NUMERICAL METHODS AND SOFTWARE TOOLS FOR MODEL REDUCTION

A. VARGA

DLR - Oberpfaffenhofen
Institute for Robotics and System Dynamics
P.O.B. 1116, D-82230 Wessling, Germany

Abstract. An overview of numerically reliable algorithms for model reduction is presented. The covered topics are the reduction of stable and unstable linear systems as well as the computational aspects of frequency weighted model reduction. The presentation of available software tools focuses on a recently developed Fortran library RASP-MODRED implementing a new generation of numerically reliable algorithms for model reduction.

1. INTRODUCTION

Model reduction is of fundamental importance in many modeling and control applications. The basic reduction algorithms discussed in this paper belong to the class of methods based on or related to balancing techniques [1, 2, 3, 4] and are primarily intended for the reduction of linear, stable, continuous- or discrete-time systems. All methods rely on guaranteed error bounds and have particular features which recommend them for use in specific applications. The basic methods combined with coprime factorization or spectral decomposition techniques can be used to reduce unstable systems [5] or to perform *frequency-weighted model reduction* (FWMR) [6, 7].

The surveyed algorithms represent the latest developments of various procedures for solving computational problems appearing in the context of model reduction. Most algorithms possess desirable attributes as generality, numerical reliability, enhanced accuracy, and thus are completely satisfactory to serve as bases for robust software implementations. Such implementations are available in a recently developed Fortran 77 library for model reduction called RASP-MODRED [8]. The implementations of routines are based on the new linear algebra standard package LAPACK [9]. It is worth mentioning that the implemented algorithms are generally superior to those implemented in the model reduction tools of commercial packages [10, 11, 12].

2. MODEL REDUCTION ALGORITHMS

Consider the n-th order original state-space model $G := (A, B, C, D)$ with the *transfer-function matrix* (TFM) $G(\lambda) = C(\lambda I - A)^{-1}B + D$, and let $G_r := (A_r, B_r, C_r, D_r)$ be an r-th order approximation of the original model ($r < n$), with the TFM $G_r = C_r(\lambda I - A_r)^{-1}B_r + D_r$. A large class of model reduction methods can be interpreted as performing a similarity transformation Z yielding

$$\left[\begin{array}{c|c} Z^{-1}AZ & Z^{-1}B \\ \hline CZ & D \end{array}\right] := \left[\begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & D \end{array}\right],$$

and then defining the reduced model (A_r, B_r, C_r, D_r) as the leading diagonal system (A_{11}, B_1, C_1, D). When writing $Z := [\,T\ U\,]$ and $Z^{-1} := [\,L^T\ V^T\,]^T$, then $\Pi = TL$ is a projector on T along L and $LT = I_r$. Thus the reduced system is $(A_r, B_r, C_r, D_r) = (LAT,\ LB,\ CT,\ D)$. Partitioned forms as above can be used to construct a so-called *singular perturbation approximation* (SPA). The matrices of the reduced model in this case are given by

$$\begin{aligned} A_r &= A_{11} + A_{12}(\gamma I - A_{22})^{-1}A_{21}, \\ B_r &= B_1 + A_{12}(\gamma I - A_{22})^{-1}B_2, \\ C_r &= C_1 + C_2(\gamma I - A_{22})^{-1}A_{21}, \\ D_r &= D + C_2(\gamma I - A_{22})^{-1}B_2. \end{aligned}$$

where $\gamma = 0$ for a continuous-time system and $\gamma = 1$ for a discrete-time system. Note that SPAs preserve the DC-gains of stable original systems.

Specific requirements for model reduction algorithms are formulated and discussed in [13]. Such requirements are: (1) applicability of methods regardless the original system is minimal or not; (2) emphasis on enhancing the numerical accuracy of computations; (3) relying on numerically reliable procedures.

The first requirement can be fulfilled by computing L and T directly, without determining Z or Z^{-1}. In

particular, if the original system is not minimal, then L and T can be chosen to compute an *exact* minimal realization of the original system [14].

The emphasis on improving the accuracy of computations led to so-called algorithms with *enhanced accuracy*. In many model reduction methods, the matrices L and T are determined from two positive semi-definite matrices P and Q, called generically *gramians*. The gramians can be always determined in Cholesky factorized forms $P = S^T S$ and $Q = R^T R$, where S and R are upper-triangular matrices. The computation of L and T can be done by computing the *singular value decomposition* (SVD)

$$SR^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \mathrm{diag}(\Sigma_1, \Sigma_2) \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T$$

where

$$\Sigma_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_r), \quad \Sigma_2 = \mathrm{diag}(\sigma_{r+1}, \ldots, \sigma_n),$$

and $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} \geq \ldots \geq \sigma_n \geq 0$.

