A more reliable reduction algorithm for behavioral model extraction

Dmitry Vasilyev
Department of Electrical Engineering and Computer Science,
Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139
email: vasilyev@mit.edu

Jacob White
Department of Electrical Engineering and Computer Science,
Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139
email: white@mit.edu

***Abstract*— In this paper we are concerned with developing more reliable model reduction algorithms. We have focused on less common, but real, examples that fail to be effectively reduced by almost all of the currently popular model reduction methods. The failure of these popular methods is due to the fact that they all separately examine controllability and observability. We then present a new method based on several modifications and extensions of the recently developed AISIAD algorithm. The modified AISAID method is demonstrated on a wide variety of examples, including electrical interconnnect, micromachined devices, and heat flow to show that the method is reliable. Our modified AISAID method is either nearly equivalent or far superior to any of the other reduction methods suitable for large scale problems.**

I. Introduction

Model order reduction is becoming a widely used technique for generating differential-equation based behavioral models from more detailed ordinary and partial differential equation based descriptions. The most mature such application is that of interconnect modeling, but these techniques are now being used to generate models of circuit blocks, micromachined devices, integrated circuit heat flow, photonic devices, and even turbine engines [1], [2], [3]. The emerging general strategy for generating low-order models is a combination approach. First, a fast but non-optimal technique is used to reduce a large model to a medium-sized model. Then a reduction method with an optimality property but a higher computational cost, like Truncated Balanced Realization (TBR) or convex optimization [4], is used to generate a very efficient low order model from the medium-sized model.

The above strategy breaks down when the fast reduction method is not able to generate a medium-sized model of reasonable accuracy, and there are some cases where this breakdown occurs for almost all of the currently popular fast reduction methods. Although this breakdown is not common, the widening application of these methods suggests that it is now important to focus on reliability.

In this paper, we examine the difficulty that the most widely used techniques for generating medium sized models, whether based on Krylov subspaces[5] or collections of frequency domain solutions, effectively separately obtain low-rank approximations of controllability (P) and observability (Q) gramians [6], [7], [8], [9], [3]. Separate controllability and observability approaches can be ineffective when applied to problems with an innate asymmetry, such as very lightly damped interconnect problems or micromachined devices with a directional input-output behavior. The difficulty is that the dominant eigenspaces of P and Q are very different from what is needed, the dominant spaces of the PQ product. In order to overcome this difficulty, the AISIAD method was proposed in [10]. However, as we shall see later (section II-F.1), the original AISIAD method has a limited applicability and there are ways of dramatically increasing its accuracy.

In this paper we present a modification of the AISIAD algorithm, which overcomes the limited applicability of the original AISIAD [10] and provides much better accuracy. In addition, we provide generalization of this method to the descriptor systems with nonsingular descriptor matrix E. Our numerical experiments show that our modified AISIAD algorithm works much better than both the original AISIAD and than the methods based on the separate approximations of gramians. We show that approximating gramians separately doesn't necessarily lead to an accurate approximation to TBR.

The rest of the paper is structured as follows. In section II we review TBR model reduction itself and its most widely used approximations. We present a motivating example which highlights the problem with such approximations. We also introduce the immediate precursor of our method, the AISIAD approximation and describe its key algorithm - solving Sylvester equation. In section III we propose a modification to the solution of Sylvester equation, which expands the class of applicable models. In section IV we propose the second improvement of the AISIAD algorithm by approximating the terms which were neglected in the original paper. We summarize the resulting algorithm in the algorithm 5. In section VI we generalize our algorithm for descriptor systems with nonsingular descriptor matrix E. In the subsequent sections we discuss applicability and accuracy of the method, along with numerical results. In section IX we present our conclusions.

II. Background and prior work

A. MOR setup and notations used

In this paper we consider approximation of the linear time-invariant (LTI) continuous-time(CT) stable causal dynamical systems (for background see [11] or [6]) in two kinds of state-space forms:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \quad \text{and} \quad \begin{cases} E\dot{x}(t) = Ax(t) + Bu(t) \\ y = Cx(t) \end{cases} \tag{1}$$

The first kind of system description will be referred as system in normal form (A,B,C), the second one - as a descriptor form or (E,A,B,C). Vector $x(t) \in \mathbb{R}^n$ is a vector of states at time t, B is an $n \times m$ input matrix, $u : \mathbb{R} \to \mathbb{R}^m$ is an input signal, C is an $l \times n$ output matrix and $y : \mathbb{R} \to \mathbb{R}^l$ is the output signal, A is an $n \times n$ system matrix, and matrix $E \in \mathbb{R}^{n \times n}$ is a nonsingular descriptor matrix. As an example, the modified nodal analysis algorithm generates descriptor systems when applied to general RLC circuits, but generates a system in normal form for RC circuits with only unit-valued grounded capacitors.

In the large-scale setting the descriptor form cannot be trivially converted to the (A,B,C) form by multiplication on E^{-1}, because this computation is unacceptably expensive and sometimes numerically ill-conditioned.

The goal of the model order reduction problem is to construct a *reduced system* having the order q much less than initial order n, while the input-output relationship for the reduced system should be as close as possible to that of the original system.

Throughout the paper we denote by P and Q the symmetric controllability and observability gramians of the system. These gramians satisfy the following generalized Lyapunov equations (in case of descriptor systems):

$$\begin{aligned} APE^T + EPA^T + BB^T = 0 \\ A^T QE + E^T QA + C^T C = 0 \end{aligned} \tag{2}$$

and have interpretations in terms of energy. Assuming zero input and initial state x_0, the output energy $\int_0^\infty (y^T y)dt = x_0^T (E^T QE)x_0$. The minimal energy in the input, needed to drive a system from the zero state at $t = -\infty$ to state x_0 at $t = 0$ is given by $\int_0^\infty (u^T u)dt = x_0^T P^{-1} x_0$.

