

Basic principles and aims of model order reduction in compliant mechanisms

M. Rösner and R. Lammering

Institute for Mechanics, Helmut-Schmidt-University/University of the Federal Armed Forces Hamburg, Holstenhofweg 85, 22043 Hamburg, Germany

Received: 25 February 2011 – Revised: 23 May 2011 – Accepted: 26 July 2011 – Published: 17 October 2011

Abstract. Model order reduction appears to be beneficial for the synthesis and simulation of compliant mechanisms due to computational costs. Model order reduction is an established method in many technical fields for the approximation of large-scale linear time-invariant dynamical systems described by ordinary differential equations. Based on system theory, underlying representations of the dynamical system are introduced from which the general reduced order model is derived by projection. During the last years, numerous new procedures were published and investigated appropriate to simulation, optimization and control. Singular value decomposition, condensation-based and Krylov subspace methods representing three order reduction methods are reviewed and their advantages and disadvantages are outlined in this paper. The convenience of applying model order reduction in compliant mechanisms is quoted. Moreover, the requested attributes for order reduction as a future research direction meeting the characteristics of compliant mechanisms are commented.

1 Introduction

A new approach to develop a feed unit of small machine tools for small workpieces is based on the application of compliant mechanisms. Currently, non-intuitive design and optimization techniques are in progress as well as controlling, measuring and calibration strategies. To describe the mechanical behavior of a feed unit in an accurate way, very large and sparse finite element models arise. This leads to numerical simulations which require an unacceptable amount of time and memory space and motivates the introduction and application of model order reduction (MOR) methods. In this paper, MOR approaches are reviewed taking a first step towards its implementation in compliant mechanisms. To the authors' knowledge, applications and investigations of MOR in CMs are not existent.

In the present case, the feed unit consists of two major parts: (a) the compliant mechanism (CM) is a mechanical structure consisting of flexure hinges and rigid regions, including piezo-electric actuators and (b) the measure and control system supplying appropriate input signals for the mechanical structure via the actuators. Figure 1 illustrates a prototype of a CM appropriate to novel machine tools (the

actuators are not shown), as described by Wulfsberg et al. (2010), embedded in the mechatronic system.

The utilization of CMs assuring the trajectory of the tool is constituted by their performance. CMs distinguish from previous standard machine tools being potentially more accurate, better scalable, cleaner, low-maintenance, less noisy and cheaper in manufacturing which makes them particularly suitable for small-scale applications.

2 System theoretical background

Assuming a linear time invariant system (LTI), the performance of the compliant mechanism can be specified by a system of second order ordinary differential equations (ODE) derived from linear finite element discretization, see e.g. Koutsovasilis and Beitel Schmidt (2008). This system describes the reaction of the mechanical structure in answer to inputs from the control system and can be written in the time domain by means of the state space form in vectorial representation as:

$$\begin{aligned} \mathbf{M}_e \ddot{\mathbf{q}}(t) + \mathbf{D}_e \dot{\mathbf{q}}(t) + \mathbf{K}_e \mathbf{q}(t) &= \mathbf{B}_e \mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{C}_e \mathbf{q}(t), \end{aligned} \quad (1)$$

where $\mathbf{M}_e \in \mathbb{R}^{n \times n}$ mass matrix, $\mathbf{D}_e \in \mathbb{R}^{n \times n}$ damping matrix, $\mathbf{K}_e \in \mathbb{R}^{n \times n}$ stiffness matrix, $\mathbf{B}_e \in \mathbb{R}^{n \times p}$ input matrix with $\mathbf{u}(t) \in \mathbb{R}^p$ input vector forming an external force vector, $\mathbf{C}_e \in \mathbb{R}^{m \times n}$



Correspondence to: M. Rösner
(malte.roesner@hsu-hh.de)

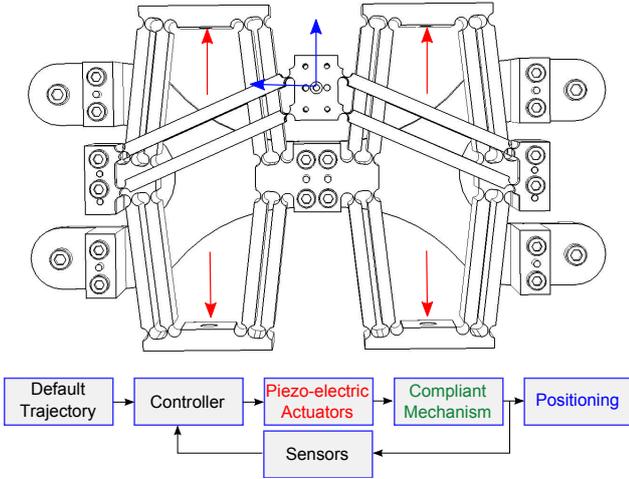


Figure 1. Prototype of a compliant mechanism used in a feed unit for novel machine tools as component of the mechatronic system.

output matrix selecting the $y(t) \in \mathbb{R}^m$ output vector of interest of the $q(t) \in \mathbb{R}^n$ internal state vector corresponding to the structural characteristics of the CM. Moreover, n is the order of the system, also referred to as state space dimension. In case $m = p = 1$ and therefore B_e and C_e switch over to vectors as well as $u(t)$ and $y(t)$ to scalars, the systems is called single input – single output (SISO) otherwise multiple input – multiple output (MIMO).

To accurately reproduce the structural behavior, in most instances, the set of ODEs stated in Eq. (1) reaches a numerousness of degrees of freedom (DOF), stated by n in the present case. These large scale systems entail prohibitive computational time and memory space for calculation making them impractical for simulation, optimization, analysis and control. Therefore, MOR techniques are implemented to gain a low dimensional approximation of the high order system. MOR can be arranged in the simulation process, as shown in Fig. 2.