The so-called *square-root* (**SR**) methods determine L and T as [15]

$$L = \Sigma_1^{-1/2} V_1^T R, \qquad T = S^T U_1 \Sigma_1^{-1/2}.$$

If r is the order of a minimal realization of G then the gramians corresponding to the resulting realization are diagonal and equal. In this case the minimal realization is called *balanced*. The **SR** approach is usually very accurate for well-equilibrated systems. However if the original system is highly unbalanced, potential accuracy losses can be induced in the reduced model if either L or T is ill-conditioned.

In order to avoid ill-conditioned projections, a *balancing-free* (**BF**) approach has been proposed in [16] in which always well-conditioned matrices L and T can be determined. These matrices are computed from orthogonal matrices whose columns span orthogonal bases for the right and left eigenspaces of the product PQ corresponding to the first r largest eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$. Because of the need to compute explicitly P and Q as well as their product, this approach is usually less accurate for moderately ill-balanced systems than the **SR** approach.

A *balancing-free square-root* (**BFSR**) algorithm which combines the advantages of the **BF** and **SR** approaches has been introduced in [14]. L and T are determined as

$$L = (Y^T X)^{-1} Y^T, \qquad T = X,$$

where X and Y are $n \times r$ matrices with orthogonal columns computed from the QR decompositions $S^T U_1 = XW$ and $R^T V_1 = YZ$, while W and Z are non-singular upper-triangular matrices. The accuracy of the **BFSR** algorithm is usually better than either of **SR** or **BF** approaches.

The SPA formulas can be used directly on a balanced minimal order realization of the original system computed with the **SR** method. A **BFSR** method to compute SPAs has been proposed in [17]. The matrices L and T are computed such that the system (LAT, LB, CT, D) is minimal and the product of corresponding gramians has a block-diagonal structure which allows the application of the SPA formulas.

Provided the Cholesky factors R and S are known, the computation of matrices L and T can be done by using exclusively numerically stable algorithms. Even the computation of the necessary SVD can be done without forming the product SR^T. Thus the effectiveness of the **SR** or **BFSR** techniques depends entirely on the accuracy of the computed Cholesky factors of the gramians. In the following sections we discuss the computation of these factors for several concrete model reduction techniques.

3. ALGORITHMS FOR STABLE SYSTEMS

In the *balance & truncate* (B&T) method [1] P and Q are the controllability and observability gramians satisfying a pair of continuous- or discrete-time Lyapunov equations

$$AP + PA^T + BB^T = 0, \quad A^T Q + QA + C^T C = 0;$$

$$APA^T + BB^T = P, \quad A^T QA + C^T C = Q.$$

These equations can be solved directly for the Cholesky factors of the gramians by using numerically reliable algorithms proposed in [18]. The **BFSR** version of the B&T method is described in [14]. Its **SR** version [15] can be used to compute balanced minimal representations. Such representations are also useful for computing reduced order models by using the SPA formulas [2] or the *Hankel-norm approximation* (HNA) method [4]. A **BFSR** version of the SPAs method is described in [17]. Note that the B&T, SPA and HNA methods belong to the family of absolute error methods which try to minimize $\|\Delta_a\|_\infty$, where Δ_a is the absolute error $\Delta_a = G - G_r$.

The *balanced stochastic truncation* (BST) method [3] is a relative error method which tries to minimize $\|\Delta_r\|_\infty$, where Δ_r is the relative error defined implicitly by $G_r = (I - \Delta_r)G$. In the BST method the gramian Q satisfies a Riccati equation, while the gramian P still satisfies a Lyapunov equation. Although the determination with high accuracy of the Cholesky factor of Q is computationally involved, it is however necessary to guarantee the effectiveness of the **BFSR** approach. Iterative refinement techniques are described for this purpose in [13].

Both the **SR** and SRBF versions of the B&T, SPA and BST algorithms are implemented in the RASP-MODRED library. The implementation of the HNA

method uses the **SR** version of the B&T method to compute a balanced minimal realization of the original system. All implemented routines are applicable to both continuous- and discrete-time systems. It is worth mentioning that implementations provided in commercial software [10, 11, 12] are only for continuous-time systems.

