Cross-Gramian-Based Dominant Subspaces

Peter Benner*   Christian Himpe†

Abstract

A standard approach for model reduction of linear input-output systems is balanced truncation, which is based on the controllability and observability properties of the underlying system. The related dominant subspace projection model reduction method similarly utilizes these system properties, yet instead of balancing, the associated subspaces are directly conjoined. In this work we extend the dominant subspace approach by computation via the cross Gramian for linear systems, and describe an a-priori error indicator for this method. Furthermore, efficient computation is discussed alongside numerical examples illustrating these findings.

Keywords: Controllability, Observability, Cross Gramian, Model Reduction, Dominant Subspace, HAPOD

MSC: 93A15, 93B11, 93B20

1 Introduction

Input-output systems map an input function to an output function via a dynamical system. The input excites or perturbs the state of the dynamical system and the output is some transformation of the state. Typically, these input and output functions are low-dimensional while the intermediate dynamical system is high(er)-dimensional. In applications from natural sciences and engineering, the dimensionality of the dynamical system may render the numerical computation of outputs from inputs excessively expensive or at least demanding.

Model reduction addresses this computational challenge by algorithms that provide surrogate systems, which approximate the input-output mapping of the original system with a low(er)-dimensional intermediate dynamical system. Practically, the trajectory of the dynamical system's state is constrained to a subspace of the original system's state-space, for example by using truncated projections. A standard approach for projection-based model reduction of input-output systems is balanced truncation [25], which transforms the state-space unitarily to a representation that is sorted (balanced) in terms of the input's

*Computational Methods in Systems and Control Theory, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany ORCID: 0000-0003-3362-4103, benner@mpi-magdeburg.mpg.de

†Computational Methods in Systems and Control Theory, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany ORCID: 0000-0003-2194-6754, himpe@mpi-magdeburg.mpg.de

effect on the state (controllability) as well as the state's effect on the output (observability) and discards (truncates) the least important states according to this measure.

Instead of balancing, this work investigates a dominant subspaces approach [28], that conjoins instead of balances, the most controllable and most observable subspaces into a projection. This unbalanced model reduction method may yield less accurate reduced-order systems, yet allows a computationally advantageous formulation while also preserving stability and providing an error quantification. The method proposed in this work, combines the dominant subspace projection model reduction (DSPMR) method [28] with the cross Gramian (matrix) [12], which encodes controllability and observability information of an underlying input-output system. For this cross-Gramian-based dominant subspace method, an a-priori error indicator is developed, and the numerical issues arising in the wake of large-scale systems are addressed, specifically by utilizing the hierarchical approximate proper orthogonal decomposition (HAPOD) [17]. Compared to other cross Gramian and SVD model reduction techniques such as [20], the proposed method does not need multiple decompositions, but a single HAPOD.

The considered class of input-output systems are generalized linear (time-invariant) systems[1], mapping input $u : \mathbb{R} \to \mathbb{R}^M$ via the state $x : \mathbb{R} \to \mathbb{R}^N$ — a solution to an ordinary differential equation — to the output $y : \mathbb{R} \to \mathbb{R}^Q$:

$$
\begin{aligned}
E\dot{x}(t) &= Ax(t) + Bu(t), \\
y(t) &= Cx(t),
\end{aligned}
\tag{1}
$$

with a system matrix $A \in \mathbb{R}^{N \times N}$, an input matrix $B \in \mathbb{R}^{N \times M}$, an output matrix $C \in \mathbb{R}^{Q \times N}$ and a mass matrix $E \in \mathbb{R}^{N \times N}$. In the scope of this work, we assume E to be non-singular as well as the matrix pencil (A, E) to be asymptotically stable, meaning the eigenvalues of the associated generalized eigenproblem lie in the open left half-plane. This type of system arises, for example, in spatial discretizations of partial differential equations using the finite element method. In Section 2 the cross Gramian for generalized linear systems is introduced, followed by Section 3, briefly describing projection-based model reduction, and extending the dominant subspace projection method to the cross Gramian together with an error indicator. The proposed model reduction technique is then tested numerically in Section 4 and a summary is given in Section 5.

2   Generalized Cross Gramian

In this section, the cross Gramian matrix, introduced in [12], is briefly reviewed from the point of view of generalized linear time-invariant (LTI) systems (1). Fundamental to system-theoretic model reduction are the controllability and observability operators [1], which are given for (1) by the generalized controllability

---

[1]Sometimes, the term descriptor system is used for this type of system, yet typically descriptor systems explicitly allow a singular mass matrix. Hence, we decided to use the term *generalized linear system*.

operator $\mathcal{C} : L_2 \to \mathbb{R}^N$ and the generalized observability operator $\mathcal{O} : \mathbb{R}^N \to L_2$:

$$\mathcal{C}(u) := \int_0^\infty e^{E^{-1}At} E^{-1} B u(t) dt,$$

$$\mathcal{O}(x_0) := C e^{E^{-1}At} E^{-1} x_0.$$

The (generalized) cross Gramian[2] is then defined as a composition of the generalized controllability and observability operators:

$$W_X := \mathcal{C} \circ \mathcal{O} = \int_0^\infty e^{E^{-1}At} E^{-1} BC e^{E^{-1}At} E^{-1} dt \in \mathbb{R}^{N \times N}, \qquad (2)$$

and jointly quantifies controllability and observability of square systems – systems with the same number of inputs and outputs $M = Q$. As for linear, square systems with $E = I$, the generalized cross Gramian is a solution to a matrix equation; in this case a Sylvester-type equation:

$$A W_X E + E W_X A = -BC,$$

which can be shown using integration-by-parts of (2):

$$
\begin{aligned}
W_X &= \int_0^\infty e^{E^{-1}At} E^{-1} BC e^{E^{-1}At} E^{-1} dt \\
&= (E^{-1}A)^{-1} e^{E^{-1}At} E^{-1} BC e^{E^{-1}At} E^{-1} \Big|_0^\infty \\
&\quad - (E^{-1}A)^{-1} \int_0^\infty e^{E^{-1}At} E^{-1} BC e^{E^{-1}At} (E^{-1}A) E^{-1} dt \\
\Rightarrow A W_X &= E e^{E^{-1}At} E^{-1} BC e^{E^{-1}At} E^{-1} \Big|_0^\infty - E W_X A E^{-1} \\
\Rightarrow A W_X E + E W_X A &= E e^{E^{-1}At} E^{-1} BC e^{E^{-1}At} \Big|_0^\infty = -BC.
\end{aligned}
$$

Besides the cross Gramian, the (generalized) controllability Gramian $W_C := \mathcal{C}\mathcal{C}^*$ and (generalized) observability Gramian $W_O := \mathcal{O}^*\mathcal{O}$ are defined accordingly [31, 35]. For systems with a symmetric Hankel operator $H := \mathcal{O}\mathcal{C}$, $H = H^*$, the (generalized) cross Gramian has the property:

$$W_X W_X = \mathcal{C}\mathcal{O}\mathcal{C}\mathcal{O} = \mathcal{C}(\mathcal{O}\mathcal{C})^*\mathcal{O} = \mathcal{C}\mathcal{C}^*\mathcal{O}^*\mathcal{O} = W_C W_O.$$

