Model Order Reduction for Coupled Systems using Low-rank Approximations

Picture at the back cover – courtesy of NASA

Model Order Reduction for Coupled Systems using Low-rank Approximations

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College
voor Promoties in het openbaar te verdedigen
op woensdag 21 maart 2012 om 16.00 uur

door

Agnieszka Lutowska

geboren te Zielona Góra, Polen

Dit proefschrift is goedgekeurd door de promotor:

prof.dr. W.H.A. Schilders

Copromotor:
dr. M.E. Hochstenbach

Contents

Contents

Chapter 1

Introduction

This thesis focuses on the development of a model reduction methodology for coupled multi-physical models to serve the efficient simulation-based design of the underlying coupled systems. Examples of coupled systems are larger systems such as magnetic resonance imaging (*MRI*) scanners, printers/copiers, precision motion stages, foldable solar panels of a space-telescope, down to very small systems such as very large scale integrated (*VLSI*) systems (see for instance [87] and [45]) and microelectromechanical systems (*MEMS*) (see for instance [74]). Figure 1.1 shows such examples. To explain the idea of *reduction* of such systems later on, we first explain the terms "system", "model", "multi-physics", "coupled", "interconnected", and "discretization".

The word *system*, which originates from the Greek word $\sigma\acute{\upsilon}\sigma\tau\eta\mu\alpha$ and the Latin word *sustema*, stands for "a set of interacting or interdependent components forming an integrated whole". In this thesis, the integrated whole is called the *system* or *coupled system* and its individual components are called *sub-systems*.

The word *model* as in "physical model" stands for a "representation" for the system under consideration, usually in terms of a set of physical quantities and relations. The Oxford dictionary explains that a model is a noun which can mean "a simplified description, especially a mathematical one, of a system or process, to assist calculations and predictions". A coupled system's model consists of the coupled sub-systems' models.
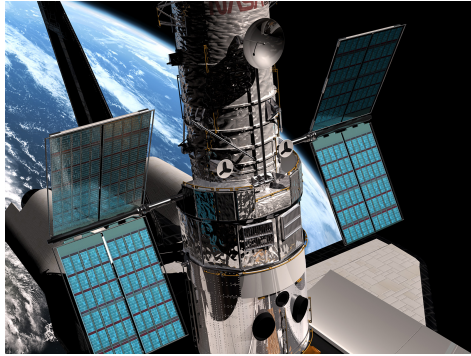
(a) Foldable solar panels (courtesy ESA)   (b) A MEMS comb drive
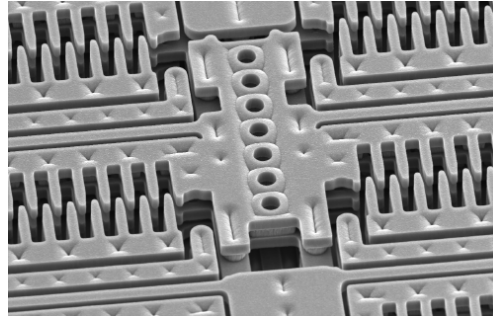
Figure 1.1: Coupled systems

A *multi-physical model* is a model which is represented by multiple physical quantities such as temperature, structural mechanical displacements ([32]), electro-magnetic fields, and so forth. Simple systems in an insulated environment can often be described with few physical quantities and relations, interacting systems frequently require more of such quantities and relations.

This thesis is about sub-systems which *interact*. When the interaction takes place inside a domain of interest or through the boundary which seperates such a domain of interest from the outside world such a system is called a *coupled system*. If the physical quantities interact through a discrete amount of *inputs* and *outputs* in space then the system is said to be an *interconnected system* (see for instance [89]) rather than a *coupled system*.

To explain the envisioned reduction, first note that most physical models can not be solved exactly with contemporary computers. To calculate an approximate solution, the involved physical quantities such as an electromagnetic field are first discretized, i.e., represented by a finite amount of *degrees of freedom*, after which the physical equations are reformulated for the discretized physical quantities, leading to a *discrete system* of equations. This process is called *discretization* of the model. An accurate representation of physical quantities such as an electromagnetic field can require millions of degrees of freedom and consume a considerable amount of data storage and computation time. Therefore, an analysis of a coupled system's dynamic behavior can require excessive amounts of data storage and computation time.

In this thesis, the concept of *model reduction* stands for *a reduction of the size of the related discrete system* which directly leads to *a reduction of the amount of degrees of freedom* of the related physical quantities – but the new degrees of freedom are perhaps no longer related to just one physical entity. We do not reduce the amount of physical quantities nor do we simplify or reduce the amount of the relations. To indicate that

we reduce the *size* of the discrete system, we write *model order reduction* rather than model reduction. We aim at the reduction of the discrete system in such a manner that we (approximately) preserve the *input-output behavior* which characterizes the coupled system.

We focus on state-of-the-art reduction techniques which *reduce the system as a whole based on available reduction techniques for the individual sub-systems.* Such methods are scarcely available and mostly in development. They have as an advantage that the individual sub-systems can be reduced in parallel (see [9]) with the method best suited for each of them. This can save a considerable amount of data storage and computational time since these systems are also smaller than the system as a whole. On the other hand, one must figure out how to couple the individually reduced models to a reduced model for the whole, i.e., need to figure out how to effectively deal with interior couplings/interconnections.

Our reduction methods are primarily for coupled time-invariant linear models. Time-dependent linear models, *affine models* (such as presented in [19]) and *non-linear models* (see for instance [63] and [88]) require other than the presented reduction techniques. Furthermore, we restrict ourselves to Krylov subspace projection techniques (see [42]).

More in detail, without loss of generality, we focus at systems which consist of two coupled subsystems. We suggest a method for the parallel reduction of the individual sub-systems, call it the *Separate Bases Reduction* algorithm (SBR), and show how to create a reduced model for the whole system based on the reduced parts. Furthermore, we show that this algorithm applied to coupled systems matches at least the same amount of moments as a standard method applied to the whole system would (see [89] for interconnected systems). We establish that a large amount of internal couplings leads to large and hence undesirable reduced models and show that this can be overcome with the use of a generalized singular value decomposition (GSVD) based reduction of the coupling blocks However, the use of a GSVD-based approximation leads to an approximation of the moments – which as benchmark examples show can still be quite accurate.

The remainder of this thesis is focused on the presentation of the SBR algorithm and the GSVD reduction of the internal couplings. It is organized as follows. First, Chapter 2, taken from [8] presents a comparison of model reduction techniques from structural dynamics, numerical mathematics and systems and control. An overview of the recent history is provided and *Krylov methods* in Section 2.2.2 are compared with *balanced truncation* in Section 2.2.3 and *mode displacement methods* in Section 2.2.1.

Next, Chapter 3 provides the basic mathematical tools for standard Krylov subspace reduction techniques existing in literature. First the time-invariant linear system and the related reduced systems are introduced, together with their transfer functions which describe the *input-output behavior*, their *moments* (scaled derivatives), and the Krylov subspace projection technique. Secondly, it is shown that the transfer functions and some of their moments are identical at exactly those *expansion points* used for the creation of

the Krylov subspace. Thus, it is established that a reduction of the discrete model does not lead to different input-output behavior at a discrete (small) amount of expansion points – more of such points invariantly leads to larger reduced models.

Chapter 4 continues with Krylov subspace techniques, but now focuses on *coupled* and *interconnected* time-invariant linear systems. First, it shows what happens if the standard techniques of the previous chapter would be applied to the coupled system as a whole – it shows that the *block structure* is lost. Next, it introduces existing techniques from the literature such as [25, 5, 39], still based on Krylov subspace methods for the coupled system as a whole, which preserve the block-structure and the number of matched moments. At the end of this chapter, we show an alternative method to efficiently calculate the second Krylov projector and extend the proof of [25] to a more general case, under assumptions.

In Chapter 5 we assume that Krylov subspace reduction methods are already available for the individual sub-systems and based thereon, we focus at the construction of a reduced-order model for the system as a whole. We show that this is possible (and also that moments are matched) in Theorem 5.6.1 and call the approach the Separate Bases Reduction algorithm (SBR). We conclude with examples and discuss for which systems this new algorithm is advantageous.

In Chapter 6 we show that the SBR algorithm also matches the standard double amount of moments if one uses two Krylov subspace projectors instead of one.

Since our SBR method can create large reduced models if the sub-systems have many internal inputs and outputs Chapter 7 focuses on the possible reduction of these internal connections – it focuses on the coupling blocks. We show that one can approximate the coupling blocks by a generalized singular value decomposition (*GSVD*) in a manner that preserves sparsity (Corollary 7.2.1). Then, by means of example, we show that one can approximate the coupling blocks with the use of a few most dominant modes of the GSVD. Effectively, this reduces the amount of internal inputs and outputs and therefore ensures that the SBR algorithm is efficient for systems composed of sub-systems with many inputs.

In Chapter 8 we show that the replacement of the coupling blocks by an explicitly rank-revealing GSVD based components leads to the same Krylov subspaces and hence matched moments. Approximations based on a few of the dominant modes lead to quite accurate moment approximation.

Finally, in Chapter 9 we apply the SBR algorithm to a benchmark system. The system under consideration is scaled in a specific manner such that it is numerically better conditioned. We conclude with some remarks and recommendations for further research in Chapter 10.

The STW project

The research presented in this thesis has been conducted within the project *Model reduction for complex high-tech systems: efficient analysis of active multi-physical systems* and is financially supported by the Dutch technology foundation STW under project number 07788.

The objective of this project is the development of model reduction procedures for coupled systems, where the coupling results from the interaction of the dynamics in multiple physical domains. This thesis addresses the mathematical aspect of coupled problem's reduction, i.e., analysis of the underlying linear systems and extension of the Krylov subspace based reduction techniques.

A second objective of the STW project is to bridge the conceptual gap between existing model reduction techniques in the fields of structural dynamics, numerical mathematics and systems and control. A collaborative effort of the members of the project team has led to a detailed overview and comparison of these methods, which is presented in Chapter 2.

Chapter 2

A comparison of model order reduction techniques

This chapter contains the paper "A comparison of model reduction techniques from structural dynamics, numerical mathematics and systems and control", by B. Besselink, A. Lutowska, U. Tabak, N. van de Wouw, M. E. Hochstenbach, H. Nijmeijer, D.J. Rixen and W.H.A. Schilders. The paper is a common work of the researchers involved in the STW project described in Chapter 1 and gives an overview and comparison of the model order reduction methods used in the three fields of expertise of the project sub-groups. A more detailed overview and a theoretical introduction to model reduction techniques based on projection onto Krylov subspaces, which are the topic of this thesis, the reader can find in Chapters 3 and 4.

2.1   Introduction

An important tool in the design of complex high-tech systems is the numerical simulation of predictive models. However, these dynamical models are typically of high order, i.e.

they are described by a large number of ordinary differential equations. This results from either the inherent complexity of the system or the discretization of partial differential equations. Model reduction can be used to find a low-order model that approximates the behavior of the original high-order model, where this low-order approximation facilitates both the computationally efficient analysis and controller design for the system to induce desired behavior.

The earliest methods for model reduction belong to the field of structural dynamics, where the dynamic analysis of structures is of interest. Typical objectives are the identification of eigenfrequencies or the computation of frequency response functions. Besides the mode displacement reduction method and extensions thereof (see e.g. [67, 31]), important techniques are given by component mode synthesis techniques [44, 13], which started to emerge in the 1960s.

The model reduction problem has also been studies in the systems and control community, where the analysis of dynamic systems and the design of feedback controllers are of interest. The most important contributions were made in the 1980s by the development of balanced truncation [57, 20] and optimal Hankel norm approximation [34].

Finally, numerically efficient methods for model reduction have been developed in the field of numerical mathematics in the 1990s. Important techniques are asymptotic waveform evaluation [66], Padé-via-Lanczos [21] and rational interpolation [36]. These methods are often applied in the design and analysis of large electronic circuits.

Despite the fact that the above techniques essentially deal with the same problem of model reduction, the results in the fields of structural dynamics, systems and control and numerical mathematics have largely been developed independently. This paper aims at providing a thorough comparison between the model reduction techniques from these three fields, facilitating the choice of a suitable reduction procedure for a given reduction problem. To this end, the most popular methods from the fields of structural dynamics, systems and control and mathematics will be reviewed. Then, the properties of these techniques will be compared, where both theoretical and numerical aspects will be discussed. In addition, these differences and commonalities will be illustrated by means of application of the model reduction techniques to a common example.

Reviews of model reduction techniques exist in literature. However, these reviews mainly focus on methods from the individual fields, i.e. they focus on methods from structural dynamics [12, 14], systems and control [38] or numerical mathematics [5, 27] only. Nonetheless, methods from systems and control and numerical mathematics are reviewed and compared in [3, 4, 30], where the comparison is mainly performed by the application of the methods to examples. In the current paper, popular model reduction techniques from all the three fields mentioned above will be reviewed. Additionally, both a qualitative and quantitative comparison will be provided. The focus of this paper is on this comparison; it does not aim at presenting a full comprehensive historical review of all method in these three domains.

In this paper, the scope will be limited to *model-based* reduction techniques for *linear time-invariant systems*. Data-based model reduction procedures such as proper orthogonal decomposition [83, 7] or methods for nonlinear systems (see e.g. [69, 81]) will not be discussed.

The outline of this paper is as follows. First, the most important model reduction techniques from the fields of structural dynamics, numerical mathematics and systems and control will be reviewed in Section 2.2. In Section 2.3, a qualitative comparison between these methods will be provided, focussing on both theoretical and numerical aspects. This comparison will be illustrated by means of examples in Section 2.4, which further clarifies differences and commonalities between methods. Finally, conclusions will be stated in Section 2.5.

Notation   The field of all real numbers is denoted by $\mathbb{R}$, whereas $\mathbb{C}$ represents the field of all complex numbers. Boldface letters are used to represent vectors and matrices, where the latter are printed in upper case. For a vector $\boldsymbol{x}$, the *Euclidian norm* is denoted by $\|\boldsymbol{x}\|$, i.e. $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$. The $\mathcal{H}_\infty$ *norm* of a system is denoted by $\|\cdot\|_\infty$.

2.2   Review of model reduction techniques

In this section, popular model reduction techniques from different fields are discussed. In Section 2.2.1, methods from structural dynamics are discussed, whereas model reduction techniques from the fields of numerical mathematics and systems and control are discussed in Sections 2.2.2 and 2.2.3, respectively.

2.2.1   Mode displacement methods

In the field of structural dynamics, the design and performance evaluation of mechanical systems is of interest. Herein, the computation of deformations, internal stresses or dynamic properties are subject of analysis. Even though the goal of analysis might differ from one specific application to another, important objectives are the prediction of regions with high stress, prediction of the eigenvalues (related to resonance frequencies) and eigenvectors (related to structural eigenmodes) and the computation of the system's response to a certain excitation.

All of the above mentioned goals share a common property. Namely, the models used in the design must, generally, contain detailed information for the precise description of the response properties of the structure. The mathematical models are basically constructed in terms of partial differential equations. These equations might be solved exactly only for simple problems and one has to resort to discretization-based

approaches, such as the finite element method (FEM) or the boundary element method (BEM). In the context of this section, the finite element related concepts are of interest.

Model reduction methodologies are efficiently used in the structural field since the 19th century. The most common methods are mode superposition methods [67], in which a limited number of free vibration modes of the structure is used to represent the displacement pattern [11]. There are also improvements of the original mode super-position method by the addition of different vectors to the expansion procedure, such as the mode acceleration or modal truncation augmentation [67, 92]. Mode superposition methods are generally considered for the complete structure. However, it is common to partition the structure in some components, on which model reduction is performed individually. Then, these reduced-order component models are coupled to represent the global behavior. These methods are all together named component mode synthesis techniques. These methods are extensions of common mode superposition methods to the partition level where the forces on the partition boundaries replace the general forces on the whole structure. In [43] and [44], Hurty provided a general method for component mode synthesis techniques. Craig & Bampton, in [13], used the static deformation shapes of the structure with respect to its boundary loads and enriched this space with the internal dynamic mode shape vectors to increase the accuracy. This method is known as the *fixed-interface reduction method* because the modes of the system are found while all the boundaries are fixed. Later on, the works of MacNeal [54] and Rubin [76] extended these methods to a class of methods known as *free-interface methods*. In these methods, the dynamic mode shape vectors used in the basis are computed without the application of any restraints on the component boundaries, where in fixed-interface methods the boundary degrees of freedom are all fixed. A recent general overview on dynamic substructuring methods can be found in [14]. Another overview that summarizes the component mode synthesis approaches can be found in [12].

