

## APPROXIMATION OF LARGE-SCALE DYNAMICAL SYSTEMS: AN OVERVIEW<sup>†</sup>

ATHANASIOS C. ANTOULAS\*, DAN C. SORENSEN\*\*

In this paper we review the state of affairs in the area of approximation of large-scale systems. We distinguish three basic categories, namely the SVD-based, the Krylov-based and the SVD-Krylov-based approximation methods. The first two were developed independently of each other and have distinct sets of attributes and drawbacks. The third approach seeks to combine the best attributes of the first two.

**Keywords:** model reduction, SVD, Hankel, balancing, Krylov

### 1. Introduction and Problem Statement

Differential equations are one of the most successful means of modeling dynamical systems. In this paper we will deal with the problem of *simplification* or *model reduction* of dynamical systems described by differential equations. Model reduction has a long history in the systems and control literature. In fact, the general topic of dimension reduction in dynamical systems is pervasive in the applied mathematics literature. The systems we will consider will be denoted by  $\Sigma$ , and will be modeled by means of a set of first-order coupled differential equations, together with a set of algebraic equations:

$$\Sigma : \begin{cases} \frac{dx(t)}{dt} = f(x(t), u(t)), \\ y(t) = h(x(t), u(t)). \end{cases} \quad (1)$$

For simplicity, we will use the following notation:

$$\Sigma = (f, h), \quad u(t) \in \mathbb{R}^m, \quad x(t) \in \mathbb{R}^n, \quad y(t) \in \mathbb{R}^p.$$

In this setting,  $u$  is the input or excitation function,  $x$  is the state, and the function  $f$  describes the *dynamics* of  $\Sigma$ ; on the other hand,  $y$  is the output or the set of

---

<sup>†</sup> This work was supported in part by the following NSF grants: DMS-9972591, CCR-9988393, and ACI-0082645.

\* Dept. of Electrical and Computer Engineering, Rice University, Houston, TX 77005-1892, e-mail: [aca@ece.rice.edu](mailto:aca@ece.rice.edu)

\*\* Dept. of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892, e-mail: [sorensen@caam.rice.edu](mailto:sorensen@caam.rice.edu)

observations and  $h$  describes the way in which the observations are deduced from the state and the input. The *complexity* of  $\Sigma$  is defined as the number of states  $n$ .

The problem that will be addressed is to *simplify* or *approximate*  $\Sigma$  with another dynamical system  $\hat{\Sigma}$ ,

$$\hat{\Sigma} = (\hat{f}, \hat{h}), \quad u(t) \in \mathbb{R}^m, \quad \hat{x}(t) \in \mathbb{R}^k, \quad \hat{y}(t) \in \mathbb{R}^p.$$

The first requirement is that the number of states (i.e. the number of first-order differential equations) of the approximant be smaller (or much smaller) than that of the original system, i.e.,  $k \ll n$ . Obviously, in the absence of additional ones, this requirement is easy to satisfy by mere elimination of equations and state variables. The difficulty arises when imposing additional constraints; one set of constraints usually attached to such an approximation problem is the following:

1. The approximation error should be *small* (the existence of a *global error bound*),
2. *Stability* and *passivity* should be preserved,
3. The procedure is to be *computationally efficient*,
4. The procedure must be *automatic*, i.e., based on the error tolerance.

The first constraint above is to be understood as follows. Let the input function  $u$  be fed into both  $\Sigma$  and  $\hat{\Sigma}$ , and let the corresponding responses be  $y$  and  $\hat{y}$ , respectively. A *small approximation error* means that  $y$  is close to  $\hat{y}$  in an appropriately defined sense and for a given set of inputs; in the sequel we will use the following criterion in coming up with reduced-order approximants: the worst error  $\|y - \hat{y}\|$  should be minimized for all  $u$  (this gives rise to the so-called  $\mathcal{H}_\infty$  error criterion). In this regard it should be noted that reduced-order systems are often abstract representations which lose physical meaning.

An important *special case* is that of linear dynamical systems:

$$\Sigma : \begin{cases} \frac{dx(t)}{dt} = Ax(t) + Bu(t), \\ y(t) = Cx(t) + Du(t). \end{cases} \quad (2)$$

For simplicity, we will use the notation:

$$\Sigma = \left( \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) \in \mathbb{R}^{(n+p) \times (n+m)}. \quad (3)$$

The problem in this case consists in approximating  $\Sigma$  with

$$\hat{\Sigma} = \left( \begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array} \right) \in \mathbb{R}^{(k+p) \times (k+m)}, \quad (4)$$

where, as before,  $k < n$  (or  $k \ll n$ ) and the four conditions above are satisfied.

## 2. Motivating Examples

Models of dynamical systems are useful primarily for two reasons: *simulation* and *control*. We will illustrate these reasons with two examples, namely, modeling of *VLSI circuits*, where the effect of interconnections needs to be simulated, and control of the *International Space Station* (ISS), where low-order controllers are necessary.

**A. High-frequency, sub-micron VLSI circuits.** The scaling trends in today's VLSI circuits are a decrease in the feature size (by more than 10% per year) and an increase in the chip size (by more than 10% per year); this results in an overall increase in chip complexity (by more than 50% per year). One consequence of these trends is the multi-layered design (currently a chip may consist of up to 10 layers). Furthermore, there is an increase in operating frequency, which has reached the gigahertz range. These developments have had an impact on physical parameters, the most important being the increase in interconnect length. What used to be a piece of wire linking two components on the chip now needs to be modeled. In current sub-micron designs (a feature size in the sub-micron range), the interconnect length is of the order of several kilometers, the number of transistors approaches one billion and the operating frequency surpasses 1 GHz.

In conclusion, before settling on a particular design, it is important to be able to quickly simulate the particular chip layout and check that the signal delays, for instance from an input to a logic gate, do not exceed certain limits. Thus, interconnections *must* be modeled. The resulting models are *very high-order*:  $n \approx 10^5$  up to  $10^6$ . The usual simulation tool SPICE turns out to be inadequate for such complex circuits. For an account of interconnect issues see (Stroobandt, 2000). Model reduction was introduced into this area by Rohrer; see (Pillage and Rohrer, 1990; Gallivan *et al.*, 1994a).

**B. ISS: International Space Station.** This is a complex structure composed of many modules (it is estimated that more than 40 Shuttle flights will be required to complete the assembly). Furthermore, each module is described in terms of  $n \approx 10^3$  state variables.

In this case controllers will be needed, for instance, to reduce oscillatory motion or to fix the orientation of the space station with respect to some desired direction. It is well-known that generically a controller of a given plant has the same dimension as the plant. Since these controllers have to be implemented on-board, they must have *low* complexity due to hardware, radiation, throughput or testing limitations. Thus, reduced-order models are very useful for the development of reduced-order controllers.

**C. In general: Systems described by PDEs.** Approximate finite-element (FE) equations are obtained through spatial discretization of certain time-dependent systems which are described by PDEs, such as parabolic equations subject to boundary control. These equations are of high complexity. One such case is presented later as an example.

**Conclusion.** In all these cases, any realistic model will have high complexity, i.e., it will require many state variables to be adequately described. The resulting complexity

### Flex Structure Variation During Assembly

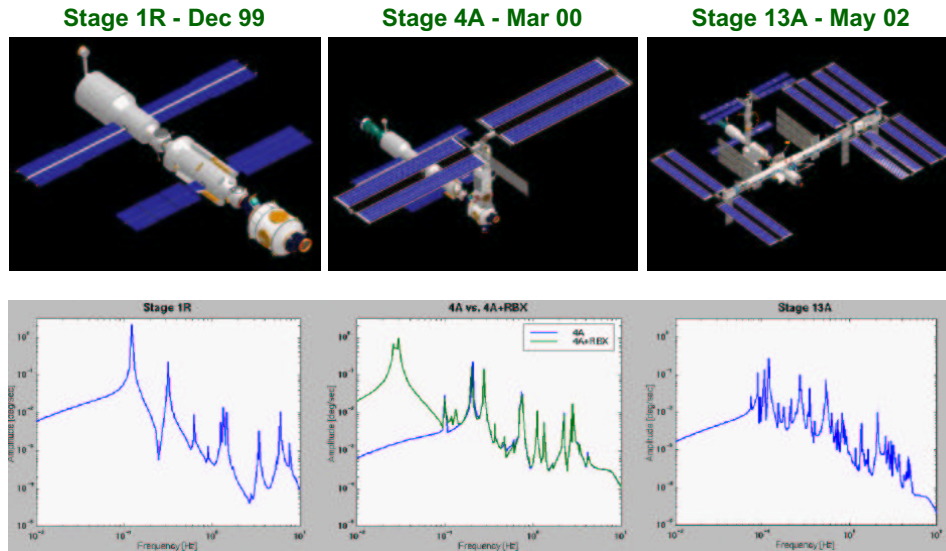


Fig. 1. ISS: Evolution of the frequency response as more components are added (Data: Draper Labs).

expressed in the number of first-order differential equations is such that a simplification or model reduction will be needed in order to perform a simulation in an amount of time which is acceptable for the application at hand, or for the design of a low-order controller which achieves desired objectives. Thus in all these cases *reduced-order models* are needed. The *issues with large-scale systems* are:

- Storage,
- Computational speed,
- Accuracy: large-scale problems are likely to be ill-conditioned,
- In addition, we need global error bounds and preservation of stability/passivity.

Due to the third item above, what works for dense computations involving low-dimensional systems may not work in a large-scale setting.

### 3. Approximation Methods

Approximation methods can be cast into three broad categories: (a) SVD-based methods, (b) Krylov-based methods, (c) iterative methods combining aspects of both the

SVD and Krylov methods. We will analyze them in the sequel and point out their strengths and weaknesses.

The *SVD-based approximation methods* have their roots in the Singular Value Decomposition and the resulting solution of the approximation of matrices by means of matrices of lower rank, which are optimal in the 2-norm (or more generally, in unitarily invariant norms). The quantities which are important in deciding to what extent a given finite-dimensional operator can be approximated by one of a lower rank are the so-called *singular values*; these are the square roots of the eigenvalues of the product of the operator and its adjoint. What is important is that the ensuing error satisfies an *a-priori* computable upper bound.

The question which arises is whether this result can be applied or extended to the case of dynamical systems. One straightforward way of applying it to a dynamical system described by (1) is as follows. Choose an input function and compute the resulting trajectory; collect samples of this trajectory at different times and compute the SVD of the resulting collection of samples. Then apply the SVD to the resulting matrix. This is a method which is widely used in computation involving PDEs; it is known as *Proper Orthogonal Decomposition (POD)*. The problem, however, in this case is that the resulting simplification heavily depends on the initial excitation function chosen and the time instances at which the measurements are taken. Consequently, the singular values obtained are not system invariants. The advantage of this method is that it can be applied to high-complexity linear as well as nonlinear systems.