Hence for symmetric systems, either, W_X or $\{W_C, W_O\}$ can be used interchangeably, if controllability and observability are to be concurrently evaluated. For non-symmetric and especially non-square systems, an approximation to the cross Gramian is defined, based on the column-wise partitioning of the input matrix B and row-wise partitioning of the output matrix C:

$$B = \begin{pmatrix} b_1 & \dots & b_M \end{pmatrix}, \quad C = \begin{pmatrix} c_1 & \dots & c_Q \end{pmatrix}^\mathsf{T}.$$

For $\bar{B} := \sum_{m=1}^M b_m$ and $\bar{C} := \sum_{q=1}^Q c_q^\mathsf{T}$, the non-symmetric cross Gramian [19] for (1) is defined as:

$$W_Z := \int_0^\infty e^{E^{-1}At} E^{-1} \bar{B}\bar{C} e^{E^{-1}At} E^{-1} dt, \qquad (3)$$

which is the cross Gramian of the average system $(A, \bar{B}, \bar{C})$.

---

[2]Note that the term *generalized cross Gramian* is used in [32] for cross Gramians of unstable systems.

The (non-symmetric) generalized cross Gramian (2) can be computed numerically, for example, using the Hessenberg-Schur algorithm [13], the alternating direction implicit (ADI) algorithm [7, 5, 6], or as an empirical cross Gramian [16].

3 Model Reduction

One of the main numerical applications of the cross Gramian is model (order) reduction, which aims to determine lower order surrogate systems for (1), with respect to the state-space dimension $N = \dim(x(t))$. The reduced order model (ROM) with $x_r : \mathbb{R} \to \mathbb{R}^n$, $n \ll N$,

$$E_r \dot{x}_r(t) = A_r x_r(t) + B_r u(t),$$
$$\tilde{y}(t) = C_r x_r(t),$$

has a reduced system matrix $A_r \in \mathbb{R}^{n \times n}$, a reduced input matrix $B_r \in \mathbb{R}^{n \times M}$, a reduced output matrix $C_r \in \mathbb{R}^{Q \times n}$ and a reduced mass matrix $E_r \in \mathbb{R}^{n \times n}$, such that the reduced system's output $\tilde{y} : \mathbb{R} \to \mathbb{R}^Q$ approximates the full order model's output:

$$\|y - \tilde{y}\| \ll 1,$$

in a suitable norm.

Following, the projection-based dominant subspace model reduction method is extended to exploit the cross Gramian for computation.

3.1 Projection-Based Model Reduction

A commonplace approach to construct reduced order models is mapping the state-space trajectory $x(t)$ to a lower dimensional subspace, using a reduction operator $V_1 : \mathbb{R}^N \to \mathbb{R}^n$ and a lifting operator $U_1 : \mathbb{R}^n \to \mathbb{R}^N$:

$$x_r(t) := V_1 x(t) \quad \to \quad x(t) \approx U_1 x_r(t).$$

In the case of (generalized) linear systems (1), the operators $U_1 \in \mathbb{R}^{N \times n}$ and $V_1 \in \mathbb{R}^{n \times N}$, can be directly applied to the system components A, B, C and E to obtain the reduced quantities:

$$A_r := V_1 A U_1, \quad B_r := V_1 B, \quad C_r := C U_1, \quad E_r := V_1 E U_1. \tag{4}$$

Hence, the aim is the computation of suitable reducing and lifting operators U_1, V_1, which are typically assumed to be bi-orthogonal $V_1 U_1 = I_n$. The dominant subspaces method considered in this work is additionally orthogonal $V_1 := U_1^\mathsf{T}$, thus, the reduction process is a Galerkin projection, which is stability preserving, if the symmetric part of the system matrix A is negative definite, and the mass matrix E positive definite [8],

$$A + A^\mathsf{T} < 0 \quad \wedge \quad E > 0. \tag{5}$$

This is a generalization of the stability preservation for systems with $E = I$, mentioned in [28]. If a system does not fulfill (5), a stabilization procedure, see for example [4, Sec. 4], can be applied to the reduced order model.

3.2 Dominant Subspaces

The *Dominant Subspace Projection Model Reduction* (DSPMR) is introduced in [28]. The idea behind DSPMR is, instead of balancing controllability and observability Gramians, to combine the associated principal subspaces obtained from approximate system Gramians. This yields a simple model reduction algorithm which is based upon low-rank factors of the controllability and observability Gramians. In [28], a low-rank Cholesky (LR Chol) factor is used, while [21] utilizes singular vectors of a truncated singular value decomposition (tSVD),

$$W_C \overset{\text{LR Chol}}{\approx} Z_C Z_C^{\intercal}, \qquad\qquad W_O \overset{\text{LR Chol}}{\approx} Z_O Z_O^{\intercal},$$
$$W_C \overset{\text{tSVD}}{\approx} U_C D_C U_C^{\intercal}, \qquad\qquad W_O \overset{\text{tSVD}}{\approx} U_O D_O U_O^{\intercal}.$$

The controllability and observability subspaces encoded in the matrix factors are now conjoined and orthogonalized, by either a SVD ([28]) or a (rank-revealing) QR-decomposition ([21, 22]). Either, the left singular vectors U, or the Q factor, can be taken as Galerkin projections, respectively:

$$QR \overset{\text{QR}}{=} [U_C, U_O] \to U_1 := Q,$$
$$UDV^{\intercal} \overset{\text{SVD}}{=} [U_C, U_O] \to U_1 := U,$$

see also [2, Sec 2.1.7].

An extension to the DSPMR method is also proposed in [28], called *Refined* Dominant Subspace Projection Model Reduction. The eponymous refinement is given by weighting factors $\omega_C, \omega_O > 0$ for the controllability and observability subspace bases respectively. The weighting factors are selected as the Frobenius norm of the respective low-rank factors, $\omega_C := \|Z_C\|_{\text{F}}^{-1}$ and $\omega_O := \|Z_O\|_{\text{F}}^{-1}$, yielding:

$$QR \overset{\text{QR}}{=} [\omega_C Z_C, \omega_O Z_O] \to U_1 := Q,$$
$$UDV^{\intercal} \overset{\text{SVD}}{=} [\omega_C Z_C, \omega_O Z_O] \to U_1 := U.$$

Obviously, this is only sensible for the Cholesky factor variant, as the norm of the (orthonormal) singular vectors is one. A similar idea for combining weighted subspaces is also used in the cotangent lift method from [27].

3.3 Cross-Gramian-Based Dominant Subspaces

Instead of the controllability and observability Gramians, also the cross Gramian can be used to obtain a dominant subspace projection. An SVD of the cross Gramian,

$$W_X \overset{\text{SVD}}{=} U_X D_X V_X^{\intercal}, \tag{6}$$

produces left and right singular vectors aggregated in matrices U_X and V_X, which induce subspaces associated to controllability (U_X) and observability (V_X) of the underlying system (A, B, C).