Discretization-based methods, such as FEM, analyze complex engineering problems by constructing piece-wise approximation polynomials over the spatial domain and solve for the unknown variables at specific locations of the discretization, known as node points [65]. This representation might already be considered to be a model order reduction process in itself. Namely, the displacement $u(z, t)$, which is dependent on the spatial variable z and time t, is represented by the finite expansion

$$u(z, t) = \sum_{j=1}^{N} \Psi_j(z) q_j(t). \qquad (2.2.1)$$

Herein, $\Psi_j(z)$ are linearly independent functions representing the displacement shape of the structure, where it is noted that they satisfy the essential boundary conditions of the problem. Next, $q_j(t)$ are the unknown functions of time, whereas N represents the number of functions exploited in the representation. Since the representation of a body consists of infinitely many points (and therefore infinitely many degrees of freedom), the

finite expansion (2.2.1) has already accomplished the task of reducing the system to a finite number of degrees of freedom.

The discretization of the differential equations of the problem results in the equations of motion of the system, which are typically of the following form:

$$\mathbf{M}\ddot{q} + \mathbf{K}q = f, \tag{2.2.2}$$

where $\mathbf{M} \in \mathbb{R}^{N \times N}$ and $\mathbf{K} \in \mathbb{R}^{N \times N}$ represent the *mass and stiffness matrices*, respectively. Furthermore, $q \in \mathbb{R}^{N}$ represents the unknown displacements of the structure and $f \in \mathbb{R}^{N}$ is the externally applied generalized force vector. Structural systems possess, most of the time, light damping and the reduction typically is based on the undamped system. Therefore, undamped systems of the form (2.2.2) are considered in this section. However, it is stressed that this is only suitable when the system is lightly damped and the eigenfrequencies are well separated [31].

In general, a detailed problem representation and the use of a high number of elements in the discretization result in large matrices and, hence, in long computation times. Model order reduction methods are used to efficiently reduce the system size and, as a consequence, achieve acceptable computation times. Reduction methods in structural dynamics may be classified into two classes, namely, methods related to *mode superposition* and methods related to component *mode synthesis* techniques. In this section, the context is limited to mode superposition methods, since they apply to the full system. This enables a comparison with methods from the fields of numerical mathematics and systems and control. More information on component mode synthesis can be found in [13].

Mode superposition methods share the common property that they use a small number of free vibration modes to represent the dynamics of the structure with some reduced number of generalized degrees of freedom. With this selection, one represents the solution vectors as a summation of free vibration modes that form a linearly independent set. This operation therefore reduces the system size to be solved and could result in important computational gains. However, there are some important points to note on the expansion procedures used in practice [62], namely:

1. the used mode shape vectors do not span the complete space;

2. the computation of eigenvectors for large systems is very expensive and time consuming;

3. the number of eigenmodes required for satisfactory accuracy is difficult to estimate a priori, which limits the automatic selection of eigenmodes;

4. the eigenbasis ignores important information related to the specific loading characteristics such that the computed eigenvectors can be nearly orthogonal to the applied loading and therefore do not participate significantly in the solution.

Three different main variants can be considered which are often used in structural dynamics community. These are the mode displacement method, mode acceleration method and modal truncation augmentation method. The latter two methods are enhancements of the mode displacement method with the addition of the contribution of the omitted parts in an expansion process.

Generally, these methods do not propose the computation of an error bound for the response studies. Consequently, the success of the methods is established on the basis of a posteriori error comparisons. Typically, either the errors on the eigenfrequencies or the errors on the input-output representation is used to show the success of the applied method. In the following sections, the mode displacement method, the mode acceleration methods and the modal truncation augmentation method will be treated in more detail.

Mode displacement method

The equation of motion of the structure (2.2.2) is recalled:

$$\mathbf{M}\ddot{\boldsymbol{q}} + \mathbf{K}\boldsymbol{q} = \boldsymbol{f}.$$

Then, the mode displacement method is based on the free vibration modes of the structure, which can be found by using a time-harmonic representation for the displacement of the unforced system (i.e. $\boldsymbol{f} = \mathbf{0}$). This leads to the generalized eigenvalue problem

$$\left(\mathbf{K} - \omega_j^2 \mathbf{M}\right)\boldsymbol{\phi}_j = \mathbf{0}, \tag{2.2.3}$$

where $\boldsymbol{\phi}_j$ is the mode shape vector corresponding to the *eigenfrequency* ω_j, with $j \in \{1,\dots,N\}$. Using the expansion concept along with the mode shape vectors $\boldsymbol{\phi}_j$, the displacement can be represented as follows:

$$\boldsymbol{q} = \sum_{j=1}^{N} \boldsymbol{\phi}_j \eta_j, \tag{2.2.4}$$

where it is recalled that N is the size of the system. Here, η_j is typically referred to as a set of modal coordinates. It is a common practice to mass-orthogonalize the mode shape vectors, resulting in

$$\boldsymbol{\phi}_i^{\mathrm{T}} \mathbf{M} \boldsymbol{\phi}_j = \delta_{ij},$$
$$\boldsymbol{\phi}_i^{\mathrm{T}} \mathbf{K} \boldsymbol{\phi}_j = \delta_{ij}\omega_j^2, \tag{2.2.5}$$

where δ_{ij} denotes the *Kronecker delta*. These orthogonality relations are used to decouple the coupled equations of motion (2.2.2). Using (2.2.5), the decoupled equations are represented in modal coordinates as

$$\ddot{\eta}_j + \omega_j^2 \eta_j = \boldsymbol{\phi}_j^{\mathrm{T}} \boldsymbol{f}, \quad j \in \{1,\dots,N\}. \tag{2.2.6}$$

An important practical point on the expansion method is related to the computation of the expansion vectors. The computation of the mode shape vectors that are used in the mode superposition methods can be an expensive task and, in practice, all the computational methods extract a limited number of vectors K of the eigenvalue problem. The general idea of the expansion procedure is to keep the first K vectors in the representation, that correspond to the lowest eigenfrequencies. This results in a *truncation*, namely,

$$q = \sum_{j=1}^{K} \phi_j \eta_j + \underbrace{\sum_{j_t=K+1}^{N} \phi_{j_t} \eta_{j_t}}_{truncated}, \qquad (2.2.7)$$

where the indices j and j_t represent the kept mode and the truncated mode indices, respectively.

Since the displacement is represented as a linear combination of K linearly independent vectors, it can also be given in matrix notation, leading to the approximation

$$q = \Phi \eta, \qquad \Phi = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_K \end{bmatrix}. \qquad (2.2.8)$$

Using (2.2.2) and (2.2.8) and projecting the resulting equations of motion on the expansion basis Φ results in the following reduced-order dynamics

$$\mathbf{M}_r \ddot{\eta} + \mathbf{K}_r \eta = f_r, \qquad (2.2.9)$$

where

$$\mathbf{M}_r = \Phi^{\mathrm{T}} \mathbf{M} \Phi = \mathbf{I}, \qquad (2.2.10)$$

$$\mathbf{K}_r = \Phi^{\mathrm{T}} \mathbf{K} \Phi = \mathrm{diag}\{\omega_1^2, \dots, \omega_K^2\}, \qquad (2.2.11)$$

$$f_r = \Phi^{\mathrm{T}} f. \qquad (2.2.12)$$

In general, the analysts are interested in the response properties of the system for the lower frequency range and therefore, the lowest modes are typically chosen. The reason behind this selection is the fact that most structures are operated at low frequencies.

The importance of a mode is mostly related to two concepts. First, the orthogonality of the mode with respect to the excitation, as given by $\phi_j^{\mathrm{T}} f$, is of importance. Secondly, the closeness of the eigenfrequency of the mode with respect to the excitation spectrum is of interest.

Mode acceleration method

The *mode acceleration* method is a computational variant of the static correction method. The static correction method aims at taking into account the contribution of the omitted

modes. The driving idea of the static correction concept is to be able to include the effects of the truncated modes statically into the summation procedure. Namely, truncated modes have a static contribution on the response for low frequencies. This results in an improvement for the response studies in the lower frequency range. The response might be represented as before but with a correction term $\boldsymbol{q}_{cor}$, such that

$$\boldsymbol{q} = \boldsymbol{\Phi}\boldsymbol{\eta} + \boldsymbol{q}_{cor}. \qquad (2.2.13)$$

To obtain the static correction term $\boldsymbol{q}_{cor}$ (with $\dot{\boldsymbol{q}}_{cor} = \ddot{\boldsymbol{q}}_{cor} = \boldsymbol{0}$), the truncated representation for the acceleration is substituted in the equation of motion (2.2.2), leading to

$$\mathbf{M} \sum_{j=1}^{K} \boldsymbol{\phi}_j \ddot{\eta}_j + \mathbf{K}\boldsymbol{q} = \boldsymbol{f}. \qquad (2.2.14)$$

Then, the use of the (reduced-order) dynamics in modal coordinates (2.2.6) leads to

$$
\begin{aligned}
\boldsymbol{q} &= \mathbf{K}^{-1}\left(\boldsymbol{f} - \mathbf{M}\sum_{j=1}^{K}\boldsymbol{\phi}_j(\boldsymbol{\phi}_j^{\mathrm{T}}\boldsymbol{f} - \omega_j^2\eta_j)\right), \\
&= \sum_{j=1}^{K}\boldsymbol{\phi}_j\eta_j + \left(\mathbf{K}^{-1} - \sum_{j=1}^{K}\frac{\boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}}}{\omega_j^2}\right)\boldsymbol{f},
\end{aligned}
\qquad (2.2.15)
$$

where the relation imposed by the eigenvalue problem (2.2.3) is used in the latter step. When comparing (2.2.15) to the truncation (2.2.7), it is observed that the correction term is given as

$$\boldsymbol{q}_{cor} = \left(\mathbf{K}^{-1} - \sum_{j=1}^{K}\frac{\boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}}}{\omega_j^2}\right)\boldsymbol{f}. \qquad (2.2.16)$$

It is noted that, by using all eigenmodes, the inverse of the stiffness matrix can be represented as [31]

$$\mathbf{K}^{-1} = \sum_{j=1}^{N}\frac{\boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}}}{\omega_j^2}, \qquad (2.2.17)$$

such that the use of (2.2.17) in (2.2.15) results in

$$\boldsymbol{q} = \sum_{j=1}^{K}\boldsymbol{\phi}_j\eta_j + \sum_{j=K+1}^{N}\frac{\boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}}}{\omega_j^2}\boldsymbol{f}. \qquad (2.2.18)$$

Even though this last form is not applicable in practice, since it requires the computation of all model vectors, it clearly shows that only the static contribution of the omitted modes $\boldsymbol{\phi}_j$, $j \in \{K+1, \ldots, N\}$, is taken into account in the correction term $\boldsymbol{q}_{cor}$

Modal truncation augmentation method

The modal truncation augmentation method is an extension of the mode acceleration method. Its main principle depends on the use of the static correction as an additional direction for the truncation expansion [17, 72, 16]. Inclusion of the correction in a modal expansion results in the modal truncation augmentation method, such that q is approximated as

$$q = \sum_{j=1}^{K} \phi_j \eta_j + q_{cor}\xi, \tag{2.2.19}$$

where q_{cor} is given by the mode acceleration method in (2.2.18) and ξ is an additional coordinate in the reduced-order system. This correction vector is included in the reduction basis, such that the new reduction basis reads

$$\Psi = \begin{bmatrix} \Phi & q_{cor} \end{bmatrix}. \tag{2.2.20}$$

Here, it is noted that Ψ is generally orthogonalized.

Modal truncation augmentation methods are mostly used when there are multiple forcing vectors acting on the system. Therefore, these correction vectors are not used a posteriori as in the mode acceleration method but they really become a part of the reduction space.

There exist also further extensions of the common mode superposition methods which include higher-order correction vectors. These methods are outlined in [72] and references therein. See Section 2.3.3 for further details.

2.2.2 Krylov subspace based model order reduction

Krylov subspace based model order reduction (MOR) methods are methods which reduce a system with many degrees of freedom (i.e. states) to a system with few(er) degrees of freedom but with similar *input-output behavior*. Typical applications are large electronic circuits with large linear subnetworks of components (see e.g. [61, 28]) and micro-electro-mechanical systems (MEMS). The main purpose of Krylov methods is the construction of an approximation of the system's transfer function which (accurately enough) describes the dependence between the input and the output of the original system, e.g. in some range of the (input) frequency domain. Methods of this type are based on projections onto a Krylov subspace and are (relatively) computationally cheap compared to other reduction techniques, for instance because they can effectively exploit parallel computing. The objective is the derivation of a smaller system with similar input-output behavior and with similar properties such as stability, passivity or a special structure of the matrices in the model description. The quality of the reduced-order approximation can be assessed by studying norms of the difference between the outputs of

the unreduced and reduced models applied for the same inputs. Preservation of properties is of importance if the reduced system has to exhibit some physical properties of the model; for instance, when the reduced system has to be a (realizable) circuit consisting out of resistors, inductors and capacitors (a RLC network), just as the original system. So far, there have been no proven a priori error-bounds for the Krylov based reduction techniques, see [41] for more details and for alternative approaches, to ensure a good (application domain dependent) approximation.

The first reduction method involving the usage of the Krylov subspace, called asymptotic waveform evaluation, was described in 1990, see [66]. However, the main focus of this paper was on finding a *Padé approximation* of the transfer function rather than on the construction of a Krylov subspace. Later, in 1995, in [21] a method called Padé via Lanczos (PVL) was proposed and the relation between the Padé approximation and Krylov subspace was shown. In 1998 the new reduction technique PRIMA was introduced in [61]. It uses the Arnoldi algorithm instead of Lanczos to build the reduction bases. These and later developments of Krylov based reduction techniques focus not only on the improvement of the accuracy of the approximation, but also on the preservation of the properties of the system to be reduced.

In this section, the basic ideas of model reduction by projection onto the Krylov subspace are explained and the application of some common reduction techniques based on Arnoldi and Lanczos algorithms (see e.g. [77] for more details), is briefly discussed.

Linear time-invariant state-space systems of the form

$$\begin{cases} \mathbf{E}\dot{x} = \mathbf{A}x + \boldsymbol{b}u \\ \quad y = \boldsymbol{c}^{\mathrm{T}}x, \end{cases} \tag{2.2.21}$$

are considered, with $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$, the input variable $u \in \mathbb{R}$, the output variable $y \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$ being a vector of the state variables. For the sake of simplicity, *SISO* systems (with scalar input and scalar output) are considered. However, the methods discussed in this section have been extended to multi-input-multi-output (*MIMO*) cases (see e.g. [26]).

If the system (2.2.21) is transformed to the Laplace domain, then, for an arbitrary $s \in \mathbb{C}$, the dependence between its input and its output is given by a transfer function $H(s)$ defined as follows

$$H(s) = \boldsymbol{c}^{\mathrm{T}}(s\mathbf{E} - \mathbf{A})^{-1}\boldsymbol{b}. \tag{2.2.22}$$

In this section, it is assumed that the pencil $(s\mathbf{E} - \mathbf{A})$ is regular, i.e. it is singular only for a finite number of $s \in \mathbb{C}$. For an arbitrary $s_0 \in \mathbb{C}$, the transfer function (2.2.22) may be rewritten in a polynomial form, using so-called *moment* expansion:

$$H(s) = \sum_{n=0}^{\infty} (-1)^n M_n(s_0)(s - s_0)^n. \tag{2.2.23}$$

Here, the coefficients $M_n(s_0)$, called moments of the transfer function, are calculated using the Taylor expansion formula and given by

$$M_n(s_0) = \boldsymbol{c}^{\mathrm{T}}[(s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}]^n(s_0\mathbf{E} - \mathbf{A})^{-1}\boldsymbol{b}. \qquad (2.2.24)$$

Expansion around $s_0 = \infty$ is evaluated based on Laurent series, and the moments then are called Markov parameters (see [3] for more details). The accuracy of the moment expansion depends on the choice of the expansion point s_0. It is also possible to use a multipoint expansion choosing multiple expansion points.