A different approach results in the case of linear systems (2). First, we observe that a linear time-invariant dynamical system  $\Sigma$  can be represented by means of an integral (convolution) operator; if, in addition,  $\Sigma$  is finite-dimensional, then this operator has finite rank and is hence compact. Consequently, it has a set of finitely many non-zero singular values. Thus, in principle, the SVD approximation method can be used to simplify such dynamical systems.

Indeed, there is a set of invariants called the *Hankel singular values* which can be attached to every linear, constant, finite-dimensional system. These invariants play the same role for dynamical systems as the singular values play for constant finite-dimensional matrices. In other words, they determine the complexity of the reduced system and at the same time they provide a global error bound for the resulting approximation. This gives rise to two model reduction methods, namely, *balanced model reduction* and *Hankel norm approximation*. Many have observed that the Hankel singular values of many systems decay extremely rapidly. Hence very low rank approximations are possible and accurate low-order reduced models will result.

The first decisive results in this direction were obtained by Adamjan *et al.* (1971; 1978). The theory as it is used today is due to Glover (1984). Independently, Moore (1981) developed the concept of balancing; this led to the very popular method of approximation by balanced truncation.

The limitation of this approach comes from the fact that the computation of the Hankel singular values involves the solution of two linear matrix equations, the *Lya-punov equations*. Subsequently, the eigenvalues of the product of two positive definite

matrices are required; these are the *reachability* and *observability* grammians. The exact solution of such equations requires dense computations and therefore can, roughly speaking, be carried out for systems of a modest dimension ( $n \approx$  a few hundred).

A modification of this method has been developed for nonlinear systems. It is the method of *empirical grammians*. Its goal is to remedy the issues arising in POD methods, at the expense of added computational complexity (Lall *et al.*, 1998).

Table 1. Overview of approximation methods.

Approximation methods for dynamical systems		
SVD		Krylov
Nonlinear Systems	Linear Systems	
<ul style="list-style-type: none"> <li>• POD methods</li> <li>• Empirical grammians</li> </ul>	<ul style="list-style-type: none"> <li>• Balanced truncation</li> <li>• Hankel approximation</li> </ul>	<ul style="list-style-type: none"> <li>• Realization</li> <li>• Interpolation</li> <li>• Lanczos</li> <li>• Arnoldi</li> </ul>
SVD-Krylov		

A different set of approximation methods has been developed. They are *iterative* in nature and hence can be applied to very high-order systems. These are the *Krylov-based* approximation methods, which do not rely on the computation of singular values. Instead, they are based on *moment matching* of the *impulse response* of  $\Sigma$ . Two widely used methods fall under this category, namely, the Lanczos and the Arnoldi procedures, which were put forward in (Lanczos, 1950) and (Arnoldi, 1951), respectively. These methods have been very influential in iterative eigenvalue computations and more recently in model reduction. Their drawbacks are that the resulting reduced-order systems have no guaranteed error bound, stability is not necessarily preserved and some of them are not automatic.

Two leading efforts in this area are Padé via Lanczos (PVL) (Bai *et al.*, 1997; Freund, 1999; Feldmann and Freund, 1995), and multipoint rational interpolation (Grimme, 1997). The PVL approach exploits the deep connection between the (nonsymmetric) Lanczos process and classical moment matching techniques, i.e., the fundamental connection of Krylov subspace projection methods in linear algebra to the partial realization problem in system theory as laid out in (Gragg, 1972; Gragg and Lindquist, 1983); see also (Boley and Golub, 1984; Feldmann and Freund, 1995; Gollivan *et al.*, 1994b; Grimme, 1997; Jaimoukha and Kasenally, 1997; Van Dooren, 1995). A closely related approach which has been applied to VLSI circuits can be found in (Kamon *et al.*, 2000); see also (Li *et al.*, 1999).

The multipoint rational interpolation approach utilizes the rational Krylov method of (Ruhe, 1984) to provide the moment matching of the transfer function at selected frequencies and hence to obtain an enhanced approximation of the transfer function over a broad frequency range. These techniques have proven to be very effective. PVL has enjoyed considerable success in circuit simulation applications. Rational interpolation achieves a remarkable approximation of the transfer function with very low-order models. Nevertheless, there are shortcomings to both the approaches. In particular, since the methods are local in nature, it is difficult to establish global error bounds. Heuristics have been developed that appear to work, but no global results exist. Secondly, the rational interpolation method requires selection of interpolation points. At present, this is not an automated process and relies on ad-hoc specification by the user.

This brings us to the third class of methods which seek to combine the best features of the SVD- and the Krylov-based methods. We will refer to these as the SVD-Krylov-based approximation methods. The essential feature of the former is the solution of Lyapunov equations and the essential feature of the latter is the use of Krylov spaces (also known in systems theory as *reachability-observability* subspaces). In particular, it uses an iterative computation of projections onto these spaces. This gives rise to various approaches which have iterative Lyapunov solvers at their center.

The first approach in the third category makes use of a recent advance in the computation of eigenvalues of large matrices called *implicit restarting* (Sorensen, 1992). This has been applied with some success to the construction of stable reduced models to known stable LTI systems (Grimme *et al.*, 1995). The approach proposed in (Jaimoukha and Kasenally, 1997) uses an implicitly restarted dual Arnoldi approach. The dual Arnoldi method runs two separate Arnoldi processes, one for the controllability subspace and the other for the observability subspace, and then constructs an oblique projection from the two orthogonal Arnoldi basis sets. The basis sets and the reduced model are updated using a generalized notion of implicit restarting. The updating process is designed to iteratively improve the approximation properties of the model. Essentially, the reduced model is reduced further, keeping the best features, and then expanded via the dual Arnoldi processes to include new information. The goal is to achieve approximation properties related to balanced realizations. Other related approaches (Boley, 1994; Li *et al.*, 1999; Penzl, 2001) work directly with projected forms of the two Lyapunov equations (5) to obtain low-rank approximations to the system grammians.

One problem in working with the two Lyapunov equations separately and then applying dense methods to the reduced equations is *consistency*. One cannot be certain that the two separate basis sets are the ones that would have been selected if the full system grammians had been available. We are not aware of a consistent implicit restarting scheme that would be effective in converging the two separate processes towards the balancing transformation corresponding to the largest Hankel singular values.

This difficulty led to the approach described in (Antoulas and Sorensen, 2001a), and makes use of a third type of grammian, called the *cross-grammian*. The primary goal of this approach is to develop *implicit restarting methods* that will iteratively

produce approximations of a specified rank  $k$  to controllability, observability, and cross grammians that are best approximations of rank  $k$  to the full grammians. These considerations lead to algorithms within the implicit restarting framework that will iteratively improve a set of  $k$  leading singular values and the corresponding  $k$ -dimensional Krylov basis instead of computing all the singular values and truncating. This reduces computational costs from  $\mathcal{O}(n^3)$  to  $\mathcal{O}(n \cdot k^2)$ , and storage costs from  $\mathcal{O}(n^2)$  to  $\mathcal{O}(n \cdot k)$ .

A second approach along the same lines was developed in (Hodel *et al.*, 1996). It proposes an approximate power method for the iterative solution of the Lyapunov equation.

In the sections that follow we will describe the methods mentioned above in detail.

### 3.1. SVD-Based Approximation Methods

#### 3.1.1. Singular Value Decomposition: SVD

Every matrix  $A \in \mathbb{R}^{n \times m}$  can be decomposed into a product of three matrices (Golub and Van Loan, 1989):

$$A = U \Sigma V^* \in \mathbb{R}^{n \times m},$$

where  $U$  and  $V$  are unitary (orthogonal) and  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n) \in \mathbb{R}^{n \times m}$ , is diagonal with nonnegative diagonal entries called *singular values*:  $\sigma_i = \sqrt{\lambda_i(A^*A)} \geq \sigma_{i+1}$ , and the 2-induced norm of  $A$  is  $\sigma_1$ . The left singular vectors are  $U = (u_1 \ u_2 \ \dots \ u_n)$ ,  $UU^* = I_n$ , and the right singular vectors are  $V = (v_1 \ v_2 \ \dots \ v_m)$ ,  $VV^* = I_m$ . Finally, the *dyadic decomposition* of  $A$  is

$$A = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \dots + \sigma_n u_n v_n^*.$$

**Optimal approximation in the 2-norm.** Given  $A \in \mathbb{R}^{n \times m}$ , find  $X \in \mathbb{R}^{n \times m}$ , rank  $X = k < \text{rank } A$ , such that an appropriate norm of the error  $E = A - X$  is minimized. This problem can be solved explicitly if we take as norm the 2-norm or any unitarily invariant norm. Although the problem is non-convex, it readily follows that a special decomposition, namely, the singular value decomposition (SVD), provides the solution. The theorem that follows is attributed to several people. Two surveys in this regard are worth noticing, namely, (Chu *et al.*, 1995) and (Hubert *et al.*, 2000).

**Theorem 1.** (Schmidt-Mirsky, Eckart-Young). *The solution of the above problem is:*

$$\min_{\text{rank } X \leq k} \|A - X\|_2 = \sigma_{k+1}(A).$$

A non-unique minimizer  $X_{\#}$  is obtained by truncating the dyadic decomposition of  $A$ :  $X_{\#} = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \dots + \sigma_k u_k v_k^*$ .

