Second-Order Structure Preserving Model Reduction

Y. Chahlaoui, D. Lemonnier, A. Vandendorpe and P. Van Dooren *

May 4, 2004

1   Introduction

There are two main categories of model reduction, namely SVD-based techniques and Krylov based techniques. The first category is used for systems of moderate order, but provides global error bounds. The second category can be applied to systems of very large order (say several thousands) but without global error bound available.

Here we consider second-order linear time-invariant systems, meaning systems of the type

$$\begin{cases} M\ddot{q}(t) + D\dot{q}(t) + Kq(t) & = & Bu(t) \\ y(t) & = & Cq(t) \end{cases}, \tag{1}$$

where $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$, $q(t) \in \mathbb{R}^n$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $M, D, K \in \mathbb{R}^{n \times n}$ and where M is assumed to be invertible. The transfer function associated with the system (1) in the Laplace domain is given by

$$T(s) \doteq C \left(Ms^2 + Ds + K \right)^{-1} B, \tag{2}$$

where $P(s) \doteq Ms^2 + Ds + K$ is called the *characteristic polynomial matrix*.

It is often advisable to construct a reduced model of size $k \ll n$ that nevertheless keeps the "second-order structure" of the system. We thus need to build a reduced model,

$$\begin{cases} \hat{M}\ddot{\hat{q}}(t) + \hat{D}\dot{\hat{q}}(t) + \hat{K}\hat{q}(t) & = & \hat{B}u(t) \\ \hat{y}(t) & = & \hat{C}\hat{q}(t) \end{cases} \tag{3}$$

where $\hat{q}(t) \in \mathbb{R}^k$, $\hat{M}, \hat{D}, \hat{K} \in \mathbb{R}^{k \times k}$, $\hat{B} \in \mathbb{R}^{k \times m}$, $\hat{C} \in \mathbb{R}^{p \times k}$, such that its transfer function is "close" to the original transfer function.

The purpose of this paper is to present SVD and Krylov based model reduction techniques that preserve the second order form of the system one wants to reduce. For SVD techniques, we define in section 3 two pairs of gramians, $(\mathcal{P}_{pos}, \mathcal{Q}_{pos})$ associated to the position $q(t)$, and $(\mathcal{P}_{vel}, \mathcal{Q}_{vel})$ associated to the velocity $\dot{q}(t)$. Each of these gramians are associated to an optimization problem. One then projects the state space by keeping the position and the velocity subspaces corresponding to the dominant eigenspaces of $\mathcal{P}_{pos}\mathcal{Q}_{pos}$ and $\mathcal{P}_{vel}\mathcal{Q}_{vel}$ respectively. Second-order structure preserving Krylov techniques are considered in section 4 where it is shown how to construct a second order transfer function that satisfies tangential interpolation conditions with respect to the original second order transfer function.

2   Model Reduction by Balanced Truncation

This section is developed in more details in [1]. So, only important points are recalled here (see [2] or [3] for a deeper study of the Balanced Truncation technique). We consider in this section linear time-invariant systems modelled by the following system of equations

$$\mathcal{S} \begin{cases} \dot{x}(t) & = & \mathcal{A}x(t) + \mathcal{B}u(t) \\ y(t) & = & \mathcal{C}x(t) \end{cases} , \quad u(t) \in \mathbb{R}^{\mathrm{m}}, x(t) \in \mathbb{R}^{\mathrm{n}}, y(t) \in \mathbb{R}^{\mathrm{p}} \tag{4}$$

which therefore have a transfer function $T(s) = \mathcal{C}(sI_n - \mathcal{A})^{-1}\mathcal{B}$ that links the inputs to the outputs in the Laplace domain. Such transfer functions are *strictly proper*, i.e. $\lim_{s \to \infty} T(s) = 0$, which happens to be the case also for the transfer function considered in (2) since M was assumed to be invertible.