Depending on the reduction method, different representations of the system are required, which are illustrated below.

Several approximation methods rely on first order systems. The transformation of Eq. (1) to a first order descriptor system is realized by applying the state vector $x(t) = [q^T | \dot{q}^T]^T$ leading to the state-space dynamical system

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \quad (2)$$

where $A \in \mathbb{R}^{n_s \times n_s}$ system matrix, $B \in \mathbb{R}^{n_s \times p}$ input matrix, $C \in \mathbb{R}^{m \times n_s}$ output matrix, $E \in \mathbb{R}^{n_s \times n_s}$ descriptor matrix and $n_s = 2n$. Since some reduction methods are not based on the representations written in Eqs. (1) and (2), respectively, but on the transfer function matrix $H(s)$. Its derivation is described in the following.

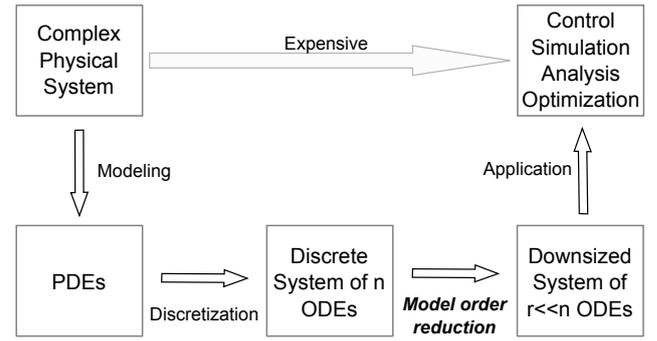


Figure 2. Classification of model order reduction in the simulation process of a dynamical system.

Using the Laplace transformation, the system specified in Eq. (1) is converted from time to frequency domain resulting in an algebraic system of equations given by

$$\begin{aligned} s^2 M_e Q(s) + s D_e Q(s) + K_e Q(s) &= B_e U(s), \\ Y(s) &= C_e Q(s), \end{aligned} \quad (3)$$

where s is the complex-valued scalar parameter and $Q(s)$, $U(s)$, $Y(s)$ are the Laplace transformed of $q(t)$, $u(t)$, $y(t)$ and provided that homogeneous initial conditions $q(t=0) = \dot{q}(t=0) = 0$ are existent.

The transfer function matrix $H(s)$ of the system can be specified by combining the two equations in Eq. (3), as

$$Y(s) = \underbrace{C_e [s^2 M_e + s D_e + K_e]^{-1} B_e}_{H(s)} U(s), \quad (4)$$

alternatively for the descriptor system in Eq. (2)

$$H(s) = C(sE - A)^{-1} B, \quad (5)$$

where the transfer function matrix $H(s)$ has a dimension of $(m \times p)$ in the MIMO case and, accordingly, is a scalar in the SISO case. The transfer function $H(s)$ describes the correlation between input $U(s)$ and output $Y(s)$, disregarding the internal states of the system.

Due to the fact that most MOR procedures are accomplished by means of projection, the approach is explained by a system represented in Eq. (2). The objective is to approximate the state vector $x(t) \in \mathbb{R}^{n_s}$ to a low-dimensional subspace. This is achieved by the substitution

$$x(t) = Vx_r(t) + \epsilon(t), \quad (6)$$

where $V \in \mathbb{R}^{n_s \times q}$ is the transformation matrix, $x_r(t) \in \mathbb{R}^q$ the reduced state vector, $\epsilon(t)$ the residual and $q \ll n_s$. Inserting the projection Eq. (6) into the linear system of ODEs in Eq. (2) results in a low-dimensional approximation

$$\begin{aligned} EV\dot{x}_r(t) &= AVx_r(t) + BVu(t) + \epsilon(t), \\ y(t) &= CVx_r(t). \end{aligned} \quad (7)$$

Table 1. Characteristics of condensation-based methods.

Advantages	Disadvantages
– stability is preserved	– not automatable
– invariant to system representation	– unnecessarily large reduced order models
– physical information is preserved	– not expandable to nonlinear systems
– independent from the number of the systems inputs and outputs	– no presetting of a minimal dimension of the reduced system possible
– error estimation possible	– computationally intense
– implemented in software (Matlab, ANSYS)	– inappropriate for large scale systems

In general, this system is overdetermined having q unknowns but n_s equations. To solve this problem and find a unique solution, Eq. (7) is pre-multiplied by a second transformation matrix $\mathbf{W} \in \mathbb{R}^{n_s \times q}$ such that $\mathbf{W}^T \boldsymbol{\epsilon}(t) = 0$, the residual is equal to zero and

$$\underbrace{\mathbf{W}^T \mathbf{E} \mathbf{V}}_{\mathbf{E}_r} \dot{\mathbf{x}}_r(t) = \underbrace{\mathbf{W}^T \mathbf{A} \mathbf{V}}_{\mathbf{A}_r} \mathbf{x}_r(t) + \underbrace{\mathbf{W}^T \mathbf{B} \mathbf{u}(t)}_{\mathbf{B}_r},$$

$$\mathbf{y}(t) = \underbrace{\mathbf{C} \mathbf{V}}_{\mathbf{C}_r} \mathbf{x}_r(t). \quad (8)$$

The system in Eq. (8) is called general *reduced order model* (ROM) by projection. The number of inputs $\mathbf{u}(t) \in \mathbb{R}^p$ and outputs $\mathbf{y}(t) \in \mathbb{R}^m$ remains the same though the order of the system, expressed by the dimension of the reduced system matrices $\mathbf{E}_r \in \mathbb{R}^{q \times q}$, $\mathbf{A}_r \in \mathbb{R}^{q \times q}$, $\mathbf{B}_r \in \mathbb{R}^{q \times p}$, $\mathbf{C}_r \in \mathbb{R}^{q \times m}$ and the reduced state vector $\mathbf{x}_r(t) \in \mathbb{R}^q$ decreases from n_s to q .