**Example 1.** In Fig. 2 a supernova and the clown pictures from Matlab are considered. They can be represented by means of a  $406 \times 406$  and a  $200 \times 320$  matrix, respectively,



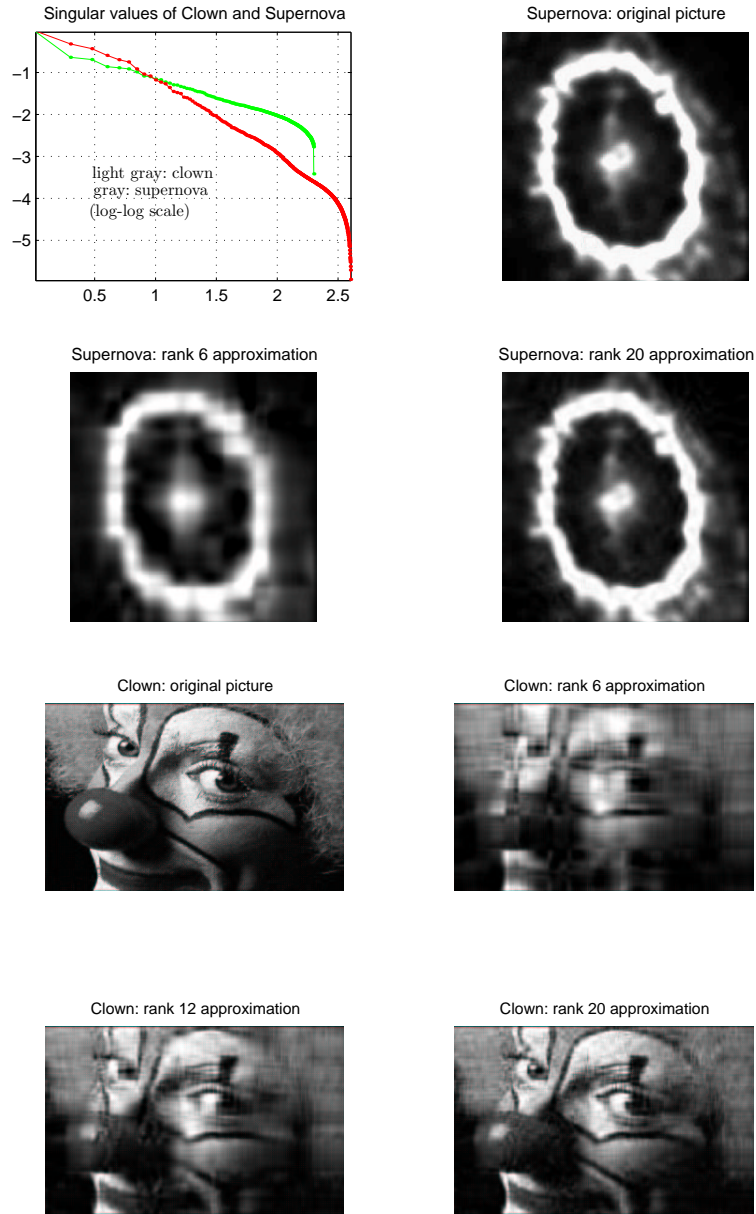


Fig. 2. Approximation of the “Supernova” and “Clown” by images of lower rank.

each entry (pixel) having a value between 0 (white) and 1 (black) in 64 gray levels. Both the matrices have full rank, i.e. 404 and 200, respectively. Their numerical ranks, however, are much lower. The singular values of these two images are shown

in the top-left subplot on a log-log scale; both sets of singular values fall off rapidly, and hence low-rank approximations with a small error (distortion) are possible. By superimposing the two singular value plots, we can also draw conclusions on the relative error for approximants of the same rank. Thus for approximants of rank up to 10, the clown image is easier to approximate, while for approximants of rank higher than 10, the converse is true. For the top plots, we have: the singular values, the original, and approximants of rank 20 and 6, respectively. For the bottom plots, clockwise from top: original, and approximants of rank 6, rank 20, and rank 12, respectively.  $\blacklozenge$

### 3.1.2. Proper Orthogonal Decomposition (POD) Methods

Consider the nonlinear system given by (1). For a fixed input  $u$ , the state trajectory at certain instances of time  $t_k$  is measured:

$$\mathcal{X} = [x(t_1) \ x(t_2) \ \cdots \ x(t_N)] \in \mathbb{R}^{n \times N}.$$

This is called a matrix of snapshots of the state. In general,  $N \gg n$ . Next, the singular value decomposition of  $\mathcal{X}$  is computed. If the singular values of this matrix fall off rapidly, a low-order approximation of this system can be computed,

$$\mathcal{X} = U\Sigma V^* \approx U_k \Sigma_k V_k^*, \quad k \ll n.$$

Equivalently, we can compute the eigenvalue decomposition of the *grammian*

$$\mathcal{X}\mathcal{X}^* = \sum_{t_i} x(t_i)x(t_i)^* = U\Sigma\Sigma^*U^* \approx U_k \Sigma_k \Sigma_k^* U_k^*, \quad k \ll n.$$

Let  $x(t) \approx U_k \xi(t)$ ,  $\xi(t) \in \mathbb{R}^k$ . Then  $U_k \dot{\xi}(t) = f(U_k \xi(t), u(t))$ , which implies the reduced-order state equation:

$$\dot{\xi}(t) = U_k^* f(U_k \xi(t), u(t)), \quad y(t) = h(U_k \xi(t), u(t)).$$

Thus the approximation  $\xi(t)$  of the state  $x(t)$  evolves on a *low-dimensional* space which is spanned by the  $k$  leading columns of  $U$ , i.e., by the leading left singular vectors of  $\mathcal{X}$ . For details, see (Berkooz *et al.*, 1996; Romo *et al.*, 1995; Volkwein, 1999). An improved version of this method is given in (Lall *et al.*, 1998).

### 3.1.3. Approximation by Balanced Truncation

Consider the system  $\Sigma$  given by (3), with  $A$  stable. Following the POD method, we consider the output to a particular input, namely, the impulse  $u = \delta$ ; in this case the output is called the *impulse response* and is denoted by  $h(t) = Ce^{At}B$ ,  $t \geq 0$ . This can be decomposed into an input-to-state map  $x(t) = e^{At}B$ , and a state-to-output map  $\eta(t) = Ce^{At}$ . Thus the input  $\delta(t)$  causes the state  $x(t)$ , while the initial condition  $x(0)$  causes the output  $y(t) = \eta(t)x(0)$ . The grammians corresponding to  $x$  and  $\eta$  are

$$\mathcal{P} = \sum_t x(t)x(t)^* = \int_0^\infty e^{At}BB^*e^{A^*t} dt$$

and

$$\mathcal{Q} = \sum_t \eta(t)^* \eta(t) = \int_0^\infty e^{A^*t} C^* C e^{At} dt.$$

Under the state transformation  $\hat{x} = Tx$ , the two grammians are transformed by congruence:  $\hat{\mathcal{P}} = T\mathcal{P}T^*$ ,  $\hat{\mathcal{Q}} = T^{-*}\mathcal{Q}T^{-1}$ ; thus  $\lambda_i(\mathcal{P}\mathcal{Q})$  are *input-output invariants* of  $\Sigma$ , called the *Hankel singular values* of  $\Sigma$ . These are fundamental invariants which determine how well  $\Sigma$  can be approximated by a reduced-order  $\hat{\Sigma}$ . They play the same role for dynamical systems that the singular values play for finite-dimensional matrices. Their *computation* is therefore of primary importance.

### 3.1.4. Computation of the Singular Values

The key for the computation of the Hankel singular values is the observation that the grammians  $\mathcal{P}$  and  $\mathcal{Q}$  are the unique Hermitian positive definite solutions to the following linear matrix equations, which are known as *Lyapunov equations*:

$$A\mathcal{P} + \mathcal{P}A^* + BB^* = 0, \quad A^*\mathcal{Q} + \mathcal{Q}A + CC^* = 0 \Rightarrow \sigma_i(\Sigma) = \sqrt{\lambda_i(\mathcal{P}\mathcal{Q})}. \quad (5)$$

Then the Hankel singular values  $\sigma_i(\Sigma)$  of the system  $\Sigma$  are the square roots of the eigenvalues of the product  $\mathcal{P}\mathcal{Q}$ .

Numerically stable algorithms to solve these equations are well known (see, e.g., (Varga, 1999)). One can directly compute an upper triangular  $U$  with  $\mathcal{P} = UU^*$  and a lower triangular  $L$  with  $\mathcal{Q} = LL^*$ . If  $A, B, C$  are real, then so are  $U$  and  $L$ . The Hankel singular values are computed as the singular values of the product  $U^*L$  and, with care, this may be computed directly without actually forming the product (Bartels and Stewart, 1972; Heath *et al.*, 1986; Laub *et al.*, 1987).

Denoting the SVD of  $U^*L$  by  $U^*L = Z\Sigma Y^*$ , a balanced realization can be constructed through a balancing transformation  $T_b$  defined by

$$T_b = \Sigma^{\frac{1}{2}} Z^* U^{-1} = \Sigma^{-\frac{1}{2}} Y^* L^* \quad \text{and} \quad T_b^{-1} = U Z \Sigma^{-\frac{1}{2}} = L^{-*} Y \Sigma^{\frac{1}{2}}.$$

Thus, in order to avoid inverting possibly not so well conditioned matrices, the formulas for  $T_b$  and  $T_b^{-1}$  involving only the inverse of  $\Sigma$  are used. Under this transformation, the solutions to the two Lyapunov equations become equal and diagonal:  $\mathcal{P} = \mathcal{Q} = \Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ , and

$$A_b \Sigma + \Sigma A_b^* + B_b B_b^* = 0, \quad A_b^* \Sigma + \Sigma A_b + C_b^* C_b = 0,$$

where  $A_b = T_b A T_b^{-1}$ ,  $B_b = T_b B$ ,  $C_b = C T_b^{-1}$ . Model reduction is now possible through simple truncation. Let  $\Sigma_2$  contain the small Hankel singular values and partition (in the balanced basis)

$$A_b = \left( \begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right), \quad B = \left( \begin{array}{c} B_1 \\ B_2 \end{array} \right), \quad C = (C_1 \mid C_2), \quad \Sigma = \left( \begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & \Sigma_2 \end{array} \right).$$

The reduced-order model is obtained by simple *truncation*

$$\hat{\Sigma} = \left( \begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & \end{array} \right),$$

that is, by taking the  $k \times k$ ,  $k \times m$ ,  $p \times k$  leading blocks of  $A_b$ ,  $B_b$ ,  $C_b$ , respectively; this system satisfies  $k$ -th order Lyapunov equations with diagonal solution  $\Sigma_1$ . This truncation leads to a balanced reduced-order system. If this truncation is such that the resulting grammians contain the largest  $k$  singular values  $\sigma_1$  through  $\sigma_k$ , then the  $\mathcal{H}_\infty$ -norm of the error system has the following upper (and lower) bound (Enns, 1984; Glover, 1984):

$$\sigma_k \leq \|\Sigma - \hat{\Sigma}\|_\infty \leq 2(\sigma_{k+1} + \sigma_{k+2} + \cdots + \sigma_n). \quad (6)$$

The  $\mathcal{H}_\infty$  norm  $\|\Sigma\|_\infty$ , of a system  $\Sigma$  is defined as the maximum of the highest peak of the frequency response, i.e., as the largest singular value of the transfer function evaluated on the imaginary axis (i.e., of the frequency response)  $\sigma_{\max}[D + C(j\omega - A)^{-1}B]$ . This result provides the rigorous justification for using the reduced-order model to predict behavior and enact control of the full system. The upper bound for the error system given above translates to an upper bound for the energy of the output error signal.