The goal of the Krylov subspace model order reduction is to find a projection-based approximation of the original transfer function, that matches the first k moments of the original transfer function. In other words, the objective is to calculate the reduced-order transfer function $\hat{H}(s)$, whose moment expansion is given by

$$\hat{H}(s) = \sum_{n=0}^{\infty}(-1)^n \hat{M}_n(s_0)(s - s_0)^n, \qquad (2.2.25)$$

with

$$\hat{M}_n(s_0) = M_n(s_o), \qquad \text{for} \quad n = 1, \ldots, k, \qquad (2.2.26)$$

and $M_n(s_0)$ being the moments of the original transfer function defined in (2.2.24). This is called the moment matching property of the reduction method.

In case of the reduction methods studied in this section, the reduced-order model is calculated using a projection $\boldsymbol{\Pi} = \mathbf{V}\mathbf{W}^{\mathrm{T}} \in \mathbb{R}^{n \times n}$, with $\mathbf{V}, \mathbf{W} \in \mathbb{R}^{n \times k}$ being biorthogonal matrices, i.e. $\mathbf{W}^{\mathrm{T}}\mathbf{V} = \mathbf{I}$. Application of the projection $\boldsymbol{\Pi}$ to the original system (2.2.21) gives

$$\begin{cases} \mathbf{W}^{\mathrm{T}}\mathbf{E}\mathbf{V}\dot{\hat{\boldsymbol{x}}} = \mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V}\hat{\boldsymbol{x}} + \mathbf{W}^{\mathrm{T}}\boldsymbol{b}u, \\ \hat{y} = \boldsymbol{c}^{\mathrm{T}}\mathbf{V}\hat{\boldsymbol{x}}, \end{cases} \qquad (2.2.27)$$

where the reduced-order state vector $\hat{\boldsymbol{x}} \in \mathbb{R}^k$ results from the state transformation

$$\boldsymbol{x} \approx \mathbf{V}\hat{\boldsymbol{x}}. \qquad (2.2.28)$$

The choice of the spaces $\mathbf{V}$ and $\mathbf{W}$ depend on the goal of the reduction procedure. In case of the Krylov subspace based methods, the aim is to approximate the input-output behavior of the system. This is done by matching the moments of the original transfer function. This means that the reduced-order transfer function corresponding to system (2.2.27), which results from applying matrices $\mathbf{V}$ and $\mathbf{W}$ to the original system matrices, has the property (2.2.26). To ensure the satisfaction of the moment matching property (2.2.26), one can choose $\mathbf{V}$ and $\mathbf{W}$ such that the columns of these matrices span so-called Krylov subspaces. The k-th Krylov subspace induced by a matrix $\mathbf{P}$ and a vector $\boldsymbol{r}$ is defined as

$$\mathcal{K}_k(\mathbf{P}, \boldsymbol{r}) = \mathrm{span}\{\boldsymbol{r}, \mathbf{P}\boldsymbol{r}, \ldots, \mathbf{P}^{k-1}\boldsymbol{r}\}. \qquad (2.2.29)$$

The choice of the starting matrix $\mathbf{P}$ and the starting vector $\boldsymbol{r}$ depends on the value s_0 around which the transfer function should be approximated. If the approximation of the transfer function (2.2.22) around $s_0 = 0$ is to be found, the matrices $\mathbf{V}$ and $\mathbf{W}$ are chosen as follows:

$$\mathbf{V} \text{ is a basis of } \mathcal{K}_{k_1}(\mathbf{A}^{-1}\mathbf{E}, \mathbf{A}^{-1}\boldsymbol{b}), \qquad (2.2.30)$$

$$\mathbf{W} \text{ is a basis of } \mathcal{K}_{k_2}(\mathbf{A}^{-\mathrm{T}}\mathbf{E}^{\mathrm{T}}, \mathbf{A}^{-\mathrm{T}}\boldsymbol{c}). \qquad (2.2.31)$$

The sizes of the subspaces, k_1 and k_2, should assure that $\mathbf{V}$ and $\mathbf{W}$ are both of rank k. If $\mathbf{V}$ and $\mathbf{W}$ are built in the way defined in (2.2.30-2.2.31), the model reduction method is called a two-sided method. If only one of the projection matrices ($\mathbf{V}$ or $\mathbf{W}$) is built in that way, the method is called one-sided. Application of the two-sided method results in a reduced model that matches the first $2k$ moments of the original transfer function. In case of one-sided methods, k moments are matched.

The general proof of the moment matching property can be found in [36]. To illustrate the idea behind this proof, the matching of the zeroth moment of the system (2.2.21) for $s_0 = 0$ is shown after [51].

According to the formula (2.2.24), the zeroth moment for $s_0 = 0$ is equal to

$$M_0(0) = -\boldsymbol{c}^{\mathrm{T}}\mathbf{A}^{-1}\boldsymbol{b}. \qquad (2.2.32)$$

With $\mathbf{V}$ chosen as in (2.2.30) and the fact that $\mathbf{A}^{-1}\boldsymbol{b}$ belongs to the Krylov subspace $\mathcal{K}_{k_1}(\mathbf{A}^{-1}\mathbf{E}, \mathbf{A}^{-1}\boldsymbol{b})$, one can find a vector $\boldsymbol{r}_0$ such that $\mathbf{V}\boldsymbol{r}_0 = \mathbf{A}^{-1}\boldsymbol{b}$. Then, using the reduction procedure defined in (2.2.27), it can be shown that

$$\hat{M}_0(0) = -\boldsymbol{c}^{\mathrm{T}}\mathbf{V}(\mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V})^{-1}\mathbf{W}^{\mathrm{T}}\boldsymbol{b} = -\boldsymbol{c}^{\mathrm{T}}\mathbf{V}(\mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V}\boldsymbol{r}_0$$

$$= -\boldsymbol{c}^{\mathrm{T}}\mathbf{V}\boldsymbol{r}_0 = -\boldsymbol{c}^{\mathrm{T}}\mathbf{A}^{-1}\boldsymbol{b} = M_0(0). \qquad (2.2.33)$$

In case the approximation around $s_0 \neq 0$ or for $s_0 = \infty$ is needed, the starting matrix and vector for building the Krylov subspace have to be modified. One can also build a subspace using different values of s_0 at the same time. More details on how to do this and suggestions for starting values for different s_0 can be found in [36].

Besides the difference in the number of moments matched, the choice to use either one- or two-sided methods influences also some other properties of the reduced system. Two-sided methods may lead to better approximations of the output y and deliver a reduced-order model, whose input-output behavior does not depend on the state space realization of the original model. In case of the one-sided techniques with $\mathbf{W} = \mathbf{V}$ and $\mathbf{V}$ defined as in (2.2.30), for certain original models, one can also prove the preservation of the passivity property.

The process of constructing the reduction matrices, $\mathbf{V}$ and $\mathbf{W}$, is not straightforward and requires the use of special techniques. Because of round-off errors, the vectors building a Krylov subspace may quickly become linearly dependent. To avoid this problem,

one usually constructs an orthogonal basis of the appropriate Krylov subspace. This can be achieved using e.g. Arnoldi or Lanczos algorithms (explanation of these algorithms and implementation details are given in [77]). The classical Arnoldi algorithm generates a set $\mathbf{V}$ of orthonormal vectors, i.e.

$$\mathbf{V}^T\mathbf{V} = \mathbf{I}, \tag{2.2.34}$$

that form a basis for a given Krylov subspace. The Lanczos algorithm finds two sets of basis vectors, $\mathbf{V}$ and $\mathbf{W}$, that span an appropriate Krylov subspace and have property

$$\mathbf{W}^T\mathbf{V} = \mathbf{I}. \tag{2.2.35}$$

Two sets of basis vectors $\mathbf{V}$ and $\mathbf{W}$ for Krylov subspaces may also be computed using a two-sided Arnoldi algorithm (see [51]). In this case, both $\mathbf{V}$ and $\mathbf{W}$ are orthonormal,

$$\mathbf{V}^T\mathbf{V} = \mathbf{I}, \qquad \mathbf{W}^T\mathbf{W} = \mathbf{I}. \tag{2.2.36}$$

As a result, each of the above mentioned techniques generates a Krylov subspace. The choice of the subspace depends on the type of algorithm and the expansion point s_0 around which the approximation is of interest. A more detailed explanation on how to choose the proper subspaces can be found in [36].

The ideas of the Krylov subspace based reduction presented in this section can be further modified, depending on e.g. the application or the specific criteria that the reduced-order model should fulfill. In electronic circuit design, there exist methods especially suited for reducing specific types of systems that exploit the characteristic structure of the underlying matrices, see e.g. [6]. In case of coupled or interconnected systems, the goal may be to preserve the interpretation of the different physical domains. More details on this topic can be found in [28, 89]. There exist also modifications that aim at preserving other properties of the original system, such as stability or passivity. In case of symmetric matrices, the algorithm SyPVL was proposed in [29] that guarantees stability. A stability and passivity preserving technique, PRIMA, is presented in [61].

2.2.3   Balanced truncation

The field of systems and control focusses on the analysis of dynamical systems and design of feedback controllers for these systems. Herein, the objective of controller design is to change the dynamics of the system to induce desired behavior. Typical examples are the stabilization of unstable systems, tracking of a reference trajectory or the rejection of external disturbances on a system.

These control strategies are applied in a broad range of practical engineering problems, such as control of mechanical or electrical systems. These applications have in common that they deal with systems with inputs and outputs. On the one hand, a dynamical system can often only be influenced by a limited number of actuators, which

are represented as inputs. Additionally, external disturbances, such as e.g. measurement noise, often act only locally as well. On the other hand, only a limited number of sensors (i.e. outputs) is available in practical engineering systems. For these systems, it is thus particularly relevant to have an accurate model for their input-output behavior. Even though this model does not need to describe the global behavior of a system, complex dynamics can still yield large models of orders up to $O(10^3)$. To facilitate controller design and/or analysis for these systems, model reduction is needed. Here, it is noted that a controller needs to be implemented in real-time, which also requires a controller realization of relatively low-order.

Model reduction procedures in the field of systems and control therefore aim at approximating the *input-output behavior* of a high-order model. The quality of the reduced-order model can thus be assessed by comparing the outputs of the high-order and reduced-order models for given inputs, where the magnitude of the output error is measured using some signal norm.

Balanced truncation is the most popular method in systems and control addressing this model reduction problem. It mainly owes its popularity due to the fact that it preserves stability of the high-order model and provides an error bound, which gives a direct measure of the quality of the reduced-order model.

The *balanced truncation* method was first presented by Moore [57], where results of Mullis and Robberts [58] were exploited. Later, the stability preservation property was found by Pernebo and Silverman [64], whereas the error bound was derived by Enns [20] and Glover [34].

Linear dynamical models with inputs and outputs in state-space form

$$\begin{cases} \dot{x} = \mathbf{A}x + \mathbf{B}u \\ y = \mathbf{C}x + \mathbf{D}u \end{cases} \tag{2.2.37}$$

are considered. Here, $u \in \mathbb{R}^m$ denotes the input whereas $y \in \mathbb{R}^p$ represents the output. The internal state is given by $x \in \mathbb{R}^n$ and the system matrices are of corresponding dimensions. Throughout this section, it is assumed that the model (2.2.37) is asymptotically stable (i.e. all eigenvalues of $\mathbf{A}$ have negative real part) and is a minimal realization, where the latter guarantees that all state components contribute to the *input-output behavior*. The transfer function of (2.2.37) is given as

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}, \quad s \in \mathbb{C}. \tag{2.2.38}$$

In balanced truncation, a reduced-order model is obtained in two steps. First, a so-called balanced realization is found, in which the states are ordered according to their contribution to the input-output behavior. Second, a reduced-order model is obtained on the basis of this balanced realization by discarding the states with the smallest influence.

In order to find the balanced realization, the input-output behavior of the system (2.2.37) has to be quantified. To this end, the so-called controllability and observability functions are defined. First, the controllability function $\mathcal{E}_c(x_0)$ gives the smallest input energy required to reach the state x_0 from the zero state in infinite time, given as

$$\mathcal{E}_c(x_0) = \inf_{\substack{u \in \mathcal{L}_2(-\infty,0) \\ x(-\infty)=0,\, x(0)=x_0}} \int_{-\infty}^0 \|u(t)\|^2 \, dt, \qquad (2.2.39)$$

where $\mathcal{L}_2(-\infty, 0)$ denotes the space of square integrable functions, defined on the domain $(-\infty, 0)$. Second, the observability function $\mathcal{E}_o(x_0)$ is defined by

$$\mathcal{E}_o(x_0) = \int_0^\infty \|y(t)\|^2 \, dt, \quad x(0) = x_0,\ u(t) = 0 \ \ \forall t \in [0, \infty), \qquad (2.2.40)$$

and gives the future output energy of the system when released from an initial condition x_0 for zero input. It is well-known (see e.g. [57, 94]) that for linear systems as in (2.2.37) the controllability and observability functions in (2.2.39) and (2.2.43) can be written as the quadratic forms

$$\mathcal{E}_c(x_0) = x_0^{\mathrm{T}} P^{-1} x_0, \qquad \mathcal{E}_o(x_0) = x_0^{\mathrm{T}} Q x_0, \qquad (2.2.41)$$

where P and Q are the controllability and observability gramian, given by

$$P = \int_0^\infty e^{At} B B^{\mathrm{T}} e^{A^{\mathrm{T}} t} \, dt \qquad (2.2.42)$$

and

$$Q = \int_0^\infty e^{A^{\mathrm{T}} t} C^{\mathrm{T}} C e^{At} \, dt, \qquad (2.2.43)$$

respectively. From (2.2.42) and (2.2.43), it is easily observed that the controllability and observability gramians are only finite when the system is asymptotically stable, which explains the assumption stated before. In addition, P and Q are symmetric and positive definite, where the latter is guaranteed by the assumption that the system (2.2.37) is minimal, i.e. controllable and observable. The controllability and observability gramian can be obtained as the unique solutions of the respective *Lyapunov equations* (see for instance [94])

$$AP + PA^{\mathrm{T}} + BB^{\mathrm{T}} = 0 \qquad (2.2.44)$$

and

$$A^{\mathrm{T}} Q + QA + C^{\mathrm{T}} C = 0, \qquad (2.2.45)$$

which makes balanced truncation computationally feasible. Nonetheless, solving the Lyapunov equations is computationally costly, such that balanced truncation is limited to systems of orders up to $\mathcal{O}(10^3)$.

Since both the controllability and observability gramian characterize the in- or output energy associated to a state x_0, they are dependent on the realization of the system (2.2.37). Stated differently, a change of coordinates $\bar{x} = \mathbf{T}x$, with $\mathbf{T} \in \mathbb{R}^{n \times n}$ a nonsingular matrix, results in a realization with system matrices

$$\bar{\mathbf{A}} = \mathbf{TAT}^{-1}, \quad \bar{\mathbf{B}} = \mathbf{TB}, \quad \bar{\mathbf{C}} = \mathbf{CT}^{-1}, \quad \bar{\mathbf{D}} = \mathbf{D}. \tag{2.2.46}$$

Then, the new *controllability gramian* and *observability gramian* are given as

$$\bar{\mathbf{P}} = \mathbf{TPT}^{\mathrm{T}}, \qquad \bar{\mathbf{Q}} = \mathbf{T}^{-\mathrm{T}}\mathbf{QT}^{-1}. \tag{2.2.47}$$

Nonetheless, the product of $\bar{\mathbf{P}}$ and $\bar{\mathbf{Q}}$ yields

$$\bar{\mathbf{P}}\bar{\mathbf{Q}} = \mathbf{TPQT}^{-1}, \tag{2.2.48}$$

indicating that the eigenvalues of the product of the controllability and observability gramian are independent of the set of coordinates and thus system invariants. These eigenvalues equal the (squared) Hankel singular values σ_i [34], such that

$$\sigma_i = \sqrt{\lambda_i(\mathbf{PQ})}, \quad i = 1, \ldots, n, \tag{2.2.49}$$

where $\lambda_i(\mathbf{X})$ denotes the i-th eigenvalue of the matrix $\mathbf{X}$.