For most of the methods, the aim of model order reduction is to provide the projection matrices \mathbf{W} and \mathbf{V} such that their calculation is computationally efficient as well as automatable and the systems characteristics are preserved. The existence of a predefined error bound is desired in addition to the applicability in large-scale systems with an order up to a few hundred thousand. Details are specified e.g. in Eid (2009) and Rudnyi and Korvink (2006).

3 Reduction methods based on condensation

The reduction of a given system using condensation methods is the most commonly applied reduction method. Applying the time independent transformation matrices \mathbf{W}^T and \mathbf{V} to the first equation in Eq. (1) gives the reduced system

$$\mathbf{M}_r \ddot{\mathbf{x}}_r(t) + \mathbf{D}_r \dot{\mathbf{x}}_r(t) + \mathbf{K}_r \mathbf{x}_r(t) = \mathbf{B}_r \mathbf{u}(t), \quad (9)$$

with $\mathbf{M}_r = \mathbf{W}^T \mathbf{M}_e \mathbf{V}$, $\mathbf{D}_r = \mathbf{W}^T \mathbf{D}_e \mathbf{V}$, $\mathbf{K}_r = \mathbf{W}^T \mathbf{K}_e \mathbf{V}$ and $\mathbf{B}_r = \mathbf{W}^T \mathbf{B}_e$ where the transformation matrices have to ensure the conservation of the potential and kinetic energy of the system. The reduced damping matrix \mathbf{D}_r may be modelled as linear combination of the mass and stiffness matrices (Rayleigh damping) with $\mathbf{D}_r = \alpha \mathbf{M}_r + \beta \mathbf{K}_r$.

The construction of the transformation matrix \mathbf{V} is carried out either by static, modal or mixed condensation which are explained briefly below, see Siedl (2008), Gasch and Knothe

(1989), Bennini (2005) and Gugel (2009). In Table 1, the main advantages and disadvantages are listed for MOR using condensation-based reduction methods.

3.1 Static condensation

Static condensation is based on the partition of the systems degrees of freedom into dependent (MDOF – master degree of freedom, index m) and independent (SDOF – slave degree of freedom, index s) ones in a physical expedient way. This procedure is also the basis for sub-structure modeling and the generation of super elements. For the static case, $\dot{\mathbf{q}}(t) = \dot{\mathbf{q}}(t) = 0$, an external force vector is introduced by $\mathbf{B}_e \mathbf{u}(t) = \mathbf{F}(t)$. The aforementioned partition into MDOF and SDOF applied to Eq. (1) results in

$$\begin{bmatrix} \mathbf{K}_{mm} & \mathbf{K}_{ms} \\ \mathbf{K}_{sm} & \mathbf{K}_{ss} \end{bmatrix} \begin{bmatrix} \mathbf{q}_m(t) \\ \mathbf{q}_s(t) \end{bmatrix} = \begin{bmatrix} \mathbf{F}_m(t) \\ \mathbf{F}_s(t) \end{bmatrix}. \quad (10)$$

The second row in Eq. (10) provides the so called static correlation between the MDOF and SDOF given by

$$\mathbf{q}_s(t) = -\mathbf{K}_{ss}^{-1} \mathbf{K}_{sm} \mathbf{q}_m(t) + \mathbf{K}_{ss}^{-1} \mathbf{F}_s(t). \quad (11)$$

One obtains the reduced system by substituting $\mathbf{q}_s(t)$ in the first row of Eq. (10) by Eq. (11). Rearranging ensues

$$\mathbf{K}_{red} \mathbf{q}_m(t) = \mathbf{F}_{red}, \quad (12)$$

where $\mathbf{K}_{red} = [\mathbf{K}_{mm} - \mathbf{K}_{ms} \mathbf{K}_{ss}^{-1} \mathbf{K}_{sm}^{-1}]$ and $\mathbf{F}_{red} = \mathbf{F}_m(t) - \mathbf{K}_{ms} \mathbf{K}_{ss}^{-1} \mathbf{F}_s(t)$. Therewith, the transformation matrices in Eq. (8) have the following structure

$$\mathbf{V} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{K}_{ss}^{-1} \mathbf{K}_{sm} \end{bmatrix}, \mathbf{W} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{K}_{ms} \mathbf{K}_{ss}^{-1} \end{bmatrix}, \quad (13)$$

where \mathbf{I} is the identity matrix. The identification of \mathbf{K}_{red} in this way is exact. In case of approximately reducing the mass matrix \mathbf{M}_e and damping matrix \mathbf{D}_e also with the transformation matrices in Eq. (13) from the static approach, the procedure is called Guyan reduction, as stated by Stelzmann et al. (2002) and Lienemann (2006).

3.2 Modal condensation

Using modal condensation, only a selected number of eigenmodes is taken into account. The involved reduction is based

Table 2. Characteristics of Krylov subspace methods.