### 3.1.5. Other Types of Balancing

Besides the balancing based on Lyapunov equations, there are others. The principle is to find a transformation which would make solutions to appropriate Lyapunov and/or Riccati equations equal and diagonal (Ober, 1991). Here is a short description of each of them:

- **Lyapunov balancing.** Solution to Lyapunov equations equal and diagonal:  $A\mathcal{P} + \mathcal{P}A^* + BB^* = 0$ ,  $A^*Q + QA + C^*C = 0$ . *Reduced model:* preserves stability; there exists an error bound.
- **LQG balancing.** Solution to Riccati equations equal and diagonal:  $A\mathcal{P} + \mathcal{P}A^* + BB^* - \mathcal{P}C^*C\mathcal{P} = 0$ ,  $A^*Q + QA^* + C^*C - QBB^*Q = 0$ . It can be applied to unstable systems. *Reduced system:* leads to feedback design for a full system.
- **Stochastic balancing.** Solution to Lyapunov-Riccati equations equal and diagonal:  $A\mathcal{P} + \mathcal{P}A^* + BB^* = 0$ ,  $A^*Q + QA + C_W^*C_W = 0$ , where  $C_W = D^{-1}(C - B_W^*Q)$  and  $B_W = \mathcal{P}C^* + BD^*$ . This leads to model reduction which belongs to those satisfying a *relative error bound*. Truncation preserves the minimum-phase property; if  $\sigma_i$ 's denote the singular values obtained in this case, the relative error bound is

$$\|H^{-1}(H - H_k)\|_\infty \leq \prod_{i=k+1}^n \left[ \frac{1 + \sigma_i}{1 - \sigma_i} \right] - 1.$$

- **Positive real balancing.** The (smallest) positive definite solution to two Riccati equations below, equal and diagonal:  $A^*Q + QA + (C^* - QB)(D + D^*)^{-1}(C - B^*Q) = 0$ ,  $AP + PA^* + (PC^* - B)(D + D^*)^{-1}(CP - B^*) = 0$ . *Passivity:*  $\sigma_i < 1$ . *Reduced model:* preserves passivity.

### 3.1.6. Hankel Norm Approximation

A similar method is *Hankel norm approximation*. This is a direct generalization of the Schmidt-Mirsky, Eckart-Young result regarding integral operators resulting from LTI (linear, time-invariant) systems. We will briefly outline the construction method in what follows. For details we refer to (Adamjan *et al.*, 1971; 1978; Glover, 1984); also see (Antoulas, 1998; 1999a) as well as the books (Antoulas, 2002; Obinata and Anderson, 2000).

#### Construction of Hankel norm approximants:

1. Find  $\tilde{\Sigma}$  such that  $\Sigma - \tilde{\Sigma}$  is all-pass<sup>1</sup> with norm  $\sigma_k$ .
2. Project  $\tilde{\Sigma}$  onto its stable part  $\tilde{\Sigma}_+$ ; the optimal approximant is  $\hat{\Sigma} = \tilde{\Sigma}_+$ . As in the balanced truncation case, a Hankel norm approximation has the following **guaranteed properties**:
  - (a) stability is preserved, and
  - (b) the global error bound (6) holds.

Its **drawbacks** are

- *dense* computations, since matrix factorizations/inversions are required; these properties imply that the computations may be ill-conditioned.
- The whole transformed system is needed in order to truncate the desired subsystem. This leads to a number of operations of the order  $\mathcal{O}(n^3)$  and storage of the order  $\mathcal{O}(n^2)$ .

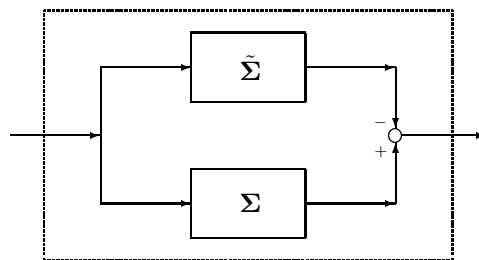


Fig. 3. Construction of Hankel norm approximations.

<sup>1</sup> A system  $\Sigma$  is all-pass or unitary if for all input/output pairs  $u, y$  the ratio  $\|y\|/\|u\|$  is constant.

We conclude this subsection with a series of remarks concerning balanced and Hankel norm approximations.

**Remark 1.**

(a) To enhance accuracy, one can obtain the balanced reduced-order model without explicitly computing the balancing transformation; this was first introduced in (Varga, 1991).

(b) Due to the importance of the Lyapunov or, more generally, of the Sylvester equation in systems theory many authors have studied various aspects of the solution of this equation in detail. For some sample papers we refer besides the classical paper (Lancaster, 1970) to more recent works (Avrachenkov and Lasserre, 2000; Ghavimi and Laub, 1996; Simoncini, 1996); the Sylvester and Lyapunov equations were also studied in the papers (Datta and Saad, 1991; de Souza and Bhattacharyya, 1981; Ghavimi and Laub, 1996; Hodel and Misra, 1997; Hodel *et al.*, 1996; Hu and Reichel, 1992; Penzl, 2001; Saad, 1990).

(c) *Ill-conditioning in balanced truncation.* In (Roth, 1952) the following equivalence was developed:

$$AX + XA^* + Q = 0 \Leftrightarrow \begin{pmatrix} A & Q \\ 0 & -A^* \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & -A^* \end{pmatrix}.$$

Thus with  $V_1^*V_1 + V_2^*V_2 = I$ ,  $\det V_2 \neq 0$ ,  $H = -V_2^{-1}A^*V_2$  and

$$\begin{pmatrix} A & Q \\ 0 & -A^* \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} H \Rightarrow X = V_1V_2^{-1}.$$

This means that the solution can be computed by determining the eigenspace of the composite matrix belonging to  $A$ . This requires the decoupling of the spectra of  $A$  and  $-A^*$ . The *conclusion* is that ill-conditioning occurs when  $A$  has eigenvalues *close* to the imaginary axis, i.e.,  $\Sigma$  has dominant poles *close* to the imaginary axis.

(d) *Canonical form.* There is a canonical form associated with balanced representations. In the single-input single-output (SISO) case, with distinct singular values, the three matrices have the form

$$A_{i,j} = -\frac{b_i b_j}{s_i s_j \sigma_i - \sigma_j}, \quad B_{i,1} = b_i, \quad C_{1,i} = s_i b_i,$$

where  $s_i$  are  $\pm 1$ ,  $b_i > 0$  and  $\sigma_i$  are the Hankel singular values. For details see (Ober, 1991).

(e) *Interpretation of balanced basis.* Given the state  $x$ , let  $\mathcal{E}_r$  denote the minimal amount of the input energy needed to steer  $0 \rightarrow x$ . Let also  $\mathcal{E}_o$  be the maximal output observation energy obtained by observing the output caused by the initial condition  $x: x \rightarrow 0$ . It follows that  $\mathcal{E}_o^2 = x^* Q x$ ,  $\mathcal{E}_r^2 = x^* P^{-1} x$ , and in a balanced basis the product is equal to one:  $\mathcal{E}_r \cdot \mathcal{E}_o = 1$ .

(f) As has already been mentioned, balancing for linear systems is equivalent to POD on the *impulse response*.

(g) Model reduction of unstable or descriptor systems can be carried out by various means, e.g., LQG balancing, frequency domain grammians (Zhou, 1999), or normalized coprime factorizations (Varga, 1998).

(h) *Weighted approximation* is also possible: for instance, frequency weighted balanced truncation with explicit  $\mathcal{L}_\infty$  error bounds, or a weighted Hankel approximation with anti-stable weighting (Enns, 1984; Kim *et al.*, 1995; Zhou, 1995).

(i) We conclude these remarks with a comment on the tightness of the bound (6). In fact, equality holds for systems  $\Sigma$  having zeros which interlace the poles; an example of such systems are electric circuits which are composed of resistors and capacitors (RC circuits).

### 3.2. Krylov-Based Approximation Methods

The *moments* associated with the linear system  $\Sigma$  (3), at some point  $s_0 \in \mathbb{C}$ , are defined as the moments of its impulse response  $h$  weighted by the exponential  $e^{s_0 t}$ , namely,

$$\eta_k = \int_0^\infty (-t)^k h(t) e^{-s_0 t} dt = \frac{d^k}{ds^k} H(s) \Big|_{s=s_0} .$$

These numbers can be expressed in terms of the series expansion of the transfer function  $H(s) = D + C(sI - A)^{-1}B$  of  $\Sigma$  around  $s_0 \in \mathbb{C}$ ; the transfer function is the Laplace transform of the impulse response:

$$H(s) = \eta_0 + \eta_1(s - s_0) + \eta_2(s - s_0)^2 + \eta_3(s - s_0)^3 + \dots .$$

The problem now consists in finding  $\hat{\Sigma}$  of (4), with

$$\hat{H}(s) = \hat{\eta}_0 + \hat{\eta}_1(s - s_0) + \hat{\eta}_2(s - s_0)^2 + \hat{\eta}_3(s - s_0)^3 + \dots ,$$

such that for appropriate  $\ell$

$$\eta_j = \hat{\eta}_j, \quad j = 1, 2, \dots, \ell.$$

The key to the success of these methods is that they can be implemented *iteratively*. These implementation methods are closely related to *iterative eigenvalue computations* (Sorensen, 1999). Thus there is a strong interaction between model reduction for large-scale systems and large-scale eigenvalue computations; this is explored in (Antoulas, 2002).

If the expansion is around infinity, the moments are known as *Markov parameters*, and the problem is known as *partial realization*; its solution is computed through the *Lanczos* and *Arnoldi* procedures. If the expansion is around zero, the problem becomes *Padé approximation*. In general, for arbitrary  $s_0 \in \mathbb{C}$ , we have the problem of *rational interpolation*, which amounts to matching moments at the specified point  $s_0$ . This

interpolation is obtained indirectly using the so-called *rational Lanczos* procedure which operates with  $(A - s_0 I)^{-1}$  in place of  $A$ . The inverse operation is applied by solving triangular systems obtained from an initial LU decomposition (using a sparse direct method if  $A$  is sparse).