Two matrices are associated with the linear system (4). These are the "controllability gramian" $\mathcal{P}$ and the "observability gramian" $\mathcal{Q}$. If $\mathcal{A}$ is Hurwitz, they are the unique solutions of the following Lyapunov equations :

$$\mathcal{A}\mathcal{P} + \mathcal{P}\mathcal{A}^T + \mathcal{B}\mathcal{B}^T = 0 \quad , \quad \mathcal{A}^T \mathcal{Q} + \mathcal{Q}\mathcal{A} + \mathcal{C}^T\mathcal{C} = 0. \tag{5}$$

A physical interpretation of the gramians is the following. The controllability matrix arises from the following optimization problem. Let

$$J(v(t), a, b) \doteq \int_a^b v(t)^T v(t) dt$$

be the *energy* of the vector function $v(t)$ in the interval $[a, b]$. Then (see [3])

$$\min_{u(t)} J(u(t), -\infty, 0)|_{x(0)=x_0} = x_0^T \mathcal{P}^{-1} x_0, \tag{6}$$

and, symmetrically, we have the dual property (for the dual system $(\mathcal{C}^T, \mathcal{A}^T, \mathcal{B}^T)$ evolving backward in time) :

$$\min_{y(t)} J(y(t), -\infty, 0)|_{x_0=x_0} = x_0^T \mathcal{Q}^{-1} x_0. \tag{7}$$

The more $x_0^T \mathcal{P}^{-1} x_0$ is small, the more "controllable" the state x_0 is, because it can be reached with a input of small energy. By duality, the more $x_0^T \mathcal{Q}^{-1} x_0$ is small, the more "observable" the state x_0 is. Thus when both gramians are equal and diagonal, the order of magnitude of a diagonal value of the product $\mathcal{P}\mathcal{Q}$ is a good measure for the influence of the corresponding state variable in the mapping from past inputs $u(t) \in \mathcal{L}_2[-\infty, 0]$ to future outputs $y(t) \in \mathcal{L}_2[0, +\infty]$ passing via that particular state at time $t = 0$.

Given a transfer function $T(s)$, the popular balanced truncation model reduction method consists in finding a state-space realization $(\mathcal{C}_{bal}, \mathcal{A}_{bal}, \mathcal{B}_{bal})$ of $T(s)$ such that the gramians are equal and diagonal and then the reduced model is constructed by keeping the states corresponding to the largest eigenvalues of the product $\mathcal{P}\mathcal{Q}$ in it. The idea of the balanced truncation technique thus consists in keeping those states that are most controllable and observable according to the gramians defined in (6) and (7).

Second-order systems can be seen as a particular class of linear systems. Indeed, by rewriting the system (1) as follows

$$\begin{cases} \dot{x}(t) & = & \begin{bmatrix} 0 & I \\ -K_M & -D_M \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ B_M \end{bmatrix} u(t) \\ \\ y(t) & = & \begin{bmatrix} C_M & 0 \end{bmatrix} x(t) \end{cases} \tag{8}$$

where the state $x(t)$ is $\begin{bmatrix} q(t)^T & \dot{q}(t)^T \end{bmatrix}^T$, and where we have chosen a coordinate system in which the mass matrix M is the identity, one recovers the form (4). We can thus rewrite the transfer function defined in (2) as

$$H(s) = \mathcal{C}(sI_{2n} - \mathcal{A})^{-1}\mathcal{B} \tag{9}$$

by defining

$$\mathcal{A} \doteq \begin{bmatrix} 0 & I \\ -K_M & -D_M \end{bmatrix} \quad , \quad \mathcal{B} \doteq \begin{bmatrix} 0 \\ B_M \end{bmatrix} \quad , \quad \mathcal{C} \doteq \begin{bmatrix} C_M & 0 \end{bmatrix}. \tag{10}$$

Unfortunately the classic balanced truncation technique cannot be applied directly to the state-space realization $(\mathcal{C}, \mathcal{A}, \mathcal{B})$ (8) of the second-order system since, in general, the resulting reduced system is not a second-order system anymore. The objective of the next section is to find a method that performs balanced truncation in some sense *and* preserves the second-order structure of the system.

3   Balanced Truncation for Second-Order Systems

The idea of our balance and truncate technique for second-order systems (called SOBT for Second-Order Balanced Truncation) is the following. First, we need to define *two* pairs of $n \times n$ gramians ("second-order gramians") that have to change according to contragradient transformations, and that must have some energetic interpretation. (Only then a balance and truncate process makes sense). The first pair $(\mathcal{P}_{pos}, \mathcal{Q}_{pos})$ will correspond to an energy optimization problem depending only on the *positions* $q(t)$ and not on the *velocities* $\dot{q}(t)$. Reciprocally, the second pair $(\mathcal{P}_{vel}, \mathcal{Q}_{vel})$ will be associated to an optimization problem depending only on the velocities $\dot{q}(t)$ and not the on the positions $q(t)$. By analogy to the first order case, the gramians $\mathcal{Q}_{pos}$ and $\mathcal{Q}_{vel}$ will be defined from the dual systems. After these definitions we then come to the balancing part of the method. For this we transform to a balanced coordinate system in which the second-order gramians are equal and diagonal : $\bar{\mathcal{P}}_{pos} = \bar{\mathcal{Q}}_{pos} = \Sigma_{pos}, \quad \bar{\mathcal{P}}_{vel} = \bar{\mathcal{Q}}_{vel} = \Sigma_{vel}$. Their diagonal values will enable us to point out what the *important* positions and the *important* velocities are, i.e. those with (hopefully) large effect on the I/O map. Hence to get a reduced second-order model we keep only the part of the system that depends on these variables. This is the truncation part of the method.