Advantages	Disadvantages
– iterative method	– no support in Matlab
– easy to implement	– no global error bound
– suitable for very large scale systems	– order depends on number of inputs/outputs
– computational efficient	– result may depend on systems representation
– stationary exact	– no guarantee for preserving stability
– robust calculation	– expansion point required

on the limitation of the explorative frequency range. Resolving the eigenvalue problem, neglecting damping and external forces, results in the modal matrix Φ assembled by the eigenvectors φ_i

$$\Phi = [\varphi_1, \varphi_2, \dots, \varphi_l] \quad (14)$$

where $\Phi \in \mathbb{R}^{l \times l}$ is getting downsized by choosing a minor frequency range resulting in the reduced modal matrix $\Phi_r \in \mathbb{R}^{l \times k}$. Using an orthogonal projection $\mathbf{W} = \mathbf{V}$ as mentioned by Rickelt-Rolf (2009), the transformation matrices are chosen such that

$$\mathbf{W} = \mathbf{V} = \Phi_r. \quad (15)$$

3.3 Mixed condensation

Mixed condensation, known as reduction by Craig and Bampton (1968), combines static and modal reduction methods. In a first step, the system in Eq. (9) is similarly re-sorted with respect to SDOF and MDOF as described before. Secondly, the MDOF are blocked out by setting $\mathbf{q}_m = 0$ resulting in an auxiliary system

$$\mathbf{M}_{ss}\ddot{\mathbf{q}}_s(t) + \mathbf{K}_{ss}\mathbf{q}_s(t) = 0, \quad (16)$$

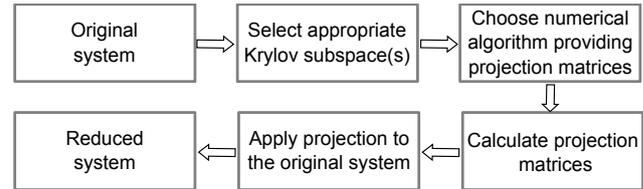
of which the eigenmodes φ_i are calculated subsequently. Analogue to the modal condensation, the reduced modal matrix Φ_r is established by selecting certain eigenmodes as described by Siedl (2008). The SDOF is again linked to the MDOF via the static correlation and the transformation matrices are complemented to

$$\mathbf{W} = \mathbf{V} = \begin{bmatrix} \Phi_r & \Phi_s \\ 0 & \mathbf{I} \end{bmatrix} \quad (17)$$

with $\Phi_s = -\mathbf{K}_{ss}^{-1}\mathbf{K}_{sm}$ and $\Phi_r = [\varphi_1, \varphi_2, \dots, \varphi_k]$ as mentioned in Sect. 3.1 and Sect. 3.2, respectively.

4 Krylov subspace methods

Originally, Krylov subspace methods were developed for iteratively solving large and sparse linear systems of equations. Multiple methods are available,

**Figure 3.** Key steps of dimensional reduction using Krylov subspaces.

for example Generalized-Minimal-Residual-Method (GMRES), Biconjugate-Gradient (BiCG), Conjugate-Gradient-Square (CGS) and Transpose-Free-Quasi-Minimal-Residual (TFQMR), which can be found e.g. in Meister (2005), Kanzow (2005), Van der Vorst (2003) and Bai (2000) to name only a few.

In case of MOR, Krylov subspaces are used to form the projection matrices \mathbf{V} and \mathbf{W} in Eq. (8). The general approach is based on the approximation of the transfer behavior of the original system by means of the transfer function $\mathbf{H}(s)$ in the frequency domain given in Eq. (4) and Eq. (5), respectively. MOR using Krylov subspaces is also well-known as moment matching, meaning that a certain number of moments of the reduced and the original system is equal. Hence, the moments of the transfer function and their matching in the sense of an approximation play a decisive role and are described below. Figure 3 illustrates the key steps using Krylov subspaces.

In Table 2, the main advantages and disadvantages are listed for MOR using Krylov subspaces, given in Salimbahrami (2006) and Eid (2009).

4.1 Transfer function and its moments

Considering the transfer function in Eq. (5) for the MIMO case, rearranging results in

$$\mathbf{H}(s) = -\mathbf{C}(\mathbf{I} - s\mathbf{A}^{-1}\mathbf{E})^{-1}\mathbf{A}^{-1}\mathbf{B}. \quad (18)$$

Taking into account the Neumann series, as described by Werner (2009) and Eid (2009),

$$(\mathbf{I} - s\mathbf{A}^{-1}\mathbf{E})^{-1} = \sum_{j=0}^{\infty} (\mathbf{A}^{-1}\mathbf{E}s)^j, \quad (19)$$

one obtains the following Taylor series

$$\mathbf{H}(s) = \sum_{j=0}^{\infty} -\mathbf{C}(\mathbf{A}^{-1}\mathbf{E})^j \mathbf{A}^{-1}\mathbf{B} s^j = \sum_{j=0}^{\infty} -\mathbf{M}_j^0 s^j. \quad (20)$$

This power series is called MacLaurin series, with $\mathbf{M}_j^0 = \mathbf{C}(\mathbf{A}^{-1}\mathbf{E})^j \mathbf{A}^{-1}\mathbf{B}$ called moments of the transfer function and its expansion point being $s = 0$. Depending on the expansion point, the moments and its corresponding MOR schemes are defined as:

- Expansion point $s = 0$ results in a MacLaurin series with its moments defined by

$$\mathbf{M}_j^{s=0} = \mathbf{C}(\mathbf{A}^{-1}\mathbf{E})^j \mathbf{A}^{-1}\mathbf{B}. \quad (21)$$

The corresponding MOR procedure is called *Padé-Approximation*.

- Expansion point $s = s_k$ results in a Taylor series with its moments defined by

$$\mathbf{M}_j^{s=s_k} = \mathbf{C}((\mathbf{A} - s_k\mathbf{E})^{-1}\mathbf{E})^j (\mathbf{A} - s_k\mathbf{E})^{-1}\mathbf{B}. \quad (22)$$

The corresponding MOR procedure is called *Shifted Padé-Approximation*, *Rational Interpolation* or *Multi-point Padé-Approximation*.

- Expansion point $s \rightarrow \infty$ results in a Markov series with its moments defined by

$$\mathbf{M}_j^{s \rightarrow \infty} = \mathbf{C}(\mathbf{E}^{-1}\mathbf{A})^j \mathbf{E}^{-1}\mathbf{B}. \quad (23)$$

The corresponding MOR procedure is called *Partial Realisation*.

