

On stochastic balancing related model reduction

A. Varga

German Aerospace Center, DLR - Oberpfaffenhofen
 Institute of Robotics and Mechatronics
 D-82234 Wessling, Germany.
 Andras.Varga@dlr.de

Abstract

We propose a general method based on the *balanced stochastic truncation* (BST) approach for the model reduction of stable linear systems. The new method relies on a recent general inner-outer factorization result and extends the applicability of the BST method to systems with infinite zeros. A computational algorithm with enhanced accuracy for the new BST model reduction approach is presented. The capabilities and advantages of the new approach are illustrated on an example.

1 Introduction

Consider $G(s)$ a $p \times m$ stable *transfer-function matrix* (TFM) with a minimal state space realization $G := (A, B, C, D)$ of order n satisfying

$$G(s) = C(sI - A)^{-1}B + D.$$

If $G(s)$ is square and D nonsingular it is possible to compute the left spectral factor $W(s)$ of $G(s)G^\sim(s)$ satisfying

$$W^\sim(s)W(s) = G(s)G^\sim(s),$$

where we denoted $G^\sim(s) := G^T(-s)$. A state space realization of $W(s)$ can be obtained as $W = (A, B_W, C_W, D^T)$ (see e.g. [15]) with

$$\begin{aligned} B_W &= PC^T + BD^T \\ C_W &= D^{-1}(C - B_W^T X), \end{aligned}$$

where P is the controllability Gramian of G given by

$$AP + PA^T + BB^T = 0 \quad (1)$$

and X is the observability Gramian of W , being the solution of the Riccati equation

$$XA + A^T X + C_W^T C_W = 0. \quad (2)$$

The state space realization of G is called a *balanced stochastic realization* [3] if $P = X = \Sigma$ with $\Sigma =$

$\text{diag}(\sigma_1, \dots, \sigma_n)$, where σ_i is the i -th Hankel singular value of the stable part of the so-called "phase matrix" $F(s) := (W^\sim(s))^{-1}G(s)$.

We assume now that G is already stochastically balanced and the singular values are ordered decreasingly. Partition Σ as

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \quad (3)$$

where $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_r)$, $\Sigma_2 = \text{diag}(\sigma_{r+1}, \dots, \sigma_n)$, and $\sigma_r > \sigma_{r+1}$. Partition similarly the system matrices

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = [C_1 \ C_2].$$

The truncated model $G_r = (A_{11}, B_1, C_1, D)$ is a stochastically balanced approximation of the original system G , is stable, and satisfies the relative error bound

$$\|G^{-1}(G - G_r)\|_\infty \leq \prod_{i=r+1}^n \frac{1 + \sigma_i}{1 - \sigma_i} - 1. \quad (4)$$

This model reduction approach, called *balanced stochastic truncation* (BST), has been introduced in [1], and further analysed and extended in [3, 4].

The BST approach belongs to the family of relative error methods [15]. In contrast to absolute error methods like the *balanced truncation* (BT) method [7] or the *singular perturbation approximation* (SPA) method [6], the BST method has the main advantage to provide a uniform approximation of the frequency response of the original system over the whole frequency domain, and even more important, to preserve phase information. For example, for a minimum-phase original system, the BST-approximation is also minimum-phase. However, this is not generally true for the absolute error methods.

Computational approaches for the BST model reduction approach are restricted to systems with full rank D . If D has no full rank, it was suggested in [2] to replace $G(s)$ by $[G(s) \ \alpha I]$ with $\alpha > 0$ and solve the

BST problem for the modified system. The BST approach applied to $[G(s) \ \alpha I]$ leads for small α values to a BST-type relative error approximation, while for large α values to an BT-type additive error approximation. However, small values of α lead to ill-conditioned Riccati equations, thus preventing the computation of accurate "pure" relative error approximations.

In this paper we propose an extension of the BST method to arbitrary stable systems. The basis of this extension is the possibility to compute the left spectral factor $W(s)$ for an arbitrary $G(s)$ in a numerically reliable way by using a recent algorithm for inner-outer factorization [9]. In this context, we propose a new, completely general computational algorithm with enhanced numerical accuracy for the BST model reduction approach. We illustrate the capabilities and advantages of the extended BST method by reducing a 120-th order finite element model of a CD-player.