Let us first define a pair of second-order gramians measuring the contribution of the position coordinates (independently of the velocities) with respect to the I/O map. A natural optimization problem (see [4]) associated with the second-order form is the following

$$\min_{\dot{q}_0 \in \mathbb{R}^n} \min_{u(t)} \ J(u(t), -\infty, 0), \tag{11}$$

subject to

$$M\ddot{q}(t) + D\dot{q}(t) + Kq(t) = Bu(t), \ q(0) = q_0.$$

It can be shown that the optimum is $q_0^T \mathcal{P}_{11}{}^{-1}q_0$, where $\mathcal{P}_{11}$ is the $n \times n$ left upper block of $\mathcal{P}$ (see equation (5)).

The solution of the dual problem will correspond to $q_0^T \mathcal{Q}_{11}{}^{-1}q_0$, where $\mathcal{Q}_{11}$ is the $n \times n$ left upper block of $\mathcal{Q}$ (5). Under the change of coordinates $q(t) = \Phi\bar{q}(t)$, it is immediate to verify that this pair of gramians undergoes a contragradient transformation :

$$(\bar{\mathcal{P}}_{11}, \bar{\mathcal{Q}}_{11}) = (\Phi^{-1}\mathcal{P}_{11}\Phi^{-T}, \Phi^T\mathcal{Q}_{11}\Phi).$$

This implies that there exists a new coordinate system such that both $\mathcal{P}_{11}$ and $\mathcal{Q}_{11}$ are equal end diagonal. Their energetic interpretation is given by looking at the underlying optimization problem. In (11), one minimizes the necessary energy to reach the given position q_0 over all past inputs and initial velocities. Hence these gramians really describe how the I/O energy is distributed among the positions.

Analogously, let us define a pair of second-order gramians that would give the contribution of the velocities with respect to the I/O map. The optimization problem associated is the following

$$\min_{q_0 \in \mathbb{R}^n} \min_{u(t)} \ J(u(t), -\infty, 0) \tag{12}$$

subject to

$$M\ddot{q}(t) + D\dot{q}(t) + Kq(t) \ = \ Bu(t), \quad \dot{q}(0) \ = \ \dot{q}_0.$$

By exactly following the same reasoning as in [4] for the optimization problem (11), one can show that the solution of (12) is $\dot{q}_0^T \mathcal{P}_{22}{}^{-1} \dot{q}_0$, where $\mathcal{P}_{22}$ is the $n \times n$ right lower block of $\mathcal{P}$. The solution of the dual problem will correspond to $\dot{q}_0^T \mathcal{Q}_{22}{}^{-1} \dot{q}_0$, where $\mathcal{Q}_{22}$ is the $n \times n$ right lower block of $\mathcal{Q}$. Again under the change of coordinates $\dot{q}(t) = \Phi \dot{\bar{q}}(t)$ one can check that this pair of gramians perform a contragradient transformation. In (12), one minimizes the necessary energy to reach the given velocity $\dot{q}_0$ over all past inputs and initial positions. Hence these gramians really describe how the I/O energy is distributed among the velocities.