In [34, Sec. 4.3], it is noted, that the sole use of either, U_X or V_X, as a Galerkin projection, will largely omit observability or controllability information respectively. Hence, both subspaces should be incorporated in the reducing and lifting operator. Balanced truncation, for example, determines a suitable Petrov-Galerkin projection[3], where $U_1 \neq V_1$, by simultaneous diagonalization of the controllability and observability Gramians.

For the proposed variant of the dominant subspace method (for an algorithmic description see Section 3.4.1), the left and right singular vectors are conjoined as before, but also scaled *column-wise* by the singular vector's associated singular values:

$$[U_X D_X, V_X D_X] \stackrel{\text{SVD}}{=} U_{CO} D_{CO} V_{CO}^\intercal \to U_1 := U_{CO}.$$

So, instead of normalizing the controllability and observability subspaces (as a whole), as in refined DSPMR, based on the common controllability-observability measure, the singular values of the cross Gramian, the vectors spanning the compound subspace are scaled individually. Here explicitly a (rank-revealing) SVD is used, instead of a QR decomposition, as the singular values D_{CO} will be used for an error indicator. An advantage of the cross-Gramian-based dominant subspace projection method is this common measure of minimality, the singular values $\sigma_i = D_{CO,ii}$ associated jointly to the "controllability" and "observability" subspaces.

3.4 Error Indicator

In this section an error indicator for the cross-Gramian-based dominant subspace method is developed. Previous works, such as [33, 39, 29, 38], already introduced error *bounds* for the Hardy H_2-norm. Here, an H_2-error *indicator* of simple structure using time-domain quantities is proposed, which is loosely related to the simplified balanced gains approach from [11]. The H_2-norm is particularly interesting, since an error estimation has relevance for the frequency-domain and the time-domain [37, Ch. 2], since it also describes the energy (L_2-norm) of the system's impulse response. Before this error indicator is derived, a straightforward property of the matrix exponential is presented.

Lemma 1
Given matrices $A \in \mathbb{R}^{N \times N}$ and $U \in \mathbb{R}^{N \times n}$, $n \leq N$, the following holds:
$$U e^{U A U^\intercal} U^\intercal = U U^\intercal e^{A U U^\intercal} = e^{U U^\intercal A} U U^\intercal.$$

Proof. The proof is a trivial consequence on the associativity of the matrix product.

$$\begin{aligned}
U e^{U^\intercal A U} U^\intercal &= U \left(\sum_{k=0}^{\infty} \frac{1}{k!} (U^\intercal A U)^k \right) U^\intercal \\
&= U (I + (U^\intercal A U) + \frac{1}{2} (U^\intercal A U)(U^\intercal A U) + \dots) U^\intercal \\
&= U U^\intercal (I + A U U^\intercal \frac{1}{2} A U U^\intercal A U U^\intercal + \dots) \\
&= U U^\intercal e^{A U U^\intercal}. \qquad \square
\end{aligned}$$

---

[3]Balanced truncation yields a Galerkin projection for state-space symmetric systems, $A = A^\intercal, B = C^\intercal, E = E^\intercal$ [9].

Next, the error indicator is constructed, which is derived from the L_2-norm of the impulse response error system, for which we assume for ease of exposition $E = I$:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{x}_r(t) \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & A_r \end{pmatrix} \begin{pmatrix} x(t) \\ x_r(t) \end{pmatrix} + \begin{pmatrix} B \\ B_r \end{pmatrix} u(t)$$

$$y_e(t) = \begin{pmatrix} C & -C_r \end{pmatrix} \begin{pmatrix} x(t) \\ x_r(t) \end{pmatrix}.$$

We consider only SISO (Single-Input-Single-Output) systems for this error indicator and impulse inputs $u(t) = \delta(t)$.

First, the H_2-norm of the error system, in impulse response form, is transformed in a manner so that Lemma 1 can be applied. Note, that the error system of a SISO system is also a SISO system with a scalar and thus symmetric impulse response:

$$\begin{aligned} \|y_e\|_{L_2}^2 &= \operatorname{tr} \Big(\int_0^\infty ((C \quad -C_r) \begin{pmatrix} e^{At} & 0 \\ 0 & e^{A_r t} \end{pmatrix} \begin{pmatrix} B \\ B_r \end{pmatrix})^2 dt \Big) \\ &= \operatorname{tr} \Big(\int_0^\infty (Ce^{At}B - C_r e^{A_r t} B_r)^2 dt \Big) \\ &= \operatorname{tr} \Big(C \int_0^\infty e^{At} B C e^{At} - e^{At} B C_r e^{A_r t} U_1^{\mathsf{T}} \\ &\qquad\qquad\qquad - U_1 e^{A_r t} B_r C e^{At} + U_1 e^{A_r t} B_r C_r e^{A_r t} U_1^{\mathsf{T}} dt B \Big), \end{aligned}$$

applying the definition of the reduced quantities (4), and subsequently the result of Lemma 1, gives:

$$\begin{aligned} \|y_e\|_{L_2}^2 &= \operatorname{tr} \Big(C \int_0^\infty e^{At} B C e^{At} - e^{At} B C e^{U_1 U_1^{\mathsf{T}} A t} U_1 U_1^{\mathsf{T}} \\ &\qquad\qquad - U_1 U_1^{\mathsf{T}} e^{A U_1 U_1^{\mathsf{T}} t} B C e^{At} \\ &\qquad\qquad + U_1 U_1^{\mathsf{T}} e^{A U_1 U_1^{\mathsf{T}} t} B C e^{U_1 U_1^{\mathsf{T}} A} U_1 U_1^{\mathsf{T}} dt B \Big). \end{aligned}$$

The next step is approximating the matrix exponentials $e^{A U_1 U_1^{\mathsf{T}} t}$ and $e^{U_1 U_1^{\mathsf{T}} A t}$ by the homogeneous system's solution operator,

$$e^{A U_1 U_1^{\mathsf{T}} t} \approx e^{At}, \quad e^{U_1 U_1^{\mathsf{T}} A t} \approx e^{At},$$

which allows to factor the previous representation to:

$$\|y_e\|_{L_2}^2 \approx \operatorname{tr} \Big(C \int_0^\infty (I - U_1 U_1^{\mathsf{T}})(e^{At} B C e^{At})(I - U_1 U_1^{\mathsf{T}}) dt B \Big).$$