2 Computation of left/right spectral factors

The basis for a general BST approach is the following theorem.

Theorem 1 *Let $G(s)$ be a stable TFM of arbitrary rank with a minimal realization $G = (A, B, C, D)$. Then there exist $W(s)$ outer and $V(s)$ co-outer with state space realizations of the form $V = (A, B_V, C, D_V)$ and $W = (A, B_W, C_W, D_W)$ such that*

$$G(s)G^\sim(s) = V(s)V^\sim(s) = W^\sim(s)W(s).$$

Proof. The proof is constructive, providing procedures to compute the two factors.

Computation of $V(s)$. Applying the result of [9] to $G^T(s)$ we can determine the inner-outer factorization

$$G^T(s) = G_i^T(s)V^T(s),$$

where $G_i^T(s)$ is the inner factor and $V^T(s)$ is the outer factor. Thus $V(s)$ is co-outer. Moreover, the procedure of [9] determines $V^T(s)$ with a state space realization of the form (A^T, C^T, B_V^T, D_V^T) with B_V and D_V appropriate matrices.

Computation of $W(s)$. Consider the conjugate system $G^\sim = (-A^T, -C^T, B^T, D^T)$. A left coprime factorization of $G^\sim(s)$ in the form $G^\sim(s) = M_i^{-1}(s)N(s)$ with M_i co-inner can be determined with N given by [15, Theorem 13.36]

$$N = (-A^T + KB^T, -C^T + KD^T, B^T, D^T),$$

where $K = -P^{-1}B$ and P satisfies (1). However

$$-A^T + KB^T = -A^T - P^{-1}BB^T = P^{-1}AP$$

thus, N has the equivalent state space realization

$$N = (A, PC^T + BD^T, -B^T P^{-1}, D^T). \quad (5)$$

We apply once again the inner-outer factorization algorithm of [9] to factorize $N(s)$ in the form $N(s) = N_i(s)W(s)$, with N_i inner and $W(s)$ outer. The state space representation of $W(s)$ is obtained as

$$W = (A, PC^T + BD^T, C_W, D_W),$$

where C_W and D_W are appropriate matrices. Overall we have $G^\sim(s) = M_i^{-1}(s)N_i(s)W(s)$ and thus $G(s)G^\sim(s) = W^\sim(s)W(s)$. \square

We can now define the balanced stochastic realization of an arbitrary stable $G(s)$ using the computed left spectral factor $W(s)$ in Theorem 1 by requiring that the two Gramians P and X in (1) and (2) are equal and diagonal. This can be achieved by computing an appropriate balancing transformation matrix [7]. In the next section, we discuss a numerically reliable approach to compute the matrices of the reduced model which exploits the structure of the model reduction problem and is suitable for robust software implementation.

3 A BST algorithm with enhanced accuracy

To determine the stochastic balanced state space representation of $G(s)$ we need to compute an appropriate transformation matrix Z such that the transformed Gramians are equal and diagonal, satisfying thus

$$Z^{-1}PZ^{-T} = Z^T X Z = \Sigma.$$

If we partition Z and Z^{-1} in accordance with Σ in (3)

$$Z^{-1} := \begin{bmatrix} L \\ V \end{bmatrix}, \quad Z := [T \ U],$$

then $LT = I_r$ and $\Pi = TL$ is a projection matrix. Thus the reduced system G_r is given by

$$G_r := (A_r, B_r, C_r, D_r) = (LAT, LB, CT, D).$$

The matrices L and T are called *truncation* matrices.

