5

Model Reduction of Time-Varying Systems

Younes Chahlaoui[1] and Paul Van Dooren[2]

[1] School of Computational Science, Florida State University, Tallahassee, U.S.A.
younes.chahlaou@laposte.net
[2] CESAME, Université catholique de Louvain, Louvain-la-Neuve, Belgium
vdooren@csam.ucl.ac.be

Summary. This paper presents new recursive projection techniques to compute reduced order models of time-varying linear systems. The methods produce a low-rank approximation of the Gramians or of the Hankel map of the system and are mainly based on matrix operations that can exploit sparsity of the model. We show the practical relevance of our results with a few benchmark examples.

5.1 Introduction

The basic idea of model reduction is to represent a complex linear dynamical system by a much simpler one. This may refer to many different techniques, but in this paper we focus on projection-based model reduction of linear systems. It can be shown in the time-invariant case [GVV03] that projection methods allow to generate almost all reduced order models and that they are in that sense quite general. Here we construct the projection based on the dominant invariant subspaces of products of the Gramians, which are energy functions for ingoing and outgoing signals of the system. When the system matrices are large and sparse, the Gramians are nevertheless dense and efficient methods will therefore have to approximate these dominant spaces without explicitly forming the Gramians themselves.

Balanced Truncation [Moo81] is probably the most popular projection-based method. This is mainly due to its simplicity: the construction is based on simple linear algebra decompositions and there is no need to first choose a set of essential parameters. Moreover an a priori upper bound is given for the $\mathcal{H}_\infty$-norm of the error between the original plant and the reduced-order model [Enn81].

An important issue in model reduction is the choice of the order of the approximation, since it affects the quality of the approximation. One would like to be able to choose this during the construction of the reduced order model, i.e. without having to evaluate in advance quality measures like the Hankel singular values (computing them all would become prohibitive for large-scale

systems). The use of iterative methods seem appealing in this context since they may offer the possibility to perform order selection during the computation of the projection spaces and not in advance.

The approach that we propose in this paper is iterative and applies as well to time-varying systems. Earlier work on model reduction of time-varying systems was typically based on the explicit computation of the time-varying solution of a matrix difference (or differential) equation [SSV83, IPM92, SR02] and such results were mainly used to prove certain properties or bounds of the reduced order model. They were in other words not presented as an efficient computational tool. We propose to update at each step two sets of basis vectors that allow to identify the dominant states. The updating equations are cheap since they only require sparse matrix vector multiplications. The ideas are explained in Chapter 24 and [CV03a, CV03b, Cha03], to which we refer for proofs and additional details. Another recent approach is to use fast matrix decomposition methods on matrices with particular structure such as a Hankel structure. Such an approach is presented in [DV98] and could be competitive with the methods presented here.

5.2 Linear Time-Varying Systems

Linear discrete *time-varying systems* are described by systems of difference equations:

$$\mathcal{S}: \quad \begin{cases} x_{k+1} = A_k x_k + B_k u_k \\ y_k = C_k x_k + D_k u_k \end{cases} \tag{5.1}$$

with input $u_k \in \mathbb{R}^m$, state $x_k \in \mathbb{R}^N$ and output $y_k \in \mathbb{R}^p$. In this paper we will assume $m, p \ll N$, the input sequence to be square-summable (i.e. $\sum_{-\infty}^{\infty} u_k^T u_k \leq \infty$), $D_k = 0$, and the matrices $\{A_k\}_{-\infty}^{\infty}$, $\{B_k\}_{-\infty}^{\infty}$, and $\{C_k\}_{-\infty}^{\infty}$ to be bounded for all k. Using the recurrence (5.1) over several time steps, one obtains the state at step k in function of past inputs over the interval $[k_i, k-1]$:

$$x_k = \Phi(k, k_i) x_{k_i} + \sum_{i=k_i}^{k-1} \Phi(k, i+1) B_i u_i$$

where $\Phi(k, k_i) := A_{k-1} \dots A_{k_i}$ is the discrete *transition matrix* over time period $[k_i, k-1]$. The transition matrix has the following properties:

$$\begin{cases} \Phi(k_2, k_0) = \Phi(k_2, k_1)\Phi(k_1, k_0), \ k_0 \leq k_1 \leq k_2 \\ \Phi(k, k) \ = I_N \qquad \forall k. \end{cases}$$

We will assume the time-varying system $\mathcal{S}$ to be *asymptotically stable*, meaning

$$\forall k \geq k_i \quad \|\Phi(k, k_i)\| \leq c \cdot a^{(k-k_i)}, \text{ with } c > 0, \ 0 < a < 1.$$

The *Gramians* over intervals $[k_i, k-1]$ and $[k, k_f]$ are then defined as follows:

$$\mathcal{G}_c(k) = \sum_{i=k_i}^{k-1} \Phi(k, i+1) B_i B_i^T \Phi^T(k, i+1),$$

$$\mathcal{G}_o(k) = \sum_{i=k}^{k_f} \Phi^T(i, k) C_i^T C_i \Phi(i, k),$$

where k_i may be $-\infty$ and k_f may be $+\infty$. It follows from the identities

$$\Phi(k_1, k_2) = \Phi(k_1, k_2+1) A_{k_2} \quad \text{and} \quad \Phi(k_1+1, k_2) = A_{k_1} \Phi(k_1, k_2),$$

where $k_1 \geq k_2$, that these Gramians can also be obtained from the Stein recurrence formulas:

$$\mathcal{G}_c(k+1) = A_k \mathcal{G}_c(k) A_k^T + B_k B_k^T \quad \text{and} \quad \mathcal{G}_o(k) = A_k^T \mathcal{G}_o(k+1) A_k + C_k^T C_k, \tag{5.2}$$

with respective initial conditions

$$\mathcal{G}_c(k_i) = 0, \quad \mathcal{G}_o(k_f + 1) = 0.$$

Notice that the first equation evolves "forward" in time, while the second one evolves "backward" in time.