At this point, it is recalled that the observability gramian $\mathbf{Q}$ characterizes the output energy associated to a given initial state x_0 and thus provides a measure of the importance of state components with respect to the output y. Hence, states generating high output energy can be considered more important than states generating little output energy, since the former are easy to observe. On the other hand, the controllability gramian $\mathbf{P}$ gives a measure of the importance of state components x_0 with respect to the input u, in the sense that states that require little input energy to reach are more relevant than states that require high input energy. States that require little energy to reach are thus easy to control. Clearly, the combination of the controllability and observability gramians gives a characterization of the importance of state components from an input-output perspective. However, in an arbitrary coordinate system, a state $\bar{x}_0^1$ that requires little energy to reach might also generate little output energy. On the other hand, a different state $\bar{x}_0^2$ might exist that requires a lot of energy to reach, but generates high output energy. In this case, it is not easy to decide which of $\bar{x}_0^1$ and $\bar{x}_0^2$ is the most important component from an input-output perspective. To facilitate this analysis, the balanced realization is introduced. Namely, there exists a state-space realization such that the corresponding controllability and observability gramians are equal and diagonal, where the entries on the diagonal are given by the Hankel singular values [57]:

$$\bar{\mathbf{P}} = \bar{\mathbf{Q}} = \mathbf{\Sigma} := \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix}. \tag{2.2.50}$$

In addition, the *Hankel singular values* are ordered as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$. In this balanced realization, the controllability and observability function are given as

$$\mathcal{E}_c(\bar{x}_0) = \bar{x}_0^{\mathrm{T}} \Sigma^{-1} \bar{x}_0, \qquad \mathcal{E}_o(\bar{x}_0) = \bar{x}_0^{\mathrm{T}} \Sigma \bar{x}_0. \tag{2.2.51}$$

Now, the form (2.2.51) allows for a clear interpretation. Namely, the realization is balanced in the sense that states that are easy to control are also easily observed. In fact, due to the ordering of the Hankel singular values, the state $\bar{x}_0 = e_1 := [1, 0, \cdots, 0]^{\mathrm{T}}$ requires the least energy to reach ($\mathcal{E}_c(e_1) = \sigma_1^{-1}$ is small) and gives the highest output energy ($\mathcal{E}_o(e_1) = \sigma_1$ is large). Stated differently, this state is easy to control and easy to observe. Hence, $\bar{x}_0 = e_1$ has the largest contribution to the input-output behavior of the system. On the other hand, the state $\bar{x}_0 = e_n := [0, \cdots, 0, 1]^{\mathrm{T}}$ is both difficult to control and difficult to observe, such that it has the smallest contribution to the input-output behavior.

The coordinate transformation $\mathbf{T}$ to obtain the balanced realization can be obtained on the basis of the controllability and observability gramians (2.2.42-2.2.43). Thereto, the *Cholesky factor* $\mathbf{U}$ of $\mathbf{P}$ is used, as well as the eigenvalue decomposition of $\mathbf{U}^{\mathrm{T}} \mathbf{Q} \mathbf{U}$:

$$\mathbf{P} = \mathbf{U}\mathbf{U}^{\mathrm{T}}, \qquad \mathbf{U}^{\mathrm{T}} \mathbf{Q} \mathbf{U} = \mathbf{K}\mathbf{S}\mathbf{K}^{\mathrm{T}}. \tag{2.2.52}$$

In the latter, it is noted that $\mathbf{U}^{\mathrm{T}} \mathbf{Q} \mathbf{U}$ is a positive definite symmetric matrix, such that the matrix of eigenvectors $\mathbf{K}$ is orthonormal. Additionally, the eigenvalues are real and, when ordered, are equal to the squared Hankel singular values such that $\mathbf{S} = \Sigma^2$ with Σ as in (2.2.50). Then, the balancing transformation and its inverse are given as

$$\mathbf{T} = \Sigma^{\frac{1}{2}} \mathbf{K}^{\mathrm{T}} \mathbf{U}^{-1}, \qquad \mathbf{T}^{-1} = \mathbf{U}\mathbf{K}\Sigma^{-\frac{1}{2}} \tag{2.2.53}$$

as can be checked by substitution of (2.2.53) in (2.2.47), while using the relations (2.2.52). An overview of alternative algorithms to obtain the balanced realization can be found in [3].

So far, a balanced realization is found, but no model reduction has been performed yet. However, the balanced realization gives a representation in which the states are ordered according to their contribution to the input-output behavior. Hence, a reduced-order model of order k can be obtained by partitioning the state $\bar{x}$ of the balanced realization as $\bar{x}^1 = [\bar{x}_1, \ldots, \bar{x}_k]^{\mathrm{T}} \in \mathbb{R}^k$ and $\bar{x}^2 = [\bar{x}_{k+1}, \ldots, \bar{x}_n]^{\mathrm{T}} \in \mathbb{R}^{n-k}$, such that $\bar{x}^1$ contains the state components with the largest influence on the input-output behavior. When the system matrices are partitioned accordingly,

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}, \quad \bar{\mathbf{A}} = \begin{bmatrix} \bar{\mathbf{A}}_{11} & \bar{\mathbf{A}}_{12} \\ \bar{\mathbf{A}}_{21} & \bar{\mathbf{A}}_{22} \end{bmatrix}, \quad \bar{\mathbf{B}} = \begin{bmatrix} \bar{\mathbf{B}}_1 \\ \bar{\mathbf{B}}_2 \end{bmatrix}, \quad \bar{\mathbf{C}} = \begin{bmatrix} \bar{\mathbf{C}}_1 & \bar{\mathbf{C}}_2 \end{bmatrix}, \quad \bar{\mathbf{D}} = \mathbf{D}, \tag{2.2.54}$$

a reduced-order system can be obtained by truncation, i.e. by setting $\bar{x}_2 = \mathbf{0}$. The resulting reduced-order model (with $\hat{x} \in \mathbb{R}^k$ an approximation of $\bar{x}^1 \in \mathbb{R}^k$) is given by the state-space realization

$$\begin{cases} \dot{\hat{x}} = \bar{\mathbf{A}}_{11}\hat{x} + \bar{\mathbf{B}}_1 u, \\ \hat{y} = \bar{\mathbf{C}}_1 \hat{x} + \bar{\mathbf{D}}u. \end{cases} \qquad (2.2.55)$$

Here, it can be observed that the reduced-order state-space system (2.2.55) is itself a balanced realization, with the controllability and observability gramians given by $\mathbf{\Sigma}_1$ (see [64]). In addition, when $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ have no diagonal entries in common (i.e. when $\sigma_k > \sigma_{k+1}$), the reduced-order system is asymptotically stable [64].

The reduced-order system thus preserves stability of the original model, and its output $\hat{y}$ serves as an approximation for the output of the high-order system y. The quality of this approximation can be assessed by means of a bound on the error. Namely, an error bound can be expressed in terms of the discarded Hankel singular values [20, 34] as

$$\|\mathbf{H}(s) - \hat{\mathbf{H}}(s)\|_\infty \leq 2 \sum_{i=k+1}^{n} \sigma_i, \qquad (2.2.56)$$

where $\mathbf{H}(s)$ and $\hat{\mathbf{H}}(s)$ are the transfer functions of the full-order system (2.2.37) and the reduced-order system (2.2.55), respectively. Furthermore, $\| \cdot \|_\infty$ denotes the $\mathcal{H}_\infty$ norm defined as

$$\|\mathbf{H}(s)\|_\infty = \sup_{\omega \in \mathbb{R}} \bar{\sigma}(\mathbf{H}(j\omega)), \qquad (2.2.57)$$

with $\bar{\sigma}(\cdot)$ the largest singular value. The error bound (2.2.56) confirms the intuitive idea that the states corresponding to the largest Hankel singular values are the most important from the perspective of input-output behavior. Namely, a good approximation (i.e. a low error bound) will be obtained when the Hankel singular values in $\mathbf{\Sigma}_2$ are small. Since these Hankel singular values are only dependent on the high-order model (2.2.37), they can be computed a priori and allow for control over the reduction error by selection of the order k. Finally, it is noted that in (2.2.56) it is assumed that all Hankel singular values are distinct. When Hankel singular values with multiplicity larger than one occur, they only need to be counted once, leading to a tighter bound (see e.g. [34]).

In the model reduction procedure presented here, a reduced-order system is obtained by truncation (i.e. setting $\bar{x}^2 = 0$) of a balanced realization. An alternative approach is given by singular perturbation [24] of this realization. Herein, it is assumed that the dynamics describing the evolution of $\bar{x}^2$ is very fast (and asymptotically stable). Then, this dynamics can be assumed to be in its equilibrium position at all time, which is obtained by setting $\dot{\bar{x}}^2 = 0$ and solving for $\bar{x}^2$ as a function of $\bar{x}^1$ and u. Contrary to balanced truncation, the singular perturbation approach guarantees that the steady-state gains of the high-order system are matched in the reduced-order system. The reduced-order model is controllable, observable, asymptotically stable and the error bound (2.2.56) also holds [49].

Balanced truncation aims at approximating a high-order system by selecting the state components that have the largest contribution in the input-output behavior, according to the energy in the input and output signals. The entire frequency range is considered in this approach. However, in many practical applications, a good approximation is only required in a specific frequency range. To this end, frequency-weighted balanced truncation can be used [20], which is an extension of the method discussed in the previous paragraphs. In frequency-weighted balanced truncation, the objective is to find a reduced-order system such that the error

$$\|\mathbf{H}_o(s)(\mathbf{H}(s) - \hat{\mathbf{H}}(s))\mathbf{H}_i(s)\|_\infty \tag{2.2.58}$$

is small, where $\mathbf{H}_i(s)$ and $\mathbf{H}_o(s)$ denote the transfer functions of an input and output frequency weight, respectively. These weights can be designed by the user to emphasize specific regions in the frequency domain. To obtain the frequency-weighted reduced-order model, controllability and observability gramians are computed on the basis of the frequency weighted high-order system, which are simultaneously diagonalized. Details can be found in [20].

When the original system is asymptotically stable, observable and controllable, and only one-sided weighting is applied (i.e. either $\mathbf{H}_i(s) = \mathbf{I}$ or $\mathbf{H}_o(s) = \mathbf{I}$), asymptotic stability of the reduced-order system is guaranteed. However, in the case of general two-sided weighting, stability of the reduced-order approximant can not be guaranteed. Nonetheless, when the reduced-order model is stable, an error bound is given in [47].

In the preceding paragraphs, the standard balanced truncation technique for asymptotically stable systems as well as an extension to frequency-weighted balanced truncation is presented. Several extensions of balanced truncation exists. For example, balanced truncation of the coprime factorization applies to unstable systems [56, 60], whereas a method preserving passivity is given in [15, 59].

Besides these methods based on balanced truncation, a popular alternative is optimal Hankel norm approximation [34], which is also based on the balanced realization. For an overview of model reduction in systems and control, see e.g. [3, 38].

2.3 Qualitative comparison on model reduction methods

In this section, the methods as discussed in Section 2.2 will be compared. First, the common feature of projection is presented in Section 2.3.1. Then, a general comparison will be given in Section 2.3.2. A close connection between moment matching and modal truncation augmentation is discussed in Section 2.3.3. Computational aspects and the preservation of properties will be discussed in Sections 2.3.4 and 2.3.5, respectively.

2.3.1 Projection

Before discussing differences between the methods from the fields of structural dynamics, numerical mathematics and systems and control, an important similarity is discussed. Namely, the methods discussed in Section 2.2 have in common that the reduced-order models are obtained by projection. Here, the reduced-order model is obtained by application of the projection $\Pi = \mathcal{V}\mathcal{W}^{\mathrm{T}}$ to the original model. In numerical mathematics, the projection matrices might be chosen as $\mathcal{V} = \mathbf{V}$ and $\mathcal{W} = \mathbf{W}$ with $\mathbf{V}$ and $\mathbf{W}$ as in (2.2.30) and (2.2.31), respectively. This specific choice ensures moment matching around $s_0 = 0$. For balanced truncation, as used in systems and control, the matrices $\mathcal{V}$ and $\mathcal{W}$ are given as the first k columns of the transformation matrices $\mathbf{T}^{-1}$ and $\mathbf{T}^{\mathrm{T}}$ as in (2.2.53), respectively. Hence, they project on the subspace of $\mathbf{PQ}$ corresponding to the largest Hankel singular values (see (2.2.48)), which yields the subspace with the largest contribution in the input-output behavior. Similarly, in the mode superposition techniques in structural dynamics, the projection is given as $\Pi = \mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}}$. Here, the projection basis $\mathbf{\Phi}$ forms a basis for the space spanned by the k most relevant eigenvectors (see (2.2.3)), which are typically chosen as the eigenvectors corresponding to the lowest eigenfrequencies. Here, it is noted that the state-space form is used in the fields of numerical mathematics and systems and control, whereas a second-order form is exploited in structural dynamics when no damping is present or when the damping can be considered as small.

2.3.2 General comparison

Besides the common feature of projection, the reduction techniques in Section 2.2 have important differences, as listed below.

- **First-order form versus second-order form**
  The most apparent distinction is the type of model under consideration. In the field of structural dynamics, models of second-order form are usually studied, whereas first-order models are examined in the fields of numerical mathematics and systems and control. Even though the use of this symmetric second-order form seems limiting, it is noted that the mechanical structures studied in this field can indeed be modeled as second-order systems. In addition, these mechanical structures typically have little damping, which motivates the exploitation of undamped vibration modes for model reduction. Nonetheless, due to the specific structure of these models, the model reduction techniques from structural dynamics can in general not be applied to other application domains. On the other hand, any model that can be written in the first-order form can be handled by the reduction techniques from numerical mathematics and systems and control, although asymptotic stability is assumed in the latter.

- **Input-output behavior versus global behavior**

  A second difference is given by the objective of the approximation. In numerical mathematics and systems and control, a reduced-order model is sought which approximates the input-output behavior of the original system. On the other hand, this input-output behavior is of less relevance in the field of structural dynamics, where the approximation of the global dynamics is of interest. Again, this results from the specific objectives in structural dynamics. Namely, typical interest is in the identification of the regions where the highest stresses or maximum displacements occur, whose locations are not known beforehand. Hence, the modeling of the global dynamical behavior is the main goal. However, extensions to mode superposition methods (see e.g. Section 2.2.1) provide techniques of incorporating the (static) influence of input forces in the reduction basis, partially taking input-output behavior into account.

  In numerical mathematics and systems and control, the internal behavior of the model is of little interest. In control design, the system behavior from the control input to the measured outputs is of relevance and this directly forms the basis for the model reduction procedure. In the analysis of large-scale electrical circuits, where moment matching methods from the field of numerical mathematics are typically applied, interest is in the reduction of linear subcircuits. Here, its influence on the total circuit is described by the inputs and outputs, such that the approximation of input-output behavior is of interest. Nonetheless, input-output behavior is only truly taken into account in the latter when two-sided projection techniques are used. Namely, when using one-sided projection techniques, either the input matrix $\mathbf{B}$ or output matrix $\mathbf{C}$ is discarded, such that the focus of the reduction is limited to the state-to-output or input-to-state behavior, respectively. In this case, the number of moments matched is independent of the choice of input or output matrix. Nonetheless, the number of moments matched for a given reduction order k is doubled in the two-sided case, when input-output behavior is fully taken into account.

- **Interpretation of reduction space**

  Model reduction techniques from structural dynamics are largely based on physical properties of mechanical systems. Therefore, the reduction space resulting from modal approximation has a useful engineering interpretation. Namely, it consists of the modes of the system, which represent the typical vibration pattern of a structure at a given frequency. The most important modes and the corresponding eigenfrequencies are preserved in the reduced-order system. Since these modes are obtained via an eigenvalue decomposition, the system in modal coordinates is in diagonal form, as discussed in Section 2.2.1. Here, it is recalled that this only holds when the system is undamped or has proportional (Rayleigh) damping or modal damping. In this diagonal form, the equations describing the dynamics of the modes are uncoupled, which means that no error is introduced in

the dynamics of the modes that are kept in the reduced-order model. In fact, the reduction error is due to the deletion of modes, rather than errors in the dynamics of the modes themselves.

In the reduction techniques from numerical mathematics and systems and control, the reduction space does not have a clear physical interpretation. Of course, this is largely due to the fact that these procedures are not limited to mechanical systems and are based on system-theoretic properties instead, as discussed in Section 2.2.3.