The users choice of the expansion point concerning its location and number in case of the Shifted Padé-Approximation, which is also suitable for multiple expansion points, affects the approximating characteristics. In the context of MOR, Padé-Approximation matches the behavior in the low frequency range whereas Partial Realisation fullfils in the high frequency range and Rational Interpolation in the user specified one. The usage of moment matching in sparse, large scale systems is due to the simplicity of the involved operations to compute these moments, namely matrix vector multiplication and matrix inversion. Details may be found in Eid (2009), Lehner and Eberhard (2006), Grimme (1997) and Salimbahrami (2006).

4.2 Krylov subspaces and moment matching

To provide the projection matrices \mathbf{V} and \mathbf{W} , Krylov subspaces are employed to match a certain number of the previously defined moments. A Krylov subspace is spanned by a succession of vectors as

$$\mathbf{K}_q(\mathbf{A}, \mathbf{b}) = \text{span}\{\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \dots, \mathbf{A}^{q-1}\mathbf{b}\}, \quad (24)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a matrix, $\mathbf{b} \in \mathbb{R}^n$ is called starting vector and $\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \dots, \mathbf{A}^{q-1}\mathbf{b}$ are the generated basic vectors spanning the q -th Krylov subspace. The first linearly independent basic vectors form the basis of this Krylov subspace. In case multiple starting vectors have to be considered, a q -th order block Krylov subspace can be rendered precisely by

$$\mathbf{K}_q(\mathbf{A}, \mathbf{B}) = \text{span}\{\mathbf{B}, \mathbf{A}\mathbf{B}, \mathbf{A}^2\mathbf{B}, \dots, \mathbf{A}^{q-1}\mathbf{B}\}, \quad (25)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_p] \in \mathbb{R}^{n \times p}$ is the starting matrix containing the p linearly independent starting vectors. If $p = 1$, one obtains the standard Krylov subspace in Eq. (24) with the starting vector \mathbf{b} .

For the reduced order system in Eq. (8), the basic vectors of a suitable Krylov subspace can be utilized to find the projection matrices \mathbf{V} and \mathbf{W} . An explicit calculation of the moments to match is not necessary. As carried out in Koutsovasilis (2009) and Salimbahrami (2006) an input block Krylov subspace $\mathbf{K}_{q1}(\mathbf{A}^{-1}\mathbf{E}, \mathbf{A}^{-1}\mathbf{B})$ and output block Krylov subspace $\mathbf{K}_{q2}(\mathbf{A}^{-T}\mathbf{E}^T, \mathbf{A}^{-T}\mathbf{C}^T)$ are to be used for an expansion point $s = 0$ to match some moments in Eq. (21).

For $s = s_k$, the input block Krylov subspace is $\mathbf{K}_{q3}((\mathbf{A} - s_k\mathbf{E})^{-1}\mathbf{E}, (\mathbf{A} - s_k\mathbf{E})^{-1}\mathbf{B})$ and the output block Krylov subspace $\mathbf{K}_{q4}((\mathbf{A} - s_k\mathbf{E})^{-T}\mathbf{E}^T, (\mathbf{A} - s_k\mathbf{E})^{-T}\mathbf{C}^T)$. If the columns of the matrix \mathbf{V} form a basis of \mathbf{K}_{q1} and \mathbf{K}_{q3} , respectively, and the matrix \mathbf{W} is chosen such that $\mathbf{A}_r = \mathbf{W}^T\mathbf{A}\mathbf{V}$ is non-singular, then the first $\frac{q1}{m}$ and accordingly $\frac{q3}{m}$ moments match. A typical choice is $\mathbf{W} = \mathbf{V}$, named Galerkin (orthogonal) projection, otherwise $\mathbf{W} \neq \mathbf{V}$ Petrov-Galerkin (oblique) projection. Using only the input Krylov subspace is known as one-sided Krylov subspace method. Two-sided methods applying input and output Krylov subspaces match $\frac{q1}{m} + \frac{q2}{m}$ and accordingly $\frac{q3}{m} + \frac{q4}{m}$ moments.

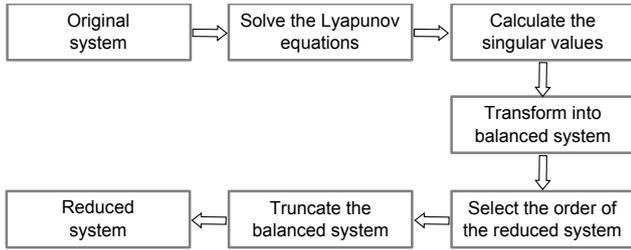
To calculate the desired projection matrices \mathbf{V} and \mathbf{W} , a wide range of numerical, mostly iterative algorithms is available. The most popular ones are Arnoldi algorithm, two-sided Arnoldi algorithm and Lanczos algorithm which are not further investigated in this paper. Details and further algorithms can be found in Bechtold et al. (2007), Meister (2005), Grimme (1997), Krohne (2007), Salimbahrami (2006), Antoulas et al. (2001), Eid (2009) and Bai (2002).

5 Control theory methods based on singular value decomposition

Methods using the singular value decomposition (SVD) aim for an approximation of matrices by lower ranked matrices, as carried out by Hackbusch (2009). The reduced order model (ROM) is found by deleting those system states that are both least controllable and observable. Figure 4 gives an overview of the key steps using singular value decomposition. In Table 3, the main advantages and disadvantages are listed for MOR using SVD methods.

Table 3. Characteristics of SVD methods.