It should be noted that the computation of moments is *numerically problematic*, because the powers of eigenvalues of  $A$  grow exponentially fast in general (the eigenvalues of  $A$  have negative real parts, many of them are larger in magnitude than one). Thus the key fact for numerical reliability is that if  $\Sigma$  is given by means of  $(A, B, C)$ , Krylov methods allow moment matching *without* explicit computation of moments. For an overview as well as details on the material presented in this section we refer to (Antoulas *et al.*, 1996; Gallivan *et al.*, 1994b; 1995; Gragg, 1972; Gragg and Lindquist, 1983; Gutknecht, 1994; Grimme, 1997; Ruhe, 1984; Sorensen, 1999; Van Dooren, 1995).

### 3.2.1. Lanczos and Arnoldi Processes

In this section we will briefly review the Lanczos and Arnoldi processes. For simplicity, we will do so in the SISO (Single-Input Single-Output) case, i.e., no block versions will be considered. There are two ways of looking at these algorithms: the conventional one and the system-theoretic one, the latter of which is less common.

**Conventional representation.** Given  $A \in \mathbb{R}^{n \times n}$  and vectors  $B, C^* \in \mathbb{R}^n$ , the two-sided Lanczos and the Arnoldi processes construct at each step matrices  $W_k, V_k \in \mathbb{R}^{n \times k}$  such that the following relationships are satisfied:

Lanczos and Arnoldi algorithms	
Lanczos	Arnoldi
Initialization	
$V_1 = \frac{1}{\sqrt{ CB }} B, \quad W_1 = \frac{\text{sign}(CB)}{\sqrt{ CB }} C^*$	$V_1 = \frac{1}{\ B\ } B$
$t$ -th step	
$AV_t = V_t T_t + r_t e_t^*$ $A^* W_t = W_t T_t^* + q_t e_t^*$ $W_t^* V_t = I_t, \quad t = 2, 3, \dots,$	$AV_t = V_t H_t + g_t e_t^*$  $V_t^* V_t = I_t, \quad t = 2, 3, \dots,$

where  $e_t \in \mathbb{R}^n$  is the  $t$ -th unit vector. The columns of  $V_t$  in the Arnoldi case are *orthogonal*, while in the Lanczos case the columns of  $V_t$  and  $W_t$  are called *bi-orthogonal*. Furthermore, in the Lanczos case  $T_t$  is *tridiagonal* and in the Arnoldi case  $H_t$  is *upper Hessenberg*.



**System theoretic representation.** We will first define the  $t \times n$  *observability* matrix  $\mathcal{O}_t$ , the  $n \times t$  *reachability* matrix  $\mathcal{R}_t$ , and the  $t \times t$  *Hankel* matrix  $\mathcal{H}_t$  associated with  $\Sigma$ :

$$\mathcal{O}_t = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{t-1} \end{bmatrix}, \quad \mathcal{R}_t = \begin{bmatrix} B & AB & \cdots & A^{t-1}B \end{bmatrix},$$

$$\mathcal{H}_t := \begin{bmatrix} \eta_1 & \eta_2 & \cdots & \eta_t \\ \eta_2 & \eta_3 & \cdots & \eta_{t+1} \\ \vdots & & \ddots & \\ \eta_t & \eta_{t+1} & \cdots & \eta_{2t-1} \end{bmatrix}.$$

It is well-known that  $\mathcal{H}_t = \mathcal{O}_t \mathcal{R}_t$ . The *key step* in the Lanczos procedure is to compute the *LU factorization* of  $\mathcal{H}_k$ :

$$\mathcal{H}_t = LU = \mathcal{O}_t \mathcal{R}_t,$$

where  $L$  is lower triangular and  $U$  is upper triangular, with  $L(i, i) = \pm U(i, i)$ . Such a factorization exists provided that  $\det \mathcal{H}_t \neq 0$ , for all  $t$ . We now define the projection  $\pi_R \pi_L$ , where

$$\pi_L = L^{-1} \mathcal{O}_t, \quad \pi_R = \mathcal{R}_t U^{-1}.$$

Similarly, the *key step* in the Arnoldi algorithm is the computation of the *QR factorization* of  $\mathcal{R}_t$ :

$$\mathcal{R}_t = VU,$$

where  $V \in \mathbb{R}^{n \times t}$ ,  $V^* V = I_t$ , and  $U$  is upper triangular. This yields the projection  $\pi_R \pi_L$ , where  $\pi_R = \pi_L^* = V$ . These two projections are used to define reduced order systems  $\hat{\Sigma}$ :

$$\hat{\Sigma} = \left( \begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \end{array} \right), \quad \hat{A} := \pi_L A \pi_R, \quad \hat{B} = \pi_L B, \quad \hat{C} = C \pi_R.$$

*Properties:* (a)  $\pi_L \pi_R = 1$ , and (b)  $\pi_R \pi_L$  is a projection. In the Lanczos case for the reduced system, let the number of moments matched be  $2k$ . Then: (i)  $\hat{\Sigma}$  matches  $2k$  Markov parameters; (ii)  $\hat{A}$  is tridiagonal; (iii)  $\hat{B}, \hat{C}^*$  are multiples of the unit vector  $e_1$ . In the Arnoldi case, let the number of moments matched be  $k$ ; the reduced system (i) matches  $k$  Markov parameters, (ii)  $\hat{A}$  is in upper Hessenberg form, and (iii)  $\hat{B}$  is a multiple of  $e_1$ .

Lanczos/Arnoldi: Iterative Implementation	
<p style="text-align: center;"><b>Two-sided Lanczos algorithm</b></p> <ol style="list-style-type: none"> <li>1. <math>\beta_1 := \sqrt{ B^*C^* }</math>, <math>\gamma_1 := \text{sgn}(B^*C^*)\beta_1</math>, <math>v_1 := B/\beta_1</math>, <math>w_1 := C^*/\gamma_1</math></li> <li>2. For <math>j = 1, \dots, k</math>, set <ol style="list-style-type: none"> <li>(a) <math>\alpha_j := w_j^*Av_j</math></li> <li>(b) <math>r_j := Av_j - \alpha_jv_j - \gamma_jv_{j-1}</math>, <math>q_j := A^*w_j - \alpha_jw_j - \beta_jw_{j-1}</math></li> <li>(c) <math>\beta_{j+1} = \sqrt{ r_j^*q_j }</math>, <math>\gamma_{j+1} = \text{sgn}(r_j^*q_j)\beta_{j+1}</math></li> <li>(d) <math>v_{j+1} = r_j/\beta_{j+1}</math>, <math>w_{j+1} = q_j/\gamma_{j+1}</math></li> </ol> </li> </ol>	<p style="text-align: center;"><b>The Arnoldi algorithm</b></p> <ol style="list-style-type: none"> <li>1. <math>v_1 := \frac{v}{\ v\ }</math>, <math>w := Av_1</math>; <math>\alpha_1 := v_1^*w</math> <math>f_1 := w - v_1\alpha_1</math>; <math>V_1 := (v_1)</math>; <math>H_1 := (\alpha_1)</math></li> <li>2. For <math>j = 1, 2, \dots, k</math> <math display="block">\beta_j := \ f_j\ , v_{j+1} := \frac{f_j}{\beta_j}</math> <math display="block">V_{j+1} := (V_j \ v_{j+1}), \hat{H}_j = \begin{pmatrix} H_j \\ \beta_j e_j^* \end{pmatrix}</math> <math display="block">w := Av_{j+1}, h := V_{j+1}^*w,</math> <math display="block">f_{j+1} = w - V_{j+1}h, H_{j+1} := (\hat{H}_j \ h)</math> </li> </ol>

The results quoted above are based on the following result.

**Lemma 1. (Degree Lemma).**

**Arnoldi process.** Consider  $V \in \mathbb{R}^{n \times k}$  and  $H \in \mathbb{R}^{k \times k}$ , which are generated at the  $k$ -th step of the Arnoldi process. For any polynomial  $\phi$  of degree less than or equal to  $k$ , there holds

$$\phi(A)Ve_1 = V\phi(H)e_1, \quad \deg \phi < k,$$

$$\phi(A)Ve_1 = V\phi(H)e_1 + f\gamma, \quad \deg \phi = k, \quad \gamma \in \mathbb{R}.$$

**Lanczos process.** Consider  $V, W \in \mathbb{R}^{n \times k}$ ,  $T \in \mathbb{R}^{k \times k}$ , which are generated at the  $k$ -th step by the Lanczos process. For any polynomial  $\phi$  of degree less than or equal to  $k$ , there hold:

$$\phi(A)Ve_1 = V\phi(T)e_1, \quad \phi(A^*)We_1 = W\phi(T^*)e_1, \quad \deg \phi < k,$$

$$\phi(A)Ve_1 = V\phi(T)e_1 + r\gamma, \quad \phi(A^*)We_1 = W\phi(T^*)e_1 + q\eta, \quad \deg \phi = k, \quad \gamma, \eta \in \mathbb{R}.$$

From this simple property it follows that

$$cA^j b = cA^j v_1 = cV_k H_k^j e_1 = e_1^* H_k^j e_1, \quad 1 \leq j < k$$

for the Arnoldi process, and

$$cA^{\ell+j} b = w_1^* A^\ell A^j v_1 = (e_1^* T_k^\ell W_k^*) (V_k T_k^j e_1) = e_1^* T_k^{\ell+j} e_1, \quad 0 \leq \ell \leq k-1, \quad 0 \leq j \leq k$$

for the Lanczos process. Thus, Arnoldi matches  $k$  moments while Lanczos matches  $2k$  moments. In the following table we summarize the three different moment matching methods, together with the moments matched, the points in the complex plane where they are matched, and the number of moments matched.

Krylov methods			
Method	Moments matched	Points in $\mathbb{C}$	Number of moments
1. $k$ steps Arnoldi/Lanczos	$cA^j b$	$\infty$	$k/2k$
2. Shift-Invert	$c(A - \sigma_0 I)^{-j-1} b$	$\sigma_0$	$k/2k$
3. Rational Krylov	$c(A - \sigma_i I)^{-j-1} b$	$\sigma_1, \dots, \sigma_i$	$k = \ell_1 + \dots + \ell_i$

**Implicitly restarted Arnoldi and Lanczos procedures.** The issue here is to get a better approximation of some desired set of preferred eigenvalues, e.g., those eigenvalues that have largest modulus or largest/smallest positive/negative real part. An idea due to Sorensen (Sorensen, 1992; Lehoucq *et al.*, 1998) is to accomplish this by projecting onto the space of desired features, by *restarting* Lanczos and Arnoldi, while staying in the Lanczos/Arnoldi framework. An overview of applications of implicit restarting can be found in (Sorensen, 1995).

**Remark 2.**

(a) In the Arnoldi algorithm, if  $A$  is symmetric, then  $H_j$  is tridiagonal, and the Arnoldi algorithm coincides with the one-sided or symmetric Lanczos algorithm.