These Gramians can also be related to the input/output map in a particular window $[k_i, k_f]$. Let us at each instant k ($k_i < k < k_f$) restrict inputs to be nonzero in the interval $[k_i, k)$ (i.e. "the past") and let us consider the outputs in the interval $[k, k_f]$ (i.e. the "future"). The state-to-outputs and inputs-to-state maps on this window are then given by :

$$\underbrace{\begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k_f} \end{bmatrix}}_{Y} = \begin{bmatrix} C_k \\ C_{k+1} A_k \\ \vdots \\ C_{k_f} \Phi(k_f, k) \end{bmatrix} \underbrace{\begin{bmatrix} B_{k-1} & A_{k-1} B_{k-2} & \dots & \Phi(k, k_i+1) B_{k_i} \end{bmatrix}}_{x(k)} \underbrace{\begin{bmatrix} u_{k-1} \\ u_{k-2} \\ \vdots \\ u_{k_i} \end{bmatrix}}_{U}.$$

The finite dimensional *Hankel matrix* $\mathcal{H}(k_f, k, k_i)$ mapping U to Y is defined as

$$\mathcal{H}(k_f, k, k_i) = \begin{bmatrix} C_k B_{k-1} & C_k A_{k-1} B_{k-2} & \dots & C_k \Phi(k, k_i+1) B_{k_i} \\ C_{k+1} A_k B_{k-1} & C_{k+1} A_k A_{k-1} B_{k-2} & & C_{k+1} \Phi(k+1, k_i+1) B_{k_i} \\ \vdots & & \ddots & \vdots \\ C_{k_f} \Phi(k_f, k) B_{k-1} & C_{k_f} \Phi(k_f, k-1) B_{k-2} & \dots & C_{k_f} \Phi(k_f, k_i+1) B_{k_i} \end{bmatrix}.$$

Notice that this matrix has at most rank N since $x(k) \in \mathbb{R}^N$ and that it factorizes as

$$\mathcal{H}(k_f,k,k_i)=\underbrace{\begin{bmatrix} C_k \\ C_{k+1}A_k \\ \vdots \\ C_{k_f}\Phi(k_f,k)\end{bmatrix}}_{\mathcal{O}(k_f,k)}\underbrace{\begin{bmatrix} B_{k-1} & A_{k-1}B_{k-2} & \dots & \Phi(k,k_i+1)B_{k_i}\end{bmatrix}}_{\mathcal{C}(k,k_i)} \quad (5.3)$$

where $\mathcal{O}(k_f,k)$ and $\mathcal{C}(k,k_i)$ are respectively the *observability* and the *controllability* matrices at instant k over the finite window $[k_i,k_f]$. They satisfy the recurrences

$$\mathcal{O}(k_f,k)=\begin{bmatrix} C_k \\ \mathcal{O}(k_f,k+1)A_k\end{bmatrix}, \quad \mathcal{C}(k+1,k_i)=\begin{bmatrix} B_k & A_k\mathcal{C}(k,k_i)\end{bmatrix} \quad (5.4)$$

evolving forward and backward in time, respectively. From these matrices one then constructs the Gramians and Hankel map via the identities

$$\begin{aligned}\mathcal{H}(k_f,k,k_i)&=\mathcal{O}(k_f,k)\mathcal{C}(k,k_i),\\ \mathcal{G}_c(k)&=\mathcal{C}(k,k_i)\mathcal{C}(k,k_i)^T,\\ \mathcal{G}_o(k)&=\mathcal{O}(k_f,k)^T\mathcal{O}(k_f,k).\end{aligned}$$

Notice that in the time-invariant case the above matrices become function only of the differences $k-k_i$ and k_f-k. In this case one typically chooses both quantities equal to $\tau:=(k_f-k_i)/2$, i.e. half the considered window length. In the time-invariant case it is also typical to consider the infinite window case, i.e. where $k_f=-k_i=\infty$.

5.3 Balanced Truncation

The method of *Balanced Truncation* is a very popular technique of model reduction for stable linear time-invariant systems because it has several appealing properties related to sensitivity, stability and approximation error [Moo81, ZDG95]. The extension to time-varying systems is again based on the construction of a new state-space coordinate system in which both Gramians are diagonal and equal [SSV83, VK83, SR02]. This is always possible when the system is uniformly controllable and observable over the considered interval [SSV83, VK83], meaning that the Gramians are uniformly bounded and have uniformly bounded inverses. It is then known that there exists a time-varying state space transformation T_k such that the Gramians $\hat{\mathcal{G}}_c(k):=T_k^{-1}\mathcal{G}_c(k)T_k^{-T}$ and $\hat{\mathcal{G}}_o(k):=T_k^T\mathcal{G}_o(k)T_k$ of the transformed system $\{T_{k+1}^{-1}A_kT_k, T_{k+1}^{-1}B_k, C_kT_k\}$, satisfy

$$T_k^{-1}\mathcal{G}_c(k)\mathcal{G}_o(k)T_k=\hat{\mathcal{G}}_c(k)\hat{\mathcal{G}}_o(k)=\Sigma^2(k), \quad 0<\Sigma(k)<\infty I.$$

One then partitions the matrix $\Sigma(k)$ into $\mathrm{diag}\{\Sigma_+(k),\Sigma_-(k)\}$ where $\Sigma_+(k)$ contains the n largest singular values of $\Sigma(k)$ and $\Sigma_-(k)$ the smallest ones. In

that coordinate system the truncated system $\{\hat{A}_k, \hat{B}_k, \hat{C}_k\}$ is just the system corresponding to the leading n columns and rows of the transformed system $\{T_{k+1}^{-1} A_k T_k, T_{k+1}^{-1} B_k, C_k T_k\}$. If we denote the first n columns of T_k by X_k and the first n rows of T_k^{-1} by Y_k^T then $Y_k^T X_k = I_n$ and

$$\{\hat{A}_k, \hat{B}_k, \hat{C}_k\} := \{Y_{k+1}^T A_k X_k, Y_{k+1}^T B_k, C_k X_k\}. \tag{5.5}$$

If for all k there is also a gap between the singular values of $\Sigma_+(k)$ and those of $\Sigma_-(k)$, then similar properties to the time-invariant case can be obtained, namely asymptotic stability and uniform controllability and observability of the truncated model [SSV83] and an error bound for the truncation error between both input/output maps in terms of the neglected singular values $\Sigma_-(k)$ or of related matrix inequalities (see [LB03, SR02] for a more detailed formulation).

Rather than computing the complete transformations T_k, one only needs to compute the matrices $X_k, Y_k \in \mathbb{R}^{N \times n}$ whose columns span the "dominant" left and right eigenvector spaces of the product $\mathcal{G}_c(k)\mathcal{G}_o(k)$ and normalize them such that $Y_k^T X_k = I_n$ to obtain the reduced model as given above. One can show that both Gramians are no longer required to be non-singular, and this can therefore be applied as well to the finite window case. In general, one can not even guarantee the gap property of the eigenvalues of the product of the Gramians.

In order to reduce the complexity of the model reduction procedure one can try to approximate the dominant left invariant subspaces X_k and Y_k by an iterative procedure which possibly exploits the sparsity of the original model $\{A_k, B_k, C_k\}$. The projection matrices will hopefully be close to invariant subspaces and one can hope to derive bounds for the approximation error between both systems. Such a procedure is explained in the next two sections and is inspired by efficient approximation techniques found in the time-invariant case [GSA03]. Bounds will be derived for the time-invariant version of this algorithm.