Forming the cross Gramian by moving the projection error terms $(I - U_1 U_1^{\mathsf{T}})$ out of the integral, applying VON NEUMANN's trace inequality [24], and replacing the cross Gramian by its SVD (6) yields:

$$\begin{aligned} \|y_e\|_{L_2}^2 &\approx \operatorname{tr}(C(I - U_1 U_1^{\mathsf{T}}) W_X (I - U_1 U_1^{\mathsf{T}}) B) \\ &\leq \sum_{k=1}^N \sigma_k((I - U_1 U_1^{\mathsf{T}}) W_X (I - U_1 U_1^{\mathsf{T}})) \sigma_k(BC) \\ &= \sum_{k=1}^N \sigma_k((I - U_1 U_1^{\mathsf{T}}) U_X D_X V_X (I - U_1 U_1^{\mathsf{T}})) \sigma_k(BC). \end{aligned}$$

We note that U_1 is the orthogonalized concatenation of U_X and V_X, and BC is of rank one due to the SISO nature of the system, hence,

$$\|y_e\|_{L_2}^2 \lesssim \sigma_{n+1}(W_X)\sigma_1(BC)$$
$$\leq \sigma_1(BC) \sum_{k=n+1}^{N} \sigma_k^2(W_X). \tag{7}$$

The last upper estimate results from the relation of the spectral norm and the Frobenius norm ($\|A\|_2 \leq \|A\|_F$). This derivation yields the following error indicator:

The L_2 impulse response model reduction error for a cross-Gramian-based dominant subspaces reduced order model is approximated by:

$$\|y - \tilde{y}\|_{L_2} \approx \sqrt{\sigma_1(BC) \sum_{k=n+1}^{N} \sigma_k^2(W_X)}.$$

One might assume that the spectral norm representation of the error indicator would be more convenient, yet in Section 4 we will show the advantage of the final Frobenius norm form.

Remark 1
Using the Cauchy-Schwarz inequality as in [14], this impulse response error indicator can be extended to squarely integrable inputs $u \in L_2$.

3.4.1 Algorithmic and Low-Rank Computation

The computation of the proposed cross-Gramian-based dominant subspace projection, as well as the classic dominant subspace projection consists of two phases: First, the computation of the system Gramians, either the cross Gramian, or the controllability and observability Gramians. And second, the assembly of the reducing (and lifting) operator.

For large-scale systems, the computation of dense system Gramians, which are of dimension $N \times N$, may be infeasible or at least inefficient. To this end, low-rank representations of the Gramians can be computed, for the cross Gramian, in example by the implicitly restarted Arnoldi algorithm [34], the factorized iteration [3], a factored ADI [5] or a low-rank empirical cross Gramian [18].

Overall, the *cross-Gramian-based dominant subspace* algorithm is summarized by:

1. Compute (low-rank) cross Gramian.

2.a Compute SVD of the cross Gramian.

2.b Weight left (controllability) singular vectors with singular values.

2.c Weight right (observability) singular vectors with singular values.

2.d Compute POD of conjoined and weighted left and right singular vectors.

The resulting POD modes (left singular vectors) represent the Galerkin projection. In principle, a similar procedure can be conducted using controllability and observability Gramians, yet it is not immediately clear if the SVD of the (weighted) conjoined singular vectors yields in general an equally useful measure.

4   Numerical Results

Following, two numerical examples are presented to illustrate the previous findings. These numerical experiments are conducted using MATLAB 2018a [23]. The system Gramians needed for the dominant subspace methods, the controllability and observability Gramian for (refined) DSPMR, and the cross Gramian for the cross-Gramian-based dominant subspaces, are computed as empirical Gramians [16] using `emgr` – empirical Gramian framework in version 5.5 [15]. All simulated trajectories for the construction of these *empirical dominant subspaces* are computed using the implicit Euler method. Before we present the numerical results, the practical computation of the cross-Gramian-based dominant subspaces is discussed.

4.1   Fused Computation

Even for moderately sized systems, the computation of the (cross) Gramian's singular vectors may be a computationally challenging task[4]. To compute the dominant subspace projections from the cross Gramian, or the controllability and observability Gramians, the hierarchical approximate proper orthogonal decomposition (HAPOD) [17] is used. Particularly, the HAPOD allows the computation of the dominant subspace directly, which means the Gramian SVD(s) and the concatenation SVD in one.

The HAPOD enables a swift computation of left singular vectors of arbitrary partitioned data sets, based on a selected projection error (on the input data) $\varepsilon > 0$ and a tree hierarchy with the data (Gramian) partitions as leafs. The tree hierarchy utilized for the experiments in this work is given by a combination of special topologies discussed in [17], the incremental HAPOD (maximally unbalanced binary tree) and the distributed HAPOD (star). Two incremental HAPODs are performed for the Gramian partitions respectively and subsequently a distributed HAPOD of the resulting singular vectors from both sub-trees yields the dominant subspace projection. Fig. 1 illustrates the overall HAPOD tree.

Since the HAPOD computes only left singular vectors, but the right singular vectors of the cross Gramian are also needed, the HAPOD of the cross Gramian (left singular vectors) and the transposed cross Gramian (right singular vectors) is computed. In the following numerical examples, the (full-order) empirical linear cross Gramian [16, Sec. 3.1.3] is used, as in-memory storage of the Gramian(s) is possible. For settings, where only parts of the cross Gramian can be kept in memory, the low-rank empirical cross Gramian [18] for the left singular vectors, and the low-rank empirical cross Gramian of the **adjoint system** for the right

---

[4]For the presented numerical examples the SVDs of system Gramians comprises the dominant fraction of computation time.
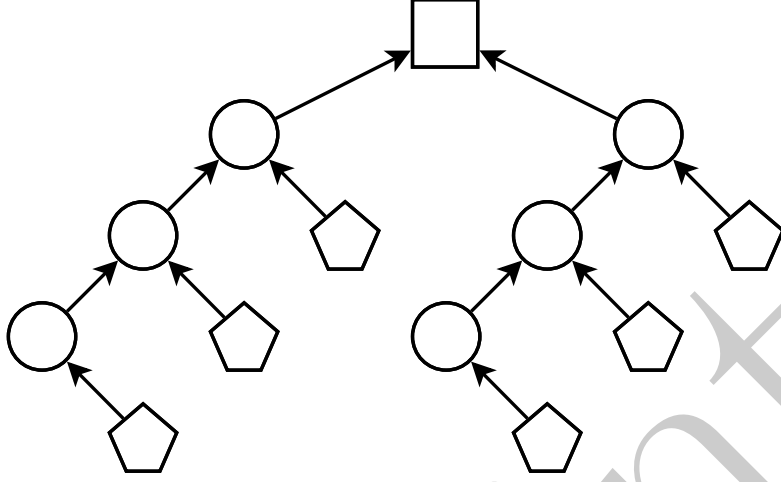
Figure 1: HAPOD tree topology for the cross-Gramian-based dominant subspaces method. Pentagons symbolize partitions of the cross Gramian (left) and the adjoint cross Gramian (right), respectively. Circles mark sub-PODs, while a square represents the root-POD returning the overall (HA)POD described in Section 4.1.