- **Global versus local approximation in frequency domain**
  The modal truncation and moment matching model reduction techniques from structural dynamics and numerical mathematics have in common that they can be considered as frequency-domain-based (or Laplace-domain-based) techniques. Therefore, they give a good approximation in some part of the frequency-domain only. This is directly apparent in the modal reduction techniques from structural dynamics, where the modes as used in the reduced-order model are selected by their corresponding eigenfrequency. Here, the modes are typically selected from the lower end of the frequency spectrum. On the other hand, moment matching in numerical mathematics is based on the Taylor series expansion of a transfer function at a specific point (or multiple points) in Laplace domain. Since the moments around this expansion point form the basis for the reduced-order model, this approximation can only be expected to be accurate around the expansion point, leading to a local approximation in frequency domain.

  In balanced truncation, as used in systems and control, the behavior in frequency domain does not form the basis of the model reduction procedure. Instead, the transfer of energy from the input to the output is used as a tool for model reduction, which can be considered as a time-domain approach. Nonetheless, specific regions in frequency domain can be emphasized by the extension to frequency-weighted balanced truncation, as noted in Section 2.2.3.

- **Automatic versus user-dependent model reduction**
  A final general difference can be found in the level of automation of the model reduction techniques from the different fields. Here, only the balanced truncation method in systems and control is fully automatic when a requirement on the quality of the reduced-order model is given. Namely, the existence of an a priori error bound (2.2.56) allows for the automatic choice of the reduction order. The methods from structural dynamics and numerical mathematics lack such an error bound.

  Even when the reduction order is chosen beforehand, the methods from structural dynamics and numerical mathematics are heuristic. That is, the mode superposition techniques from structural dynamics are dependent on the frequency range of interest, which needs to be specified a priori. Herein, typically the modes corresponding to the lowest frequencies are chosen. Similarly, the reduction procedure

in the moment matching techniques from numerical mathematics is dependent on the choice of expansion points. However, the influence of this choice on the properties of the reduced-order system is largely an open problem and very few guidelines for this selection exists. Therefore, these expansion points are typically chosen as $s_0 = 0$ or $s_0 = \infty$. Of course, the computational procedure in mode superposition and moment matching is fully automatic as soon as a choice is specified for the frequency range of interest and the expansion point, respectively.

2.3.3 Moment matching and model truncation augmentation

A close link exists between modal truncation augmentation techniques used in structural dynamics (Section 2.2.1) and the moment matching methods (Section 2.2.2). This can be understood by considering the series expansion of the (non-damped) structural equations (2.2.2) in the Laplace domain for s^2:

$$q = (\mathbf{K} + s^2\mathbf{M})^{-1}f = \sum_{i=0}^{\infty} \left((\mathbf{K} + s_0^2\mathbf{M})^{-1}\mathbf{M}\right)^i (\mathbf{K} + s_0^2\mathbf{M})^{-1}f, \qquad (2.3.1)$$

where s_0 is a chosen expansion point. Clearly this expansion is similar to the moment matching expansion (2.2.23) except that here it is written for the second-order form. The reduction basis suggested by this expansion is the Krylov series

$$\mathcal{K}_k((\mathbf{K} + s_0^2\mathbf{M})^{-1}\mathbf{M}, (\mathbf{K} + s_0^2\mathbf{M})^{-1}f) \qquad (2.3.2)$$

In the modal truncation augmentation approach the reduction basis consists of some eigenmodes of the system and modal truncation corrections as described in (2.2.20). Recalling the definition (2.2.16) of the correction vectors, it can be seen that the reduction basis (2.2.20) for the modal truncation augmentation is

$$\text{span}\left\{\left[\begin{array}{cc}\mathbf{\Phi} & q_{cor}\end{array}\right]\right\} = \text{span}\left\{\left[\begin{array}{cc}\mathbf{\Phi} & \mathbf{K}^{-1}f\end{array}\right]\right\}, \qquad (2.3.3)$$

indicating that it includes the zero-order expansion term around $s_0 = 0$. Thus, it conserves the zero-order moment of the second-order problem around $s_0 = 0$, which is a direct consequence of the fact that the reduction basis includes the exact static solution. Through a similar reasoning one could say that substructuring methods that include the interface static modes (like the Craig-Bampton, the Rubin/MacNeal and the Dual Craig-Bamtpon methods) are matching the zero-order moments for the interface forces.

The modal truncation augmentation form presented in Section 2.2.1 includes only the zeroth-order correction as indicated by the basis (2.3.3). Higher-order corrections as suggested in the Krylov sequence (2.3.2) can also be included in the reduction space as proposed in [18, 92, 48, 1, 50, 72], which guarantees matching higher-orders moments and thus leads to an approach similar to the moment matching technique. Higher-order

correction modes have also been used in the context of substructuring and mode component synthesis [17, 70, 73]. Note that the high-order corrections for structural problems can be obtained as a by-product of the Lanczos algorithm used to compute the eigenmodes [71] and that one can also consider quasi-static corrections (i.e. for $s_0 \neq 0$) in case one is interested in a specific frequency range [93].

It is important to observe that the modal truncation augmentation uses a reduction basis that, in addition to the moments, also includes true eigenmodes of the system. In that sense this method differs from the usual moment matching techniques and it accounts both for the global behavior of the system (through its eigenmodes) and for input-specific components (through the moments).

2.3.4 Computational aspects

A general comparison of properties of model reduction techniques from the fields of structural dynamics, numerical mathematics and systems and control was given in Section 2.3.2. Computational aspects are addressed in the current section.

From a computational point of view, the methods from systems and control have the highest cost. In these methods, the computational complexity is mainly due to the solution of two Lyapunov equations (see (2.2.44) and (2.2.45)), which are of the size of the original high-order model. This seriously hinders the applicability of balanced truncation to systems of very high order. In addition, a full eigenvalue decomposition (see (2.2.52)) is required, such that the total computational cost associated to balanced truncation is high. Finally, a full coordinate transformation has to be computed, before reduction can be performed by means of truncation.

The computational cost for reduction techniques from the fields of structural dynamics and numerical mathematics is significantly lower. First, these methods do not require the computation of a full coordinate transformation. Instead, only the reduction space is computed, which is given by only k basis vectors. Furthermore, the computations are less costly since the matrix operations that are required are relatively cheap when compared to those needed for the solution of Lyapunov equations. In the mode displacement techniques from structural dynamics, only the most important eigenvalues and eigenvectors need to be computed. Since the frequency domain of interest is typically known beforehand, efficient iterative methods can be used to find the eigenfrequencies in this range. For the Krylov-subspace based moment matching techniques from numerical mathematics, the numerical cost is even less. Namely, the application of the Arnoldi or Lanczos methods only requires the solutions of linear sets of equations or matrix-vector multiplications. Therefore, moment matching methods by Krylov subspaces can be applied to systems of very high order.

Even though the application of balanced truncation seems limited from a computational point of view, it is remarked that the perception of "high-order" differs in the three

different fields. Especially, models of very low order (i.e. $O(10^0 - 10^1)$) are of interest in the field of systems and control. This is mainly due to the fact that controllers have to be implemented in real-time, which provides a limit on the order of the controller. Furthermore, low-order controllers are preferred because of their limited complexity. Hence, even though the computation of Lyapunov equations limits the applicability of balanced truncation to systems of order $O(10^3)$, it still provides a solution to relevant model reduction problems in practice. On the other hand, the models describing mechanical structures in the field of structural dynamics typically result from finite element procedures, leading to models of orders up to $O(10^6)$. Similarly, the moment matching techniques from numerical mathematics typically find application in the analysis of large-scale electrical circuits, leading to models of order up to $O(10^6)$. From these applications, the need for numerically efficient model reduction procedures is clear.

2.3.5 Preservation of properties

In model reduction, the objective is the construction of a reduced-order model that gives a good approximation of the original high-order model. Herein, it is of crucial importance that the reduced-order model preserves some properties of the original system, among which stability is the most important. If the high-order system is asymptotically stable, balanced truncation (see Section 2.2.3) indeed preserves this property, which is due to the fact that the (diagonal) gramians act as Lyapunov equations. The moment matching techniques from Section 2.2.2 do not satisfy such a property, such that stability of the reduced-order model can not be guaranteed in general. Nonetheless, methods exist that preserve stability for classes of linear systems (see e.g. [61]).

In the mode superposition techniques outlined in Section 2.2.1, stability of the reduced-order model can not be guaranteed when the original high-order system exhibits general damping. However, in the important cases of undamped systems or systems with positive definite symmetric damping matrix (which includes the cases of proportional (Rayleigh) and modal damping), the stability properties are indeed preserved. In fact, since the reduced-order model is based on the computation of the undamped vibration modes, reduction of an undamped system leads to an undamped reduced-order system, where the most important eigenfrequencies are preserved. Stated differently, the pole locations of the most important poles remain unchanged. This property does in general not hold for balanced truncation and moment matching techniques.

Furthermore, it is obvious that modal superposition techniques preserve the second-order form in the reduced-order model. Nonetheless, this is an important feature in the field of structural dynamics since it implies that the kinematic relation between displacement and velocity is preserved. This does not hold for balanced truncation and moment matching, even if the models stem from a second-order form.

Next, it is remarked that the existence of an error bound, as discussed in Section 2.3.2, is closely related to stability preservation. Namely, a bound on the difference
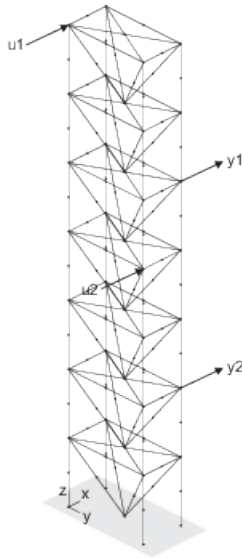
Figure 2.1: Truss frame system.

of solutions from the high-order and reduced-order systems can only be expected to exist
when both systems are stable, making stability a prerequisite for the existence of an error
bound.

Finally, it is often important to preserve other system properties besides stability.
Herein, passivity and bounded realness are the most notable. Even though the methods
as discussed in Section 2.2 do not generally preserve these properties, it is noted that
extensions exist that do. For the different fields, some references to the literature are
given in the corresponding parts in Section 2.2.

2.4   Illustrative example

To illustrate the differences between methods as discussed in Section 2.3, the model
reduction procedures of Section 2.2 are applied to a common benchmark example.

The benchmark example is chosen from the domain of structural dynamics, to allow
for application of all model reduction techniques discussed in Section 2.2. More speci-
fically, the truss frame system as depicted in Figure 2.1 is considered. Here, the nodes,

as indicated by points, are connected by beam elements, leading to a model of the form

$$\mathbf{M}\ddot{q} + \mathbf{D}\dot{q} + \mathbf{K}q = \tilde{b}_1 u_1 + \tilde{b}_2 u_2, \tag{2.4.1}$$

$$y_1 = \tilde{c}_1^\mathrm{T} q, \tag{2.4.2}$$

$$y_2 = \tilde{c}_2^\mathrm{T} q, \tag{2.4.3}$$

where $q \in \mathbb{R}^N$ is the vector of the displacements and rotations of the nodes, with $N = 714$. The truss frame is subject to forces acting (in x-direction) on the positions as shown in Figure 2.1, which are modeled as the inputs u_1 and u_2 in (2.4.1). In addition, the displacement (in x-direction) of the truss frame is measured at two locations, leading to the outputs y_1 and y_2. Finally, the truss frame model is lightly damped, which is modeled using modal damping.

In order to apply model reduction techniques from the fields of numerical mathematics and systems and control, the truss frame model (2.4.1) has to be written in state-space form. By choosing the state vector as $x^\mathrm{T} = [q^\mathrm{T} \ \dot{q}^\mathrm{T}]$, the dynamics is given by

$$\dot{x} = \mathbf{A}x + b_1 u_1 + b_2 u_2, \tag{2.4.4}$$

$$y_i = c_i^\mathrm{T} x, \quad i \in \{1, 2\}, \tag{2.4.5}$$

where it is noted that $x \in \mathbb{R}^n$ with $n = 2N$. The system matrices read

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{D} \end{bmatrix}, \quad b_i = \begin{bmatrix} \mathbf{0} \\ \mathbf{M}^{-1}\tilde{b}_i \end{bmatrix}, \quad c_i = \begin{bmatrix} \tilde{c}_i \\ \mathbf{0} \end{bmatrix}, \quad i \in \{1, 2\}. \tag{2.4.6}$$

Alternatively, the moment matching methods discussed in Section 2.2.2 can also be applied using system descriptions of the form

$$\mathbf{E}\dot{x} = \mathbf{A}x + b_1 u_1 + b_2 u_2, \tag{2.4.7}$$

$$y_i = c_i^\mathrm{T} x, \quad i \in \{1, 2\}, \tag{2.4.8}$$

with matrices

$$\mathbf{E} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{K} & -\mathbf{D} \end{bmatrix}, \quad b_i = \begin{bmatrix} \mathbf{0} \\ \tilde{b}_i \end{bmatrix}, \quad i \in \{1, 2\}, \tag{2.4.9}$$

which avoids the need for inversion of the matrix $\mathbf{M}$. In this form, the output vectors c_i remain unchanged.

The model reduction techniques from Section 2.2 are applied to this example. From the field of structural dynamics, the mode displacement method is used. Here, it is recalled that this method is based on the undamped system (i.e. $\mathbf{D} = \mathbf{0}$) and the location of the inputs and outputs (i.e. knowledge on $\tilde{b}_i$ and $\tilde{c}_i$) is not taken into account. Nonetheless, the (static) influence of the locations of the inputs can be taken into account

by the extensions given by mode acceleration and modal truncation augmentation (see Section 2.2.1) On the other hand, model reduction with respect to input u_1 and output y_1 is performed using moment matching and balanced truncation. Since these methods are based on the state-space form (2.4.5) (or (2.4.8)), damping can be included. In moment matching, the expansion point is chosen as $s_0 = 0$ and a one-sided projection is used, based on the input only (see (2.2.30)). In reduction, the reduced-order size of the first-order models k is chosen as $2K$, with K the number of modes taken into account in the mode displacement methods. This choice is motivated by the fact that the representation of a second-order model in first-order form doubles the number of equations. Hereby, the reduced-order models in first-order and second-order form have a similar approximation accuracy. Choosing $K = 2$ ($k = 4$) leads to the frequency response functions, with input u_1 and output y_1, as depicted in Figures 2.2 and 2.3.

In the mode displacement method (MD), the K lowest eigenvectors are chosen in the reduction basis. Therefore, the first two resonance peaks are captured by this reduced-order model. On the other hand, the method of balanced truncation (BT) does not capture the second resonance peak. Instead, the third resonance is approximated. This is caused by the fact that balanced truncation takes the location of the input and output into account. Specifically, the second and third resonance peak correspond to a bending mode around the y-axis and a torsional mode, respectively, where the latter has a larger influence on the input-output behavior from input u_1 to output y_1. Finally, the moment matching technique (MM) gives a good approximation at low frequencies, which originates from the choice of the expansion point as $s_0 = 0$. Therefore, for $k = 4$, only the first resonance peak is captured. This is most clearly shown in the error magnitude in Figure 2.3, where the moment matching techniques gives the best approximation for low frequencies. However, moment matching gives the largest $\mathcal{H}_\infty$ norm of the error system. The lowest norm is obtained for balanced truncation, which outperforms the mode displacement method by the selection of the third rather than the second resonance peak.

To illustrate the influence of the locations of the inputs and outputs on the reduced-order model, the frequency response functions for input u_2 and output y_2 are depicted in Figure 2.4, whereas the corresponding error is given in Figure 2.5. Here, the same reduction bases were used as in Figure 2.2. Hence, the input u_2 and output y_2 are not taken into account in the model reduction procedure. Since the mode displacement method is based on the global dynamics rather than specific inputs and outputs, it also gives a good approximation for these new inputs. On the contrary, balanced truncation and moment matching are dependent on the input-output behavior, where it is recalled that reduction was based on input u_1 and output y_1. Therefore, they do not give a good approximation for the input-output behavior from input u_2 to output y_2, as is clear from the large errors in Figure 2.5.