4. REDUCTION OF UNSTABLE SYSTEMS

The reduction of unstable systems can be performed by using the methods for stable systems in conjunction with two imbedding techniques. The first approach consists in reducing only the stable projection of G and then including the unstable projection unmodified in the resulting reduced model. The following is a simple procedure for this computation:

1. Decompose additively G as
$$G = G_1 + G_2$$
such that G_1 has only stable poles and G_2 has only unstable poles.
2. Determine G_{1r}, a reduced order approximation of the stable part G_1.
3. Assemble the reduced model G_r as
$$G_r = G_{1r} + G_2.$$

The second approach is based on computing a stable *rational coprime factorization* (RCF) of G say in the form $G = M^{-1}N$, where M, N are stable and proper rational TFMs, and then to reduce the stable system $[\,N\ M\,]$. From the resulting reduced model $[\,N_r\ M_r\,]$ we obtain $G_r = M_r^{-1}N_r$.

The coprime factorization approach used in conjunction with the B&T or BST methods fits in the general projection formulation introduced in Section 2. The gramians necessary to compute the projection are the gramians of the system $[\,N\ M\,]$. The computed matrices L and T by using either the **SR** or **BFSR** methods can be directly applied to the matrices of the original system. The main computational problem is how to compute the RCF to allow a smooth and efficient imbedding which prevents computational overheads. Two factorization algorithms proposed recently compute particular RCFs which fulfill these aims: the RCF with prescribed stability degree [19] and the RCF with inner denominator [20]. Both are based on a numerically reliable Schur technique for pole assignment. The use of other RCFs is presently under consideration.

RASP-MODRED provides all necessary tools to perform the reduction of unstable system. Routines are provided to compute left/right RCFs with prescribed stability degree or with inner denominators, to compute additive spectral decompositions, or to perform the back transformations. A modular implementation allows arbitrary combinations between various factorization and model reduction methods.

5. ALGORITHMS FOR FWMR

The FWMR methods try to minimize a weighted error of the form $\|W_1(G - G_r)W_2\|_\infty$, where W_1 and W_2 are suitable weighting TFMs. Many controller reduction problems can be formulated as FWMR problems [21]. Two basic approaches can be used to solve such problems. The approach proposed in [7] can be easily imbedded in the general formulation of Section 2. Provided G and the weights W_1 and W_2 are all stable TFMs, then P and Q are the frequency-weighted controllability and observability gramians of GW_2 and W_1G, respectively (for details see [21]). Unfortunately no proof of stability of the two-sided weighted approximation exists unless either $W_1 = I$ or $W_2 = I$.

In the second approach we assume that G is stable and W_1, W_2 are invertible, having only unstable poles and zeros. The technique proposed in [6] to solve the FWMR problem computes first G_1 the n-th order stable projection of W_1GW_2 and then computes the r-th order approximation G_{1r} of G_1 by using one of methods for stable systems. Finally G_r results as the r-th order stable projection of $W_1^{-1}G_{1r}W_2^{-1}$.

RASP-MODRED provides all necessary tools to perform FWMR. Special routines based on algorithms proposed in [22] are provided to compute efficiently the stable projections for the second approach.

6. THE RASP-MODRED LIBRARY

RASP-MODRED is one of the first numerical libraries developed by using the new linear algebra package LAPACK [9]. The library provides a rich set of computational facilities for model reduction. Besides the already mentioned functions, routines to evaluate Hankel- and L^2-norms of TFMs, to perform bilinear transformations, to compute systems couplings, are also available. Many lower level computational routines can have a special importance for other applications areas. In its present state of development the library consists of 77 routines and is continuously extended. Routines for alternative FWMR methods, for computing normalized RCF, or for evaluation of L_∞-norm are presently under development.

The implementation of the library has been done in accordance with the newly established RASP/SLICOT mutual compatibility concept [23]. Thus the implemented routines belong simultaneously to both RASP [24] and SLICOT [25] libraries. This software sharing strategy is meant to save future efforts in developing both libraries.

7. CONCLUSIONS

We presented an up to date overview of numerically reliable algorithms and associated software tools for model reduction. The algorithmic richness and the complexity of the model reduction problems require efficient and robust software implementations which can exploit efficiently all structural aspects of the underlying computational problems. This is possible only in high level languages such as Fortran. In contrast, implementations in MATLAB, although much more compact than the corresponding Fortran codes, are generally less efficient with respect to both operation count and memory usage. Moreover, many MATLAB implementations, done unfortunately by people with insufficient numerical expertise, are unsatisfactory with respect to requirements as generality, numerical reliability, accuracy.

8. ACKNOWLEDGMENTS

Many of presented algorithms as well as the development of the RASP-MODRED library are the results of a two-years stay of the author at the Ruhr-University Bochum as an Alexander von Humboldt research fellow under the direction of Prof. K.H. Fasol. The author also acknowledges the support of Prof. G. Grübel (DLR-Oberpfaffenhofen). Without his support the development of the RASP-MODRED library would have been not possible.