singular vectors, can be utilized, since the cross Gramian of the adjoint system is equal to the system's transposed cross Gramian:

$$
\begin{aligned}
\widetilde{W}_X &:= \int_0^\infty e^{E^{-\intercal} A^\intercal t} E^{-\intercal} C^\intercal B^\intercal e^{E^{-\intercal} A^\intercal t} E^{-\intercal} dt \\
&\overset{\text{Lemma 1}}{=} \int_0^\infty (C e^{E^{-1} A t} E^{-1})^\intercal (e^{E^{-1} A t} E^{-1} B)^\intercal dt \\
&= \int_0^\infty (e^{E^{-1} A t} E^{-1} B C e^{E^{-1} A t} E^{-1})^\intercal dt = W_X^\intercal.
\end{aligned}
$$

Since a projection-error-driven POD method is used, the following error bound holds[5], for a given mean projection error $\varepsilon > 0$:

$$
\frac{1}{N} \sum_{k=1}^N \| (I - U_1 U_1^\intercal) W_{X,*k} \|^2 = \frac{1}{N} \sum_{k=n+1}^N \sigma_k^2(W_X) \le \varepsilon^2.
$$

This means the error indicator (7) can be bounded using the prescribed cross Gramian's projection error,

$$
\| y - \tilde{y} \|_{L_2} \lessapprox \sqrt{\sigma_1(BC) \sum_{k=n+1}^N \sigma_k^2(W_X)} \le \varepsilon \sqrt{N \sigma_1(BC)},
$$

thus making it an *a-priori* error indicator.

---

[5]This is shown for the HAPOD in [17].

4.2 FOM Benchmark

The first numerical example compares the cross-Gramian-based dominant sub-space method with the classic refined dominant subspace method as well as (empirical) balanced truncation[6] for the "FOM" example in [28], which is also part of the SLICOT Benchmark Collection [10]. This linear SISO system (with $E = I$) of the structure:

$$\dot{x}(t) = Ax(t) + Bu(t),$$
$$y(t) = Cx(t),$$

is of order $N = 1006$, and the system components are given by:

$$A_1 = \begin{pmatrix} -1 & 100 \\ -100 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -1 & 200 \\ -200 & -1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} -1 & 400 \\ -400 & -1 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} -1 & & & \\ & -2 & & \\ & & \ddots & \\ & & & -1000 \end{pmatrix}, \quad A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & A_3 & \\ & & & A_4 \end{pmatrix},$$

$$C = \begin{pmatrix} C_1 & C_2 \end{pmatrix}, \quad C_1 = \begin{pmatrix} 10 & \dots & 10 \end{pmatrix} \in \mathbb{R}^6, \quad C_2 = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{1000},$$

$$B = C^\mathsf{T}.$$

The empirical Gramians are constructed using impulse input $u(t) = \delta(t)$ and the reduced systems are also tested with this input, to evaluate the error indicator. In Fig. 2, the (empirical) cross-Gramian-based dominant subspaces method, the (empirical) refined dominant subspaces method, (empirical) balanced truncation and the error indicator from Section 3.4 are compared, for a given projection error $\varepsilon \in \{10^{-3}, \dots, 10^{-12}\}$. The same projection error is selected for the controllability and observability Gramians used by the refined DSPMR and low-rank empirical balanced truncation.

In Fig. 2a the prescribed projection error of the respective Gramians is plotted against the resulting L_2 model reduction error. For a given projection error the refined DSPMR method produces the lowest model reduction error, and low-rank balanced truncation the largest, while the proposed cross-Gramian-based dominant subspace method is in between. The error indicator overestimates the error for larger and underestimates for smaller projection errors. Note, that the error indicator just scales the projection error by a constant, hence it appears as a line in the log-log plot.

Fig. 2b depicts the resulting reduced order of the tested methods against the model reduction error. Balanced truncation produces the smallest, and DSPMR the largest reduced models, again the cross-Gramian-based method is in between. These results follow intuitions that DSPMR produces the most accurate, but largest subspaces, while balanced truncation may have a smaller, but less accurate subspaces. Hence, the cross-Gramian-based dominant subspace method appears as a compromise. The error indicator is rather conservative, which is due to its simple structure.

---

[6]In the numerical experiments at hand, low-rank Gramians are balanced, whereas the rank is determined by the projection error of the POD compression of the empirical controllability and observability Gramian. In this sense, this method is related to balanced POD [30].
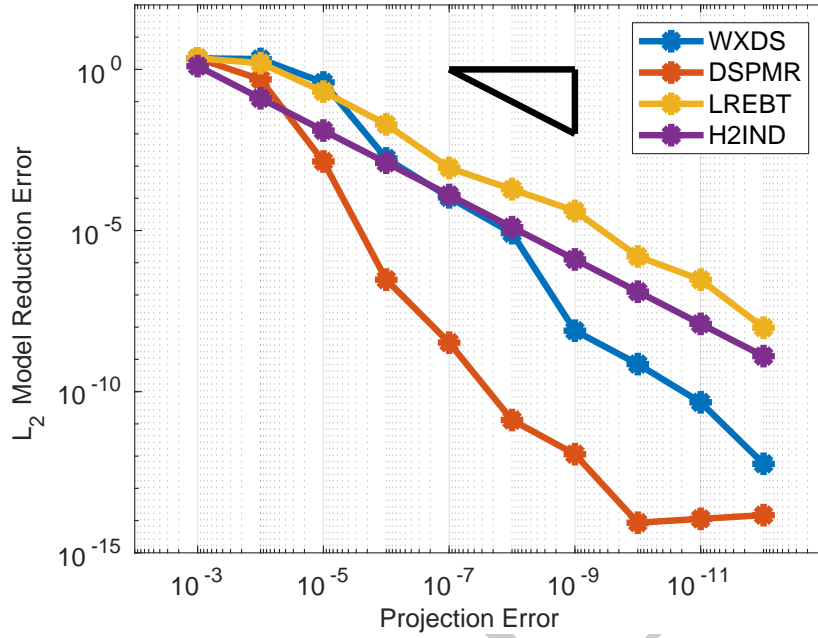
(a) Cross Gramian projection error versus L_2 model reduction error.
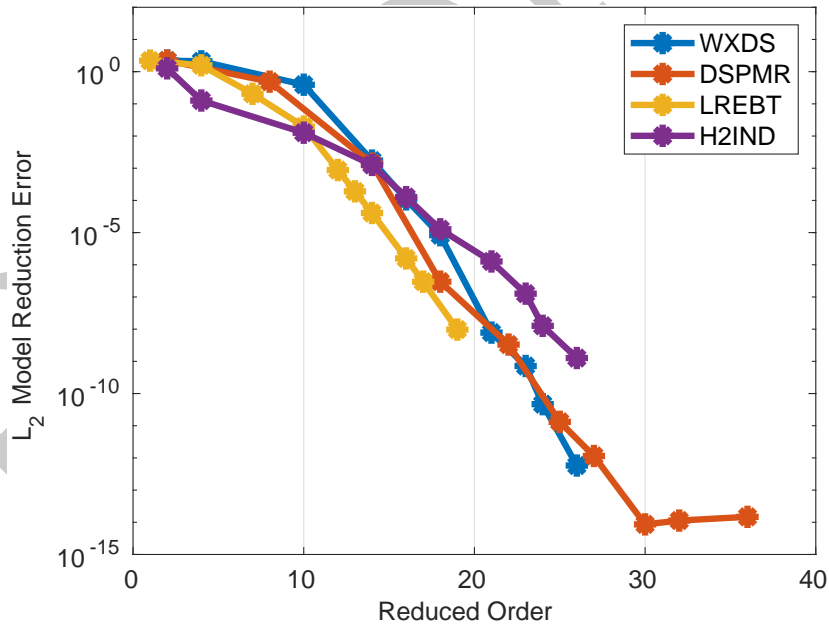
(b) Reduced order versus L_2 model reduction error.