In the balancing related model reduction methods, the truncation matrices L and T can be determined knowing only the Cholesky factors of the Gramians P and X . Let $P = S^T S$ and $X = R^T R$ be in Cholesky factored forms and compute the singular value decomposition

$$RS^T = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} [V_1 \ V_2]^T, \quad (6)$$

where $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_r)$, $\Sigma_2 = \text{diag}(\sigma_{r+1}, \dots, \sigma_n)$, and $\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} \geq \dots \geq \sigma_n \geq 0$. Then the truncation matrices L and T can be determined as [11]

$$L = \Sigma_1^{-1/2} V_1^T R, \quad T = S^T U_1 \Sigma_1^{-1/2}. \quad (7)$$

The emphasis on improving the accuracy of computations by exploiting the problem structure led to so-called algorithms with *enhanced accuracy*. The above approach to compute the truncation matrices using exclusively the Cholesky factors of the Gramians is called the *square-root* BST method. We can exploit another structural feature of the problem to increase the computational efficiency of the BST method. Since G and W share the same state matrix A we can first reduce A to a *real Schur form* (RSF). This reduction is always necessary for the efficient and accurate computation of the Gramians in Cholesky factored forms $P = S^T S$ and $X = R^T R$ by employing the algorithm of Hammarling [5]. This preliminary reduction is included in the following BST model reduction algorithm:

Square-root BST model reduction method.

1. Compute orthogonal Q to reduce A to a RSF

$$A \leftarrow Q^T A Q, \quad B \leftarrow Q^T B, \quad C \leftarrow C Q.$$

2. Solve for the Cholesky factor S of P the Lyapunov equation

$$A(S^T S) + (S^T S)A^T + B B^T = 0$$

3. With $P = S^T S$, form $N = (A, P C^T + B D^T, -B^T P^{-1}, D^T)$ and compute, using the algorithm in [9], the outer factor of the inner-outer factorization of N as

$$W = (A, P C^T + B D^T, C_W, D_W).$$

4. Solve for the Cholesky factor R of X the Lyapunov equation

$$(R^T R)A + A^T (R^T R) + C_W^T C_W = 0.$$

5. Compute the singular value decomposition (6) and determine the truncation matrices as

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

6. Compute the reduced system

$$G_r = (L A T, L B, C T, D).$$

An alternative "balancing-free" approach to compute the truncation matrices has been proposed in [10]. This approach avoids basically balancing and was primarily intended to improve the numerical accuracy of computations when ill-conditioned (i.e., nearly rank deficient) truncation matrices L and T can result because the original system is poorly scaled. It relies on the observation that, for any invertible \tilde{Z} , the pair $(\tilde{Z}^{-1} L, T \tilde{Z})$ defines the same projection $\Pi = T L$ as the pair (L, T) in (7), but in a different coordinate form. A *balancing-free square-root* approach combining the *square-root*

method of [11] and the *balancing-free* technique of [10] has been proposed for the BST method in [12]. This involves the computation of two QR-decompositions

$$S^T V_1 = T J, \quad R^T U_1 = F H,$$

where J and H are nonsingular matrices and T and F are matrices with orthonormal columns. With the computed F and T , L is computed as

$$L = (F^T T)^{-1} F^T.$$

The resulting truncation matrices L and T are always well conditioned.

The key computation in the BST model reduction approach is the computation of the inner-outer factorization of N at step 3 using the algorithm of [9]. This computation consists in two main steps: (1) the full row rank compression of N involving the solution of a reduced order standard Riccati equation, and (2) the dislocation of unstable zeros by solving a reduced order Lyapunov equation. A delicate computation is a preliminary orthogonal reduction of the system pencil to isolate a reduced order subsystem with special properties. This reduction, involving repeated rank determinations, can be sensible to the scaling of the system.

It would be ideal to perform the BST algorithm on an already balanced minimal realization, with the balanced Gramian P having a reasonably small condition number. However, for a nearly non-minimal system, the controllability Gramian P can be almost singular. Therefore, the need to invert P to form N at step 3 can lead to an unacceptable overall accuracy loss. In such cases, it is advisable to apply first an alternative method (e.g., the additive error BT approach) to perform a preliminary order reduction which guarantees a well conditioned P . Such a preliminary reduction could also be beneficial when the applicability of the BST-method is restricted by a too high order of the original system.