5.4 Recursive Low-Rank Gramian Algorithm (RLRG)

Large scale system models $\{A_k, B_k, C_k\}$ are often sparse and since the construction of a good approximate time-varying system model $\{\hat{A}_k, \hat{B}_k, \hat{C}_k\}$ requires an approximation at every time step k it seems crucial to find a method that is of low complexity at every time step and therefore exploits the sparsity of the original model.

If the Gramians $\mathcal{G}_c(k)$ and $\mathcal{G}_o(k)$ of the system $\{A_k, B_k, C_k\}$ were of rank $n << N$, for all $k \in [k_i, k_f]$ then the system would be actually of degree n. The idea is thus to replace

$$\mathcal{G}_c(k) = \mathcal{C}(k, k_i)\mathcal{C}(k, k_i)^T \quad \text{and} \quad \mathcal{G}_o(k) = \mathcal{O}(k_f, k)^T \mathcal{O}(k_f, k)$$

by semi-definite rank n_k approximations

$$\mathcal{P}_k := S_k S_k^T \quad \text{and} \quad \mathcal{Q}_k := R_k R_k^T,$$

respectively (for simplicity, we will assume n_k constant and equal to n). If such a *factorized approximation* is available, then

$$\mathcal{G}_c(k)\mathcal{G}_o(k) \approx S_k S_k^T R_k R_k^T$$

and the right hand side has clearly $X_k := S_k$ as right invariant subspace, and $Y_k := R_k$ as left invariant subspace. Normalizing X_k and Y_k such that $Y_k^T X_k = I_n$ will then yield an appropriate projected system (5.5) at each step k.

Note that the Gramian recurrences (5.2) evolve forward and backward in time and so will the recurrences for the approximations. We introduce the indices

$$l := k_i + i, \quad r := k_f + 1 - i$$

to simplify the indexing of the low-rank updating equations. At step i we compute the singular value decompositions of the matrices

$$\left[\, B_{l-1} \middle| A_{l-1} S_{l-1} \,\right] \quad \text{and} \quad \left[\frac{C_r}{R_{r+1}^T A_r}\right],$$

which yield transformation matrices $U := \left[\, U_+ \middle| U_- \,\right]$ and $V := \left[\, V_+ \middle| V_- \,\right]$ defining

$$\left[\, S_l \middle| E_c(l) \,\right] := \left[\, B_{l-1} \middle| A_{l-1} S_{l-1} \,\right] \left[\, V_+ \middle| V_- \,\right], \tag{5.6}$$

$$\left[\, R_r \middle| E_o(r) \,\right] := \left[\, C_r^T \middle| A_r^T R_{r+1} \,\right] \left[\, U_+ \middle| U_- \,\right], \tag{5.7}$$

where $V_+ \in \mathbb{R}^{(m+n)\times n}$ and $U_+ \in \mathbb{R}^{(p+n)\times n}$. These iterations are initialized at step $i = 0$ with

$$S_{k_i} = 0 \quad \text{and} \quad R_{k_f+1} = 0.$$

A MATLAB-like procedure corresponding to these recurrences would be as follows.

Algorithm RLRG
$l = k_i; r = k_f + 1; \tau = r - l - 1; S_l = 0; R_r = 0;$
for $i = 1 : \tau;$
  $l = l + 1; M = \left[\, B_{l-1} \middle| A_{l-1} S_{l-1} \,\right];$
  $[U, \Sigma, V] = svd(M, 0); S_l = M * V(:, 1 : n);$
  $r = r - 1; M = \left[\, C_r^T \middle| A_r^T R_{r+1} \,\right];$
  $[V, \Sigma, U] = svd(M, 0); R_r = M * U(:, 1 : n);$
end

At each iteration, we need to multiply $A_{l-1}S_{l-1}$ and $R_{r+1}^T A_r$ (which requires $4Nn\alpha$ flops, where α is the average number of nonzero elements in

each row or column of the sparse matrices A_i) and perform the transformations U and V (which require $O(N(n+m)^2)$ flops and $O(N(n+p)^2)$ flops, respectively [GV96]). When $N \gg n > m, p, \alpha$ this is altogether linear in the largest dimension N. Notice that the matrices S_{l-1} and R_{r+1} are multiplied at each step by time-varying matrices, which seems to preclude adaptive SVD updating techniques such as those used in [GSA03].

At each iteration step, $E_c(l)$ and $E_o(r)$ are neglected, which corresponds to the best rank n approximations at that step. But we would like to bound the global errors

$$\mathcal{E}_c(l) := \mathcal{G}_c(l) - \mathcal{P}_l = \mathcal{G}_c(l) - S_l S_l^T, \quad \text{and} \quad \mathcal{E}_o(r) := \mathcal{G}_o(r) - \mathcal{Q}_r = \mathcal{G}_o(r) - R_r R_r^T.$$

The following lemma [CV02] is proven in [Cha03] and leads to such bounds.

Lemma 5.4.1. *At each iteration, there exists orthogonal matrices*

$$V^{(i)} \in \mathbb{R}^{(n+im)\times(n+im)} \quad \textit{and} \quad U^{(i)} \in \mathbb{R}^{(n+ip)\times(n+ip)},$$

satisfying:

$$\mathcal{C}(l, k_i) V^{(i)} = \left[\, S_l \middle| E_c(l) \middle| A_{l-1} E_c(l-1) \middle| \ldots \middle| \Phi(l, k_i+1) E_c(k_i+1) \,\right],$$

and

$$\mathcal{O}(k_f, r)^T U^{(i)} = \left[\, R_r \middle| E_o(r) \middle| A_r^T E_o(r+1) \middle| \ldots \middle| \Phi(k_f, r)^T E_o(k_f) \,\right],$$

where $E_c(i)$ and $E_o(i)$ are the neglected parts at each iteration.

The above identities then lead to expressions for the errors:

$$\mathcal{E}_c(l) = \sum_{j=1}^{i} \Phi(l, k_i+j) E_c(k_i+j) E_c(k_i+j)^T \Phi(l, k_i+j)^T, \tag{5.8}$$

$$\mathcal{E}_o(r) = \sum_{j=0}^{i-1} \Phi(k_f - j, r)^T E_o(k_f - j) E_o(k_f - j)^T \Phi(k_f - j, r). \tag{5.9}$$

It is shown in [CV02, Cha03] that the norms of $\mathcal{E}_c(l)$ and $\mathcal{E}_o(r)$ can then be bounded in terms of

$$\eta_c(l) = \max_{k_i+1 \le j \le l} \|E_c(j)\|_2, \quad \text{and} \quad \eta_o(r) = \max_{r \le j \le k_f} \|E_o(j)\|_2,$$

which we refer to as the "noise" levels η_c and η_o of the recursive singular value decompositions (5.6,5.7).