Figure 2: Model reduction error of the **FOM benchmark** example from Section 4.2 for cross-Gramian-based dominant subspaces (WXDS), refined dominant subspaces (DSPMR), low-rank empirical balanced truncation (LREBT) and the H_2-error indicator (H2IND).

12

4.3 Convection Benchmark

The second numerical example evaluates the convection benchmark [36, Convection][7] from the Oberwolfach Benchmark Collection [26]. This is a two-dimensional computational fluid dynamics application of thermal flow modelled by a convection-diffusion partial differential equation:

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T - v \nabla T + \dot{q}$$

with the solution temperature $T(x, t)$, the thermal conductivity κ, the fluid speed v of fixed direction, and the heat generation rate $\dot{q}$. The model is discretized in space using the finite element method, yielding a generalized linear system (1) of order $N = 9669 \approx 10^4$, a single input $M = 1$ and five outputs $Q = 5$. For a more detailed description of this benchmark see [26] and references therein. This model is tested in two variants: First, in a diffusion-dominant setting with zero flow speed $v = 0$, and second, in a convection-dominant setting with a flow speed $v = 0.5$. Due to the MIMO nature of the system we use the average system (see (3)) for the error indicator computation.

This set of experiments is organized in the same manner as Section 4.2, but conducted for the prescribed projection errors $\varepsilon \in \{10^{-2}, \ldots, 10^{-8}\}$. As indicated in Section 3.4, the average system (averaged over outputs) $\bar{C} := \sum_{q=1}^{Q} c_q^\intercal$ is used for the computation of the error indicator. The resulting reduced order models are tested with impulse input $u(t) = \delta(t)$.

4.3.1 Diffusion-Dominant Variant

The experimental results of the diffusion-dominant variant ($v = 0$) is depicted in Fig. 3. In Fig. 3a the prescribed projection error for the (empirical) system Gramians versus the resulting L_2 model reduction error is plotted. As in Section 4.2, the DSPMR method produces the reduced order models with the lowest model reduction error. Reduced systems from (empirical) balanced truncation and the (empirical) cross-Gramian-based dominant subspace method result in similar errors, while the error indicator behaves behaves like a conservative upper bound to the cross-Gramian-based model reduction error.

Fig. 3b shows the model reduction error for the reduced orders resulting from the prescribed projection error. Balanced truncation achieves the smallest and DSPMR the largest reduced models, the cross-Gramian-based dominant subspace method reduced order model dimension lies in between, and the error indicator shows a similar behavior as the latter.
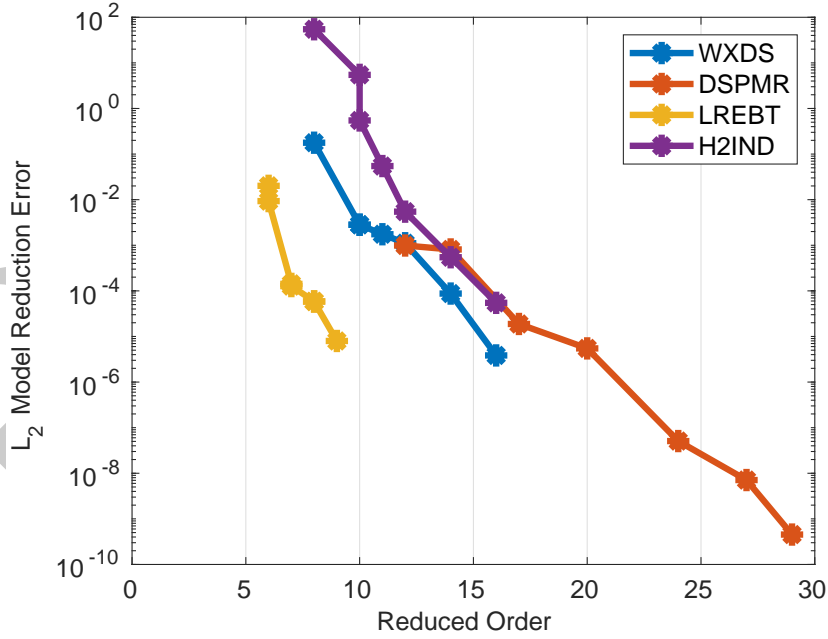
4.3.2 Convection-Dominant Variant

The experimental results of the convection-dominant variant ($v = 0.5$) is presented in Fig. 4. Overall, the plots Fig. 4a and Fig. 4b are similar to the diffusion-dominant variant, yet the model reduction errors are higher and flatten out for the cross-Gramian-based method for $\varepsilon < 10^{-4}$. This means that the model reduction error does not follow the error indicator, which is equal for diffusion- and convection-dominant benchmark variants.

---

[7]http://modelreduction.org/index.php/Convection
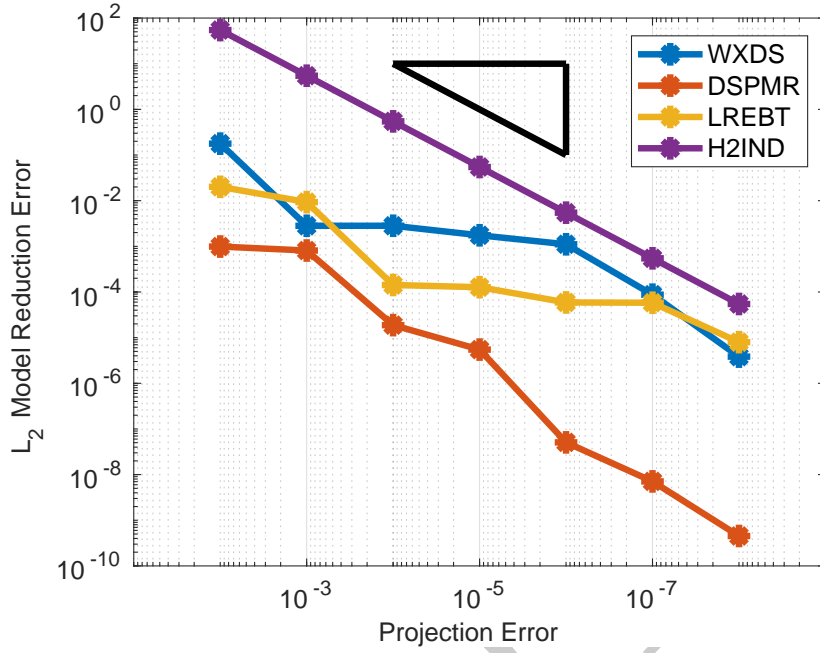
(a) Cross Gramian projection error versus L_2 model reduction error.