In the following, we denote order of the initial (large) system by n, and the order of a reduced system by q. By the notation *orthogonal basis* we refer to any matrix $V \in \mathbb{R}^{n\times q}$, such that $V^T V = I_{q\times q}$. By $\Lambda(A) = \{\lambda_i(A)\}$ we denote a spectrum of A. By $\Lambda_k^{dom}(A)$ we shall denote k largest eigenvalues of A. By $\mathcal{K}_q(A,B)$ we denote the order-q Krylov subspace of A corresponding to B:

$$\mathcal{K}_q(A,B) = span\{B, AB, \ldots, A^{q-1}B\} \tag{3}$$

B. TBR model reduction algorithm

Truncated balanced model order reduction algorithm was originally proposed in [12], and we provide it here as algorithm 1. The properties of this algorithm are well described elsewhere [12] [13].

Algorithm 1: Balanced-truncation reduction algorithm (TBR)
Input: Initial system (A,B,C), desired reduced order q
Output: Reduced-order system $(\hat{A},\hat{B},\hat{C})$
(1) Find observability grammian P:
$AP + PA^T + BB^T = 0$;
(2) Find controllability grammian Q:
$A^T Q + QA + C^T C = 0$;
(3) Compute q left and right dominant eigenvectors of PQ:
$(PQ)S_c = S_c D_{PQ}$, where $D_{PQ} = diag(\Lambda_q^{dom}(PQ))$
$S_o^T(PQ) = D_{PQ} S_o^T$ and scale columns of S_o such that
$S_o^T S_c = I_{q\times q}$
(4) Project the system matrices
$\hat{A} = S_o^T A S_c, \quad \hat{B} = S_o^T B, \quad \hat{C} = CS_c$
(5) **return** $(\hat{A},\hat{B},\hat{C})$

The TBR model reduction algorithm produced models which have a bound on the worst-case frequency response error, and are typically good wideband models.

Note that TBR reduction algorithm uses the dominant eigenspaces of PQ, roughly the modes with the best product of controllability and observability.

C. Approximations to the TBR algorithm

Currently the most widely used approximations to the TBR algorithm are based on low-rank approximations of gramians $\hat{P} \approx P$ and $\hat{Q} \approx Q$. They either use $\hat{P}$ and $\hat{Q}$ instead of P and Q in the balancing procedure via SVD, or they merge the dominant eigenspaces of $\hat{P}$ and $\hat{Q}$ in one orthogonal projection matrix and use this orthogonal basis for subsequent projection of system matrices (step 4 of algorithm 1). The first method is referred as *low-rank square-root method* and the second option is referred as *dominant gramian eigenspaces method (DGE)* (for detailed description see [6], pp. 37-39).

It is well-known that such approximations produce accurate reduced models only if the dominant eigenspaces of P and Q coincide [14], as, for example, in symmetric systems ($A = A^T, C = B^T$). There have been attempts to use cross-gramian approaches [14] in order to overcome this shortcoming, however this approach can be used only for square symmetric LTI systems. The most promising attempt was made in the work [10] where the method called AISIAD (Approximate Implicit Subspace Iteration with Alternating Directions) was developed. However, AISIAD method is applicable to the class of systems with matrix A, which has the property of being Hurwitz under orthogonal projections, which narrows the class of applicable dynamical systems.

D. Motivating example: simple RLC line

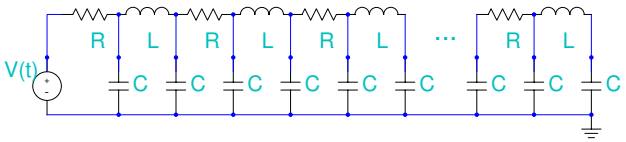

Fig. 1. RLC transmission line model as a motivating example.

The fact that computing dominant approximations of P and Q separately does not approach accuracy of TBR reduction, becomes apparent if one performs a modified nodal analysis (MNA) of the simple RLC transmission line depicted on figure 1 and then considers a very lightly damped case. In the MNA formulation, the state space consists of the voltages on the capacitors and currents through inductors of a circuit. Let the input to the system be voltage applied to the first node, and the output be the the current through the first resistor in the chain. The MNA analysis results in the descriptor system (E,A,B,C) with the positive semidefinite matrices E and $(-A)$ and $C = B^T$. We used a lightly damped line, with the parameters $R = 0.05, L = 10^{-10}, C = 10^{-15}$, the number of inductors $N = 100$. If we convert this system to the state-space model (A,B,C), the first N dominant eigenvectors of P and first N dominant eigenvectors of Q span almost completely orthogonal subspaces! This gives an approximation of PQ being almost zero.

This means that in order to get a good approximation of a product PQ one needs to get a low-rank approximations of P and Q essentially greater than N, and is consequently not applicable in a large-scale setting.

This example illustrates a fundamental problem: capturing dominant controllable and dominant observable modes separately is not sufficient to get a good approximation to TBR, and can lead to arbitrarily large errors in the frequency domain. Even the PRIMA algorithm [15], which guarantees passivity of the reduced model[1] produces quite bad approximations in the lightly damped cases in the H_∞ norm, due to the fact that it approximates only dominant controllable states (see Section VIII for numerical results).

The method below is different in the sense that it directly approximates the product of PQ and therefore takes into account the fact that the separately determined most controllable and most observable modes may be different than the modes with the highest "controllability times observability" measure.

E. Original AISIAD algorithm

It is known that the transfer function of the reduced system depends only on the column spans of the projection matrices S_o and S_c (see

[1]For considerations on passivity enforcement read further sections

[6], p. 23), therefore for the approximation of the TBR we need to approximate the dominant eigenspaces of PQ and QP.

The AISIAD algorithm roughly approximates the dominant eigenspaces of the products PQ and QP using a power method, and then constructs projection matrices using these approximations.

