigma_i E)^{-1} E V_k K_{k,k} = W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) V_{k+1} H_{k+1,k} \\ = W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) V_k H_{k,k} + W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) r_k e_k^T.$$

Rewriting  $W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) V_k H_{k,k}$  as in (43) yields (41).  $\square$

**Lemma 3** Let  $W_{k+1}$ ,  $V_{k+1}$ ,  $K_{k+1,k}$  and  $H_{k+1,k}$  be the results of Algorithm 2 with  $k = \bar{k}$  (i.e., the algorithm is run to completion), then

$$W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) r_k \{e_k^T (J_{k,k} H_{k,k})^{j-1} J_{k,k} e_1\} = 0 \quad (45)$$

and

$$e^T V_k \{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j-1} W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) r_k = 0 \quad (46)$$

for  $i = 1, 2, \dots, \bar{i}$ ,  $j = 1, 2, \dots, \bar{j}$  and where  $J_{k,k} = [K_{k,k} + H_{k,k} (\sigma_1 - \sigma_i)]^{-1}$ .

*Proof:* For the case  $i = 1$ , both (45) and (46) are trivial since  $W_k^T (A - \sigma_i E)^{-1} (A - \sigma_1 E) r_k = W_k^T r_k = 0$  by the imposed biorthogonality condition.

For  $i > 1$ , (45) can be demonstrated by first noting that the leading  $(i-1)\bar{j} \times (i-1)\bar{j}$  submatrix of  $[K_{k,k} + H_{k,k} (\sigma_1 - \sigma_i)]$  is upper-Hessenberg while from (11), columns  $l = (i-1)\bar{j} + 1, \dots, i\bar{j} - 1$  of  $[K_{k,k} + H_{k,k} (\sigma_1 - \sigma_i)]$  are

$$\left( \begin{bmatrix} \gamma_{1,l} \\ \vdots \\ \gamma_{l+1,l} \\ 0 \end{bmatrix} (\sigma_i - \sigma_1) + e_l \right) - \begin{bmatrix} \gamma_{1,l} \\ \vdots \\ \gamma_{l+1,l} \\ 0 \end{bmatrix} (\sigma_1 - \sigma_i) = e_l.$$

Thus  $[K_{k,k} + H_{k,k}(\sigma_1 - \sigma_i)]$  takes the form

$$\left[ \begin{array}{c|c} X_1 & 0 \\ \hline 0 & I \\ \hline 0 & 0 \end{array} \middle| X_2 \right] \quad (47)$$

where  $X_1 \in \mathbb{R}^{(i-1)\bar{j} \times (i-1)\bar{j}}$ ,  $X_2 \in \mathbb{R}^{k \times (\bar{i}-i)\bar{j}+1}$ , and  $I \in \mathbb{R}^{(\bar{j}-1) \times (\bar{j}-1)}$  is an identity matrix.

By simple inspection,  $J_{k,k}$  must also take on the general form of (47). Thus for  $i > 1$ ,

$$e_l^T J_{k,k} = \begin{cases} \sum_{t=l}^k \alpha_t e_t^T & \text{for } (i-1)\bar{j} < l < i\bar{j} \\ \sum_{t=i\bar{j}}^k \alpha_t e_t^T & \text{for } l \geq i\bar{j} \end{cases} \quad (48)$$

because columns  $(i-1)\bar{j} + 1$  through  $i\bar{j} - 1$  of  $J_{k,k} = [K_{k,k} + H_{k,k}(\sigma_1 - \sigma_i)]^{-1}$  are standard unit vectors. Since  $H_{k,k}$  is upper-Hessenberg,

$$e_l^T H_{k,k} = \sum_{t=l-1}^k \alpha_t e_t^T \quad (49)$$

for any value of  $l$ . Through repeated use of (48) and (49), one obtains

$$e_k^T (J_{k,k} H_{k,k})^{j-1} J_{k,k} e_1 = \left( \sum_{t=i\bar{j}-j+1}^k \alpha_t e_t^T \right) e_1 = 0$$

for  $i > 1$  and  $j \leq \bar{j}$  so that (45) is shown.

To show (46) for  $i > 1$ , we begin by rewriting the left side of (46) as (similar to (44))

$$\begin{aligned} c^T V_k \{W_k^T (A - \sigma_i E)^{-1} E V_k\}^{j-1} W_k^T \{I + (A - \sigma_i E)^{-1} E (\sigma_i - \sigma_1)\} r_k = \\ (\sigma_i - \sigma_1) c^T V_k \{W_k (A - \sigma_i E)^{-1} E V_k\}^{j-1} W_k^T (A - \sigma_i E)^{-1} E r_k. \end{aligned} \quad (50)$$

Now since  $V_k W_k^T$  is a biorthogonal projector with the column spaces of  $V_k$  and  $W_k$  defined in Theorem 1,  $c^T \{(A - \sigma_i E)^{-1} E\}^{j-1} V_k W_k^T = c^T \{(A - \sigma_i E)^{-1} E\}^{j-1}$  for  $j = 1, 2, \dots, \bar{j}$ . Repeated use of this last fact on (50) yields that the left side of (46) is equivalent to

$$(\sigma_i - \sigma_1) c^T \{(A - \sigma_i E)^{-1} E\}^j r_k. \quad (51)$$

But (51) is zero for  $1 \leq j \leq \bar{j}$  since  $c^T \{(A - \sigma_i E)^{-1} E\}^j \in \text{rowsp}(W_k^T)$  and  $W_k^T r_k = 0$ .  $\square$