Advantages	Disadvantages
– invariant to representation	– only for small systems
– applicable to nonlinear systems	– new methods may have convergence troubles
– comprising all relevant system parts	– computationally expensive (Lyapunov equations)
– existence of an error bound	
– preservation of stability	
– fully automatable	
– SVD supported in Matlab	

**Figure 4.** Key steps of dimensional reduction using singular value decomposition (SVD).

5.1 Controllability and observability of a system

Considering the system in Eq. (2) represented by

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \hat{\mathbf{A}}\mathbf{x}(t) + \hat{\mathbf{B}}\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t),\end{aligned}\quad (26)$$

the two important control theory quantities controllability \mathcal{P} and observability \mathcal{Q} can be introduced, as done by Liemann (2006), via the Gramians

$$\mathcal{P} = \int_0^{\infty} e^{\hat{\mathbf{A}}t} \hat{\mathbf{B}}\hat{\mathbf{B}}^T e^{\hat{\mathbf{A}}^T t} dt \quad (27)$$

and

$$\mathcal{Q} = \int_0^{\infty} e^{\hat{\mathbf{A}}^T t} \mathbf{C}^T \mathbf{C} e^{\hat{\mathbf{A}} t} dt. \quad (28)$$

The concept of controllability examines how the states $\mathbf{x}(t)$ of a system are linked to its inputs $\mathbf{u}(t)$ whereas observability deals with the connection between the states $\mathbf{x}(t)$ and its outputs $\mathbf{y}(t)$, as delineated by Eid (2009).

Under certain conditions, described e.g. by Bechtold et al. (2007) and Koutsovasilis (2009), the Gramians can be found by solving the two system related Lyapunov equations given by

$$\begin{aligned}\hat{\mathbf{A}}\mathcal{P} + \mathcal{P}\hat{\mathbf{A}}^T + \hat{\mathbf{B}}\hat{\mathbf{B}}^T &= 0, \\ \hat{\mathbf{A}}^T \mathcal{Q} + \mathcal{Q}\hat{\mathbf{A}} + \mathbf{C}^T \mathbf{C} &= 0.\end{aligned}\quad (29)$$

5.2 The singular value decomposition

As known from linear algebra and described by Gugel (2009) and Hackbusch (2009), any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be decomposed into the product of three matrices, more precisely an orthogonal matrix \mathbf{U} , a diagonal matrix $\mathbf{\Sigma}$ and the transpose of an orthogonal matrix \mathbf{V} such that

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (30)$$

where the matrix

$$\hat{\mathbf{\Sigma}} = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_k \end{bmatrix}, \mathbf{\Sigma} = \begin{bmatrix} \hat{\mathbf{\Sigma}} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times n} \quad (31)$$

comprises the singular values $\sigma_i = \sqrt{\lambda_i(\mathbf{A}^T \mathbf{A})} \geq \sigma_{i+1} \geq 0$ on its diagonal in descending order with $\text{rank}(\mathbf{A}) = k$ and $k = \min\{m, n\}$. The left singular vectors $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_m] \in \mathbb{R}^{m \times m}$ are the orthonormal vectors of $\mathbf{A}\mathbf{A}^T$, the right singular vectors $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n] \in \mathbb{R}^{n \times n}$ are the orthonormal vectors of $\mathbf{A}^T \mathbf{A}$. The term in Eq. (30) is called a singular value decomposition (SVD) of \mathbf{A} and provides a basis for finding a lower ranked matrix \mathbf{A}_r which best approximates \mathbf{A} such that

$$\mathbf{A}_r = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \quad (32)$$

The task is to find an optimal approximation in a certain system norm, eg. the spectral norm (\mathcal{H}_2 -norm), as specified by Antoulas and Sorensen (2001), Hackbusch (2009) and Dahmen and Reusken (2008), via the Schmidt-Mirsky theorem which results in the minimisation of the approximation error such that

$$\min_{\text{rank}(\mathbf{A}_r) \leq \text{rank}(\mathbf{A})} \|\mathbf{A} - \mathbf{A}_r\|_{\mathcal{H}_2} = \sigma_{k+1}(\mathbf{A}). \quad (33)$$

In the following, the most commonly used SVD procedures are specified, namely balanced truncation approximation, singular perturbation approximation and Hankel norm approximation. They distinguish in the rating of the approximation error and the procedure of its minimization, see Benner et al. (2004).

5.3 Balanced truncation approximation

In case of a stable system, balanced truncation approximation (BTA) assumes that the system is balanced first and afterwards truncated. Based on the system representation given in Eq. (26), the balanced partition of the system matrices and state vector can be quoted as

$$\begin{aligned} \begin{bmatrix} \dot{\mathbf{x}}_1(t) \\ \dot{\mathbf{x}}_2(t) \end{bmatrix} &= \begin{bmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} \\ \hat{\mathbf{A}}_{21} & \hat{\mathbf{A}}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{B}}_1 \\ \hat{\mathbf{B}}_2 \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) &= \begin{bmatrix} \hat{\mathbf{C}}_1 & \hat{\mathbf{C}}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} \end{aligned} \quad (34)$$

with $\hat{\mathbf{A}}_{11} \in \mathbb{R}^{r \times r}$, $\hat{\mathbf{B}}_1 \in \mathbb{R}^{r \times m}$ and $\hat{\mathbf{C}}_1 \in \mathbb{R}^{p \times r}$. At all times, a balancing transformation is feasible for minimal order systems. The McMillan degree of the system is specified as the order of any minimal state-space realization of the transfer-function matrix, as stated by Sasane (2002) and Baur and Benner (2008).