The AISIAD algorithm was originally proposed in [10], and we present it here as algorithm 2. It does not use low-rank approximations of gramians at all, however as we show below, it is crucial to use low-rank approximations of P and Q in order to produce accurate reduced models.

Algorithm 2: Original AISIAD algorithm

Input: System matrices (A,B,C), reduced order q, initial orthogonal basis $V \in \mathbb{R}^{n\times q}$

Output: Order-q reduced model $(\hat{A},\hat{B},\hat{C})$.

(1) **repeat**

(2) Approximate $X_i \approx PV_i$ by solving $AX_i + X_iH_i^T + \hat{M}_i = 0$, where $H_i = V_i^TAV_i, \quad \hat{M}_i = BB^TV_i$

(3) Obtain orthogonal basis which spans the same subspace as X_i: $[W_i,S_i] = qr(X_i,0)$

(4) Approximate $Y_i \approx QW_i$ by solving $A^TY_i + Y_iF_i + \hat{N}_i = 0$, where $F_i = W_i^TAW_i, \quad \hat{N}_i = C^TCW_i$

(5) Obtain orthogonal basis for the approximation of QW_i and make it the next approximation of V: $[V_{i+1},R_{i+1}] = qr(Y_i,0)$

(6) **until** convergence

(7) Obtain an oblique projection out of V_{i+1} and W_i:

$$V_L \leftarrow V_{i+1}, \quad W_R \leftarrow W_i$$
$$\begin{bmatrix} U_e & \Sigma & V_e \end{bmatrix} = svd(V_L^TW_R)$$
$$S_o = V_LU_L\Sigma^{-1/2}, \quad S_c = W_RV_L\Sigma^{-1/2}$$

Project the initial system using S_o and S_c:

$$\hat{A} = S_o^TAS_c, \quad \hat{B} = S_o^TB, \quad \hat{C} = CS_c$$

(8) **return** $(\hat{A},\hat{B},\hat{C})$

Consider the steps 3 and 5 of algorithm 2 in more detail. We present derivations for approximation of PV here, the derivations for QW are similar. From Lyapunov equation for P:

$$AP + PA^T + BB^T = 0 \tag{4}$$

Multiplying from the right-hand side by V, we get the following equation:

$$A\underbrace{PV_i}_{X_i} + \underbrace{PV_i}_{X_i}\underbrace{V_i^TAV_i}_{H_i} + \underbrace{P(I - V_iV_i^T)A^TV_i + BB^TV_i}_{M_i} = 0 \tag{5}$$

As we see, in the original AISIAD algorithm the term $P(I - V_iV_i^T)A^TV_i$ is neglected. In this paper we claim that neglecting this term results in a very poor approximation quality, and this term should be instead approximated, using a low-rank approximant for the gramian.

F. Solution of a specialized Sylvester equation

The most important routine in the algorithm 2 is obtaining a solution of the Sylvester equation

$$AX + XH + M = 0, \quad A \in \mathbb{R}^{n\times n}, H \in \mathbb{R}^{q\times q}, M \in \mathbb{R}^{n\times q} \tag{6}$$

where $q \ll n$ (hence the name "specialized").

1) Original solver of Sylvester equation: Consider the following matrix:

$$S = \begin{bmatrix} A & M \\ 0 & -H \end{bmatrix} \tag{7}$$

and assume that we have found the matrices $V_1 \in \mathbb{R}^{n\times q}, Z \in \mathbb{R}^{q\times q}$ and nonsingular $V_2 \in \mathbb{R}^{q\times q}$ such that

$$\begin{bmatrix} A & M \\ 0 & -H \end{bmatrix}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} Z \tag{8}$$

Then one can clearly see that the matrix $V_1V_2^{-1}$ satisfies (6). Moreover, for the purposes of the algorithm 2 it is sufficient to find only V_1, since we make use only of the column span of X.

Evidently, $\Lambda(S) = \Lambda(A) \cup \Lambda(-H)$, and since $-HV_2 = V_2Z$, matrix V_2 is nonsingular if and only if $\Lambda(Z) = \Lambda(-H)$. Note that the matrices A and $(-H)$ should not have common eigenvalues, otherwise (6) does not have a unique solution.

In the original AISIAD the use of Implicitly Restarted Arnoldi (IRA) method [16] is proposed as the means of solving (8). This way, one can obtain a partial Schur decomposition of S with upper-triangular matrix Z in (8). In order to impose the condition $\Lambda(Z) = \Lambda(-H)$ using IRA, authors [16] needed to restrict H to be a Hurwitz matrix. **Therefore, original algorithm is applicable only to the cases where H is Hurwitz (i.e. the initial matrix A satisfies the condition of V^TAV being Hurwitz for all choices of orthogonal basis V). This imposes a severe constraint on matrix A and narrows the class of applicable systems for the whole original method**.

III. Improvement I: Algorithm for specialized Sylvester equation

It is possible to solve equation (6) in the following way.

Let's consider a complex Schur decomposition of matrix $H = USU'$, where S is upper-triangular, and U is unitary. Since matrix H is small $q \times q$, this Schur decomposition is inexpensive. Multiplying (6) from right by U yields:

$$A(XU) + (XU)S + MU = 0, \tag{9}$$

Backsolving for each column of the matrix $(\tilde{X} = XU)$ starting from the first one:

$$(A + s_{jj}I_n)\tilde{x}_j = (MU)_j - \sum_{i=1}^{j-1} s_{ij}\tilde{x}_i, \tag{10}$$

Here $\tilde{x}_j$ denotes j^{th} column of the matrix XU, and s_{ij} denotes (i,j)-th element of the matrix S. The speed of these q solutions of linear system of equations determines the overall speed of the proposed algorithm. We can employ a sparse solver if matrix A is sparse. Alternatively, if fast matrix-vector products can be computed, one can employ an iterative Krylov-subspace solver such as GMRES in order to solve (10).