However, the moment matching technique used here is one-sided. Hence, only the input matrix is taken into account and, in this case, moment matching (of k moments) can be proven for any output. To illustrate this, the input-output behavior from input u_1
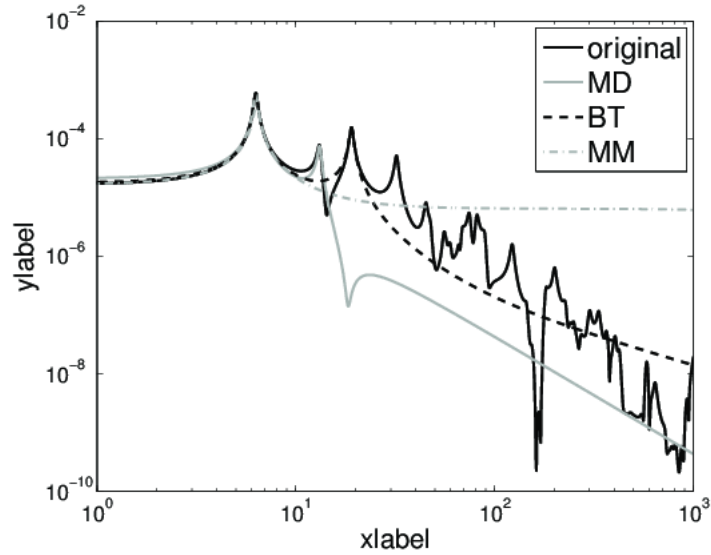
Figure 2.2: Comparison of the modal displacement method (MD), balanced truncation (BT) and moment matching (MM) for reduction to $K = 2$ ($k = 4$): magnitude of the frequency response function for input u_1 and output y_1.
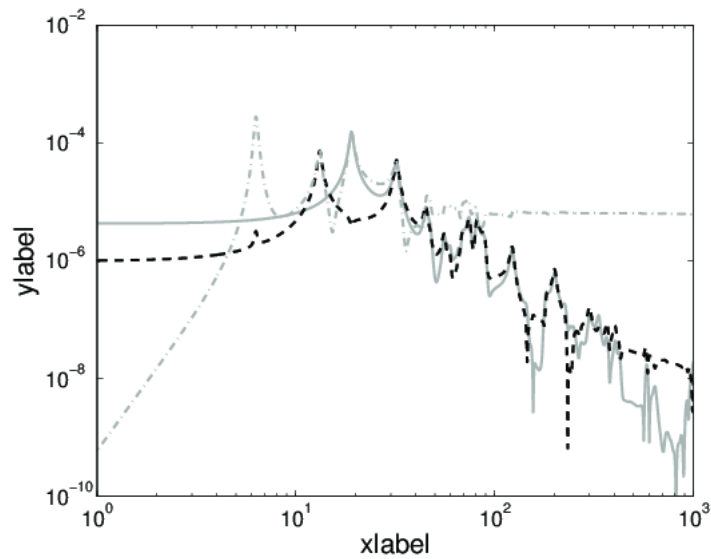
Figure 2.3: Magnitude of the error for reduction to $K = 2$ ($k = 4$) for input u_1 and output y_1. Line styles as in Figure 2.2.
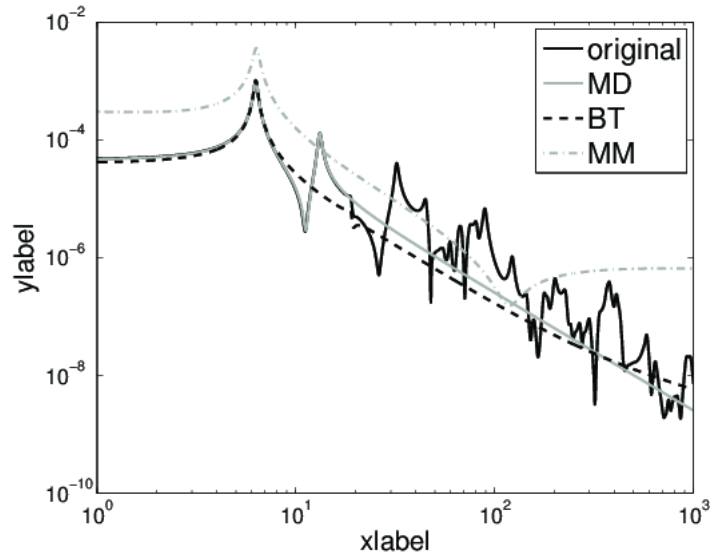
Figure 2.4: Comparison of the modal displacement method (MD), balanced truncation (BT) and moment matching (MM) for reduction to $K = 2$ ($k = 4$): magnitude of the frequency response function for input u_2 and output y_2. For balanced truncation and moment matching, the reduced-order model is based on input u_1 and output y_1.
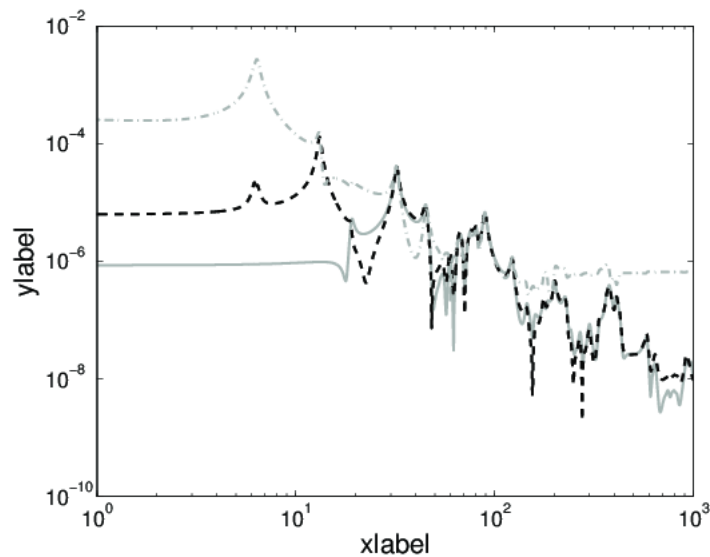
Figure 2.5: Magnitude of the error for reduction to $K = 2$ ($k = 4$) for input u_2 and output y_2. Line styles as in Figure 2.4. For balanced truncation and moment matching, the reduced-order model is based on input u_1 and output y_1.
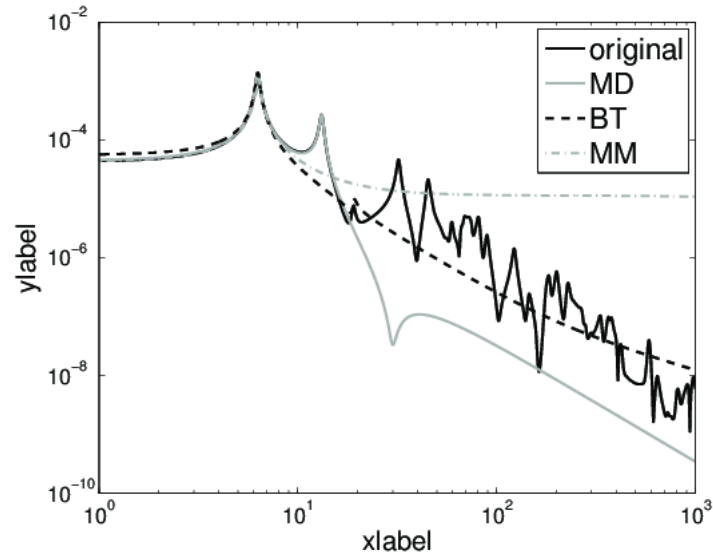
Figure 2.6: Comparison of the modal displacement method (MD), balanced truncation (BT) and moment matching (MM) for reduction to $K = 2$ ($k = 4$): magnitude of the frequency response function for input u_1 and output y_2. For balanced truncation and moment matching, the reduced-order model is based on input u_1 and output y_1.
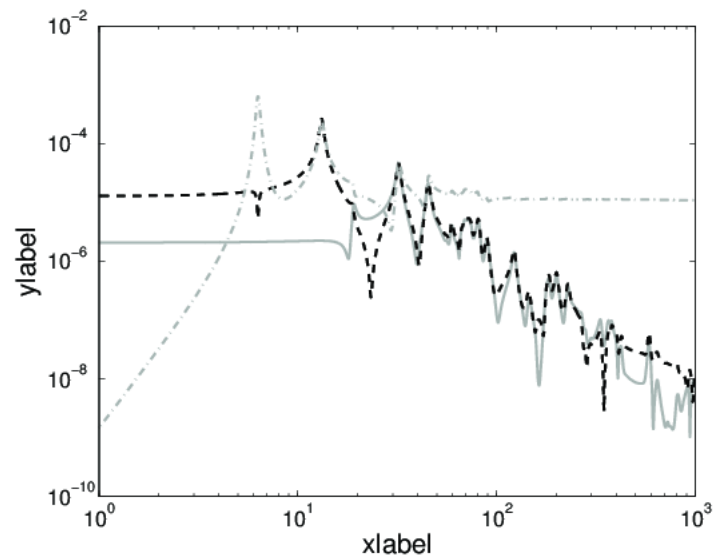
Figure 2.7: Magnitude of the error for reduction to $K = 2$ ($k = 4$) for input u_1 and output y_2. Line styles as in Figure 2.6. For balanced truncation and moment matching, the reduced-order model is based on input u_1 and output y_1.

to output y_2 is shown in Figures 2.6 and 2.7, again using the reduction basis generated for input u_1 and output y_1. As expected, the performance obtained by moment matching is similar to that of Figure 2.2 (and Figure 2.3), where the same input u_1 was used. For the mode displacement method and balanced truncation, the conclusions as stated before hold.

Finally, stability of the reduced-order models is checked. Since the truss frame system exhibits modal damping, the reduced-order model obtained by the mode displacement technique is guaranteed to be stable. Stability is also guaranteed in the case of balanced truncation. For moment matching, stability can not be guaranteed a priori. However, for $k = 4$, the reduced-order model obtained by moment matching is stable, as follows from an a posteriori check. On the other hand, for $k = 6$, stability is not preserved using moment matching.

2.5   Conclusions

In this paper, an overview and comparison of popular model reduction methods from the fields of structural dynamics, numerical mathematics and systems and control are provided. A detailed review is given on mode displacement techniques, moment matching methods and balanced truncation, whereas important extensions are outlined briefly.

The differences and similarities between presented methods are discussed, both qualitatively and quantitatively. Here, an important difference is the fact that the global dynamics is taken into account in the mode displacement methods, whereas moment matching and balanced truncation aim at the approximation of input-output behavior. Moreover, the computational cost of the methods differs, which limits the application of balanced truncation to systems of moderate size. On the other hand, balanced truncation has an a priori error bound, which is not the case for the mode displacement and moment matching techniques. Also, balanced truncation and the mode displacement method preserve stability of the high-order model, whereas stability is not guaranteed when applying moment matching.

The overview of the differences and commonalities between the different reduction method facilitates the choice of the reduction technique with the desirable properties for a given reduction problem.

Finally, these differences are illustrated by means of application of the different methods to a common benchmark example.

Chapter 3

Model Order Reduction

A brief introduction to the Krylov subspace projection based methods was given already in the comparison paper presented in Chapter 2. This chapter provides all Krylov-related theorems, i.e., all Krylov-related results which are used in the upcoming chapters. The results themselves have been published in different works and papers, for instance in [36, 26, 51], what we add are a uniform notation and comprehensive proofs based on the fewest possible assumptions.

3.1 Introduction

In this section, we introduce the linear transient system, that is the subject of reduction, and its transfer function. We also define the reduced-order linear transient system and the reduced-order transfer function. Finally, we introduce some basic notations and properties, that will be used in the reminder of the chapter and this thesis.

Definition 3.1.1. *(The linear time-invariant system and its transfer function* $\mathbf{H}$*)*
Let $k, l, n \in \mathbb{N}$*, matrices* $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{n \times n}$*,* $\mathbf{B} \in \mathbb{R}^{n \times k}$*,* $\mathbf{C} \in \mathbb{R}^{n \times l}$*, and* $\mathbf{D} \in \mathbb{R}^{l \times k}$*. Matrices* $\mathbf{B}, \mathbf{C}$ *are called input map resp. output map. According to [3, (4.11), (4.12)] the linear transient system in the time-domain*

$$
\begin{aligned}
\mathbf{E}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\
\mathbf{y}(t) &= \mathbf{C}^{\mathrm{T}}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)
\end{aligned}
\tag{3.1.1}
$$

is (after a Laplace transform, assuming $\mathbf{x}(0) = \mathbf{0}$) related to the system

$$s\mathbf{E}\mathbf{X}(s) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s)$$
$$\mathbf{Y}(s) = \mathbf{C}^{\mathrm{T}}\mathbf{X}(s) + \mathbf{D}\mathbf{U}(s), \tag{3.1.2}$$

which has its linear relation between input(s) $\mathbf{U}(s)$ and output(s) $\mathbf{Y}(s)$ determined by the transfer function

$$\mathbf{H}(s) = \underbrace{\mathbf{C}^{\mathrm{T}}}_{l\times n}(s\underbrace{\mathbf{E}}_{n\times n} - \underbrace{\mathbf{A}}_{n\times n})^{-1}\underbrace{\mathbf{B}}_{n\times k} + \underbrace{\mathbf{D}}_{l\times k}, \tag{3.1.3}$$

assuming that pencil $s \mapsto \mathbf{F}(s)$, $\mathbf{F}(s) = s\mathbf{E} - \mathbf{A}$, is regular (only singular for a finite number of $s \in \mathbb{C}$ ([28, below (9)]). Matrix $\mathbf{E}$ does not need to be non-singular.

Definition 3.1.2. *(The reduced linear transient system and its transfer function $\hat{\mathbf{H}}$)*
Let $k, l, n \in \mathbb{N}$, matrices $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times k}$, $\mathbf{C} \in \mathbb{R}^{n\times l}$, and $\mathbf{D} \in \mathbb{R}^{l\times k}$. Let $V, W \subset \mathbb{C}^n$ be linear vector spaces and let $m := \dim(V) = \dim(W)$ be the dimension of V. Assume that $V = \mathrm{colspan}\mathbf{V}$, $W = \mathrm{colspan}\mathbf{W}$ and that $\mathbf{V}$ and $\mathbf{W}$ are of full column rank.
The linear reduced transient system in the time-domain

$$\mathbf{W}^{\mathrm{T}}\mathbf{E}\mathbf{V}\dot{\mathbf{x}}(t) = \mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V}\mathbf{x}(t) + \mathbf{W}^{\mathrm{T}}\mathbf{B}\mathbf{u}(t)$$
$$\hat{\mathbf{y}}(t) = \mathbf{C}^{\mathrm{T}}\mathbf{V}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \tag{3.1.4}$$

is (after a Laplace transform, assuming $\mathbf{x}(0) = \mathbf{0}$) related to the reduced Laplace system

$$s\mathbf{W}^{\mathrm{T}}\mathbf{E}\mathbf{V}\mathbf{X}(s) = \mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V}\mathbf{X}(s) + \mathbf{W}^{\mathrm{T}}\mathbf{B}\mathbf{U}(s)$$
$$\hat{\mathbf{Y}}(s) = \mathbf{C}^{\mathrm{T}}\mathbf{V}\mathbf{X}(s) + \mathbf{D}\mathbf{U}(s), \tag{3.1.5}$$

which has a linear relation between input(s) $\mathbf{U}(s)$ and output(s) $\hat{\mathbf{Y}}$ determined by the reduced transfer function $\hat{\mathbf{H}}$ which depends on the linear spaces V and W:

$$\hat{\mathbf{H}}(s) = \mathbf{C}^{\mathrm{T}}\mathbf{V}(s\underbrace{\mathbf{W}^{\mathrm{T}}}_{m\times n}\mathbf{E}\underbrace{\mathbf{V}}_{n\times m} - \mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{B} + \mathbf{D}$$
$$= \underbrace{\mathbf{C}^{\mathrm{T}}\mathbf{V}}_{l\times m}(\underbrace{s\mathbf{W}^{\mathrm{T}}\mathbf{E}\mathbf{V} - \mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V}}_{m\times m})^{-1}\underbrace{\mathbf{W}^{\mathrm{T}}\mathbf{B}}_{m\times k} + \mathbf{D}$$
$$= \hat{\mathbf{C}}^{\mathrm{T}}(s\hat{\mathbf{E}} - \hat{\mathbf{A}})^{-1}\hat{\mathbf{B}} + \hat{\mathbf{D}}, \tag{3.1.6}$$

$$V = \{\mathbf{V}\mathbf{x} \colon \mathbf{x} \in \mathbb{R}^m\},$$
$$W = \{\mathbf{W}\mathbf{x} \colon \mathbf{x} \in \mathbb{R}^m\},$$

where $\hat{\mathbf{D}} = \mathbf{D}$, $\hat{\mathbf{C}}^{\mathrm{T}} = \mathbf{C}^{\mathrm{T}}\mathbf{V} \implies \hat{\mathbf{C}} = \mathbf{V}^{\mathrm{T}}\mathbf{C}$, $\hat{\mathbf{E}} = \mathbf{W}^{\mathrm{T}}\mathbf{E}\mathbf{V}$, $\hat{\mathbf{A}} = \mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V}$, and $\hat{\mathbf{B}} = \mathbf{W}^{\mathrm{T}}\mathbf{B}$, assuming that $s \mapsto \hat{\mathbf{F}}(s)$, $\hat{\mathbf{F}}(s) = s\mathbf{W}^{\mathrm{T}}\mathbf{E}\mathbf{V} - \mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V}$, is non-singular for all $s \in \mathbb{C}$, (see remark 3.1.1 below). The space V consists of vectors in $\mathbb{R}^n$. Since it is assumedly spanned by m linearly independent column-vectors, the related matrix is $\mathbf{V} \in \mathbb{R}^{n\times m}$ which implies that if $\mathbf{v} \in V$ then there exists an $\mathbf{x} \in \mathbb{R}^m$ such that $\mathbf{v} = \mathbf{V}\mathbf{x}$.