4 Discrete-time systems

Let $G(z)$ be a discrete-time TFM with a minimal state space realization $G = (A, B, C, D)$ of order n satisfying $G(z) = C(zI - A)^{-1}B + D$. Let $G^\sim(z) := G^T(z^{-1})$ denote the conjugate TFM for $G(z)$. We have a result similar to Theorem 1 for the spectral factorization of discrete-time systems.

Theorem 2 *Let $G(z)$ be a stable TFM of arbitrary rank with a minimal realization $G = (A, B, C, D)$. Then there exist $W(z)$ outer and $V(z)$ co-outer with state space realizations of the form $V = (A, B_V, C, D_V)$ and $W = (A, B_W, C_W, D_W)$ such that*

$$G(z)G^\sim(z) = V(z)V^\sim(z) = W^\sim(z)W(z).$$

Proof. The proof is immediate by using bilinear transformation formulas as in [13]. \square

The bilinear transformation technique can serve as basis for a discrete-time BST algorithm. We apply the continuous-time BST algorithm to the continuous-time system $G_c(s) := G(\frac{s+1}{s-1})$ resulted by a bilinear transformation. If $G_{c,r}$ is the computed continuous-time BST approximation for G_c , then the discrete-time BST approximation is obtained by the inverse bilinear transformation $G_r(z) = G_{c,r}(\frac{z+1}{z-1})$. The disadvantage of this approach is that the resulting reduced system G_r has possibly a nonzero D_r matrix even if the original system is strictly proper.

It is also possible to develop a discrete-time analog of the continuous-time BST-algorithm. The main difficulty is that, in general, G could have poles in the origin. In such a case, A is singular and thus a standard state space realization for G^\sim can not be formed. Even for a nonsingular A , the explicit computation of $G^\sim := (A^{-T}, -A^{-T}C^T, B^T A^{-T}, D^T - B^T A^{-T}C^T)$ is not advisable, when A is ill-conditioned. However, explicit formulas for the numerator factor N in $G^\sim(z) = M_i^{-1}(z)N(z)$ can still be derived with help of bilinear transformations formulas. By applying the continuous-time coprime factorization algorithm to $G_c(s) := G(\frac{s+1}{s-1})$ we get the corresponding continuous-time N_c (see (5)). Then we compute $N(z) := N_c(\frac{z+1}{z-1})$ and we obtain explicit expressions for the matrices of the state space representation

$$N = (A, B_N, C_N, D_N),$$

with

$$\begin{aligned} B_N &= APC^T + BD^T, \\ C_N &= B^T(A^T - I)^{-1}P^{-1}(A - I), \\ D_N &= D^T - B^T(A^T - I)^{-1}(C^T + P^{-1}B_N), \end{aligned}$$

where P is the discrete-time controllability Gramian satisfying

$$APA^T - P + BB^T = 0. \quad (8)$$

For the inner-outer factorization at step 3 the discrete-time algorithm proposed in [8] is used, to obtain $W = (A, APC^T + BD^T, B_W, C_W)$. The observability Gramian X can be computed by solving

$$A^T X A - X + C_W^T C_W = 0. \quad (9)$$

Both Lyapunov equations (8) and (9) can be solved directly for the Cholesky factors of the Gramians S and R , respectively, using the discrete-time Hammarling algorithm [5]. Note however, that due to the rather complicated expressions of the state space matrices of the realization of N , the explicit discrete-time approach seems to have no special computational advantages over the bilinear transformation approach.

5 Numerical example

To illustrate the capabilities of the proposed approach, we apply it to a strictly proper model and compare it with two well-known absolute error approaches: the balanced truncation (BT) method [7] and the singular perturbation approximation (SPA) [6]. As example, we use the 120-th order single-input single-output finite element model of the dynamics between the lens actuator and radial arm position of a portable compact disc player discussed in [14]. This system is not minimum-phase, having besides the 113 stable zeros, also 5 unstable zeros and 2 infinite zeros.

Due to physical constraints on the size of the systems' controller, a reduced model with order $r \leq 15$ is desired. For comparison purposes we determined 15-th order reduced models G_r^{BST} , G_r^{BT} , and G_r^{SPA} with the BST, BT and SPA methods, respectively. The Hankel singular values of the original G system and the singular values computed by the BST algorithm are shown in Figures 1 and 2, respectively.