After the matrix $\tilde{X} = XU$ has been computed, the solution X can be recovered using multiplication by U' from the right.

We summarize our algorithm for solving (6) in algorithm 3.

It is evident that for single-input single-output (SISO) system the proposed approximation is equivalent to rational Krylov method [5], for the interpolation points being $(-\Lambda(H))$ (assuming H being diagonalizable). If one performs projection of the initial system onto dominant eigenspaces of these approximations of P and Q, such obtained reduced model will match the initial model at $s_{1\ldots q} = -\Lambda(V^TAV)$. This important fact unifying Krylov-subspace model reduction and low-rank approximation of gramians was first noted in [17].

Algorithm 3: Solving generalized sylvester equation
Input: Matrices A, H and M
Output: Solution X
(1) Perform a complex Schur decomposition of H: $H = USU'$
(2) $\tilde{M} \leftarrow MU$
(3) **for** j=1 **to** q
(4) Solve for $\tilde{x}_j$: $(A + s_{jj} I_n)\tilde{x}_j = \tilde{M}_j - \sum_{i=1}^{j-1} s_{ij}\tilde{x}_i$
(5) Assign j^{th} column of $\tilde{X}$ being $\tilde{x}_j$.
(6) **return** $X = \tilde{X}U'$

A. Comparison of the two Sylvester solvers

For Hurwitz H both methods are equivalent assuming exact arithmetic. The method described in the section II-F.1 ensures that matrix $\begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$ contains orthogonormal set of vectors. However, this fact does not impose any constraint on the conditioning of V_1. Matrix V_1 can have a very small condition number, whereas columns of V_1 may be almost linearly dependent. On contrary, the proposed method employs only orthogonal transformations, therefore is **more numerically favourable**.

In addition, **the proposed method eliminates the above mentioned important applicability constraint**. It can be applied to any solvable Sylvester equation, broadening the set of applicable problems.

IV. Improvement II: Employing low-rank gramian approximations

As another important modification, we do not discard terms $P(I - V_i V_i^T)A^T V_i$ in (5). We can use any well-developed method to obtain low-rank approximations of P and Q, such as Low-rank ADI [6] or LR-Smith algorithms [18], or projection-based methods [14]. In our code we use simple projection-based algorithm outlined in algorithm 4 for the example of controllability gramian approximation.

Algorithm 4: Low-rank approximation of gramians
Input: Matrices A and B, desired order of approximation k
Output: Low-rank approximation of $P \approx V_p S_p V_p^T$
(1) Compute orthogonal basis for the Krylov subspace as an initial guess: $\text{colspan}(V_0) = \mathcal{K}_k(A^{-1}, B)$
(2) **repeat**
(3) Approximate $X_i \approx PV_i$ by solving $AX_i + X_i H_i^T + \hat{M}_i = 0$, where $H_i = V_i^T A V_i, \quad \hat{M}_i = BB^T V_i$
(4) Perform SVD of X: $[V_{i+1}, S_{i+1}, G_{i+1}] = svd(X, 0)$
(5) **until** convergence
(6) **return** V_{i+1}, S_{i+1}

V. The modified AISIAD algorithm

Combining two improvements outlined above, we now describe the modified AISIAD method which we propose as a replacement for the algorithm 2. We outline it as the algorithm 5

We would like to note that if full exact gramians are known, the modified AISIAD algorithm becomes the power method for the matrices PQ and QP and therefore is guaranteed to converge to the exact TBR solution.

Algorithm 5: Proposed algorithm for approximation of TBR.
Input: System matrices (A, B, C), desired reduced order q, initial projection matrix $V \in \mathbb{R}^{n \times q}$
Output: Order-q reduced model $(\hat{A}, \hat{B}, \hat{C})$.
(1) Get a low-rank approximations of gramians $\tilde{P} \approx P$ and $\tilde{Q} \approx Q$ using any applicable algorithm.
(2) **repeat**
(3) Solve using algorithm 3 $AX_i + X_i H_i^T + \hat{M}_i = 0$, where $H_i = V_i^T A V_i$ $\hat{M}_i = BB^T V_i + \tilde{P}(I - V_i V_i^T)A^T V_i$
(4) $[W_i, S_i] = qr(X_i, 0)$
(5) Solve using algorithm 3 $A^T Y_i + Y_i F_i + \hat{N}_i = 0$, where $F_i = W_i^T A W_i$ $\hat{N}_i = C^T C W_i + \tilde{Q}(I - W_i W_i^T)AW_i$
(6) $[V_{i+1}, R_{i+1}] = qr(Y_i, 0)$
(7) **until** convergence
(8) Obtain an oblique projection out of V_{i+1} and W_i:

$$V_L \leftarrow V_{i+1}, \quad W_R \leftarrow W_i$$
$$\begin{bmatrix} U_e & \Sigma & V_e \end{bmatrix} = svd(V_L^T W_R)$$
$$S_o = V_L U_L \Sigma^{-1/2}, \quad S_c = W_R V_L \Sigma^{-1/2}$$

Project the initial system using S_o and S_c:

$$\hat{A} = S_o^T A S_c, \quad \hat{B} = S_o^T B, \quad \hat{C} = C S_c$$

(9) **return** $(\hat{A}, \hat{B}, \hat{C})$

VI. Modified AISIAD algorithm for descriptor systems

We have generalized the modified AISIAD algorithm 5 for the systems in the descriptor form with nonsingular matrix E.

The TBR algorithm 1 has been generalized for the case of descriptor systems in [19]. For the case of nonsingular E it's enough to compute gramians from (2) and obtain the projection matrices S_o and S_c as the dominant eigenspaces of $QEPE^T$ and $PE^T QE$ respectively. The reduced system is $(S_o^T E S_c, S_o^T A S_c, S_o^T B, C S_c)$.