Theorem 5.4.2. *If the system (5.1) is stable, i.e.,*

$$\|\Phi(k, k_0)\| \le c \cdot a^{(k-k_0)}, \text{ with } c > 0,\ 0 < a < 1,$$

then

$$\|\mathcal{E}_c(l)\|_2 \le \frac{\eta_c^2(l) c^2}{1-a^2}, \quad \textit{and} \quad \|\mathcal{E}_o(r)\|_2 \le \frac{\eta_o^2(r) c^2}{1-a^2}.$$

5.4.1 Time-Invariant Case

It is interesting to note that for linear time-invariant systems $\{A, B, C\}$, the differences $\mathcal{E}_c(l)$ and $\mathcal{E}_o(r)$ remain bounded for large i, and this shows the strength of Theorem 5.4.2. We then have the following result, shown in [CV02, Cha03].

Theorem 5.4.3. *Let P and Q be the solutions of*

$$P = APA^T + I, \quad \textit{and} \quad Q = A^T QA + I,$$

then

$$\begin{aligned} \|\mathcal{E}_c(l)\|_2 &\leq \eta_c^2(l)\|P\|_2 \leq \eta_c^2(l)\tfrac{\kappa(A)^2}{1-\rho(A)^2}, \\ \|\mathcal{E}_o(r)\|_2 &\leq \eta_o^2(r)\|Q\|_2 \leq \eta_o^2(r)\tfrac{\kappa(A)^2}{1-\rho(A)^2}, \end{aligned} \tag{5.10}$$

$$\|\mathcal{G}_c(l)\mathcal{G}_o(r) - \mathcal{P}_l\mathcal{Q}_r\|_2 \leq \frac{\kappa(A)^2}{1-\rho(A)^2}\left(\eta_c^2(l)\|\mathcal{G}_o(r)\|_2 + \eta_o^2(r)\|\mathcal{G}_c(l)\|_2\right), \tag{5.11}$$

where $\kappa(A)$ is the condition number and $\rho(A)$ is the spectral radius of A.

In [GSA03], bounds very similar to (5.10) were obtained but the results in that paper only apply to the time-invariant case. The bound (5.11) says that if one Gramian is not well approximated, the product of the Gramians, which is related to the Hankel singular values, will not be well approximated. Notice that this only makes sense when $l = r$. In the time-invariant case one can also estimate the convergence to the infinite horizon Gramians, which we denote by $\mathcal{G}_c$ and $\mathcal{G}_o$ and are defined by he identities

$$\mathcal{G}_c = A\mathcal{G}_c A^T + BB^T, \quad \text{and} \quad \mathcal{G}_o = A^T\mathcal{G}_o A + C^T C.$$

Theorem 5.4.4. *At each step i of (5.6,5.7) we have the following error bounds*

$$\begin{aligned} \|\mathcal{P}_{i-1} - \mathcal{G}_c\|_2 &\leq \|\mathcal{P}_i - \mathcal{P}_{i-1} + E_c(i)E_c^T(i)\|_2\|P\|_2 \\ &\leq \|\mathcal{P}_i - \mathcal{P}_{i-1} + E_c(i)E_c^T(i)\|_2\frac{\kappa(A)^2}{1-\rho(A)^2}, \\ \|\mathcal{Q}_{i+1} - \mathcal{G}_o\|_2 &\leq \|\mathcal{Q}_i - \mathcal{Q}_{i+1} + E_o(i)E_o^T(i)\|_2\|Q\|_2 \\ &\leq \|\mathcal{Q}_i - \mathcal{Q}_{i+1} + E_o(i)E_o^T(i)\|_2\frac{\kappa(A)^2}{1-\rho(A)^2}, \end{aligned}$$

where $\kappa(A)$ is the condition number and $\rho(A)$ is the spectral radius of A.

Proof. We prove the result only for $\mathcal{P}_{i-1}$ since both results are dual. Start from

$$\mathcal{P}_i + E_c(i)E_c^T(i) = A\mathcal{P}_{i-1}A^T + BB^T,$$

to obtain

$$(\mathcal{G}_c - \mathcal{P}_{i-1}) = A(\mathcal{G}_c - \mathcal{P}_{i-1})A^T + (\mathcal{P}_i - \mathcal{P}_{i-1} + E_c(i)E_c(i)^T).$$

Use the solution P of the linear system $P = APA^T + I$ and its growth factor $\frac{\kappa(A)^2}{1-\rho(A)^2}$ to obtain from there the desired bound. □

This theorem says that when convergence is observed, we can bound the accuracy of the current estimates of the Gramians in terms of quantities computed in the last step only. Using very different arguments, is was mentioned in [Cha03] that this in fact holds approximately for the time-varying case as well.

5.4.2 Periodic Case

The simplest class of time-varying models is the class of periodic systems. This is because every K-periodic system,

$$\{A_{K+k}, B_{K+k}, C_{K+k}\} = \{A_k, B_k, C_k\}$$

is in fact equivalent [MB75] to K *lifted* time-invariant systems:

$$\begin{cases} \hat{x}_{k+1}^{(h)} = \hat{A}^{(h)}\hat{x}_k^{(h)} + \hat{B}^{(h)}\hat{u}_k^{(h)} \\ \hat{y}_k^{(h)} = \hat{C}^{(h)}\hat{x}_k^{(h)} + \hat{D}^{(h)}\hat{u}_k^{(h)} \end{cases} \tag{5.12}$$

where the state $\hat{x}_k^{(h)} := x_{h+kK}$ evolves over K time steps with state transition matrix $\hat{A}^{(h)} := \Phi(h+K, h)$, where $\hat{u}_k^{(h)}$ and $\hat{y}_k^{(h)}$ are the stacked input and output vectors:

$$\hat{u}_k^{(h)} := [u_{h+kK}^T,\ u_{h+kK+1}^T,\ \ldots,\ u_{h+kK+K-1}^T]^T$$

$$\hat{y}_k^{(h)} := [y_{h+kK}^T,\ y_{h+kK+1}^T,\ \ldots,\ y_{h+kK+K-1}^T]^T$$

and where $\hat{B}^{(h)}$, $\hat{C}^{(h)}$ and $\hat{D}^{(h)}$ are defined in terms of the matrices $\{A_k, B_k, C_k\}$ (see [MB75]). Obviously, there are K such time invariant liftings for $h = 1, \ldots, K$, and each one has a transfer function. For such systems a theorem similar to Theorem 5.4.3 was obtained in [CV02, Cha03].