Remark 3.1.1. For $\hat{\mathbf{F}}$ to be square **and** nonsingular $\mathbf{V}$ and $\mathbf{W}$ need to be of the same order **and** of maximal column rank: Assume that $\mathbf{V}$ is not of maximal column rank. Then there exists a coordinate vector $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^m$ such that $\mathbf{Vx} = \mathbf{0}$ which implies that $\hat{\mathbf{F}}\mathbf{x} = \mathbf{W}^{\mathrm{T}}\mathbf{F}\mathbf{V}\mathbf{x} = \mathbf{0}$ which shows that $\hat{\mathbf{F}}$ is singular. Thus, $\hat{\mathbf{F}}$ non-singular implies $\mathbf{V}, \mathbf{W}$ must be of maximal column rank which implies that at its turn $m \leq n$. For a matrix of column-vectors $\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_k] \in \mathbb{R}^{m \times k}$ we interpret $\mathbf{Vx}$ as the element $\mathbf{Vx} = (\mathbf{Vx}_1, \ldots, \mathbf{Vx}_k) \in V^k$.

Notation and basic relations

For the remainder of this thesis, we assume, that the *pencil $s\mathbf{E} - \mathbf{A}$* is regular, i.e., that it is singular only for a finite number of $s \in \mathbb{C}$. We also assume that the matrix $\mathbf{E}$ is non-singular, whenever its inverse is explicitly used.

From now on we use abbreviations $\mathbf{F}$, $\mathbf{G}$, $\hat{\mathbf{F}}$, and $\hat{\mathbf{G}}$

$$\mathbf{F} \equiv \mathbf{F}(s) = s\mathbf{E} - \mathbf{A} \qquad \in \mathbb{C}^{n \times n}, \tag{3.1.7}$$

$$\mathbf{G} \equiv \mathbf{G}(s) = (s\mathbf{E} - \mathbf{A})^{-1} \qquad \in \mathbb{C}^{n \times n}, \tag{3.1.8}$$

$$\hat{\mathbf{F}} \equiv \hat{\mathbf{F}}(s) = s\hat{\mathbf{E}} - \hat{\mathbf{A}} \qquad \in \mathbb{C}^{m \times m}, \tag{3.1.9}$$

$$\hat{\mathbf{G}} \equiv \hat{\mathbf{G}}(s) = (s\hat{\mathbf{E}} - \hat{\mathbf{A}})^{-1} \qquad \in \mathbb{C}^{m \times m}. \tag{3.1.10}$$

To group the factors of a matrix product $\mathbf{ABC}$ we write *matrix product* $\mathbf{A} \circ \mathbf{B} \circ \mathbf{C}$. The *operator ∂* stands for *the derivative* or a partial derivative.

Because later on we compare derivatives of matrix-valued functions of a scalar complex value, we present several basic derivative relations.

Definition 3.1.3. $\mathbb{C}^d \ni \mathbf{x} \mapsto \mathbf{A}(\mathbf{x}) \in \mathbb{C}^{n \times m}$ *a matrix of which each coefficient depends on the parameters* $(x_1, \ldots, x_d) = \mathbf{x} \in \mathbb{C}^d$. *Let* $p \in \{1, \ldots, d\}$. *Then the pth partial derivative of* $\mathbf{A}$ *is*

$$\mathbf{x} \mapsto \partial_p \mathbf{A}(\mathbf{x}) = \begin{bmatrix} \partial_p a_{11}(\mathbf{x}) & \cdots & \partial_p a_{1n}(\mathbf{x}) \\ \vdots & & \vdots \\ \partial_p a_{n1}(\mathbf{x}) & \cdots & \partial_p a_{nn}(\mathbf{x}) \end{bmatrix} \in \mathbb{C}^{n \times m}. \tag{3.1.11}$$

Lemma 3.1.1. *Let* $\mathbf{x} \in \mathbb{C}^d$, *and* $p \in \{1, \ldots, d\}$. *Let* $\mathbf{x} \mapsto \mathbf{A}(\mathbf{x}) \in \mathbb{C}^{n \times m}$ *be a matrix-function and let* $\mathbf{B} \in \mathbb{C}^{k \times n}$ *and* $\mathbf{C} \in \mathbb{C}^{m \times l}$ *be independent of* $\mathbf{p}$. *Then*

$$\partial_p (\mathbf{B} \circ \mathbf{A} \circ \mathbf{C}) = \mathbf{B} \circ (\partial_p \mathbf{A}) \circ \mathbf{C}. \tag{3.1.12}$$

Let $\mathbf{x} \mapsto \mathbf{B}(\mathbf{x}) \in \mathbb{C}^{m \times l}$. *Then*

$$\partial_p (\mathbf{A} \circ \mathbf{B}) = (\partial_p \mathbf{A}) \circ \mathbf{B} + \mathbf{A} \circ (\partial_p \mathbf{B}). \tag{3.1.13}$$

Assume that $n = m$ and $\mathbf{x} \mapsto \mathbf{A}(\mathbf{x})$ is non-singular for all $\mathbf{x} \in \mathbb{C}^d$ then

$$\partial_p \left(\mathbf{A}^{-1} \right) = -\mathbf{A}^{-1} \circ \partial_p \mathbf{A} \circ \mathbf{A}^{-1}. \tag{3.1.14}$$

If in addition $\partial_p \mathbf{A}$ does not depend on variable $\mathbf{x}$ then

$$\partial_p^k \left(\mathbf{A}^{-1} \right) = (-1)^k \cdot k! \cdot \left(\mathbf{A}^{-1} \partial_p \mathbf{A} \right)^k \mathbf{A}^{-1}. \tag{3.1.15}$$

Let $s \in \mathbb{C}$, $s \mapsto \mathbf{A}(s) \in \mathbb{C}^{n \times m}$ be a matrix-function. Then

$$\partial \left(s\mathbf{A}(s) \right) = \mathbf{A}(s) + s\partial \mathbf{A}(s). \tag{3.1.16}$$

Theorem 3.1.1. *Let $s \mapsto \mathbf{A}(s) \in \mathbb{C}^{n \times n}$ be a matrix-function of one degree of freedom s. Then*

$$\partial^k (s\mathbf{A}) = k\partial^{k-1}\mathbf{A} + s\partial^k\mathbf{A}, \quad k = 0, 1, 2, \ldots. \tag{3.1.17}$$

3.2 The moments of the transfer function $\mathbf{H}(s)$

In this section, we derive formulas for the higher-order *derivatives of* $\mathbf{H}$ given by (3.1.3). Based on these results, we define so called moments of the transfer function, which will be used in the further sections as a measure of the approximation accuracy of the reduced-order model. We assume that the transfer function is an analytic function, i.e., that (3.2.2) exists for all s in a topologically open neighborhood of s_0.

Theorem 3.2.1. *Let $\mathbf{A}, \ldots, \mathbf{H}$ be as in (3.1.3), i.e., $\mathbf{H}(s) = \mathbf{C}^T(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$. Then for all integer $i > 0$*

$$\partial^i \mathbf{H}(s) = (-1)^i i! \cdot \mathbf{C}^T \left(\mathbf{G}(s)\mathbf{E} \right)^i \mathbf{G}(s)\mathbf{B} = (-1)^i i! \cdot \mathbf{C}^T \mathbf{G}(s) \left(\mathbf{E}\mathbf{G}(s) \right)^i \mathbf{B} \tag{3.2.1}$$

and (based on a Taylor series expansion around $s_0 \in \mathbb{C}$)

$$\mathbf{H}(s) = \sum_{i=0}^{\infty} \frac{(s - s_0)^i}{i!} \partial^i \mathbf{H}(s_0) = \mathbf{D} + \sum_{i=0}^{\infty} (-1)^i (s - s_0)^i \mathbf{C}^T (\mathbf{G}(s_0)\mathbf{E})^i \mathbf{G}(s_0)\mathbf{B}, \tag{3.2.2}$$

for all s in an open neighborhood of s_0, which for $s_0 = 0$ and $\mathbf{E} = \mathbf{I}$ reduces to

$$\mathbf{H}(s) = \mathbf{D} - \sum_{i=0}^{\infty} s^i \mathbf{C}^T \mathbf{A}^{-(i+1)} \mathbf{B},$$

for all s in an open neighborhood of 0.

Proof. For $i > 0$

$$\partial^i \mathbf{H}(s) \underset{(3.1.12)}{=} \mathbf{C}^{\mathrm{T}}(\partial^i \mathbf{G}(s))\mathbf{B} \underset{(3.1.15)}{=} (-1)^i i! \mathbf{C}^{\mathrm{T}}(\mathbf{G}(s)\mathbf{E})^i \mathbf{G}(s)\mathbf{B}$$

as was to be shown.

Furthermore

$$
\begin{aligned}
\mathbf{H}(s) &= \sum_{i=0}^{\infty} \frac{(s - s_0)^i}{i!} \partial^i \mathbf{H}(s_0) \\
&\underset{(3.2.1)}{=} \mathbf{D} + \sum_{i=0}^{\infty} \frac{(s - s_0)^i}{i!} (-1)^i i! \cdot \mathbf{C}^{\mathrm{T}}(\mathbf{G}(s_0)\mathbf{E})^i \mathbf{G}(s_0)\mathbf{B} \\
&= \mathbf{D} + \sum_{i=0}^{\infty} (s - s_0)^i (-1)^i \cdot \mathbf{C}^{\mathrm{T}}(\mathbf{G}(s_0)\mathbf{E})^i \mathbf{G}(s_0)\mathbf{B},
\end{aligned}
$$

which for $s_0 = 0$ and $\mathbf{E} = \mathbf{I}$ reduces to

$$
\begin{aligned}
\mathbf{H}(s) &\underset{s_0 \equiv 0}{=} \mathbf{D} + \sum_{i=0}^{\infty} (-s)^i \mathbf{C}^{\mathrm{T}} \left((0 \cdot \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}\right)^i (0 \cdot \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \\
&= \mathbf{D} + \sum_{i=0}^{\infty} (-s)^i (-1)^{i+1} \mathbf{C}^{\mathrm{T}} \mathbf{A}^{-(i+1)} \mathbf{E}^i \mathbf{B} \\
&\underset{\mathbf{E} = \mathbf{I}}{=} \mathbf{D} - \sum_{i=0}^{\infty} s^i \mathbf{C}^{\mathrm{T}} \mathbf{A}^{-(i+1)} \mathbf{B},
\end{aligned}
$$

for all s in an open neighborhood of 0. $\qquad\square$

Definition 3.2.1. *(The ith moment of the transfer function $\mathbf{H}$)*
Assuming $\mathbf{D} = \mathbf{0}$, the ith moment of $\mathbf{H}$ at $s_0 \in \mathbb{C}$ is defined as (see [46])

$$M_i(s_0) = \mathbf{C}^{\mathrm{T}}(\mathbf{G}(s_0)\mathbf{E})^i \mathbf{G}(s_0)\mathbf{B}. \tag{3.2.3}$$

3.3 Moments at infinity

To obtain the Taylor series expansion at infinity, one substitutes $1/s$ for s and calculates the Taylor series' coefficients at $s = 0$. Let $T(s) = 1/s$. One can not straightforwardly apply the chain-rule to $\mathbf{H}_\infty(s) = \mathbf{H}(1/s) = (\mathbf{H} \circ T)(s)$, since it leads to

$$\partial \mathbf{H}_\infty(s) = \partial(\mathbf{H} \circ T)(s) = \partial \mathbf{H}(T(s)) \cdot \partial T(s),$$

where the factor $\partial T(s) = -1/s^2 \to -\infty$ for $s \to 0$. Therefore we follow an alternative approach based on a rewritten form of $\mathbf{H}_\infty$:

$$\mathbf{H}_\infty(s) = \mathbf{H}(1/s) = \mathbf{C}^{\mathrm{T}} \mathbf{G}(1/s)\mathbf{B} + \mathbf{D} = \mathbf{C}^{\mathrm{T}}(s \cdot \underbrace{\frac{1}{s}\mathbf{G}(\frac{1}{s})}_{\mathbf{G}_\infty(s)})\mathbf{B} + \mathbf{D} = \mathbf{C}^{\mathrm{T}} s \cdot \mathbf{G}_\infty(s)\mathbf{B} + \mathbf{D}, \tag{3.3.1}$$

where

$$\mathbf{G}_\infty(s) = \frac{1}{s}\mathbf{G}(\frac{1}{s}) = (\mathbf{E}/s - \mathbf{A})^{-1}/s = ((\mathbf{E} - s\mathbf{A})/s)^{-1}/s = (\mathbf{E} - s\mathbf{A})^{-1} = \left(\underbrace{(s(-\mathbf{A}) - (-\mathbf{E}))}_{\mathbf{F}_\infty(s)}\right)^{-1}$$

(3.3.2)

is differentiable for $s \downarrow 0$. This implies that for i fixed

$$\partial^i\mathbf{G}_\infty(s) \underset{(3.1.15)}{=} (-1)^i i!\,(\mathbf{G}_\infty\partial\mathbf{F}_\infty)^i\,\mathbf{G}_\infty = (-1)^i i!\,(\mathbf{G}_\infty \circ -\mathbf{A})^i\,\mathbf{G}_\infty = i!\,(\mathbf{G}_\infty \circ \mathbf{A})^i\,\mathbf{G}_\infty.$$

(3.3.3)

Hence for all $i > 0$

$$\begin{aligned}
\partial^i\mathbf{H}_\infty(s) \underset{(3.3.1)}{=}\ & \partial^i\mathbf{C}^\mathrm{T} s\mathbf{G}_\infty(s)\mathbf{B} + \partial^i\mathbf{D} \\
\underset{(3.1.12)}{=}\ & \mathbf{C}^\mathrm{T}[\partial^i\,(s\mathbf{G}_\infty(s))\,]\mathbf{B} \\
\underset{(3.1.17)}{=}\ & \mathbf{C}^\mathrm{T}\left[i\partial^{i-1}\mathbf{G}_\infty(s) + s\partial^i\mathbf{G}_\infty(s)\right]\mathbf{B} \\
\underset{(3.3.3)}{=}\ & \mathbf{C}^\mathrm{T}\left[i(i-1)!\cdot(\mathbf{G}_\infty(s)\mathbf{A})^{i-1}\mathbf{G}_\infty(s) + s\cdot i!\cdot(\mathbf{G}_\infty(s)\mathbf{A})^i\mathbf{G}_\infty(s)\right]\mathbf{B} \\
=\ & \mathbf{C}^\mathrm{T}\left[i!\left((\mathbf{G}_\infty(s)\mathbf{A})^{i-1}\mathbf{G}_\infty(s) + s(\mathbf{G}_\infty(s)\mathbf{A})^i\mathbf{G}_\infty(s)\right)\right]\mathbf{B}.
\end{aligned}$$

Let $i > 0$. Since $\mathbf{G}_\infty(0) = \mathbf{E}^{-1}$ one obtains the *derivatives of* $\mathbf{H}$ *at* ∞:

$$\partial^i\mathbf{H}_\infty(0) = i!\cdot\mathbf{C}^\mathrm{T}(\mathbf{E}^{-1}\mathbf{A})^{i-1}\mathbf{E}^{-1}\mathbf{B}.$$