The conclusion is that these second-order gramians are good candidates for our problem. We make thus the choice :

$$(\mathcal{P}_{pos}, \ \mathcal{Q}_{pos}) \ = \ (\mathcal{P}_{11}, \ \mathcal{Q}_{11}) \quad \text{and} \quad (\mathcal{P}_{vel}, \ \mathcal{Q}_{vel}) \ = \ (\mathcal{P}_{22}, \ \mathcal{Q}_{22}) \ . \tag{13}$$

In the new model reduction technique that we propose here, we want to be able to balance both pairs of second-order gramians at the same time, and this is not possible with a change of coordinates of the type $q(t) = \Phi \bar{q}(t)$. For these reasons we work in a state-space context, starting with the system (8). The method SOBT proceeds then as follows : One first computes both pairs of second-order gramians $(\mathcal{P}_{pos}, \ \mathcal{Q}_{pos})$ and $(\mathcal{P}_{vel}, \ \mathcal{Q}_{vel})$. One then computes the contragradient transformations that put $\mathcal{P}_{pos} = \mathcal{Q}_{pos} = \Lambda_{pos}$ and $\mathcal{P}_{vel} = \mathcal{Q}_{vel} = \Lambda_{vel}$ where Λ_{pos} and Λ_{vel} are positive definite block diagonal matrices. On then truncates the positions corresponding to the smallest eigenvalues of Λ_{pos} and the velocities corresponding to the smallest eigenvalues of Λ_{vel}.

A deeper study of the Second-Order Balanced truncation technique and a comparaison with other techniques can be found in [5].

4 Second-Order Structure Preserving Krylov Techniques

Krylov techniques for model reduction of linear systems have first been introduced by [6]. These have been studied for generalized state space systems in [7]. Recently, a generalization of these techniques for Tangential Interpolation has been studied in [8] (see [9] for a survey).

The first Second Order structure preserving Krylov technique has been studied in [10] for interpolation at $s = \infty$. Recently, several people started to work on this problem. For instance, Z. Bai (http://www.cs.ucdavis.edu/~bai/) and coworkers studied a new Arnoldi procedure for computing Krylov subspaces associated to second order systems, R. Freund (http://netlib.bell-labs.com/who/freund/) studied Krylov techniques for model

reduction of systems of arbitrary order k (and not only of second order) and finally P. Van Dooren (http://www.auto.ucl.ac.be/~vdooren/) and A. Vandendorpe studied the tangential interpolation for second order systems.

In this section, we construct a second-order transfer function $\hat{T}(s)$ of order k that satisfies the following interpolation conditions with respect to the second-order transfer function $T(s)$ of order n:

$$x_i \left(T(s) - \hat{T}(s) \right) = O(\lambda_i - s) \quad , \quad \left(T(s) - \hat{T}(s) \right) x_{i+k} = O(\lambda_{i+k} - s), \qquad (14)$$

where $x_1, \ldots, x_k \in \mathbb{C}^{1 \times p}$ and $x_{k+1}, \ldots, x_{2k} \in \mathbb{C}^{m \times 1}$. This can be done by computing *generalized* Krylov subspaces as follows :

Algorithm 4.1 *1. Construct Z and V such that*

$$V = \left[\begin{array}{ccc} (\lambda_{k+1}I_{2n} - \mathcal{A})^{-1}\mathcal{B}x_{k+1} & \ldots & (\lambda_{2k}I_{2n} - \mathcal{A})^{-1}\mathcal{B}x_{2k} \end{array} \right]$$

$$Z^T = \left[\begin{array}{c} x_1\mathcal{C}(\lambda_1 I_{2n} - \mathcal{A})^{-1} \\ \vdots \\ x_k\mathcal{C}(\lambda_k I_{2n} - \mathcal{A})^{-1} \end{array} \right],$$

where $(\mathcal{C}, \mathcal{A}, \mathcal{B})$ are defined in (10).

2. Let V_1 and $V_2 \in \mathbb{C}^{n \times k}$ be the first n rows and the last n rows of V respectively. Let Z_1 and $Z_2 \in \mathbb{C}^{n \times k}$ be the first n rows and the last n rows of Z respectively. Construct

$$\mathcal{V} \doteq \left[\begin{array}{cc} V_1 M_1 & \\ & V_2 M_2 \end{array} \right] \quad , \quad \mathcal{Z} \doteq \left[\begin{array}{cc} Z_1 N_1 & \\ & Z_2 N_2 \end{array} \right], \qquad (15)$$

where the invertible matrices $M_1, M_2, N_1, N_2 \in \mathbb{C}^{n \times n}$ are chosen such that $\mathcal{Z}^T \mathcal{V} = I_{2k}$.