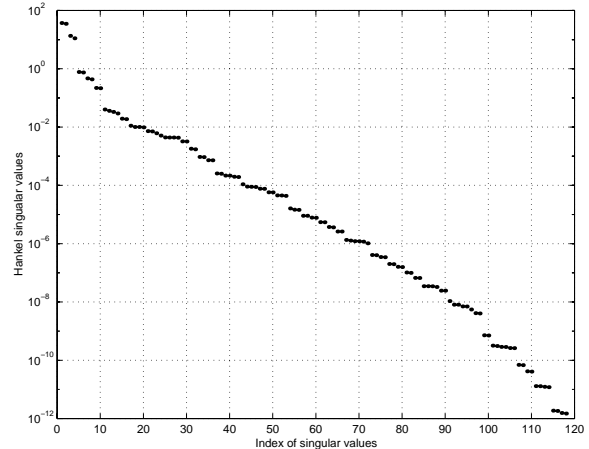


Figure 1: Hankel singular values of G .

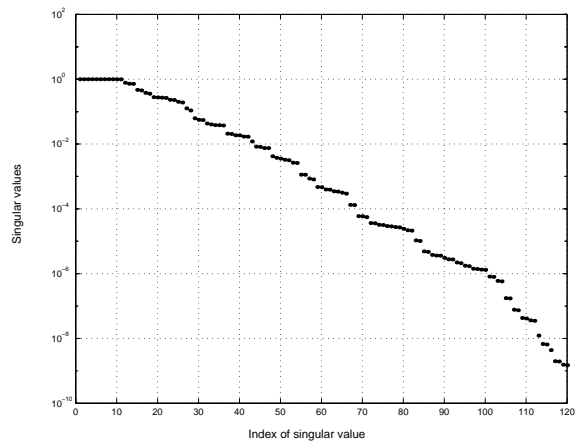


Figure 2: Stochastic singular values.

From the wide range of variation of the Hankel singular values of G in Figure 1 we expected numerical difficulties when inverting P . Indeed, the controllability Gramian P to be used at step 3 of the algorithm had a condition number of order 10^{17} . Therefore, the original system is numerically uncontrollable and the subsequent operations with P led to severe accuracy loss. To prevent excessive accuracy loss or even the failure of the inner-outer factorization algorithm an additional scaling of the model before step 3 was performed. The computed reduced system has 8 stable, 5 unstable and two infinite zeros. The 5 unstable zeros are practically the same as those of the original system. These results regarding the preservation of unstable zeros are consistent with the theory of BST model reduction developed in [3] for non-minimum phase systems.

An alternative approach which we also tested was to perform the BST method after a preliminary order reduction with the BT method to guarantee an acceptable condition number for P . The preliminary reduction with the BT method led to a system of order 72, for which the condition number of P was of order 10^7 (as expected from Figure 1). The zeros of this reduced model contain two infinite zeros and the 5 unstable zeros of the original system. This infinite and unstable zeros structure is kept also after performing the BST approach on the reduced model. Note that no additional scaling at step 3 was necessary to use this combined BT and BST approach.

We computed the absolute and relative errors from the frequency responses evaluated on a grid of 10000 frequency values in the frequency range $[10^{-8}, 10^8]$. The resulting absolute errors for the three methods are

$$\begin{aligned} \|G - G_r^{BST}\|_\infty &= 8.047, \\ \|G - G_r^{BT}\|_\infty &= 0.0423, \\ \|G - G_r^{SPA}\|_\infty &= 0.0423 \end{aligned}$$

and the relative errors are

$$\begin{aligned} \|G^{-1}(G - G_r^{BST})\|_\infty &= 1.07, \\ \|G^{-1}(G - G_r^{BT})\|_\infty &= 2.1682 \cdot 10^3, \\ \|G^{-1}(G - G_r^{SPA})\|_\infty &= 8.1742 \cdot 10^8. \end{aligned}$$

As expected, the BST method has the lowest relative error in comparison with the very poor relative errors for the BT and SPA methods. On contrary, the absolute errors for the absolute error methods are much better than that for the BST approach.