Theorem 5.4.5. *Let P and Q be the solutions of, respectively, $P = \tilde{A}P\tilde{A}^T + I_{KN}$ and $Q = \tilde{A}^T Q\tilde{A} + I_{KN}$, where*

$$\tilde{A} := \begin{pmatrix} 0 & \ldots & 0 & A_K \\ A_1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \ldots & A_{K-1} & 0 \end{pmatrix} \quad \textit{and} \quad \begin{array}{c} P := \operatorname{diag}(P_1, \ldots, P_{K-1}, P_K) \\ Q := \operatorname{diag}(Q_1, \ldots, Q_{K-1}, Q_K) \end{array}$$

then

$$\|\mathcal{E}_c(l)\|_2 \leq \eta_c^2(l)\|P\|_2 \leq \eta_c^2(l)\frac{\kappa(\tilde{A})^2}{1-\rho(\tilde{A})^2},$$

$$\|\mathcal{E}_o(r)\|_2 \leq \eta_c^2(r)\|Q\|_2 \leq \eta_c^2(r)\frac{\kappa(\tilde{A})^2}{1-\rho(\tilde{A})^2}.$$

Using multirate sampling [TAS01], we constructed in [CV02] a time-varying system model of period $K = 2$ and dimension $N = 122$ of the arm of the CD player described in Chapter 24, Section 4 of this volume. We refer to [CV02] for more details but we recall here some results illustrating the convergence of the Gramian estimates $\mathcal{P}_k = S_k S_k^T$, which were chosen of rank 20. Every two steps these should converge to the steady state solutions corresponding to the even and odd infinite horizon controllability Gramians.

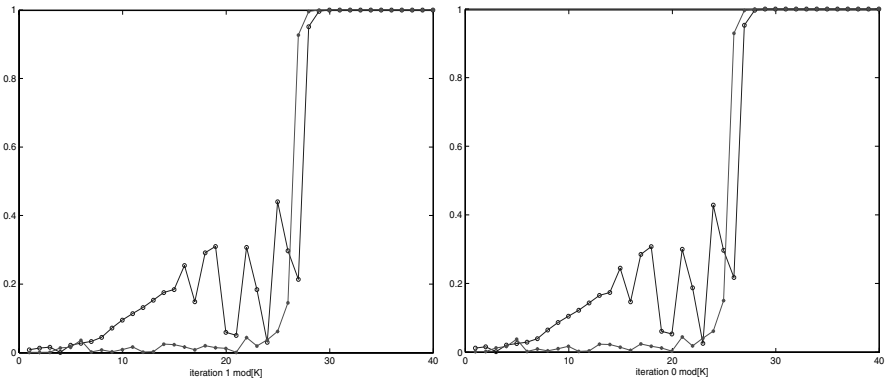


Fig. 5.1. $\circ$: $\cos(\measuredangle(S_k, S_{k-2}))$, $*$: $\cos(\measuredangle(S_k, S_\infty))$ for odd and even k

Since only the spaces matter and not the actual matrices, we show in Figure 5.1 (left) the cosine of the canonical angle between the dominant subspace of odd iterations $(k-2)$ and k, i.e. $\cos(\measuredangle(S_{k-2}, S_k))$, and the canonical angle with the exact dominant subspace, denoted as S_∞, of the controllability Gramian of the lifted LTI system (5.12), i.e. $(\cos(\measuredangle(S_k, S_\infty))$. This is repeated in Figure 5.1 (right) for the even iterates. The results for the observability Gramians are similar and are not shown here. Figure 5.1 shows the convergence and the accuracy of our algorithm. It can be seen that convergence is quick and is well predicted by the errors performed in the last updating steps.

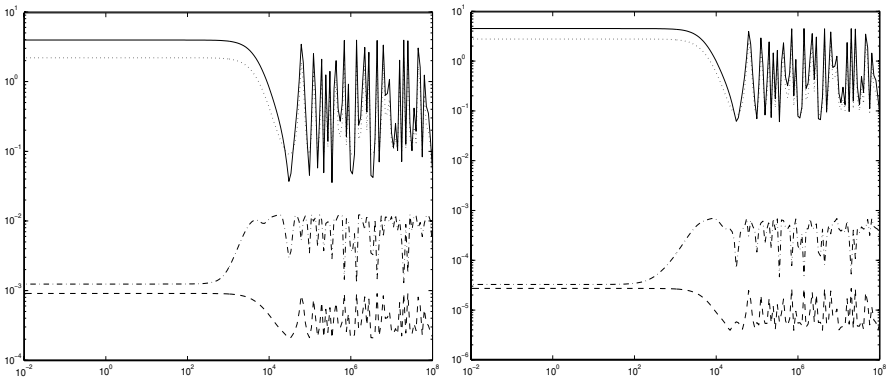


Fig. 5.2. –: full model, $\cdots$: approx. errors (20 steps), -·- approx. errors (60 steps), - - approx.errors (exact Gramian)

In Figure 5.2 we compare frequency responses of the time-invariant lifted systems (5.12) for odd and even iterates. In each figure we give the amplitude of the frequency response of the original model, the absolute errors in the frequency response of the projected systems using projectors obtained after 20 steps and 60 steps, and the absolute errors in the frequency response of the projected systems using the exact dominant subspace of the Gramians of the lifted system. The graphs show that after 60 steps an approximation comparable to Balanced Truncation is obtained.

5.5 Recursive Low-Rank Hankel Algorithm (RLRH)

The algorithm of the previous section yields an independent approximation of the two Gramians. If the original system was poorly balanced, it often happens that the approximation of the product of the two Gramians is far less accurate than that of the individual Gramians. This will affect the quality of the approximation of the reduced model since the product of the Gramians plays an important role in the frequency domain error.

In [CV03a, CV03b] an algorithm is presented which avoids this problem. The key idea is to use the underlying recurrences defining the time-varying Hankel map $\mathcal{H}(k_f, k, k_i) = \mathcal{O}(k_f, k)\mathcal{C}(k, k_i)$. Because the system order at each instant is given by the rank of the Hankel matrix at that instant, it is a good idea to approximate the system by approximating the Hankel matrix via a recursive SVD performed at each step. The technique is very similar to that of the previous section but now we perform at each step the singular value decomposition of a product similar to the products $\mathcal{O}(k_f, k)\mathcal{C}(k, k_i)$. Consider indeed the singular value decomposition of the matrix

$$\left[\frac{C_r}{R_{r+1}^T A_r}\right] \cdot \left[\, B_{l-1} \middle| A_{l-1} S_{l-1} \,\right] = U \Sigma V^T \tag{5.13}$$

and partition $U := \left[\, U_+ \middle| U_- \,\right]$, $V := \left[\, V_+ \middle| V_- \,\right]$ where $U_+ \in \mathbb{R}^{(p+n)\times n}$ and $V_+ \in \mathbb{R}^{(m+n)\times n}$. Define then

$$\left[\, S_l \middle| E_c(l) \,\right] := \left[\, B_{l-1} \middle| A_{l-1} S_{l-1} \,\right] \left[\, V_+ \middle| V_- \,\right], \tag{5.14}$$

$$\left[\, R_r \middle| E_o(r) \,\right] := \left[\, C_r^T \middle| A_r^T R_{r+1} \,\right] \left[\, U_+ \middle| U_- \,\right]. \tag{5.15}$$