(b) The realization and partial realization problems were introduced by Kalman (1979). Their generalization is a rational interpolation. A general version of this problem was solved in (Antoulas *et al.*, 1990), and with stability constraint in (Antoulas and Anderson, 1989). The recursive interpolation problem is studied in (Antoulas, 1993; Antoulas and Willems, 1993). For connections of these results with Lanczos, see (Antoulas *et al.*, 1996). A generalization of the tridiagonal form of  $A$  for MIMO systems is described in (Antoulas and Bishop, 1987). The connection between partial realizations and the Lanczos algorithm was studied in (Gragg, 1972; Gragg and Lindquist, 1983).

(c) In (numerical) linear algebra the terminology *Krylov subspace*  $\mathcal{K}_t$  is used to denote the space spanned by the first  $t$  columns of the reachability (or the observability) matrices  $\mathcal{K}_t = \text{span col } \mathcal{R}_t$ . In systems theory these subspaces are known as reachability and observability subspaces, respectively.

(d) The residual in the Arnoldi case is  $f_j = Av_j - V_j h_j$ , where  $h_j$  is chosen so that the norm  $\|f_j\|$  is minimized. It turns out that  $V_j^* h_j = 0$  and  $h_j = V_j^* Av_j$ , that is,  $f_j = (I - V_j V_j^*) Av_j$ .

(e) **Comparison between Lanczos and Arnoldi algorithms.** These two algorithms are similar but have important differences. While Lanczos yields a Padé-type approximation, this is not true for Arnoldi. Then Arnoldi is numerically stable, while Lanczos is not. The number of operations needed to compute a  $k$ -th order approximant is  $\mathcal{O}(k^2 n)$  for both (as opposed to  $\mathcal{O}(n^3)$  operations required for SVD-based methods). The storage required is  $2kn$  in both the cases. The basic operation is  $w = Av$  and  $\hat{v} = Av$ ,  $\hat{w} = A^* w$ , respectively, and the cost is  $2k$ ; this is followed by

$v_+ = (I - VV^*)w$ , and  $v_+, w_+$  which have a cost of  $n\mathcal{O}(k^2)$ . Finally, it should be mentioned that the Lanczos process satisfies a *three-term recurrence* relation, while Arnoldi does not. This implies the definition of *orthogonal polynomials* associated with Lanczos.

(f) An implementation of Krylov methods requires *matrix-vector* multiplications exclusively, and no matrix factorizations and/or inversions are needed. Furthermore, there is no need to compute the transformed  $n$ -th order model and then truncate. This means the absence of ill-conditioning.

(g) Drawbacks:

- There is no global error bound; for a local result, see (Bai and Ye, 1998).
- $\hat{\Sigma}$  may not be stable, even if  $\Sigma$  is stable. Remedy: *implicit restart* (Grimme *et al.*, 1995).
- Lanczos breaks down if  $\det \mathcal{H}_i = 0$ . Remedy: *look-ahead* methods (Gutknecht, 1994).
- $\hat{\Sigma}$  tends to approximate the high frequency poles of  $\Sigma$ . Hence the steady-state error may be significant.  
Remedy: match expansions around other frequencies (*rational Lanczos*) (Ruhe, 1984; Gallivan *et al.*, 1995; Grimme, 1997).

### 3.3. SVD-Krylov-Based Approximation Methods

Up to this point we have briefly reviewed the merits and difficulties associated with two fundamental ways of carrying out model reduction. We reached the conclusion that *SVD*-based approximation methods have a number of desirable properties, namely, there exists a global error bound and stability is preserved. The drawback, however, is that applicability is restricted to relatively low dimensions. On the other hand, the *Krylov*-based methods remedied the last issue (i.e., applicability) but had a number of important drawbacks, namely, no guaranteed stability preservation and no global error bound. These considerations give rise to the *SVD-Krylov*-based approach, which aims at combining the best attributes of the two previously described approximation methods.

The basic ingredients of *SVD*-based methods which give rise to all the nice properties of this approach are the two Lyapunov equations used to compute the grammians. On the other hand, the basic ingredients of the *Krylov*-based methods are precisely the *Krylov* subspaces, i.e., the span of the columns of the reachability and observability matrices, and the iterative way of computing projections onto these subspaces. The *SVD-Krylov* methods thus aim at combining these two sets of attributes by seeking iterative solutions of Lyapunov or, more generally, of Sylvester equations. In doing so the hope is that there will be devised a method which combines the best attributes of the *SVD* and *Krylov* methods. Such a method does not exist yet. The quest, however, continues.

Two methods in this category are (Antoulas and Sorensen, 2001a; Hodel *et al.*, 1996; Jaimoukha and Kasenally, 1997). In the sequel we will briefly describe the approach in the former reference.

The difficulty (pointed out earlier) of working with two Lyapunov equations separately, as proposed in (Jaimoukha and Kasenally, 1997), has led to the consideration of a different approach which is based upon the notion of a cross-grammian that was introduced in (Fernando and Nicholson, 1983). The cross-grammian  $X$  is defined for square systems ( $m = p$ ) as the solution to the Sylvester equation  $AX + XA + BC = 0$ . In the SISO (Single-Input Single-Output) case, the solution  $X$  is similar to a Hermitian matrix and the eigenvalues of  $X$  are actually signed singular values of  $\Sigma$ . In fact, it is possible to show in this case that  $X^2 = \mathcal{P}\mathcal{Q}$ . Our approach consists essentially in constructing low-rank  $k$  approximate solutions to this system by setting  $X = V\hat{X}W^*$  with  $W^*V = I_k$  and projecting using  $W$  and  $V$  as the reduced basis. This uses an implicit restart mechanism that allows the computation of the best-rank  $k$  approximation to  $X$ . This solution leads to an error estimate.

The primary goal of (Antoulas and Sorensen, 2001a) is to develop *implicit restarting methods* that will iteratively produce approximations of a specified rank  $k$  to controllability, observability, and cross grammians that are best approximations of rank  $k$  to the full grammians. We observe that the Hankel singular values of such systems decay extremely rapidly. Hence very low-rank approximations to the system grammians are possible and accurate low-order reduced models will result. The approach consists in investigating a non-Krylov method related to the recently developed Jacobi-Davidson method to the large scale eigenvalue problem. This has the advantage of admitting a wider variety of restarting options. However, the advantage is gained at the expense of giving up the Krylov properties which provide deep connections with the fundamental systems theory. Nevertheless, our method will be based on a quantity called the cross-grammian and because of this some important system theoretic properties are indeed retained which are not necessarily related to moment matching.

This method uses subspace projection techniques combined with variants of implicit restarting to compute a partially balanced realization directly. Instead of computing all the Hankel singular values and then casting out the smallest ones, an implicit restarting mechanism is used to compute just the largest ones and thereby greatly reduces the computational and storage costs required to produce a  $k$ -th order reduced model. Thus computational costs are reduced from  $\mathcal{O}(n^3)$  to  $\mathcal{O}(n \cdot k^2)$ , and storage costs from  $\mathcal{O}(n^2)$  to  $\mathcal{O}(n \cdot k)$ .

We conclude this section by summarizing in the table below the salient features of each one of the three methods. The features of the third column form a *wish list*, as there is currently no method which fulfills all of them.

**Remark 3.**

- An approximation method which has been popular is model reduction by *modal approximation* (Varga, 1995). This method works well when the dominant eigenfrequencies of the original system are of interest and have to be preserved.

SVD methods	Krylov methods	SVD-Krylov methods
1. preservation of stability 2. global error bound 3. applicability $n \lesssim 100$	1. numerical efficiency 2. applicability: $n \gg 100$	1. preservation of stability 2. global error bound 3. numerical efficiency 4. applicability: $n \gg 100$

- All systems which can be well approximated using SVD-based methods have Hankel singular values which drop off rapidly. Therefore an important issue in model reduction is to determine the rate of decrease in the singular values. Few results in this direction are known, and this remains an open problem (Antoulas, 1999b).
- We would like to draw the readers' attention to two new books, one of which is already available (Obinata and Anderson, 2000), and the other is in preparation (Antoulas, 2002; Antoulas and Van Dooren, 2000).
- SVD-based model reduction has been extended to time-varying systems; for an overview, see (Van Dooren, 2000; Varga, 2000).

#### 4. Numerical Experiments

Matlab codes implementing all of the methods mentioned above have already been developed; furthermore, a new software library called SLICOT has been developed for control and contains a package which implements all the dense model reduction methods; for details, we refer to (Varga, 1999). A package which is used for large-scale eigenvalue computation is ARPACK (Lehoucq *et al.*, 1998). It should be mentioned that one of the largest examples computed involves the determination of stability of a matrix resulting from the discretization of a 3D Navier Stokes equation. The size is several million (Lehoucq and Salinger, 2001). This is relevant to the paper at hand because the techniques used are the same as those outlined above.

Most of the problems we have seen exhibit the behavior that is indicative of the success for this approach. Namely, the singular values of  $\Sigma$  decay very rapidly indicating that  $\Sigma$  is typically approximated well with a reduced order system of very low dimension. In the following example we show the results of computing  $\hat{\Sigma}$ , where its dimension  $k$  is determined by a cut off. We determine  $k$  by taking the first  $k$  such that

$$\sigma_{k+1} < \tau \sigma_1.$$

A typical value of  $\tau$  is  $\tau = \sqrt{\text{machine precision}}$ . The important point is that the specification of this single tolerance determines everything in the SVD-based methods. The order of the reduced model and the eigenvalue placement are completely automatic. Moreover, the value of  $\tau$  directly determines the error of approximation for  $\|y - y_k\|$  over the entire frequency range. As has already been noted, the key parameters for model reduction are the *Hankel singular values*.