In the modified AISIAD algorithm for descriptor systems, we use the approximated power iterations in order to obtain the dominant eigenspaces of $QEPE^T$ and $PE^T QE$ by approximating $PE^T V$ and QEW. For the approximation of the first product, multiply the generalized Lyapunov equation for P from right by V:

$$\underbrace{A P E^T V}_{X} + \underbrace{E P E^T V}_{X} \underbrace{V^T E A^T V}_{H} + \underbrace{E P (I - E^T V V^T) A^T V + B B^T V}_{M} = 0 \quad (11)$$

As before, we can compute a low-rank approximation for the gramian $\hat{P} \approx P$ using methods, for example, described in [8], and therefore obtain approximation of the term $\hat{M} \approx M$.

The equation (11) leads to the following matrix equation:

$$AX + EX\hat{H} + \hat{M} = 0, \quad (12)$$

We can solve (12) analogously to solving (6) by performing a Schur decomposition of $H = USU'$ and then solving for the columns of matrix XU. In this case instead of (10) we will have to solve the

following system of equations:

$$(A+s_{jj}E)\tilde{x}_j = (MU)_j - \sum_{i=1}^{j-1} s_{ij}\tilde{x}_i \tag{13}$$

Again, this system can be solved fast if matrices A and E are sparse, or if the fast solver is available.

The calculations for approximation of QEW are analogous.

We outlined the resulting algorithm as algorithm 6.

Algorithm 6: Modified AISIAD algorithm for descriptor systems with nonsingular E.

Input: System matrices (E,A,B,C), desired reduced order q, initial projection matrix $V \in \mathbb{R}^{n\times q}$

Output: Order-q reduced model $(\hat{E},\hat{A},\hat{B},\hat{C})$.

(1) Get a low-rank approximations of proper gramians $\tilde{P} \approx P$ and $\tilde{Q} \approx Q$

(2) **repeat**

(3) Solve $AX_i + EX_iH_i + \hat{M}_i = 0$, where $H_i = V_i^T A^T V_i$ $\hat{M}_i = BB^T V_i + E\tilde{P}(I - E^T V_i V_i^T)A^T V_i$

(4) $[W_i, S_i] = qr(X_i, 0)$

(5) Solve $A^T Y_i + E^T Y_i F_i + \hat{N}_i = 0$, where $F_i = W_i^T A W_i$, $\hat{N}_i = C^T C W_i + E^T \tilde{Q}(I - EW_iW_i^T)AW_i$

(6) $[V_{i+1}, R_{i+1}] = qr(Y_i, 0)$

(7) **until** convergence

(8) Obtain an oblique projection out of V_{i+1} and W_i:

$$V_L \leftarrow V_{i+1}, \quad W_R \leftarrow W_i$$
$$\begin{bmatrix} U_e & \Sigma & V_e \end{bmatrix} = svd(V_L^T E W_R)$$
$$S_o = V_L U_L \Sigma^{-1/2}, \quad S_c = W_R V_L \Sigma^{-1/2}$$

Project the initial system using S_o and S_c:

$$\hat{E} = S_o' E S_c, \hat{A} = S_o^T A S_c, \quad \hat{B} = S_o^T B, \quad \hat{C} = C S_c$$

(9) **return** $(\hat{E},\hat{A},\hat{B},\hat{C})$

VII. Advantages and limitations of the proposed algorithm

The proposed algorithm is applicable to any stable linear system in a state-space form. We have extended it for descriptor systems with nonsingular descriptor matrix E.

Advantages of the proposed method is its extended applicability to a broad range of systems, its improved accuracy and low cost.

The major factor, which governs the accuracy of the proposed method is the accuracy of low-rank approximations of P and Q.

In addition, there is no benefit of applying AISIAD to the symmetric systems ($A = A^T, B = C^T$), since for such systems $P = Q$, and AISIAD cannot do better than dominant gramian eigenspace method (DGE).

A. Complexity of the modified AISIAD algorithm

The computational cost of the modified AISIAD algorithm is directly proportional to the cost of solving q linear systems of equations in (13). If we assume that the matrices A and E are sparse enough to compute the solution in order-n time, this will correspond to linear complexity of the whole algorithm with respect to scaling by n! Our numerical experiments on the RLC circuit example (described in the next section) fully support this statement: for RLC example the time taken to reduce the system scales linearly with n. The largest model we tried so far had the order $n = 500{,}000$.

One can employ iterative solvers for the solution of (13) if the matrices are dense.

If the sparse solver is employed, the cost of the algorithm with available low-rank approximations to P and Q is approximately

$$2N_{it}q(C_{factor} + C_{bksolve}),$$

where N_{it} is a number of modified AISIAD iterations, C_{factor} is a cost of a matrix factorization of $A + s_{jj}E$, and $C_{bksolve}$ is a cost of backward-solving for the solution.

An interesting feature of the proposed algorithm is that it uses one backward solve per one matrix factorization, therefore for each iteration $2q$ matrix factorizations and $2q$ backward solves need to be performed. The linear systems in (13) are essentially the same as in the multiple-point Padé approximation via Krylov-subspaces [9]. However, modified AISIAD algorithm uses q different shift parameters, whereas PVL method generally uses less than q, therefore for PVL the number of backward solves per one matrix factorization is usually more than one. The Arnoldi algorithm requires only one matrix factorization and q backward solves. Therefore, both PVL and Arnoldi are faster than the modified AISIAD algorithm by a constant factor.

B. Passivity preservation

The modified AISIAD method does not impose any assumptions on the physical nature of the input and output signals. In other words, this method is *generic*. However, it is very important for many model reduction problems to preserve some properties of the transfer function, like positive-realness (in case where input signals are port voltages and output signals are port currents) or bounded-reallness (in case of S-parameter modelling).