It then follows that

$$\left[\frac{R_r^T}{E_o^T(r)}\right] \left[\, S_l \middle| E_c(l) \,\right] = \left[\begin{array}{c|c} \Sigma_+ & 0 \\ \hline 0 & \Sigma_- \end{array}\right], \tag{5.16}$$

where Σ_- contains the neglected singular values at this step. For the initialization at step $i = 0$ we use again

$$S_{k_i} = 0 \quad \text{and} \quad R_{k_f+1} = 0$$

and iterate for $i = 1, \ldots, \tau$ where $\tau := (k_f - k_i)/2$ is the half interval length. The approximate factorizations that one obtains are those indicated in Figure 5.3 and the corresponding MATLAB-like algorithm is now as follows.

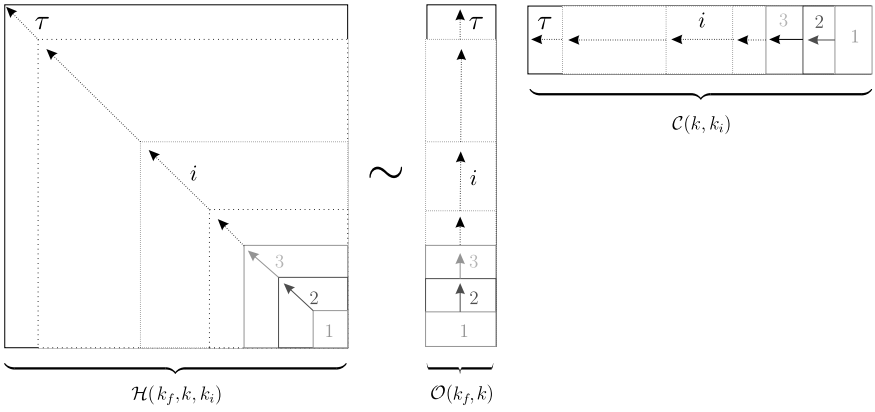


Fig. 5.3. Submatrix sequence approximated by low rank approximations

Algorithm RLRH
$l = k_i; r = k_f + 1; \tau = (r - l - 1)/2; S_l = 0; R_r = 0;$
for $i = 1 : \tau;$
$\quad l = l + 1; M = \left[\, B_{l-1} \middle| A_{l-1} S_{l-1} \,\right]; r = r - 1; N = \left[\, C_r^T \middle| A_r^T R_{r+1} \,\right];$
$\quad [U, \Sigma, V] = svd(N^T M); S_l = M * V(:, 1 : n); R_r = N * U(:, 1 : n);$
end

The amount of work involved in this algorithm is comparable to the earlier algorithm. We need to form the products $A_{l-1}S_{l-1}$ and $R_{r+1}^T A_r$, which requires $4Nn\alpha$ flops. The construction of the left hand side of (5.13) requires an additional $2N(n+m)(n+p)$ flops and the application of the transformations U and V requires $O((p+n)(m+n)(2n+p+m))$ flops, and so the complexity of this algorithm is $O(N(p+n)(m+n))$ for each iteration if $N \gg n > m, p, \alpha$.

As before we have a lemma, shown in [CV03a, CV03b, Cha03], linking the intermediate error matrices and the matrices $\mathcal{O}(k_f, r)$ and $\mathcal{C}(l, k_i)$.

Theorem 5.5.1. *At each iteration, there exist orthogonal matrices* $V^{(i)} \in \mathbb{R}^{(n+im)\times(n+im)}$ *and* $U^{(i)} \in \mathbb{R}^{(n+ip)\times(n+ip)}$ *satisfying:*

$$\mathcal{C}(l, k_i)V^{(i)} = \left[\, S_l \middle| E_c(l) \middle| A_{l-1}\mathcal{C}_e(l, k_i + 1) \,\right]$$

$$\mathcal{O}(k_f, r)^T U^{(i)} = \left[\, R_r \middle| E_o(r) \middle| A_r^T \mathcal{O}_e(k_f, r + 1) \,\right]$$

where $E_c(l)$ *and* $E_o(r)$ *are the neglected parts at each iteration, and the matrices* $\mathcal{C}_e(j, k_i)$ *and* $\mathcal{O}_e(k_f, j)$ *are defined as follows:*

$$\mathcal{C}_e(j, k_i) := \left[\, E_c(j-1) \middle| \ldots \middle| \Phi(j-1, k_i)E_c(k_i) \,\right],$$

$$\mathcal{O}_e(k_f, j)^T := \left[\, E_o(j) \middle| \ldots \middle| \Phi(k_f, j)^T E_o(k_f) \,\right].$$

As a consequence of this theorem we show in [CV03a, CV03b, Cha03] the following result which yields an approximation of the original Hankel map $\mathcal{H}(k_f, k, k_i)$.

Theorem 5.5.2. *There exist orthogonal matrices* $V^{(\tau)} \in \mathbb{R}^{(n+\tau m)\times(n+\tau m)}$ *and* $U^{(\tau)} \in \mathbb{R}^{(n+\tau p)\times(n+\tau p)}$ *such that* $U^{(\tau)T}\mathcal{H}(k_f, k, k_i)V^{(\tau)}$ *is equal to*

$$\left[\begin{array}{c|c|c} R_\tau^T S_\tau & 0 & R_\tau^T A_{\tau-1}\mathcal{C}_e(\tau, k_i) \\ \hline 0 & E_o^T(\tau)E_c(\tau) & E_o^T(\tau)A_{\tau-1}\mathcal{C}_e(\tau, k_i) \\ \hline \mathcal{O}_e(k_f, \tau+1)A_\tau S_\tau & \mathcal{O}_e(k_f, \tau+1)A_\tau E_c(\tau) & \mathcal{O}_e(k_f, \tau+1)A_\tau A_{\tau-1}\mathcal{C}_e(\tau, k_i) \end{array}\right].$$

This result enables us to evaluate the quality of our approximations by using the Hankel map without passing via the Gramians, which is exploited in [CV03a, CV03b, Cha03] to obtain bounds for the error. Notice also that since we are defining projectors for finite time windows, these algorithms could be applied to linear time-invariant systems that are unstable. One can then not show any property of stability for the reduced order model, but the finite horizon Hankel map will at least be well approximated.