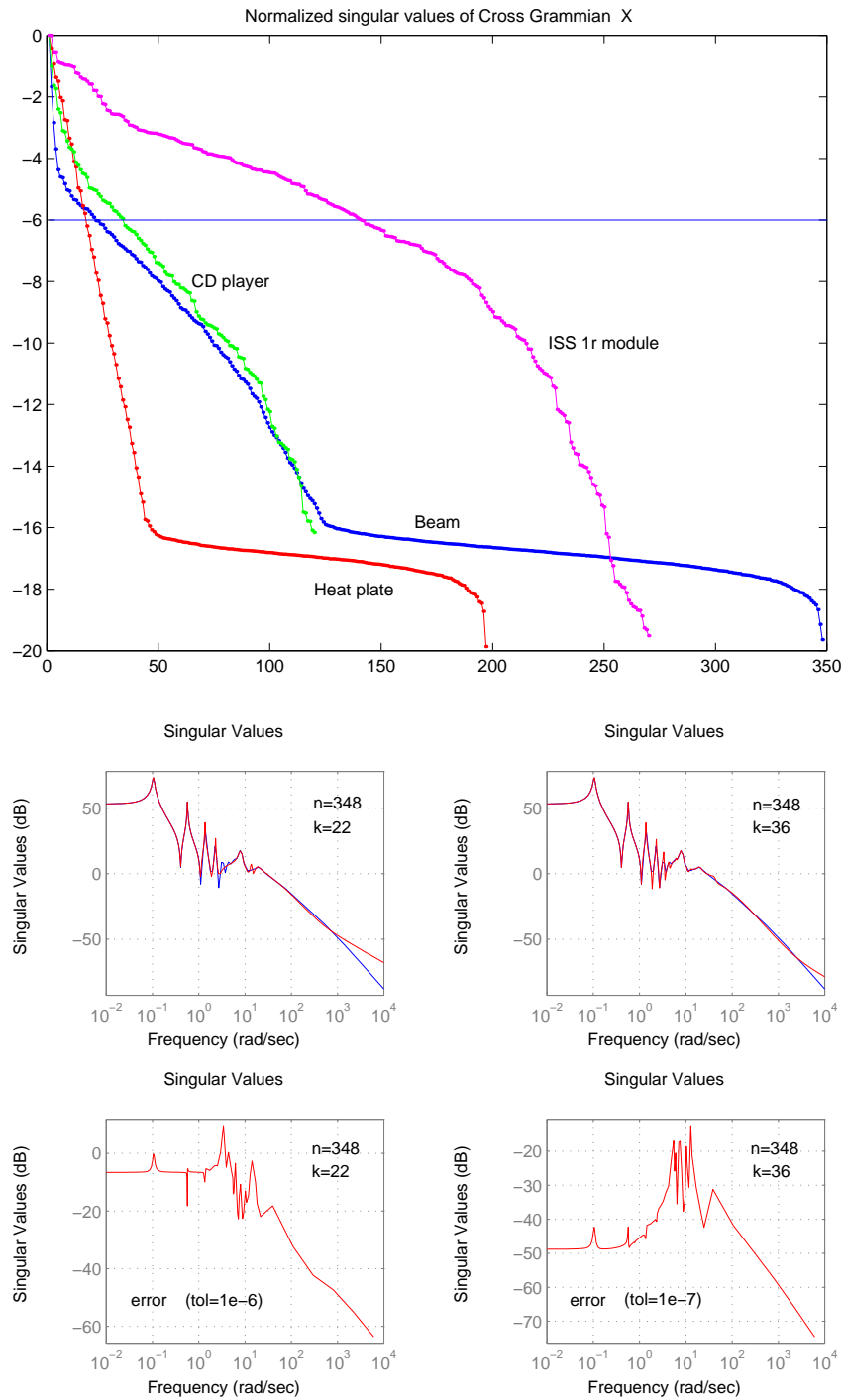


Fig. 4. Hankel singular values for four examples and results for clamped beam displacements.

In Fig. 4, the top plot side compares the Hankel singular values of four systems: a finite-element discretization of a heat plate, a finite-element discretization of a clamped beam, the model of a component of a CD player, and a model of the flex modes of the 1r module of the ISS (International Space Station). It follows from this plot that for a give tolerance the ISS model is the hardest to approximate, as its Hankel singular values decay the least rapidly.

The bottom plot shows results for a finite-element calculation of the displacement and stress in the clamped beam with Rayleigh (or proportional) damping. The input is a force at the unclamped end, while the output is the displacement. The figure shows a Bode plot of the response of the original system which is of order 348 (solid line) and the response of the reduced-order model of order 48 (dashed line). The order of the reduced model is automatically determined once the error requirement is specified.

## 5. Unifying Features and Conclusions

The unifying feature of all model reduction methods is that the main operation is *projection*. Given  $\Sigma$  in (3) and a *projection*  $\pi = \pi_L \pi_R$  (i.e.,  $\pi_R \pi_L = I_k$ ), a reduced-order model  $\hat{\Sigma}$  is obtained by projecting  $\Sigma$  as follows:

$$\hat{\Sigma} = \left( \begin{array}{c|c} \pi_L A \pi_R & \pi_L B \\ \hline C \pi_R & \end{array} \right) : \frac{d\hat{x}}{dt} = [\pi_L A \pi_R] \hat{x} + [\pi_L B] u, \quad y = [C \pi_R] \hat{x}.$$

The *quality* of the reduced model is measured in terms of the frequency response

$$H(j\omega) = D + C(j\omega I - A)^{-1} B.$$

*Choices of projectors* for the reduction methods discussed above are:

1. *Balanced truncation*: Project onto the dominant eigenspace of  $\mathcal{P}\mathcal{Q}$ , where  $\mathcal{P}$ ,  $\mathcal{Q}$  are the grammians.
2. *Optimal Hankel norm approximation*: Solve for grammians. Embed in a lossless transfer function and project onto its stable eigenspace.
3. *Krylov-based approximation*: Project onto controllability and/or observability spaces, i.e.,  $\mathcal{R}$  and  $\mathcal{O}$ .
4. *New method* (Antoulas and Sorensen, 2001a): Project onto the dominant eigenspace of the best rank- $k$  approximant  $X_k$  of the cross-grammian  $X$ .

**Complexity issues** (Antoulas and Van Dooren, 2000). The major costs associated with computing balanced and Hankel norm approximants are approximately  $30n^3$  for computing the grammians and approximately  $25n^3$  for performing the balancing. For rational Krylov at  $k$  points,  $\frac{2}{3}kn^3$  operations are needed. Of course, various iterative methods (sign, disc, Smith) can accelerate the computation of grammians, in particular on parallel machines. For approximate or sparse decompositions, we need  $c_1 \alpha kn$  operations to compute the grammians and  $\mathcal{O}(n^3)$  operations for embedding



or balancing. The corresponding number for rational Krylov is  $c_2kan$ , where  $k$  is the number of expansion points and  $\alpha$  is the average number of non-zero elements per row in  $A$ .

## 6. Conclusions

We have outlined the current trends in the computation of reduced-order models for large-scale systems. The SVD-based methods and the Krylov-based methods each enjoy disjoint sets of advantages. Therefore a third approach referred to as SVD-Krylov, which aims at combining the best attributes of these two methods, has emerged.

## References

- Adamjan V.M., Arov D.Z. and Krein M.G. (1971): *Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem*. — Math. USSR Sbornik, Vol.15, pp.31–73.
- Adamjan V.M., Arov D.Z. and Krein M.G. (1978): *Infinite block Hankel matrices and related extension problems*. — Amer. Math. Soc. Trans., Vol.111, pp.133–156.
- Antoulas A.C. (1993): *Recursive modeling of discrete-time series*, In: IMA volume on Linear Algebra for Control (P. Van Dooren and B.F. Wyman, Eds.). — pp. 1–22.
- Antoulas A.C. (1998): *Approximation of linear operators in the 2-norm*. — Lin. Alg. Applicns., Vol.278, pp.309–316.
- Antoulas A.C. (1999a): *Approximation of linear dynamical systems*, In: Wiley Encyclopedia of Electrical and Electronics Engineering, Vol.11 (J.G. Webster, Ed.). — pp.403–422.
- Antoulas A.C. (1999b): *On eigenvalues and singular values of linear dynamical systems*. — Proc. XVI Housholder Symp., Chateau Whistler.
- Antoulas A.C. (2002): *Approximation of Large-Scale Dynamical Systems*. — Philadelphia: SIAM Press (to appear).
- Antoulas A.C. and Anderson B.D.O. (1989): *On the problem of stable rational interpolation*. — Lin. Alg. Applicns., Vol.122–124, pp.301–329.
- Antoulas A.C. and Bishop R.H. (1987): *Continued-fraction decomposition of linear systems in the state space*. — Syst. Contr. Lett., Vol.9, pp.43–53.
- Antoulas A.C. and Sorensen D.C. (2001a): *Projection methods for balanced model reduction*. — Techn. Rep. ECE-CAAM Depts., Rice University.
- Antoulas A.C. and Van Dooren P.M. (2000): *Approximation of Large-Scale Dynamical Systems*. — Notes of a Short Course given at the SIAM Annual Meeting, Puerto Rico. (Notes available from the authors' websites.)
- Antoulas A.C. and Willems J.C. (1993): *A behavioral approach to linear exact modeling*. — IEEE Trans. Automat. Contr., Vol.AC-38, pp.1776–1802.
- Antoulas A.C., Ball J.A., Kang J. and Willems J.C. (1990): *On the solution of the minimal rational interpolation problem*. — Lin. Alg. Applicns., Vol.137/138, pp.511–573.