So far, the only method which is applicable for large-scale model reduction and which preserves passivity[2] is the PRIMA algorithm [15]. This method is based on Krylov-subspace projections, which can be viewed as approximating dominant controllable states [17]. As it was mentioned before, this can sometimes lead to large errors in the frequency domain, which do not necessarily decrease with increasing of the reduced order. This is fully consistent with the experimental results which we present in the next section. The same can be said about variants of PRIMA, which uses dominant eigenspaces of P and Q for the projection bases.

As a practical (and widely used) solution, we can obtain a passive model by post-processing. Since modified AISIAD produces a very accurate models in the frequency domain, we can, for example, use the poles of the reduced model, and re-fit the reduced transfer function using any convex optimization algorithms which ensure passivity [20], [21], [22]. We have tested this approach on the RLC line example and present our results in the next section.

VIII. Computational results

For the test cases we used four benchmark systems, which we describe below. For each of these systems we compared the original AISIAD, modified AISIAD, dominant gramian eigenspaces (DGE), low-rank square root (LRSQRT), Arnoldi [5], [15] and Padé via Lanczos (PVL) [9] reduction algorithms. As an error metric, we used the H_∞ norm of the difference between sufficiently accurate reduced model [3] (in the examples it was the TBR model of order 100-150) and all above mentioned approximations. Note that our error metric is essentially the maximum of the difference between the original

[2] with assumption A being positive semidefinite and $B = C^T$

[3] Using non-reduced model for computing H_∞ norm is very expensive

and reduced system's transfer functions over the entire $j\omega$ axis. We assumed an error to be infinity if the reduced model was unstable (these cases correspond to discontinuities of the lines on our error plots).

Our results showed that the modified AISIAD always outperforms all of the above mentioned methods, with the exception of LRSQRT. For example of the rail cooling and some RLC circuits, modified AISIAD performed much better than LRSQRT. However, for other cases it showed almost identical performance. For several RLC examples modified AISIAD turned out to be slightly inferior to LRSQRT method.

A. RLC transmission line

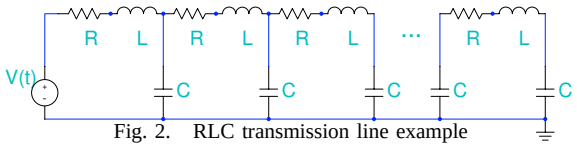

Fig. 2. RLC transmission line example

The first system is an RLC transmission line depicted on figure 2, with varying values for R, L and C. Input signal $u(t)$ is the voltage at the first node. The outputs are the voltage at the last node and current flowing through the first inductor. The state vector consists of node voltages and inductor currents, and nodal analysis equations result in a system (A, B, C) with non-symmetric, indefinite matrix A. We varied the size of this system from several hundreds to hundreds of thousands, for different values of R, L and C and different choices of output signals. The maximal order of the system was 500,000.

Our results showed that modified AISIAD method always produces more accurate results than any above mentioned reduction methods in the H_∞ error metric.

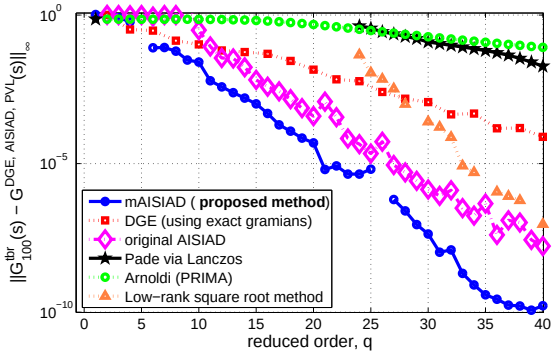

Fig. 3. Reduction errors for the RLC transmission line of order $n = 1000$, $R = 1, L = 2, C = 3$

On the figure 3 the H_∞ errors of the reductions for this RLC line are plotted versus the reduced order q. The initial order of the system was 1000. As the figure suggests, the errors for the DGE method (as well as all other methods!) is much bigger than the errors for the modified AISIAD algorithm. We'd like to stress that here we used exact low-rank approximant for DGE method, whereas for the modified AISIAD we used approximated gramians (the ones provided by algorithm 4). This way, the curve for DGE is a universal upper bound for all family of methods that approximate P and Q separately. Evidently, AISIAD is the best method for this case, significantly outperforming the original AISIAD method.

1) RLC line - MNA formulation: We have used modified nodal analysis (MNA) formulation for the transmission line depicted on figure 4. The inputs were the voltage sources either at a single end or both ends, and the outputs were either currents through the end resistors or (in the case of a single input) voltage at the other end of a line.

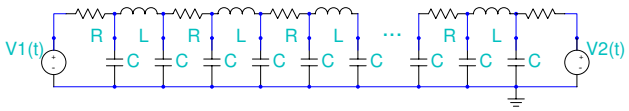

Fig. 4. RLC transmission line model as a motivating example.

The MNA formulation for this line results in a dynamical system in the form (E, A, B, C). We have observed that the modified AISIAD method always works better than PVL, DGE and Arnoldi (which is the PRIMA algorithm [15])[4]. However, low-rank square root method sometimes gives comparable results as modified AISIAD, and for two-port impedance model in some cases it even produces inferior results with respect to LRSQRT. However, the two-port impedance model is almost irreducible, it's Hankel singular values are quite high.

2) Passive post-processing: We have used the RLC transmission line model with input being the voltage at the first node and the output being current through the first resistor of the line (cf. Figure 1). The passivity constraint implies the transfer function being *positive-real*, that is, in addition to being stable, it satisfies the following condition:

$$\mathfrak{Re}(H(j\omega)) > 0, \quad \forall \omega \tag{14}$$

The parameters of RLC line were $N = 1000$, $R = 0.1, L = 2, C = 15$. For this model, the modified AISIAD reduced model of order $q = 30$ is not passive, with the H_∞ norm of error being 0.70%. We have used this model for the passive fitting algorithm from [22] and obtained a positive-real reduced model of order $q = 20$, with an H_∞ error 0.96%. The PRIMA algorithm for this order has a tremendously higher H_∞ error, which is 88.2%. Figure 5 shows the real parts of the above mentioned transfer functions.