The singular values are the eigenvalues of the observability \mathcal{P} and controllability \mathcal{Q} Gramians: $\sigma_i = \sqrt{\lambda_i(\mathcal{P}\mathcal{Q})}$ which are called Hankel singular values (HSV) of Σ , e.g. described by Gugercin and Antoulas (2004) and Antoulas and Sorensen (2001). The Gramians are equal and diagonal. The balanced truncation approximation is predicted on transforming the state-space-system into a balanced realization. Those state variables that are least observable and controllable, and therefore related to the smallest Hankel singular values, are truncated with respect to the error bound between original and reduced transfer function

$$\|\mathbf{H}(s) - \mathbf{H}_r(s)\|_{\mathcal{H}_\infty} \leq 2 \sum_{r+1}^m \sigma_i. \quad (35)$$

The ROM obtained by BTA is related to

$$\begin{aligned} \dot{\mathbf{x}}_1(t) &= \hat{\mathbf{A}}_{11} \mathbf{x}_1(t) + \hat{\mathbf{B}}_1 \mathbf{u}(t), \\ \mathbf{y}(t) &= \hat{\mathbf{C}}_1 \mathbf{x}_1(t). \end{aligned} \quad (36)$$

as mentioned by Obinata and Anderson (2000). BTA involves an approximation error in the low-frequency region as mentioned by Bechtold et al. (2007), Obinata and Anderson (2000) and Baur and Benner (2008). For details and aspects in numerical calculation see e.g. Benner and Quintanaort'i (2004), Bechtold et al. (2007), Koutsovasilis (2009), Antoulas (2005) and Gugercin and Antoulas (2004).

5.4 Singular perturbation approximation

The singular perturbation approximation (SPA) is closely related to the balanced truncation approximation and ensures zero error at zero frequency and is applicable to nonlinear systems. Referring to the expression in Eq. (34) and as explained by Liu and Anderson (1989) and Antoulas et al. (2001) the reduced system is identified by

$$\begin{aligned} \dot{\mathbf{x}}_r(t) &= \tilde{\mathbf{A}} \mathbf{x}(t) + \tilde{\mathbf{B}} \mathbf{u}(t), \\ \mathbf{y}_r(t) &= \tilde{\mathbf{C}} \mathbf{x}(t), \end{aligned} \quad (37)$$

with the reduced system matrices $\tilde{\mathbf{A}} = \hat{\mathbf{A}}_{11} - \hat{\mathbf{A}}_{12} \hat{\mathbf{A}}_{22}^{-1} \hat{\mathbf{A}}_{21}$, $\tilde{\mathbf{B}} = \hat{\mathbf{B}}_1 - \hat{\mathbf{A}}_{12} \hat{\mathbf{A}}_{22}^{-1} \hat{\mathbf{B}}_2$ and $\tilde{\mathbf{C}} = \hat{\mathbf{C}}_1 - \hat{\mathbf{C}}_2 \hat{\mathbf{A}}_{22}^{-1} \hat{\mathbf{A}}_{21}$. In contrast to BTA, SPA has an approximation error at high frequencies, but it matches at low frequencies, see Obinata and Anderson (2000).

5.5 Hankel norm approximation

The Hankel norm approximation (HNA) employs the so-called Hankel norm which is defined as the maximal Hankel singular value of the system given in Eq. (26):

$$\|\mathbf{H}(s)\|_H = \sqrt{\lambda_{\max}(\mathcal{P}\mathcal{Q})} = \sigma_{\max}. \quad (38)$$

The aim is to minimize the approximation error between the original system transfer function $\mathbf{H}(s)$ and the reduced one $\mathbf{H}_r(s)$ with order r via the Hankel norm

$$\|\mathbf{H}(s) - \mathbf{H}_r(s)\|_{\mathcal{H}_2} = \sigma_{r+1}(\mathbf{H}(s)) \leq \|\mathbf{H}(s) - \tilde{\mathbf{H}}_r(s)\|_H, \quad (39)$$

as stated by Glover (1984), Green and Limebeer (1995), Bechtold et al. (2007), Benner et al. (2004), Antoulas (2005) and Antoulas and Sorensen (2001) where also the computation of the optimal Hankel norm approximation is performed. In Eq. (39) $\tilde{\mathbf{H}}_r(s)$ denotes all transfer functions of McMillan degree less than or equal to r .

6 Conclusion: benefits and aims for model order reduction in compliant mechanisms

In this work, system theoretical basics are reviewed which constitute the origin for model order reduction (MOR). Three different procedures, namely modal reduction, Krylov subspaces and singular value decomposition (SVD) based methods, are described with their associated characteristics.

The application of MOR can be motivated by challenges arising in the numerical characterisation of compliant mechanisms (CM) which can be traced back to the numerousness of ODEs required to specify their behavior in an accurate way. In most cases, large finite element models arise to capture a detailed representation of the geometrical domain and the dynamical performance of the CM. This circumstance is unfeasible, particularly with regard to further investigations such as simulation, optimization and control. For this reason, MOR in CMs seems to be a purposeful approach to generate efficient models and would benefit the design and development as well as analysis process of CMs.

The specific feature of CMs exists in their mechanical topology as they consist of rigid parts connected by flexure hinges. This characteristic has effects on the potentiality of MOR and requires further investigations to provide a suitable algorithm. Especially, the algorithm has to notably incorporate the distinct differences between the compliant and rigid parts of the structure with regard to their mechanical properties and influence on the systems transfer behavior. Moreover, having a small error between original and reduced

system implying a global error bound, the procedure should preserve the passivity as well as stability, has to be automatable, stable and computational efficient also for high order systems.

In future work, different MOR schemes applied to CM will be investigated and reasonable combinations among them will be analysed to benefit from their advantages. Furthermore, a specific adjustment and upgrading is desired. Out of this, the aim is to provide a reduction procedure meeting the special circumstances of CMs.

Acknowledgements. This work is supported by the German Research Foundation (DFG) as part of the research project “SPP 1476 – Small machine tools for small workpieces” which is gratefully acknowledged.