(b) Reduced order versus L_2 model reduction error.

Figure 3: Model reduction error of the **diffusion-dominant** convection benchmark from Section 4.3 for cross-Gramian-based dominant subspaces (WXDS), refined dominant subspaces (DSPMR), low-rank empirical balanced truncation (LREBT) and the H_2-error indicator (H2IND).

Hence, for stiff or hyperbolic systems the cross-Gramian-based dominant subspace model reduction needs to be adapted, for example by fitting the known singular values exponentially and incorporate such an empirical decay rate.

5 Summary

In this work we revisited the dominant subspaces projection model reduction method, and presented a variant based on the cross Gramian matrix for generalized linear systems. This model reduction requires only a single low-rank decomposition of the cross Gramian, and provides an *a-priori* error indicator. The applicability of this method to control-affine nonlinear systems will be subject of future work, which is in principal possible due to the utilized empirical Gramian computation leading to empirical dominant subspaces.

Code Availability Section

The source code of the presented numerical examples can be obtained from:

http://runmycode.org/companion/view/3270

and is authored by: CHRISTIAN HIMPE.

Acknowledgement

References

[1] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*, volume 6 of *Adv. Des. Control*. SIAM Publications, Philadelphia, PA, 2005. `doi:10.1137/1.9780898718713`.

[2] U. Baur, P. Benner, and L. Feng. Model order reduction for linear and nonlinear systems: A system-theoretic perspective. *Arch. Comput. Methods Eng.*, 21(4):331–358, 2014. `doi:10.1007/s11831-014-9111-2`.

[3] P. Benner. Solving large-scale control problems. *IEEE Control Syst. Mag.*, 14(1):44–59, 2004. `doi:10.1109/MCS.2004.1272745`.

[4] P. Benner, C. Himpe, and T. Mitchell. On reduced input-output dynamic mode decomposition. *Advances in Computational Mathematics*, 2018. (Accepted). `doi:10.1007/s10444-018-9592-x`.

[5] P. Benner and P. Kürschner. Computing real low-rank solutions of Sylvester equations by the factored ADI method. *Comput. Math. Appl.*, 67(9):1656–1672, 2014. `doi:10.1016/j.camwa.2014.03.004`.
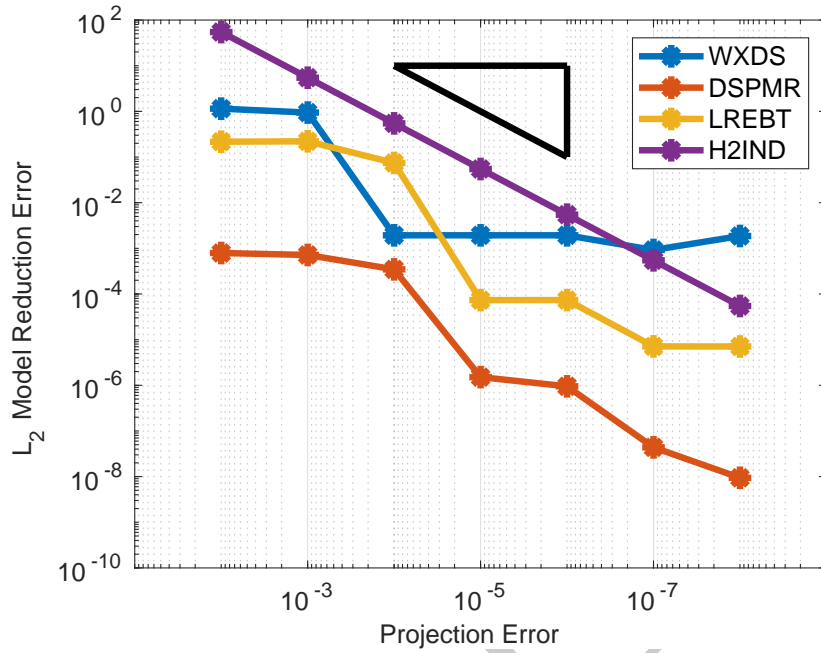
(a) Cross Gramian projection error versus L_2 model reduction error.
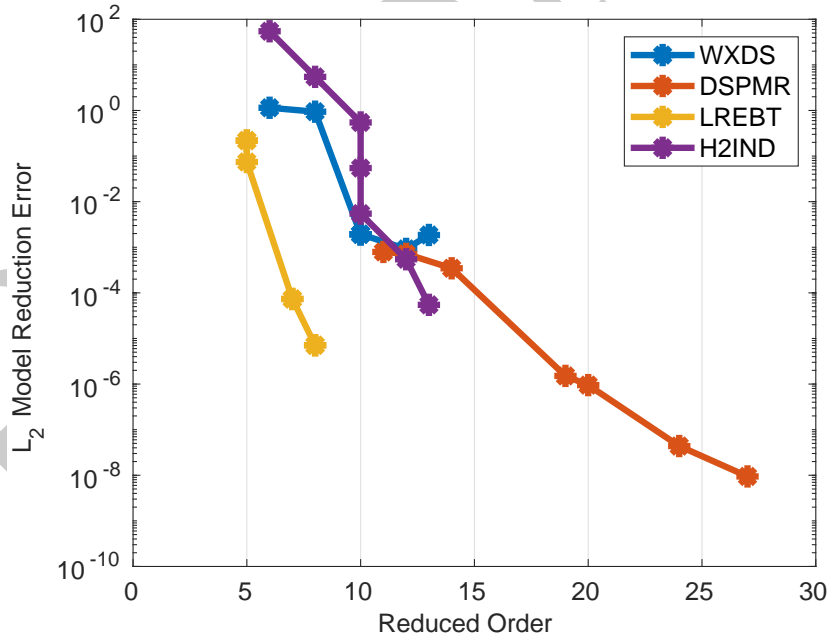
(b) Reduced order versus L_2 model reduction error.

Figure 4: Model reduction error of the **convection-dominant** convection benchmark from Section 4.3 for cross-Gramian-based dominant subspaces (WXDS), refined dominant subspaces (DSPMR), low-rank balanced truncation (LREBT) and the H_2-error indicator (H2IND).

[6] P. Benner, P. Kürschner, and J. Saak. Self-generating and efficient shift parameters in ADI methods for large Lyapunov and Sylvester equations. *Electron. Trans. Numer. Anal.*, 43:142–162, 2014. URL: `http://etna.mcs.kent.edu/volumes/2011-2020/vol43/abstract.php?vol=43&pages=142-162`.

[7] P. Benner, R.-C. Li, and N. Truhar. On the ADI method for Sylvester equations. *J. Comput. Appl. Math.*, 233(4):1035–1045, December 2009. `doi:10.1016/j.cam.2009.08.108`.