The good uniform approximation of both amplitude and phase obtained with the BST method can be seen in the Nichols plots in Figure 3. Note that while the central frequencies part is reasonably well approximated by all methods, there are significant differences at both low and high frequencies, where the BT and SPA methods provide poor approximations.

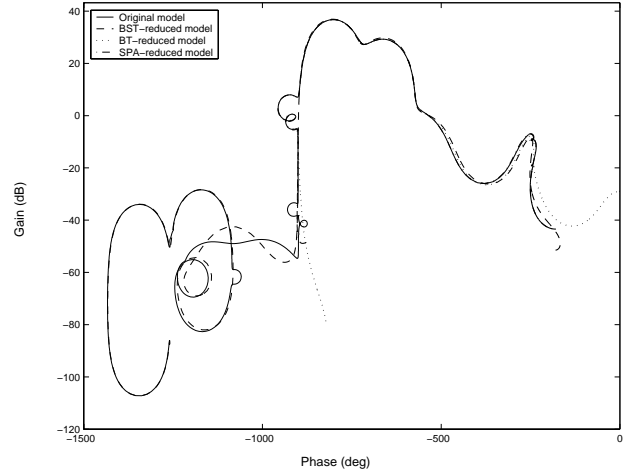


Figure 3: Nichols frequency responses.

The Figure 4 shows more in details the performance of different computed approximations on basis of Bode magnitude plots. All methods approximate satisfactorily the central peak at frequency about 120 Hz, but have different approximation properties at low and high frequencies. The SPA method has zero stationary error but appears inappropriate at high frequencies. Although the BT approximation is somewhat better at high frequencies, its performance at both low and high frequencies appears still unsatisfactory. The BST approximation appears satisfactory over the whole range of frequency values, being able to follow the shape of the gain with high fidelity over the whole frequency range (see also the zoomed region in Figure 5).

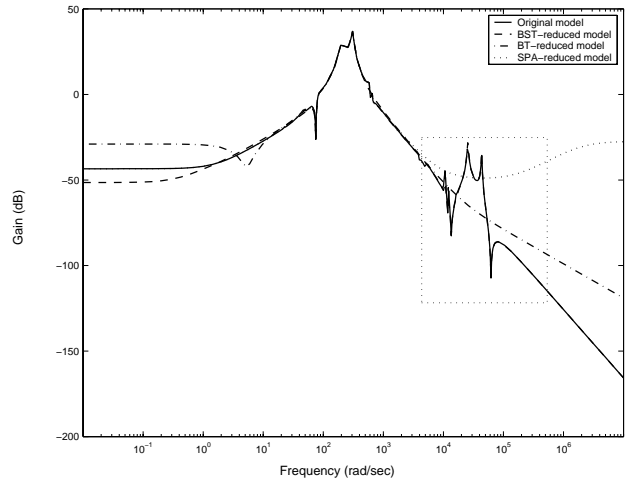


Figure 4: Bode frequency responses

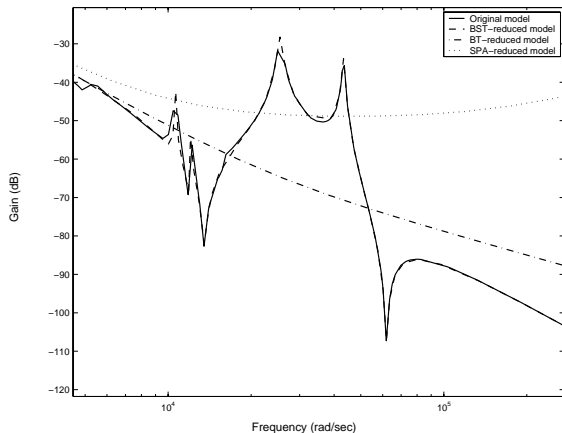


Figure 5: Zoomed area of Figure 4.