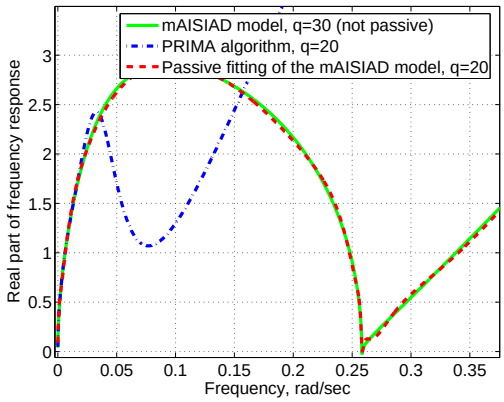

Fig. 5. Real parts of the transfer functions for the modified AISIAD reduced model (which has been used for the post-processing, solid line), PRIMA reduced model (dash-dotted line) and the model obtained after post-processing of mAISIAD model (dashed line). One can note that PRIMA algorithm poorly approximates the original transfer function away from the expansion point (which is at zero frequency). The non-reduced transfer function is almost indistinguishable from the mAISIAD model.

[4]PRIMA algorithm has it's own advantages though, because it preserves passivity of a reduced model. However, we are concerned here only with H_∞ norm as an error measure.

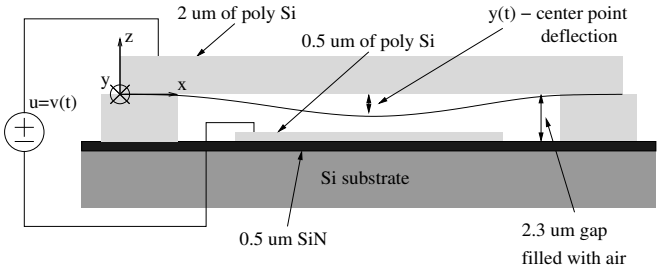

Fig. 6. Micromachined switch example

3) RC line: In order to test the modified AISIAD algorithm on a symmetric system ($A = A^T, B = C^T$), we used a simple RC line (figure 2 with $L = 0$) with input being the voltage at the first node and output being the current through the first resistor. The state vector was the vector of node voltages. For this system $P = Q$ and dominant eigenspaces of PQ and QP will be the same as the ones of P and Q separately, therefore modified AISIAD should achieve exactly the same accuracy as DGE method. Our numerical experiments fully support this statement: errors for DGE and modified AISIAD are the same for this test case.

B. Linearization of micromachined switch

The second example was the linearization of the micromachined switch (fixed-fixed beam) shown on figure 6. Following Hung et al. [2], the dynamical behavior of this coupled electro-mechanical-fluid system can be modeled with 1D Euler's beam equation and 2D Reynolds' squeeze film damping equation [2]. The linearization of this model around equilibrium leads to the nonsymmetric system (A,B,C) with indefinite system matrix A. The order $n = 880$.

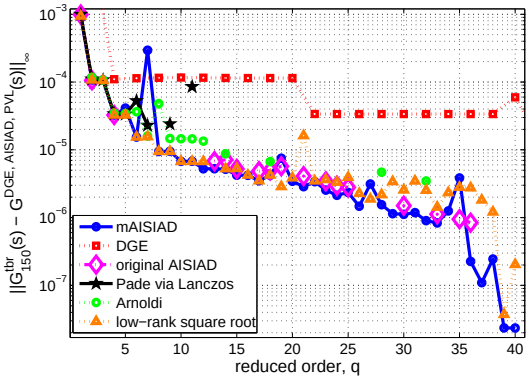

Fig. 7. Errors for the MEMS linearization, $N = 880$.

On the figure 7 the errors for the MEMS test case are presented. Here still modified AISIAD method performs better than any other method, but the difference is not as dramatic as for other examples. The LRSQRT method showed the results similar to the modified AISIAD.

C. Cooling profile of steel rail

This test case was obtained from the Oberwolfach Model Reduction Benchmark Collection web site [23]. The reader is referred to the description of *Heat transfer problem for cooling of steel profiles* benchmark on the mentioned web site for descriptions. This is the model in a descriptor form (E,A,B,C) with $n = 1357$, having 7 inputs and 6 outputs.

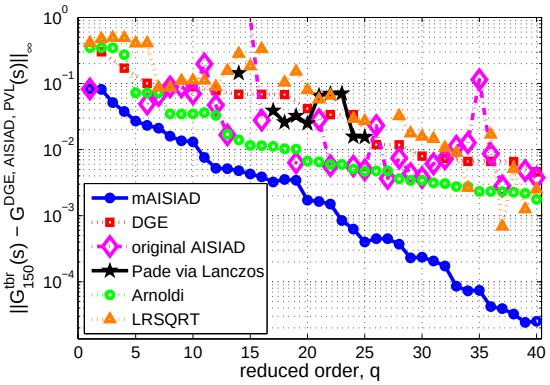

Fig. 8. Reduction errors for the rail example, $n = 1357$, 7 inputs, 6 outputs

For this example the modified AISIAD showed superior performance with respect to any other approximations, including LRSQRT method.

On the figure 8 we present the error plot for this example. Here, again, AISIAD greatly outperforms any other approximations to TBR, as well as Krylov-subspace based reductions. The reduced models of order $q = 2,3,4$ are unstable, but it's even smaller than the number of inputs. As expected, modified AISIAD outperforms original AISIAD algorithm.

D. Optical filter

This test case was obtained from the Oberwolfach Model Reduction Benchmark Collection web site [23]. The reader is referred to the description of *Tunable Optical Filter* benchmark on the mentioned web site for descriptions. This is the model in a descriptor form (E,A,B,C) with $n = 1668$, having 1 input and 5 outputs. The corresponding errors are presented on figure 9. Here the dominant gramian eigenspace projection was computed using the same approximate gramians which were used for the modified AISIAD method. The LRSQRT method showed very similar errors as the modified AISIAD.