[8] B. N. Bond and L. Daniel. Guaranteed stable projection-based model reduction for indefinite and unstable linear systems. In *2008 IEEE/ACM International Conference on Computer-Aided Design*, 2008. `doi:10.1109/ICCAD.2008.4681657`.

[9] R. Bru, C. Coll, and N. Thome. Symmetric singular linear control systems. *Applied Mathematics Letters*, 15(6):671–675, 2002. `doi:10.1016/S0893-9659(02)00026-5`.

[10] Y. Chahlaoui and P. Van Dooren. A collection of benchmark examples for model reduction of linear time invariant dynamical systems. Technical Report 2002–2, SLICOT Working Note, 2002. Available from `www.slicot.org`.

[11] A. Davidson. Balanced systems and model reduction. *Electron. Lett.*, 22(10):531–532, 1986. `doi:10.1049/el:19860362`.

[12] K. V. Fernando and H. Nicholson. On the structure of balanced and other principal representations of SISO systems. *IEEE Trans. Autom. Control*, 28(2):228–231, 1983. `doi:10.1109/TAC.1983.1103195`.

[13] J. D. Gardiner, A. J. Laub, J. J. Amato, and C. B. Moler. Solution of the Sylvester matrix equation $AXB + CXD = E$. *ACM Trans. Math. Software*, 18(2):223–231, 1992. `doi:10.1145/146847.146929`.

[14] S. Gugercin, A. C. Antoulas, and C. Beattie. $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30(2):609–638, 2008. `doi:10.1137/060666123`.

[15] C. Himpe. emgr – EMpirical GRamian framework (version 5.5). `https://gramian.de`, 2018. `doi:10.5281/zenodo.1401500`.

[16] C. Himpe. emgr – the Empirical Gramian Framework. *Algorithms*, 11(7):91, 2018. `doi:10.3390/a11070091`.

[17] C. Himpe, T. Leibner, and S. Rave. Hierarchical approximate proper orthogonal decomposition. *SIAM J. Sci. Comput.*, 2018. (Accepted for publication). URL: `https://arxiv.org/abs/1607.05210`.

[18] C. Himpe, T. Leibner, S. Rave, and J. Saak. Fast low-rank empirical cross Gramians. *Proc. Appl. Math. Mech.*, 17(1):841–842, 2017. `doi:10.1002/pamm.201710388`.

[19] C. Himpe and M. Ohlberger. A note on the cross Gramian for non-symmetric systems. *Systems Science and Control Engineering*, 4(1):199–208, 2016. `doi:10.1080/21642583.2016.1215273`.

[20] Y.-L. Jiang, Z.-Z. Qi, and P. Yang. Model order reduction of linear systems via the cross Gramian and SVD. *IEEE Transactions on Circuits and Systems II: Express Briefs*, (Early Access), 2018. `doi:10.1109/TCSII.2018.2864115`.

[21] J.-R. Li and J. White. Efficient model reduction of interconnect via approximate system Gramians. In *1999 IEEE/ACM International Conference on Computer-Aided Design. Digest of Technical Papers*, pages 380–383, 1999. `doi:10.1109/ICCAD.1999.810679`.

[22] J.-R. Li and J. White. Reduction of large circuit models via low rank approximate Gramians. *Int. J. Appl. Math. Comput. Sci.*, 11(5):1151–1171, 2001. URL: `http://eudml.org/doc/207549`.

[23] The MathWorks, Inc., `http://www.matlab.com`. *MATLAB*.

[24] L. Mirsky. A trace inequality of John von Neumann. *Monatshefte für Mathematik*, 79(4):303–306, 1975. `doi:10.1007/BF01647331`.

[25] B. C. Moore. Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Autom. Control*, AC–26(1):17–32, 1981. `doi:10.1109/TAC.1981.1102568`.

[26] C. Moosmann and A. Greiner. Convective thermal flow problems. In *Dimension Reduction of Large-Scale Systems*, volume 45, pages 341–343. Springer, 2005. `doi:10.1007/3-540-27909-1\_16`.

[27] L. Peng and K. Mohseni. Symplectic model reduction of Hamiltonian system. *SIAM J. Sci. Comput.*, 38(1):A1–A27, 2016. `doi:10.1137/140978922`.

[28] T. Penzl. Algorithms for model reduction of large dynamical systems. *Linear Algebra Appl.*, 415(2–3):322–343, 2006. (Reprint of Technical Report SFB393/99-40, TU Chemnitz, 1999.). `doi:10.1016/j.laa.2006.01.007`.

[29] M. Redmann and P. Kürschner. An output error bound for time-limited balanced truncation. *Syst. Control Lett.*, 2018. (Accepted for publication). URL: `https://arxiv.org/abs/1710.07572`.

[30] C. W. Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *Int. J. Bifurcat. Chaos*, 15(3):997–1013, 2005. `doi:10.1142/S0218127405012429`.

[31] J. Saak. *Efficient Numerical Solution of Large Scale Algebraic Matrix Equations in PDE Control and Model Order Reduction*. Dissertation, Technische Universität Chemnitz, Chemnitz, Germany, July 2009. URL: `http://nbn-resolving.de/urn:nbn:de:bsz:ch1-200901642`.

[32] H. R. Shaker. Generalized cross-Gramian for linear systems. In *Proc. IEEE Conf. Ind. Electron. Appl.*, pages 749–751, 2012. `doi:10.1109/ICIEA.2012.6360824`.

[33] G. Shi and C.-R. J. Shi. Model-order reduction by dominant subspace projection: error bound, subspace computation, and circuit applications. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 52(5):975–993, 2005. `doi:10.1109/TCSI.2005.846217`.

[34] D. C. Sorensen and A. C. Antoulas. The Sylvester equation and approximate balanced reduction. *Numer. Lin. Alg. Appl.*, 351–352:671–700, 2002. `doi:10.1016/S0024-3795(02)00283-5`.

[35] T. Stykel. Gramian-based model reduction for descriptor systems. *Math. Control Signals Systems*, 16(4):297–319, 2004. `doi:10.1007/s00498-004-0141-4`.

[36] The MORwiki Community. MORwiki - Model Order Reduction Wiki. `http://modelreduction.org`.

[37] R. Toscano. *Structured Controllers for Uncertain Systems*. Advances in Industrial Control. Springer London, 2013. `doi:10.1007/978-1-4471-5188-3`.

[38] X. Wang and M. Yu. The error bound of timing domain in model order reduction by Krylov subspace methods. *Journal of Circuits, Systems, and Computers*, 27(6):1850093, 2018. `doi:10.1142/S0218126618500937`.

[39] T. Wolf, H. Panzer, and B. Lohmann. Gramian-based error bound in model reduction by Krylov subspace methods. *IFAC Proceedings Volumes (Proceedings of the 18th IFAC World Congress)*, 44(1):3587–3592, 2011. `doi:10.3182/20110828-6-IT-1002.02809`.