Edited by: J. A. Gallego Sánchez

Reviewed by: two anonymous referees

References

- Antoulas, A.: Approximation of large-scale dynamical systems, Society for Industrial Mathematics, 2005.
- Antoulas, A. and Sorensen, D.: Approximation of large-scale dynamical systems: An Overview, *Int. J. Appl. Math. Comput. Sci.*, 11, 1093–1121, 2001.
- Antoulas, A., Sorensen, D., and Gugercin, S.: A survey of model reduction methods for large-scale systems, in: *Structured matrices in mathematics, computer science, and engineering: proceedings of an AMS-IMS-SIAM joint summer research conference*, Vol. 280, p. 193, 2001.
- Bai, Z.: Templates for the solution of algebraic eigenvalue problems, Society for Industrial Mathematics, 2000.
- Bai, Z.: Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems, *Appl. Numer. Math.*, 43, 9–44, 2002.
- Baur, U. and Benner, P.: Gramian-Based Model Reduction for Data-Sparse Systems, *SIAM J. Sci. Comput.*, 31, 776–798, 2008.
- Bechtold, T., Rudnyi, E., and Korvink, J.: Fast simulation of electro-thermal MEMS: efficient dynamic compact models, Springer Verlag, 2007.
- Benner, P. and Quintana-ort'i, E. S.: G.: Parallel model reduction of large-scale linear descriptor systems via Balanced Truncation, in: *LECT NOTES COMPUT SC*, Proc. 6th Intl. Meeting VEC-PAR04, 28–30 June 2004, 65–78, 2004.
- Benner, P., Quintana-Orti, E., and Quintana-Orti, G.: Computing optimal Hankel norm approximations of large-scale systems, in: *Decision and Control*, 2004, CDC. 43rd IEEE Conference on, Vol. 3, 3078–3083, 2004.
- Bennini, F.: Ordnungsreduktion von elektrostatisch-mechanischen Finite Elemente Modellen für die Mikrosystemtechnik, Ph.D. thesis, Universitätsbibliothek der Technischen Universität, 2005.
- Craig, R. and Bampton, M.: Coupling of substructures for dynamic analysis, *AIAA journal*, 6, 1313–1319, 1968.
- Dahmen, W. and Reusken, A.: *Numerik für Ingenieure und Naturwissenschaftler*, Springer, 2008.
- Eid, R.: Time Domain Model Reduction by Moment Matching, Ph.D. thesis, TU München, 2009.
- Gasch, R. and Knothe, K.: *Strukturdynamik*, Bd. 2, Springer, New York, NY, USA, 1989.
- Glover, K.: All optimal Hankel-norm approximations of linear multivariable systems and their L-error bounds, *Int. J. Control.*, 39, 1115–1193, 1984.
- Green, M. and Limebeer, D.: *Linear robust control*, Prentice Hall, 1995.
- Grimme, E.: Krylov projection methods for model reduction, Ph.D. thesis, Citeseer, 1997.
- Gugel, D.: Ordnungsreduktion in der Mikrosystemtechnik, Ph.D. thesis, Universitätsbibliothek der TU Chemnitz, 2009.
- Gugercin, S. and Antoulas, A.: A survey of model reduction by balanced truncation and some new results, *Int. J. Control.*, 77, 748–766, 2004.
- Hackbusch, W.: *Hierarchische Matrizen*, Springer-Verlag Berlin Heidelberg, 2009.
- Kanzow, C.: *Numerik linearer Gleichungssysteme: direkte und iterative Verfahren*, Springer, 2005.
- Koutsovasilis, P.: Model Order Reduction in Structural Mechanics Coupling the Rigid and Elastic Multi Body Dynamics, Ph.D. thesis, TU Dresden, 2009.
- Koutsovasilis, P. and Beiteltschmidt, M.: Comparison of model reduction techniques for large mechanical systems, *Multibody Syst. Dyn.*, 20, 111–128, 2008.
- Krohne, K.: Order reduction of finite-volume models and its application to microwave device optimization, Ph.D. thesis, Swiss Federal Institute of Technology Zurich, 2007.
- Lehner, M. and Eberhard, P.: Model Reduction in Flexible Multibody Systems, *At.-Autom.*, 54, 170–177, 2006.
- Lienemann, J.: Complexity reduction techniques for advanced MEMS actuators simulation, Ph.D. thesis, Universitätsbibliothek Freiburg, 2006.
- Liu, Y. and Anderson, B.: Singular perturbation approximation of balanced systems, *Int. J. Control.*, 50, 1379–1405, 1989.
- Meister, A.: *Numerik linearer Gleichungssysteme: eine Einführung in moderne Verfahren*, Vieweg + Teubner Verlag, 2005.
- Obinata, G. and Anderson, B.: *Model reduction for control system design*, Springer Verlag, 2000.
- Rickelt-Rolf, C.: Modellreduktion und Substrukturtechnik zur effizienten Simulation dynamischer, teilgeschädigter Systeme, Ph.D. thesis, TU Braunschweig, 2009.
- Rudnyi, E. and Korvink, J.: Model order reduction for large scale engineering models developed in ANSYS, in: *Lect. Notes Comput. Sc.*, 349–356, Springer, 2006.
- Salimbahrami, B.: Structure preserving order reduction of large scale second order models, Ph.D. thesis, TU München, 2006.
- Sasane, A.: Hankel norm approximation for infinite-dimensional systems, Springer, 2002.
- Siedl, D.: Simulation des dynamischen Verhaltens von Werkzeugmaschinen während Verfahrenbewegungen, Herbert Utz Verlag, 2008.
- Stelzmann, U., Groth, C., and Müller, G.: *FEM für Praktiker-Band 2: Strukturdynamik*, Expert Verlag, 2002.
- Van der Vorst, H.: *Iterative Krylov methods for large linear systems*, Cambridge University Press, 2003.
- Werner, D.: *Funktionalanalysis*, Springer, 2009.
- Wulfsberg, J., Redlich, T., and Kohrs, P.: Square Foot Manufacturing: a new production concept for micro manufacturing, *Prod. Engineer.*, 4, 75–83, 2010.