6 Conclusion

We presented a computational method for the BST model reduction which is applicable regardless the original system has infinite zeros or not, or its TFM has full rank or not. Each step of the proposed BST algorithm can be performed using numerically reliable procedures. The critical computation is the inner-outer factorization of a TFM with possible infinite zeros and/or non-full rank. This computation can be performed in a numerically reliable way by using a recently developed inner-outer factorization algorithm [9]. The proposed new BST algorithm for continuous-time systems can be used for discrete-time systems as well by means of bilinear transformation techniques.

Several open issues related to the proposed BST algorithm are still to be addressed. One important theoretical problem is the derivation of a bound for the achieved relative error. Based on limited experimental evidence, we conjecture that a bound of the form (4) could be also valid for the new BST method. From numerical point of view it is important to study the role of scaling to prevent excessive ranges for the elements of system matrices and to simultaneously reduce the condition number of controllability Gramian. This is especially important to increase the reliability of rank decisions when performing the inner-outer factorization at step 3.

In spite of intrinsic numerical difficulties, we believe that the BST approach is a viable method to obtain reduced order plant models or controllers, when a good uniform approximation over the whole frequency range is more important than small absolute errors. As model reduction algorithm, this method is best suited as final step of reducing high order models, following a preliminary order reduction with powerful additive error methods like the BT or SPA approaches. These addi-

tive error methods perform well regardless the original system is minimal or not and can compute reduced order models with a prescribed condition number of the Gramians. By combining the BT and BST method, the risk of accuracy losses in the BST method due to inversion of an ill-conditioned Gramian can be significantly reduced.

References

- [1] U. B. Desai and D. Pal. A transformation approach to stochastic model reduction. *IEEE Trans. Autom. Control*, 29:1097–1100, 1984.
- [2] K. Glover. Multiplicative approximation of linear multivariable systems with L_∞ error bounds. In *Proc. ACC'86, Seattle, WA*, pp. 1705–1709, 1986.
- [3] M. Green. Balanced stochastic realizations. *Lin. Alg. & Appl.*, 98:211–247, 1988.
- [4] M. Green. A relative error bound for balanced stochastic truncation. *IEEE Trans. Autom. Control*, 33:961–965, 1988.
- [5] S. J. Hammarling. Numerical solution of the stable, non-negative definite Lyapunov equation. *IMA J. Numer. Anal.*, 2:303–323, 1982.
- [6] Y. Liu and B. D. O. Anderson. Singular perturbation approximation of balanced systems. *Int. J. Control*, 50:1379–1405, 1989.
- [7] B. C. Moore. Principal component analysis in linear system: controllability, observability and model reduction. *IEEE Trans. Autom. Control*, 26:17–32, 1981.
- [8] C. Oară and A. Varga. The general inner-outer factorization problem for discrete-time systems. In *Proc. ECC'99, Karlsruhe, Germany*, 1999.
- [9] C. Oară and A. Varga. Computation of general inner-outer and spectral factorizations. *IEEE Trans. Autom. Control*, 45, No. 7, 2000.
- [10] M. G. Safonov and R. Y. Chiang. Model reduction for robust control: a Schur relative error method. *Int. J. Adapt. Contr.&Sign. Proc.*, 2:259–272, 1988.
- [11] M. S. Tombs and I. Postlethwaite. Truncated balanced realization of a stable non-minimal state-space system. *Int. J. Control*, 46:1319–1330, 1987.
- [12] A. Varga and K. H. Fasol. A new square-root balancing-free stochastic truncation model reduction algorithm. In *Prepr. of 12th IFAC World Congress, Sydney, Australia*, vol. 7, pp. 153–156, 1993.
- [13] W. Wang and M. G. Safonov. Comparison between continuous and discrete-time model truncation. In *Proc. 29th CDC, Honolulu, Hawaii*, pp. 494–499, 1990.
- [14] P. Wortelboer. *Frequency-weighted Balanced Reduction of Closed-loop Mechanical Servo-systems: Theory and Tools*. PhD thesis, Techn. Univ. Delft, 1994.
- [15] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, 1996.