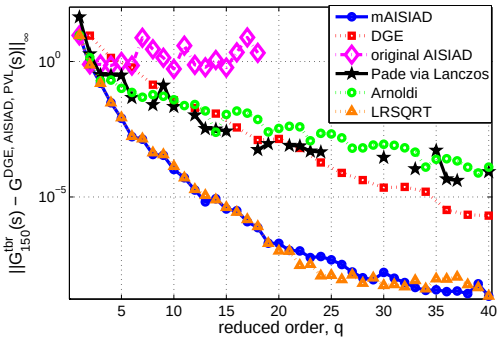

Fig. 9. Reduction errors for the optical filter example, $n = 1668$, 1 inputs, 5 outputs

IX. Conclusions

As the field of computationally efficient techniques for model reduction has matured, the range of applications has continued to expand, and in this expanded set are cases where existing reduction methods have proved to be unreliable, at least for wide-band applications. In this paper we demonstrated that there are examples where

separately examining controllability and observability does not yeild reasonably-sized reduced models. Since almost all the existing methods for reducing large scale problems effectively perform this separate examination, new methods are needed. We described such a reduction method, based on several modifications of the recently developed AISIAD algorithm. We developed a more general strategy for solving the AISIAD algorithm's Sylvester equation, we extended the method to descriptor systems (i.e. circuits with floating capacitors or coupled inductors), and we dramatically improved AISAID's accuracy by employing low-rank approximations of separate gramians. Finally, we demonstrated that our modified AISAID method is reliable using a wide variety of examples. In these examples, the modified AISAID method was either nearly equivalent or far superior to any of the other methods suitable for large scale problems. We have demonstrated that the passive post-processing of the modified AISIAD models gives significantly more accurate results than PRIMA algorithm.

REFERENCES

[1] M. Rewieński and J. White. A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 22(2):155–170, 2003.

[2] E. Hung, Y.Yang, and S. Senturia. Low-order models for fast dynamical simulation of mems microstructures. In *Proc. of IEEE Int. Conf. on Solid State Sensors and Actuators*, 1997.

[3] K.Willcox and J.Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA Journal*, 40(11):2323–2330, November 2002.

[4] A. Megretski. Optimal model order reduction for maximal real part norms. In A. Rantzer and C. I. Byrnes, editors, *New Directions in Mathematical System Theory and Optimization*, volume 286, 2003.

[5] E.J. Grimme. *Krylov projection methods for model reduction.* PhD thesis, University of Illinois at Urbana Champaign, 1997.

[6] Jing-Rebecca Li. *Model reduction of large linear systems via low rank system gramians.* PhD thesis, MIT, 2000.

[7] Joel Phillips and L. Miguel Silveira. Poor man's tbr: A simple model reduction scheme. In *Proc. of Conf. on Design, Autom. and test in Europe*, page 20938. IEEE Computer Society, 2004.

[8] Stykel Tatjana. Model reduction of descriptor systems via low rank generalized gramians. In *Proc. of BIRS Workshop on Model Reduction Problems, Banff, Canada*, 2004.

[9] P.Feldman and R. Freund. Efficient linear circuit analysis by padé approximation via lanczos process. *IEEE Transactions on Computer-Aided Design*, 14(5):639 – 649, 1995.

[10] Yunkai Zhou. *Numerical Methods for Large Scale Matrix Equations with Applications in LTI System Model Reduction*. PhD thesis, Rice University, 2002.

[11] Kemin Zhou, John C. Doyle, and Keith Glover. *Robust and optimal control*. Prentice-Hall, Inc., 1996.

[12] Moore B.C. Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Autom. Control*, AC-26(1):17–32, 1981.

[13] Keith Glover. All optimal hankel-norm approximations of linear multivariable systems and their l^∞ -error bounds. *International Journal of Control*, 39(6):1115–1193, 1984.

[14] D.C.Sorensen and A.C. Antoulas. Projection methods for balanced model reduction. Technical report, Rice University, 2001.

[15] A. Odabasioglu, M. Celik, and L. T. Pileggi. Prima: passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. Computer-Aided Design*, 17(8):645–654, 1997.

[16] D. C. Sorensen. Implicit application of polynomial filters in a k-step arnoldi method. *SIAM J. Matrix Anal. Appl.*, 13(1):357–385, 1992.

[17] K. Gallivan, A. Vandendorpe, and P. Van Dooren. Sylvester equations and projection-based model reduction. *J. Comput. Appl. Math.*, 162(1):213–229, 2004.

[18] A.C.Antoulas, D.C.Sorensen, and S.Gugercin. A modibed low-rank smith method for large-scale lyapunov equations. Technical report, Rice University, 2001.

[19] Tatjana Stykel. *Analysis and numerical solution of generalized Lyapunov equations*. PhD thesis, Technischen universität Berlin, 2002.

[20] C.P. Coelho, Phillips J., and Silveira L.M. A convex programmin approach for generating guaranteed passive approximations to tabulated frequency-data. *IEEE Trans. CAD*, 23(2):293–301, 2004.

[21] B. Gustavsen and A. Semlyen. Enforcing passivity for admittance matrices approximated by rational functions. *IEEE Trans. power systems*, 16(1):97–104, 2001.

[22] K.C. Sou, A. Megretski, and Daniel L. A quasi-convex optimization approach to parametrized model reduction. In *Proceedings of the 42th Design Automation Conferece, Anaheim CA*, pages 933–938, 2005.

[23] Oberwolfach model reduction benchmark collection, http://www.imtek.uni-freiburg.de/simulation/benchmark/.

[24] J.R. Li and J.K. White. Low rank solution of lyapunov equations. *SIAM Review, SIGEST*, 46(4):693–713, 2004.