- Antoulas A.C., Grimme E.J. and Sorensen D. (1996): *On behaviors, rational interpolation and the Lanczos algorithm*. — Proc. IFAC World Congress, San Francisco.
- Arnoldi W.E. (1951): *The principle of minimized iterations in the solution of the matrix eigenvalue problem*. — Quart. Appl. Math., Vol.9, pp.17–29.
- Avrachenkov K.E. and Lasserre J.B. (2000): *Analytic perturbation of Sylvester matrix equations*. — Proc. IEEE Conf. Decision and Control, Sydney.
- Bai Z. and Ye Q. (1998): *Error estimation of the Padé approximation of transfer functions via the Lanczos process*. — Electron. Trans. Numer. Anal., Vol.7, pp.1–17.
- Bai Z., Feldmann P. and Freund R.W. (1997): *Stable and passive reduced-order models based on partial Padé approximation via the Lanczos process*. — Numer. Anal. Manuscript 97–3–10, Bell Laboratories, Murray Hill.
- Bartels R.H. and Stewart G.W. (1972): *Solution of the matrix equation  $AX + XB = C$* . — Comm. ACM, Vol.15, pp.820–826.
- Berkooz G., Holmes P. and Lumley J.L. (1996): *Coherent Structures, Dynamical Systems and Symmetry*. — Cambridge: Cambridge University Press.
- Boley D.L. (1994): *Krylov space methods on state-space control models*. — Circ. Syst. Signal Process., Vol.13, pp.733–758.
- Boley D.L. and Golub G.H. (1984): *The Lanczos algorithm and controllability*. — Syst. Contr. Lett., Vol.4, pp.317–324.
- Chu M.T., Funderlic R.E. and Golub G.H. (1995): *A rank-one reduction formula and its applications to matrix factorizations*. — SIAM Rev., Vol.37, pp.512–530.
- Datta B.N. and Saad Y. (1991): *Arnoldi methods for large Sylvester-like observer matrix equations and an associated algorithm for partial spectrum assignment*. — Lin. Alg. Applics., Vol.154–156, pp.225–244.
- de Souza E. and Bhattacharyya S.P. (1981): *Controllability, observability and the solution of  $AX - XB = C$* . — Lin. Alg. Applics., Vol.39, pp.167–188.
- Enns D.F. (1984): *Model reduction with balanced realizations: An error bound and frequency weighted generalization*. — Proc. 23rd IEEE Conf. Decision and Control, pp.127–132.
- Feldmann P. and Freund R.W. (1995): *Efficient linear circuit analysis by Padé approximation via the Lanczos process*. — IEEE Trans. Comp.-Aided Design, Vol.14, pp.639–649.
- Fernando K.V. and Nicholson H. (1983): *On the structure of balanced and other principal representations of SISO systems*. — IEEE Trans. Automat. Contr., Vol.AC–28, pp.228–231.
- Freund R.W. (1999): *Reduced-order modeling techniques based on Krylov subspaces and their use in circuit simulation*, In: Applied and Computational Control, Signals, and Circuits (B.N. Datta, Ed.) — Boston: Birkhäuser, Vol.1, pp.435–498.
- Gallivan K., Grimme E.J. and Van Dooren P. (1994a): *Asymptotic waveform evaluation via a restarted Lanczos method*. — Appl. Math. Lett., Vol.7, pp.75–80.
- Gallivan K., Grimme E. and Van Dooren P. (1994b): *Padé Approximation of large-scale dynamic systems with Lanczos methods*. — Proc. 33rd IEEE Conf. Decision and Control, Lake Buena Vista, FL.
- Gallivan K., Grimme E. and Van Dooren P. (1995): *A rational Lanczos algorithm for model reduction*. — CSRD Rep. No.1411, University of Illinois at Urbana-Champaign, IL.

- Ghavimi A.R. and Laub A.J. (1996): *Numerical methods for nearly singular constrained matrix Sylvester equations*. — SIAM J. Matrix Anal. Appl., Vol.17, pp.212–221.
- Glover K. (1984): *All optimal Hankel-norm approximations of linear multivariable systems and their  $L^\infty$ -error bounds*. — Int. J. Contr., Vol.39, pp.1115–1193.
- Golub G.H. and Van Loan C.F. (1989): *Matrix Computations*. — Baltimore, London: The Johns Hopkins University Press.
- Gragg W.B. (1972): *The Padé table and its relation to certain algorithms of numerical analysis*. — SIAM Rev., Vol.14, pp.1–62.
- Gragg W.B. and Lindquist A. (1983): *On the partial realization problem*. — Lin. Alg. Applicns., Vol.50, pp.277–319.
- Grimme E.J. (1997): *Krylov Projection Methods for Model Reduction*. — Ph.D. Thesis, ECE Dept., Univ. of Illinois, Urbana-Champaign.
- Grimme E.J., Sorensen D.C. and Van Dooren P. (1995): *Model reduction of state space systems via an implicitly restarted Lanczos method*. — Num. Algorithms, Vol.12, pp.1–31.
- Gutknecht M.H. (1994): *The Lanczos process and Padé approximation*. — Proc. Cornelius Lanczos Intl. Centenary Conf., (J.D. Brown *et al.*, Eds.), pp.61–75, Philadelphia: SIAM.
- Heath M.T., Laub A.J., Paige C.H. and Ward R.C. (1986): *Computing the singular value decomposition of a product of two matrices*. — SIAM J. Sci. Stat. Computing, Vol.7, pp.1147–1159.
- Hodel A.S. and Misra P. (1997): *Least-squares approximate solution of overdetermined Sylvester equations*. — SIAM J. Matrix Anal. Applicns., Vol.18, pp.279–290.
- Hodel A.S., Tenison R.B. and Poola K. (1996): *Numerical solution of large Lyapunov equations by approximate power iteration*. — Lin. Alg. Applicns., Vol.236, pp.205–230.
- Hu D. and Reichel L. (1992): *Krylov-subspace methods for the Sylvester equation*. — Lin. Alg. Applicns., Vol.172, pp.283–313.
- Hubert L., Meuleman J. and Heiser W. (2000): *Two purposes for matrix factorization: A historical appraisal*. — SIAM Rev., Vol.42, pp.68–82.
- Jaimoukha I.M. and Kasenally E.M. (1997): *Implicitly restarted Krylov subspace methods for stable partial realizations*. — SIAM J. Matrix Anal. Applicns., Vol.18, pp.633–652.
- Kalman R.E. (1979): *On partial realizations, transfer functions, and canonical forms*. — Acta Polytech. Scand. Ma., Vol.31, pp.9–32.
- Kamon M., Wang F. and White J. (2000): *Generating nearly optimal compact models from Krylov-subspace based reduced-order models*. — IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Process., Vol.47, pp.239–248.
- Kim S.W., Anderson B.D.O. and Madievski A.G. (1995): *Error bound for transfer function order reduction using frequency weighted balanced truncation*. — Syst. Contr. Lett., Vol.24, pp.183–192.
- Lall S., Marsden J.E. and Glavaski S. (1998): *Empirical model reduction of controlled nonlinear systems*. — Techn. Rep. CIT-CDS-98-008, California Institute of Technology.
- Lancaster P. (1970): *Explicit solutions of matrix equations*. — SIAM Rev., Vol.12, pp.544–566.
- Lanczos C. (1950): *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.

- Laub A.J., Heath M.T., Paige C.C. and Ward R.C. (1987): *Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms.* — IEEE Trans. Automat. Contr., Vol.AC-32, pp.115–121.
- Lehoucq R.B. and Salinger A.G. (2001): *Large-scale eigenvalue calculations for stability analysis of steady flows on massively parallel computers.* — Int. J. Numer. Meth. Fluids, Vol.36, pp.309–327.
- Lehoucq R.B., Sorensen D.C. and Yang C. (1998): *ARPACK: User's Guide.* — Philadelphia: SIAM.
- Li J., Wang F. and White J. (1999): *An efficient Lyapunov equation-based approach for generating reduced-order models of interconnect.* — Proc. 36th IEEE/ACM Design Automation Conf., New Orleans, LA.
- Moore B.C. (1981): *Principal component analysis in linear systems: Controllability, observability and model reduction.* — IEEE Trans. Automat. Contr., Vol.AC-26, pp.17–32.
- Ober R. (1991): *Balanced parametrization of classes of linear systems.* — SIAM J. Contr. Optim., Vol.29, pp.1251–1287.
- Obinata G. and Anderson B.D.O. (2000): *Model Reduction for Control System Design.* — New York: Springer.
- Penzl T. (2001): *A cyclic low rank Smith method for large sparse Lyapunov equations.* — SIAM J. Sci. Comp. (to appear).
- Pillage L.T. and Rohrer R.A. (1990): *Asymptotic waveform evaluation for timing analysis.* — IEEE Trans. Comp. Aided Design, Vol.9, pp.352–366.
- Romo T.D., Clarage J.B., Sorensen D.C. and Phillips Jr. G.N. (1995): *Automatic identification of discrete substates in proteins: Singular value decomposition analysis of time-averaged crystallographic refinements.* — Proteins: Structure, Function, and Genetics, Vol.22, pp.311–321.
- Roth W.E. (1952): *The equation  $AX - YB = C$  and  $AX - XB = C$  in matrices.* — Proc. Amer. Math. Soc., Vol.3, pp.392–396.
- Ruhe A. (1984): *Rational Krylov sequence methods for eigenvalue computation.* — Lin. Alg. Applics., Vol.58, pp.391–405.
- Saad Y. (1990): *Numerical solution of large Lyapunov equations,* In: Signal Processing, Scattering and Operator Theory, and Numerical Methods, Proc. MTNS-89 — Vol.3, pp.503–511, Boston: Birkhäuser.
- Simoncini V. (1996): *On the numerical solution of  $AX - XB = C$ .* — BIT, Vol.36, pp.814–830.
- Sorensen D.C. (1992): *Implicit application of polynomial filters in a k-step Arnoldi method.* — SIAM J. Matrix Anal. Applic., Vol.13, pp.357–385.
- Sorensen D.C. (1995): *Implicitly restarted Arnoldi/Lanczos methods and large-scale SVD applications,* In: SVD and Signal Processing III (M. Moonen and B. de Moor, Eds.) — Amsterdam: Elsevier.
- Sorensen D.C. (1999): *Lecture notes on numerical methods for large scale eigenvalue problems.* — Draft of lectures delivered at the Royal Institute of Technology KTH, Stockholm, Spring.
- Stroobandt D. (2000): *Recent advances in system-level interconnect prediction.* — IEEE Circuits and Systems Newsletter, Vol.11, pp.4–20.

- Van Dooren P. (1995): *The Lanczos algorithm and Padé approximations*. — Short Course, Benelux Meeting on Systems and Control.
- Van Dooren P.M. (2000): *Gramian based model reduction of large-scale dynamical systems*, In: Numerical Analysis 1999 — London: Chapman and Hall, CRC Press, pp.231–247.
- Varga A. (1991): *Balancing-free square-root algorithms for computing singular perturbation approximations*. — Proc. 30th IEEE CDC, Brighton, UK, pp.1062–1065.
- Varga A. (1995): *Enhanced modal approach for model reduction*. — Math. Modell. Syst., Vol.1, pp.91–105.
- Varga A. (1998): *Computation of normalized coprime factorizations of rational matrices*. — Syst. Contr. Lett., Vol.33, pp.37–45.
- Varga A. (1999): *Model reduction routines for SLICOT*, Niconet Rep. No.8. Report available from: <ftp://wgs.esat.kuleuven.ac.be/pub/WGS/REPORTS/NIC1999--8.ps.Z>
- Varga A. (2000): *Balanced truncation model reduction of periodic systems*. — Proc. IEEE CDC, Session WeP05, Sydney.
- Volkwein S. (1999): *Proper orthogonal decomposition and singular value decomposition*. — Tech. Rep. SFB–153, Institut für Mathematik, Universität Graz.
- Zhou K. (1995): *Frequency-weighted  $\mathcal{L}_\infty$  norm and optimal Hankel norm model reduction*. — IEEE Trans. Automat. Contr., Vol.AC–40, pp.1687–1699.
- Zhou K. (1999): *A new balanced realization and model reduction method for unstable systems*. — Proc. 14th IFAC World Congress, Beijing, Vol.D–2b–04–4, pp.123–128.