9. REFERENCES

[1] B. C. Moore. Principal component analysis in linear system: controllability, observability and model reduction. *IEEE Trans. Autom. Contr.*, 26:17–32, 1981.

[2] Y. Liu and B. D. O. Anderson. Singular perturbation approximation of balanced systems. *Int. J. Control*, 50:1379–1405, 1989.

[3] U. B. Desai and D. Pal. A transformation approach to stochastic model reduction. *IEEE Trans. Autom. Contr.*, 29:1097–1100, 1984.

[4] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds. *Int. J. Control*, 39:1115–1193, 1974.

[5] Y. Liu and B. D. O. Anderson. Controller reduction via stable factorization and balancing. *Int. J. Control*, 44:507–531, 1986.

[6] G. A. Latham and B. D. O. Anderson. Frequency-weighted optimal Hankel norm approximation of stable transfer functions. *Syst. Contr. Lett.*, 5:229–236, 1985.

[7] D. Enns. *Model Reduction for Control Systems Design*. PhD thesis, Dept. Aeronaut. Astronaut., Stanford Univ., Stanford, CA, 1984.

[8] A. Varga. *RASP Model Order Reduction Programs*. University of Bochum and DLR-Oberpfaffenhofen, TR R88-92, August 1992.

[9] E. Anderson *et al.* *LAPACK User's Guide*. SIAM, Philadelphia, 1992.

[10] R. Y. Chiang and M. G. Safonov. *Robust Control Toolbox 2.0*. The MathWorks Inc., 1992.

[11] G. Balas, J. Doyle, K. Glover, A. Packard, and R. Smith. *μ-Analysis and Synthesis Toolbox 1.0*. The MathWorks Inc., 1991.

[12] B.D.O Anderson and B. James. *$MATRIX_X$ Model Reduction Module*. Integrated Systems Inc., Santa Clara, CA, 1991.

[13] A. Varga and K. H. Fasol. A new square-root balancing-free stochastic truncation model reduction algorithm. *Prepr. 12th IFAC World Congress, Sydney*, vol. 7, pp. 153–156, 1993.

[14] A. Varga. Efficient minimal realization procedure based on balancing. In A. El Moudni, P. Borne, and S. G. Tzafestas, editors, *Prepr. of IMACS Symp. on Modelling and Control of Technological Systems*, vol. 2, pp. 42–47, 1991.

[15] M. S. Tombs and I. Postlethwaite. Truncated balanced realization of a stable non-minimal state-space system. *Int. J. Control*, 46:1319–1330, 1987.

[16] M. G. Safonov and R. Y. Chiang. A Schur method for balanced-truncation model reduction. *IEEE Trans. Autom. Contr.*, 34:729–733, 1989.

[17] A. Varga. Balancing-free square-root algorithm for computing singular perturbation approximations. *Proc. 30th IEEE CDC, Brighton, UK*, pp. 1062–1065, 1991.

[18] S. J. Hammarling. Numerical solution of the stable, non-negative definite lyapunov equation. *IMA J. Numer. Anal.*, 2:303–323, 1982.

[19] A. Varga. Coprime factors model reduction based on accuracy enhancing techniques. *Syst. Anal. Model. Sim.*, 11:303–311, 1993.

[20] A. Varga. A Schur method for computing coprime factorizations with inner denominators and applications in model reduction. *Proc. 1993 ACC, San Francisco, CA*, pp. 2130–2131, 1993.

[21] B. D. O. Anderson and Y. Liu. Controller reduction: concepts and approaches. *IEEE Trans. Autom. Contr.*, 34:802–812, 1989.

[22] A. Varga. Explicit formulas for an efficient implementation of the frequency-weighted model reduction approach. *Proc. 1993 ECC, Groningen, NL*, pp. 693–696, 1993.

[23] G. Grübel, A. Varga, A. van den Boom, and A. J. Geurts. Towards a coordinated development of numerical CACSD software: the RASP/SLICOT compatibility concept. *Prepr. CACSD'94, Tucson*, 1993.

[24] G. Grübel and H.-D. Joos. RASP and RSYST - two complementary program libraries for concurrent control engineering. In *Prepr. 5th IFAC/IMACS Symp. CADCS'91, Swansea, UK*, pp 101–106. Pergamon Press, 1991.

[25] A. van den Boom *et al.* SLICOT, a subroutine library in control and systems theory. *Prepr. 5th IFAC/IMACS Symp. CADCS'91, Swansea, UK*, pp. 89–94. Pergamon Press, 1991.