## References

- [1] A.C. Antoulas and B.D.O. Anderson, Rational interpolation and state variable realizations, *Linear Algebra and Appl.*, vol. 137, pp. 479–509, 1990.
- [2] G.A. Baker Jr., *Essentials of Padé Approximants*, New York, NY: Academic Press, 1975.
- [3] D.L. Boley and G. Golub, The nonsymmetric Lanczos algorithm and controllability, *Systems Control Lett.*, vol. 16, pp. 97–105, 1991.
- [4] D.L. Boley, Krylov space methods on state-space control models, Report No. TR92-18, Department of Computer Science, University of Minnesota, MN, 1992.
- [5] R. R. Craig Jr., Recent literature on structural modeling, identification, and analysis, in J. L. Junkins, editor, *Mechanics and Control of Large Flexible Structures*, Washington, DC: AIAA, 1990.
- [6] E. Chiprout and M. S. Nakhla, *Asymptotic Waveform Evaluation and Moment Matching for Interconnect Analysis*, Boston, MA: Kluwer Academic Publishers, 1994.

- [7] P. Feldman and R.W. Freund, Efficient linear circuit analysis by Padé approximation via the Lanczos process, *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 639–649, 1995.
- [8] L. Fortuna, G. Nunnari and A. Gallo, *Model Reduction Techniques with Applications in Electrical Engineering*, London: Springer-Verlag, 1992.
- [9] R.W. Freund, M.H. Gutknecht and N.M. Nachtigal, An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices, *SIAM J. Sci. Comp.*, vol. 14, pp. 137–158, 1993.
- [10] K. Gallivan, E. Grimme and P. Van Dooren, Asymptotic waveform evaluation via a Lanczos method, *Appl. Math. Lett.*, vol. 7, pp. 75–80, 1994.
- [11] K. Gallivan, E. Grimme and P. Van Dooren, Padé Approximation of large-scale dynamic systems with Lanczos methods, *Proc. 33rd IEEE Conf. on Decision and Control*, (Lake Buena Vista, FL), 1994.
- [12] E. Gallopoulos and Y. Saad, Efficient solution of parabolic equations by Krylov approximation methods, *SIAM J. Sci. Statist. Comput.*, vol. 13, pp. 1236–1264, 1992.
- [13] G.H. Golub and C.F Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins University Press, 1989.
- [14] W.B. Gragg and A. Lindquist, On the partial realization problem, *Linear Algebra and Appl.*, vol. 50, pp. 277–319, 1983.
- [15] E. Grimme, D. Sorensen and P. Van Dooren, Model reduction of state space systems via an implicitly restarted Lanczos method, to appear *Numer. Algorithms*, 1995.
- [16] C. Hwang and M.-Y. Chen, A multi-point continued-fraction expansion for linear system reduction, *IEEE Trans. Automatic Control*, vol. 31, pp. 648–651, 1986.
- [17] I.M. Jaimoukha and E.M. Kasenally, Oblique projection methods for large scale model reduction, to appear *SIAM J. Matrix Anal. Appl.*, 1995.
- [18] C. Kenney, A.J. Laub and S. Stubberud, Frequency response computation via rational interpolation, *IEEE Trans. Automatic Control*, vol 38, pp. 1203–1213, 1993.
- [19] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Nat. Bureau Stan.*, vol. 45, pp. 255–282, 1950.
- [20] B. Nour-Omid and R.W. Clough, Dynamic analysis of structures using Lanczos co-ordinates, *Earthquake Eng. and Struc. Dyn.*, vol. 12, pp. 565–577, 1984.
- [21] B.N. Parlett, D.R. Taylor and Z.S. Liu, A look-ahead Lanczos algorithm for unsymmetric matrices, *Math. Comp.*, vol. 44, pp. 105–124, 1985.
- [22] B.N. Parlett, Reduction to tridiagonal form and minimal realizations, *SIAM J. Matrix Anal. Appl.*, vol. 13, pp. 567–593, 1992.
- [23] A. Ruhe, Rational Krylov sequence methods for eigenvalue computation, *Linear Algebra and Appl.*, vol. 58, pp. 391–405, 1984.
- [24] A. Ruhe, Rational Krylov algorithms for nonsymmetric eigenvalue problems II: Matrix pairs, *Linear Algebra and Appl.*, vol. 197, pp. 283–295, 1994.
- [25] A. Ruhe, The rational Krylov algorithm for nonsymmetric eigenvalue problems III: Complex shifts for real matrices, *BIT* vol. 34, pp. 165–176, 1994.
- [26] Y. Saad, Analysis of some Krylov subspace approximations to the matrix exponential operator, *SIAM J. Numer. Anal.*, vol. 29, pp. 209–228, 1992.

- [27] T.-J. Su and R.R. Craig Jr., Krylov vector methods for model reduction and control of flexible structures, in C.T. Leondes, editor, *Control and Dynamic Systems: Integrated Technology Methods in Aerospace Systems Design*, London: Academic Press, 1992.
- [28] P. Van Dooren, Numerical linear algebra techniques for large scale matrix problems in systems and control, *Proc. 31st IEEE Conf. on Decision and Control*, (Tucson, AZ), 1992.
- [29] C.D. Villemagne and R.E. Skelton, Model reduction using a projection formulation, *Internat. J. Control*, vol. 46, pp. 2141–2169, 1987.
- [30] E.L. Wilson, M.-W. Yuan and J.M. Dickens, Dynamic analysis by direct superposition of Ritz vectors, *Earthquake Eng. and Struc. Dyn.*, Vol. 10, pp. 813–821, 1982.
- [31] P. Wortelboer, Frequency-weighted balanced reduction of closed-loop mechanical servo-systems: theory and tools, PhD thesis, Technical University Delft, 1994.
- [32] H. Xiheng, FF-Padé method of model reduction in frequency domain, *IEEE Trans. Automatic Control*, vol. 32, pp. 243–246, 1987.