5.5.1 Time-Invariant Case

As for the Gramian based approximation, we can analyze the quality of this approach in the time-invariant case. Since all matrices A, B and C are then

constant, all Hankel maps are time-invariant as well and only the interval width plays a role in the obtained decomposition. We can e.g. run the RLRH algorithm on an interval $[k_i, k_f] = [-\tau, \tau]$ for $\tau \in \mathbb{N}$ and approximate the Gramians $\mathcal{G}_c(0)$ and $\mathcal{G}_o(0)$ of the original model by $S_0 S_0^T$ and $R_0 R_0^T$, respectively, at the origin of the symmetric interval $[-\tau, \tau]$. The differences between the approximate low-rank Gramians and the exact Gramians

$$\mathcal{E}_c(0) := \mathcal{G}_c(0) - \mathcal{P}_0, \quad \mathcal{E}_o(0) := \mathcal{G}_o(0) - \mathcal{Q}_0$$

then remain bounded for intervals of growing length 2τ, as indicated in the following theorem ([CV03a, CV03b, Cha03]).

Theorem 5.5.3. *Let P and Q be respectively the solutions of $P = APA^T + I$, and $Q = A^T QA + I$, then*

$$\|\mathcal{E}_c(0)\|_2 \leq \eta_c^2 \|P\|_2 \leq \eta_c^2 \frac{\kappa(A)^2}{1-\rho(A)^2}, \quad \|\mathcal{E}_o(0)\|_2 \leq \eta_o^2 \|Q\|_2 \leq \eta_o^2 \frac{\kappa(A)^2}{1-\rho(A)^2}$$

where $\eta_c := \max\limits_{-\tau \leq k \leq 0} \|E_c(k)\|_2$ *and* $\eta_o := \max\limits_{0 \leq k \leq \tau} \|E_o(k)\|_2$.

Similarly, we obtain an approximation of the Hankel map as follows (see [CV03a, CV03b, Cha03]).

Theorem 5.5.4. *Using the first n columns $U_+^{(0)}$ of $U^{(0)}$ and $V_+^{(0)}$ of $V^{(0)}$, we obtain a rank n approximation of the Hankel map:*

$$\mathcal{H}(\tau, 0, -\tau) - U_+^{(0)} R_0^T \cdot S_0 V_+^{(0)T} = \mathcal{E}_h(0),$$

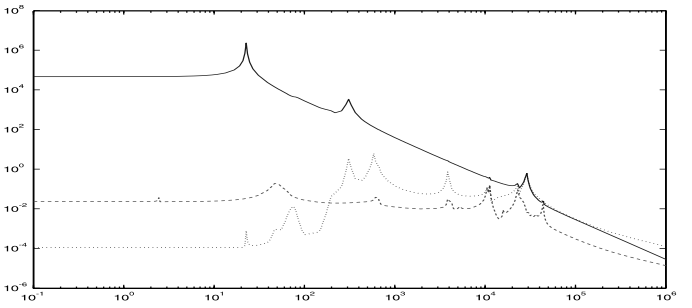
for which we have the error bound:

$$\|\mathcal{E}_h(0)\|_2 \leq \frac{\kappa(A)}{\sqrt{1-\rho(A)^2}} \max\{\eta_c \|R_0^T A\|_2, \eta_o \|A S_0\|_2\} + \frac{\kappa(A)^2}{1-\rho(A)^2} \eta_o \eta_c.$$

An important advantage of the RLRH method is that the computed projectors are independent of the coordinate system used to describe the original system $\{A, B, C\}$. This can be seen as follows. When performing a state-space transformation T we obtain a new system $\{\hat{A}, \hat{B}, \hat{C}\} := \{T^{-1}AT, T^{-1}B, CT\}$. It is easy to see that under such transformations the updating equations of R_r and S_l transform to $\hat{R}_k = T^T R_k$ and $\hat{S}_l = T^{-1} S_l$, and this is preserved by the iteration. One shows that the constructed projector therefore follows the same state-space transformation as the system model. Therefore, the constructed reduced order model does not depend on whether or not one starts with a balanced realization for the original system. For the RLRG method, on the other hand, one can lose a lot of accuracy when using a poorly balanced realization to construct a reduced order model.

5.6 Numerical Examples

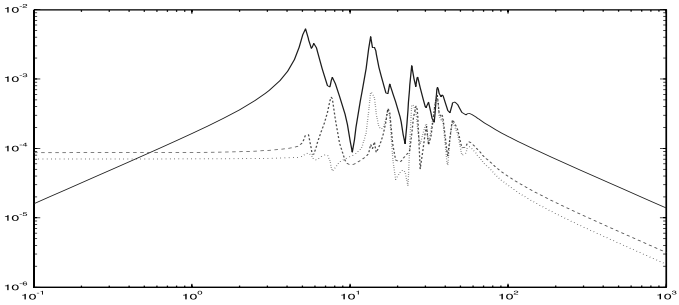
In this section we apply our algorithm to discretizations of three different dynamical systems: a Building model, a CD Player model, and the International Space Station model. These benchmarks are described in more details in Chapter 24, Sections 4, 6, 7. It was shown in [CV03a, Cha03], that for the same problem, the RLRG method gives less accurate results: as predicted by the discussion of the previous section, the RLRG method deteriorates especially when the original system is poorly balanced. Since the RLRH method is to be preferred over the RLRG method, we only compare here the RLRH method with Balanced Truncation. The approximate system $\mathcal{S}_{BT}$ for balanced truncation and $\mathcal{S}_{RLRH}$ for the recursive low rank Hankel method, are both calculated for a same degree. We show the maximal singular value of the frequency responses of the system and the maximal singular value of the two error functions.



σ_{max}-plot of the frequency responses.
— full model, - - - BT error system, ··· RLRH error system.

$cond(T)$	$\rho(A)$	$cond(A)$	$\|\mathcal{S}\|_{\mathcal{H}_\infty}$	$\|\mathcal{S}-\mathcal{S}_{BT}\|_{\mathcal{H}_\infty}$	$\|\mathcal{S}-\mathcal{S}_{RLRH}\|_{\mathcal{H}_\infty}$
40.7341	1	1.00705	$2.3198.10^6$	0.2040	6.1890

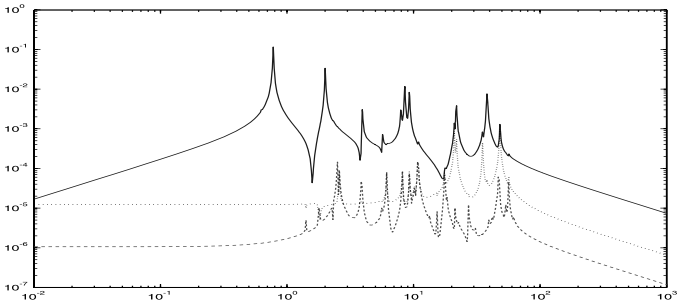
Fig. 5.4. CD-player model $N = 120, \quad m = p = 2, \quad n = 24$



σ_{max}-plot of the frequency responses.
— full model, - - - BT error system, ··· RLRH error system.