3. Construct the matrices

$$\hat{\mathcal{C}} \doteq \mathcal{C}\mathcal{V} \quad , \quad \hat{\mathcal{A}} \doteq \mathcal{Z}^T \mathcal{A}\mathcal{V} \quad , \quad \hat{\mathcal{B}} \doteq \mathcal{Z}^T \mathcal{B}.$$

4. Define the reduced order transfer function

$$\hat{T}(s) \doteq \hat{\mathcal{C}}(sI_{2k} - \hat{\mathcal{A}})^{-1}\hat{\mathcal{B}}.$$

It can be shown that $\hat{T}(s)$ is a second-order transfer function of Mc Millan degree $2k$ that satisfies the interpolation conditions (14).

In order to prove that the interpolation conditions are satisfied, one deduces from (15) that

$$Im(V) \subseteq Im(\mathcal{V}) \quad , \quad Im(Z) \subseteq Im(\mathcal{Z}).$$

As shown in [8], this is sufficient to ensure the interpolation conditions to be satisfied. The second order structure of the reduced order transfer function follows from the following lemma.

Lemma 4.1 *Let $(\mathcal{C}, \mathcal{A}, \mathcal{B})$ be the state space realization defined in (10). If one projects such a state space realization with $2n \times 2k$ bloc diagonal matrices*

$$\mathcal{Z} \doteq \left[\begin{array}{cc} Z_1 & 0 \\ 0 & Z_2 \end{array} \right] \quad , \quad \mathcal{V} \doteq \left[\begin{array}{cc} V_1 & 0 \\ 0 & V_2 \end{array} \right],$$

where $Z_1, V_1, Z_2, V_2 \in \mathbb{C}^{n \times k}$ are chosen such that $\mathcal{Z}^T V = I_{2k}$, then the reduced transfer function

$$\hat{T}(s) \doteq \mathcal{C}\mathcal{V} \left(\mathcal{Z}^T (sI_{2n} - \mathcal{A})\mathcal{V} \right)^{-1} \mathcal{Z}^T \mathcal{B}$$

is a second-order transfer function, provided the matrix $Z_1^T V_2$ is invertible.

More details are given in [11].

5   Concluding Remarks

Concerning SVD-Like techniques, a Second Order Balanced Truncation technique has been presented. If this technique preserves the structure, the drawback is that there is no guaranteed global error bound available. Even worse, stability may be lost in the reduced order system.

Concerning Krylov techniques, it has been shown in this paper that it is possible to use a Krylov technique while preserving the second-order structure, but there is a price to pay. Generically, imposing $2k$ interpolation conditions and the second order structure results in a reduced transfer function of order $2k$ rather than k if the second order structure was not imposed.

All the techniques presented here can be generalized for *interconnected systems*, as it will be shown in a subsequent paper.

References

[1] A. Vandendorpe, P. Van Dooren, Model reduction of interconnected systems, in: MTNS 2004 (16th Symp. on the Mathematical Theory of Networks and Systems), 2004.

[2] K. Zhou, J. C. Doyle, K. Glover, Robust and optimal control., Prentice Hall, Inc, Upper Saddle River, NJ, 1996.

[3] K. D. Glover, All optimal Hankel-norm approximation of linear multivariable systems and their L^∞-error bounds, Int. J. Control 39 (6) (1984) 1115–1193.

[4] D. G. Meyer, S. Srinivasan, Balancing and model reduction for second-order form linear systems, IEEE Trans. Automat. Control 41 (11) (1996) 1632–1644.

[5] Y. Chahlaoui, D. Lemonnier, A. Vandendorpe, P. Van Dooren, Second-Order Balanced Truncation, Linear Algebra Appl., Special Issue on Model Reduction Accepted.

[6] C. de Villemagne, R. E. Skelton, Model reductions using a projection formulation, Internat. J. Control 46 (6) (1987) 2141–2169.

[7] E. Grimme, Krylov projection methods for model reduction, Ph.D. thesis, Department of Electrical Engineering, University of Illinois at Urbana-Champaign (1997).

[8] K. Gallivan, A. Vandendorpe, P. Van Dooren, Model reduction of MIMO systems via tangential interpolation, SIAM Journal on Matrix Analysis and Applications Accepted, http://www.auto.ucl.ac.be/vandendorpe/.

[9] A. C. Antoulas, Lectures on optimal approximation of linear systems, Lecture Notes (October 1996).

[10] T.-J. Su, R. Craig Jr., Model Reduction and Control of Flexible Structures Using Krylov Vectors, J. Guidance 14 (2) (1991) 260–267.

[11] A. Vandendorpe, P. Van Dooren, Krylov techniques for model reduction of second-order systems, Tech. Rep. 07, CESAME, Université catholique de Louvain (2004).