$cond(T)$	$\rho(A)$	$cond(A)$	$\|\mathcal{S}\|_{\mathcal{H}_\infty}$	$\|\mathcal{S}-\mathcal{S}_{BT}\|_{\mathcal{H}_\infty}$	$\|\mathcal{S}-\mathcal{S}_{RLRH}\|_{\mathcal{H}_\infty}$
347.078	0.9988	5.8264	0.0053	$6.0251.10^{-4}$	$6.7317.10^{-4}$

Fig. 5.5. Building model $N = 48, \quad m = p = 1, \quad n = 10$



σ_{max}-plot of the frequency responses.
— full model, - - - BT error system, ··· RLRH error system.

$cond(T)$	$\rho(A)$	$cond(A)$	$\|\mathcal{S}\|_{\mathcal{H}_\infty}$	$\|\mathcal{S}-\mathcal{S}_{BT}\|_{\mathcal{H}_\infty}$	$\|\mathcal{S}-\mathcal{S}_{RLRH}\|_{\mathcal{H}_\infty}$
740178	0.9998	5.82405	0.1159	$2.3630.10^{-4}$	0.0011

Fig. 5.6. ISS model $N = 270, \quad m = p = 3, \quad n = 32$

The corresponding H_∞ norms are also given in the table following each example. Each table also contains the condition number $cond(T)$ of the balancing state-space transformation T, the spectral radius $\rho(A)$ and the condition number $cond(A)$ since they play a role in the error bounds obtained in this paper. It can be seen from these examples that the RLRH method performs reasonably well in comparison to the balanced truncation method, and this independently from whether or not the original system was poorly balanced. Even though these models are not large they are good benchmarks in the sense

that their transfer functions are not easy to approximate. Larger experiments are reported in [Cha03].

5.7 Conclusion

In this paper we show how to construct low-dimensional projected systems of time-varying systems. The algorithms proposed are based on low-rank approximations of the Gramians and of the Hankel map which defines the input-output mapping. Both methods have the advantage of exploiting sparsity in the data to yield a complexity that is linear in the state dimension of the original model.

The key idea is to compute only a finite window of the Gramians or Hankel map of the time-varying system and to compute recursively projection matrices that capture the dominant behavior of the Gramians or Hankel map. The *Recursive Low-Rank Hankel* approximation method is to be preferred over the *Recursive Low-Rank Gramian* approximation method because it is not sensitive to the coordinate system in which the original system is described.

The two algorithms are mainly meant for time-varying systems but their performance is illustrated using time-invariant and periodic systems because the quality of the methods can then be assessed by the frequency responses of the error functions.

Acknowledgments

This paper presents research supported by NSF contracts CCR-99-12415 and ITR ACI-03-24944 and by the Belgian Programme on Inter-university Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with its authors. The work of the first author has been partially carried out within the framework of a collaboration agreement between CESAME (Université Catholique de Louvain, Belgium) and LINMA of the Faculty of Sciences (Université Chouaib Doukkali, Morocco), funded by the Secretary of the State for Development Cooperation and by the CIUF (Conseil Interuniversitaire de la Communauté Française, Belgium).

References

[Cha03] Chahlaoui, Y.: *Recursive low rank Hankel approximation and model reduction.* Doctoral Thesis, Université catholique de Louvain, Louvain-la-Neuve (2003)

[CV02] Chahlaoui, Y. and Van Dooren, P.: Estimating Gramians of large-scale time-varying systems. In: *Proc. IFAC World Congress*, Barcelona, Paper 2440 (2002)

[CV03a] Chahlaoui, Y. and Van Dooren, P.: Recursive Gramian and Hankel map approximation of large dynamical systems. In: *CD-Rom Proceedings SIAM Applied Linear Algebra Conference*, Williamsburg, Paper MS14-1 (2003)

[CV03b] Chahlaoui, Y. and Van Dooren, P.: Recursive low rank Hankel approximation and model reduction. In: *CD-Rom Proceedings ECC 2003*, Cambridge, Paper 553 (2003)

[DV98] Dewilde, P. and van der Veen, A.-J.: *Time-varying systems and computations*. Kluwer Academic Publishers, Boston (1998)

[Enn81] Enns, D.: Model reduction with balanced realizations: An error bound and frequency weighted generalization. In: *Proc. of the IEEE Conference on Decision and Control*, San Diego, 127–132 (1981)

[GVV03] Gallivan, K., Vandendorpe, A. and Van Dooren, P.: Sylvester equations and projection-based model reduction. *J. Comp. Appl. Math.*, **162**, 213–229 (2003)

[GV96] Golub, G. and Van Loan, C.: *Matrix Computations*. Johns Hopkins University Press, Baltimore (1996)

[GSA03] Gugercin, S., Sorenson, D. and Antoulas, A.: A modified low-rank Smith method for large-scale Lyapunov equations. *Numerical Algorithms*, **32(1)**, 27–55 (2003)

[IPM92] Imae, J., Perkins, J.E. and Moore, J.B.: Toward time-varying balanced realization via Riccati equations. *Math. Control Signals Systems*, **5**, 313–326 (1992)

[LB03] Lall, S., and Beck, C.: Error-bounds for balanced model-reduction of linear time-varying systems. *IEEE Trans. Automat. Control*, **48(6)**, 946–956 (2003)

[MB75] Meyer, R. and Burrus, C.: A unified analysis of multirate and periodically time-varying digital filters. *IEEE Trans. Circ. Systems*, **22**, 162–168 (1975)

[Moo81] Moore, B.: Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Automat. Control*, **26**, 17–31 (1981)

[SR02] Sandberg, H., and Rantzer, H.: Balanced model reduction of linear time-varying systems. In: *Proc. IFAC02, 15th Triennial World Congress*, Barcelona (2002)

[SSV83] Shokoohi, S., Silverman, L., and Van Dooren, P.: Linear time-variable systems: Balancing and model reduction. *IEEE Trans. Automat. Control*, **28**, 810–822 (1983)

[TAS01] Tornero, J., Albertos, P., and Salt, J.: Periodic optimal control of multirate sampled data systems. In: *Proc. PSYCO2001, IFAC Conf. Periodic Control Systems*, Como, 199–204 (2001)

[VK83] Verriest, E., and Kailath, T.: On generalized balanced realizations. *IEEE Trans. Automat. Control*, **28(8)**, 833–844 (1983)

[ZDG95] Zhou, K., Doyle, J., and Glover, K.: *Robust and optimal control*. Prentice Hall, Upper Saddle River (1995)