This shows that the Taylor series at infinity of $\mathbf{H}$ is that of $\mathbf{H}_\infty$ at 0, which is

$$\begin{aligned}
\mathbf{H}_\infty(s) &= \sum_{i=0}^\infty \frac{(s-0)^i}{i!}\cdot\partial^i\mathbf{H}_\infty(0) \\
&= \mathbf{D} + \sum_{i=1}^\infty \frac{s^i}{i!}\cdot i!\mathbf{C}^\mathrm{T}\left(\mathbf{E}^{-1}\mathbf{A}\right)^{i-1}\mathbf{E}^{-1}\mathbf{B} \\
&= \mathbf{D} + \sum_{i=1}^\infty s^i\cdot\mathbf{C}^\mathrm{T}\left(\mathbf{E}^{-1}\mathbf{A}\right)^{i-1}\mathbf{E}^{-1}\mathbf{B},
\end{aligned}$$

for all s in an open neighborhood of 0, i.e.,

$$\mathbf{H}(s) = \mathbf{D} + \sum_{i=1}^\infty s^{-i}\cdot\mathbf{C}^\mathrm{T}\left(\mathbf{E}^{-1}\mathbf{A}\right)^{i-1}\mathbf{E}^{-1}\mathbf{B}, \quad s\to\infty.$$

Definition 3.3.1. *(The ith moment of the transfer function* $\mathbf{H}(s)$ *at infinity)*
Assuming $\mathbf{D} = \mathbf{0}$, *the ith moment of the transfer function* $\mathbf{H}(s)$ *at infinity is called a Markov parameter and defined as (see [3])*

$$M_{-i} = \mathbf{C}^\mathrm{T}\left(\mathbf{E}^{-1}\mathbf{A}\right)^{i-1}\mathbf{E}^{-1}\mathbf{B}.$$

(3.3.4)

3.4   Moment matching property of the reduced-order system

As already mentioned in Section 2.2.2, the goal of the Krylov subspace model order reduction is to find a projection-based approximation of the original transfer function, that matches the first k moments of the original transfer function. In other words, the objective is to calculate the reduced-order transfer function $\hat{H}(s)$, whose moment expansion is given by

$$\hat{H}(s) = \sum_{i=0}^{\infty} (-1)^i \hat{M}_i(s_0)(s - s_0)^i, \tag{3.4.1}$$

with

$$\hat{M}_i(s_0) = M_i(s_0), \qquad \text{for} \quad i = 1, \ldots, k, \tag{3.4.2}$$

and $M_i(s_0)$ being the moments of the original transfer function defined in Definition 3.2.1. This is called *the moment matching property* of the reduced-order system. By construction, if the moments of the original and reduced transfer function match, then also the derivatives of these functions match. This means that the reduced transfer function interpolates the original transfer function, i.e., in that sense approximates the original transfer function.

In this section, we prove, that under certain assumptions, one can create a reduced-order model that matches a certain number of the moments of the original system. We begin by defining the Krylov subspace.

Definition 3.4.1. *(The pth Krylov subspace)*
For $n, k \in \mathbb{N}$, $k \leq n$, square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, and vector $\mathbf{B} \in \mathbb{R}^{n \times k}$ define the pth Krylov subspace

$$\mathcal{K}_p(\mathbf{A}, \mathbf{B}) = colspan\{\mathbf{B}, \mathbf{A}\mathbf{B}, \mathbf{A}^2\mathbf{B}, \ldots, \mathbf{A}^{p-1}\mathbf{B}\}. \tag{3.4.3}$$

It is possible that $\dim(\mathcal{K}_p(\mathbf{A}, \mathbf{B})) < pk$ since since some of the columns of $\mathcal{K}_p(\mathbf{A}, \mathbf{B})$ might be linearly dependent. Let V be as in (3.1.6). Assume $m := \dim(V)$ and let in addition $\mathcal{K}_p(\mathbf{A}, \mathbf{B}) \subset V$. One finds $\dim(\mathcal{K}_p(\mathbf{A}, \mathbf{B})) \leq \dim(V) = m$.

Example 3.4.1. *As an example where some of the columns of $\mathcal{K}_p(\mathbf{A}, \mathbf{B})$ are bound to be linearly dependent, consider $\mathbf{A} = \mathbf{I}$ for which $\dim(\mathcal{K}_p(\mathbf{A}, \mathbf{B})) = \dim(\mathbf{B}) \leq k$ for every $p \in \mathbb{N}$.*

In the remainder of this chapter, when possible, we will denote the matrix $\mathbf{G}(s)$ by $\mathbf{G}$.

Lemma 3.4.1. *Let $\mathbf{A}, \ldots, \mathbf{H}$ be as in (3.1.3) and $\hat{\mathbf{A}}, \ldots, \hat{\mathbf{H}}, V, W$ be as in (3.1.6). Then*

$$\hat{\mathbf{G}}^{-1} = \mathbf{W}^\mathsf{T}\mathbf{G}^{-1}\mathbf{V}, \qquad \hat{\mathbf{G}}^{-\mathsf{T}} = \mathbf{V}^\mathsf{T}\mathbf{G}^{-\mathsf{T}}\mathbf{W}. \tag{3.4.4}$$

Proof. The last result is obtained in a similar manner to the first one:

$$\mathbf{W}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{V} = \mathbf{W}^{\mathrm{T}}\mathbf{F}\mathbf{V} = \mathbf{W}^{\mathrm{T}}\left(s\mathbf{E} - \mathbf{A}\right)\mathbf{V} = s\mathbf{W}^{\mathrm{T}}\mathbf{E}\mathbf{V} - \mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{V} = \hat{\mathbf{F}} = \hat{\mathbf{G}}^{-1}.$$

□

Theorem 3.4.1 and Theorem 3.4.2 together are used to show that the moments of the original and reduced-order transfer functions match, for instance in Theorem 3.4.3. The first of the two Theorems shows the relation between **GB** and its reduced variant, the second one the relation between a multiplication with **GE** and a multiplication with its reduced variant, illustrated using the following *commutative diagram*:

$$
\begin{array}{ccccccccc}
\hat{\mathbf{G}}\hat{\mathbf{B}} & =\!=\!= & \mathbf{x}_0 & \xrightarrow{\hat{\mathbf{G}}\hat{\mathbf{E}}} & \mathbf{x}_1 & \xrightarrow{\hat{\mathbf{G}}\hat{\mathbf{E}}} & \mathbf{x}_2 \in \mathbb{C}^m & & (\hat{\mathbf{G}}\hat{\mathbf{E}})^i\hat{\mathbf{G}}\hat{\mathbf{B}} \\
\text{Lem.3.4.1:}\mathbf{V}\downarrow & & \mathbf{V}\downarrow & & \mathbf{V}\downarrow & & \mathbf{V}\downarrow & \Longrightarrow & \mathbf{V}\downarrow \\
\mathbf{GB} & =\!=\!= & \mathbf{V}\mathbf{x}_0 & \xrightarrow[\text{The.3.4.1:}\mathbf{GE}]{} & \mathbf{V}\mathbf{x}_1 & \xrightarrow{\mathbf{GE}} & \mathbf{V}\mathbf{x}_2 \in \mathbb{C}^n & & (\mathbf{GE})^i\mathbf{GB}
\end{array}
$$

(in Theorem 3.4.1, by construction of the Krylov space **V**, there exists a **y** to each **x** such that $\mathbf{Vy} = \mathbf{GEx}$).

Theorem 3.4.1. *If* $\mathbf{A}, \dots, \mathbf{H}$ *are as in* (3.1.3)*,* $\hat{\mathbf{A}}, \dots, \hat{\mathbf{H}}, \mathbf{V}, \mathbf{W}$ *are as in* (3.1.6) *and* $\mathbf{GB} \in V^k$*, resp.* $\mathbf{G}^{\mathrm{T}}\mathbf{C} \in W^l$ *then*

$$\mathbf{V}\underbrace{\hat{\mathbf{G}}\hat{\mathbf{B}}}_{m\times k} = \mathbf{GB} \in \mathbb{C}^{n\times k}, \qquad \mathbf{W}\underbrace{\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{C}}}_{m\times l} = \mathbf{G}^{\mathrm{T}}\mathbf{C} \in \mathbb{C}^{n\times l}. \tag{3.4.5}$$

Proof. Because $\mathbf{GB} \in V^k$ there exists $\mathbf{x} \in \mathbb{R}^{m\times k}$ such that $\mathbf{Vx} = \mathbf{GB}$ and

$$\mathbf{V}\hat{\mathbf{G}}\hat{\mathbf{B}} \underset{(3.1.6)}{=} \mathbf{V}\hat{\mathbf{G}}\mathbf{W}^{\mathrm{T}}\mathbf{B} = \mathbf{V}\hat{\mathbf{G}}\mathbf{W}^{\mathrm{T}}\mathbf{G}^{-1}\underbrace{\mathbf{GB}}_{\mathbf{Vx}} \underset{(3.4.4)}{=} \mathbf{V}\hat{\mathbf{G}}\underbrace{\mathbf{W}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{V}}_{\hat{\mathbf{G}}^{-1}}\mathbf{x} = \mathbf{GB}. \tag{3.4.6}$$

Thus $\hat{\mathbf{G}}\hat{\mathbf{B}}$ are the coordinates of **GB** relative to the linearly independent columns of **V**. Next, because $\mathbf{G}^{\mathrm{T}}\mathbf{C} \in W^k$ there exists $\mathbf{x} \in \mathbb{R}^{m\times l}$ such that $\mathbf{Wx} = \mathbf{G}^{\mathrm{T}}\mathbf{C}$ and

$$\mathbf{W}\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{C}} \underset{(3.1.6)}{=} \mathbf{W}\hat{\mathbf{G}}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{C} = \mathbf{W}\hat{\mathbf{G}}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{G}^{-\mathrm{T}}\underbrace{\mathbf{G}^{\mathrm{T}}\mathbf{C}}_{\mathbf{Wx}} \underset{(3.4.4)}{=} \mathbf{W}\hat{\mathbf{G}}\underbrace{\mathbf{V}^{\mathrm{T}}\mathbf{G}^{-\mathrm{T}}\mathbf{W}}_{\hat{\mathbf{G}}^{-\mathrm{T}}}\mathbf{x} = \mathbf{G}^{\mathrm{T}}\mathbf{C}. \tag{3.4.7}$$

□

Theorem 3.4.2. *Let* $\mathbf{A}, \dots, \mathbf{H}$ *be as in* (3.1.3) *and* $\hat{\mathbf{A}}, \dots, \hat{\mathbf{H}}, V, W$ *be as in* (3.1.6)*. Let* $k \in \mathbb{N}$*. Then (independent of* **W***) for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m\times k}$

$$\mathbf{Vy} = (\mathbf{GE})\,\mathbf{Vx} \implies \mathbf{y} = \left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)\mathbf{x} \tag{3.4.8}$$

$$\implies \mathbf{V}\underbrace{\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)\mathbf{x}}_{\mathbf{y}} = (\mathbf{GE})\,\mathbf{Vx}. \tag{3.4.9}$$

Similarly (independent of $\mathbf{V}$*) for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times k}$

$$\mathbf{W}\mathbf{y} = (\mathbf{EG})^{\mathsf{T}} \mathbf{W}\mathbf{x} \implies \mathbf{y} = \left(\hat{\mathbf{E}}\hat{\mathbf{G}}\right)^{\mathsf{T}} \mathbf{x} \tag{3.4.10}$$

$$\implies \mathbf{W} \underbrace{\left(\hat{\mathbf{E}}\hat{\mathbf{G}}\right)^{\mathsf{T}} \mathbf{x}}_{\mathbf{y}} = (\mathbf{EG})^{\mathsf{T}} \mathbf{W}\mathbf{x}. \tag{3.4.11}$$

Proof. To prove (3.4.8) observe that

$$\hat{\mathbf{E}}\mathbf{x} = \mathbf{W}^{\mathsf{T}}\mathbf{E}\mathbf{V}\mathbf{x} = \mathbf{W}^{\mathsf{T}}\mathbf{G}^{-1}(\mathbf{GE})\mathbf{V}\mathbf{x} \underset{\mathbf{V}\mathbf{y}=\mathbf{GEVx}}{=} \mathbf{W}^{\mathsf{T}}\mathbf{G}^{-1}\mathbf{V}\mathbf{y} \underset{(3.4.4)}{=} \hat{\mathbf{G}}^{-1}\mathbf{y}, \tag{3.4.12}$$

after which left-multiplication with $\hat{\mathbf{G}}$ of the left- and right-hand side leads to (3.4.8). Note that for the case of interest $m < n$ the right-hand side of (3.4.8) can not be replaced by $\mathbf{y} = \mathbf{V}^{\mathsf{T}}\mathbf{GEVx}$. Note that $\mathbf{V}$ and $\mathbf{W}$ do not need to have orthogonal columns and that this theorem holds for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times k}$. To prove (3.4.10), observe that

$$\hat{\mathbf{E}}^{\mathsf{T}}\mathbf{x} = \mathbf{V}^{\mathsf{T}}\mathbf{E}^{\mathsf{T}}\mathbf{W}\mathbf{x} = \mathbf{V}^{\mathsf{T}}\mathbf{G}^{-\mathsf{T}}(\mathbf{EG})^{\mathsf{T}}\mathbf{W}\mathbf{x} \underset{\mathbf{W}\mathbf{y}=(\mathbf{EG})^{\mathsf{T}}\mathbf{Wx}}{=} \mathbf{V}^{\mathsf{T}}\mathbf{G}^{-\mathsf{T}}\mathbf{W}\mathbf{y} \underset{(3.4.4)}{=} \hat{\mathbf{G}}^{-\mathsf{T}}\mathbf{y}.$$

The reverse implication $\Leftarrow$ in (3.4.8): $\mathbf{y} = \hat{\mathbf{G}}\hat{\mathbf{E}}\mathbf{x}$ implies

$$\hat{\mathbf{E}}\mathbf{x} = \hat{\mathbf{G}}^{-1}\mathbf{y} \underset{(3.4.12)}{\Longleftrightarrow} \mathbf{W}^{\mathsf{T}}\mathbf{G}^{-1}(\mathbf{GE})\mathbf{V}\mathbf{x} = \mathbf{W}^{\mathsf{T}}\mathbf{G}^{-1}\mathbf{V}\mathbf{y} \implies \mathbf{V}\mathbf{y} - (\mathbf{GE})\mathbf{V}\mathbf{x} \in \mathbf{G}^{-\mathsf{T}}W^{\perp}.$$

$$\square$$

Theorem 3.2.1 showed that for all $s \in \mathbb{C}$ and $i = 1, 2, \ldots$, the ith derivative of the (reduced) transfer function is given by

$$\partial^{i}\mathbf{H}(s) \underset{(3.2.1)}{=} (-1)^{i} \cdot i! \cdot \mathbf{C}^{\mathsf{T}} \left(\mathbf{G}(s)\mathbf{E}\right)^{i} \mathbf{G}(s)\mathbf{B}$$
$$\partial^{i}\hat{\mathbf{H}}(s) \underset{(3.2.1)}{=} (-1)^{i} \cdot i! \cdot \hat{\mathbf{C}}^{\mathsf{T}} \left(\hat{\mathbf{G}}(s)\hat{\mathbf{E}}\right)^{i} \hat{\mathbf{G}}(s)\hat{\mathbf{B}}. \tag{3.4.13}$$

If and only if $\mathbf{D} = \mathbf{0} \in \mathbb{R}^{l \times k}$ then these relations also hold for $i = 0$. Next we show that for specific (possibly with non-orthogonal columns) choices of $\mathbf{V}$ and $\mathbf{W}$ the moments of $\mathbf{H}$ and $\hat{\mathbf{H}}$ match.

Theorem 3.4.3. *Let* $\mathbf{A}, \ldots, \mathbf{H}$ *be as in* (3.1.3) *and* $\hat{\mathbf{A}}, \ldots, \hat{\mathbf{H}}$ *be as in* (3.1.6). *Assume that* V *in* (3.1.6) *is such that* $\mathcal{K}_{p}(\mathbf{GE}, \mathbf{GB}) \subset V$. *Then* $\hat{\mathbf{H}}$ *matches moments* $0, \ldots, p - 1$ *for all* $s \in \mathbb{C}$

$$\partial^{i}\mathbf{H}(s) = \partial^{i}\hat{\mathbf{H}}(s) \qquad \forall i = 0, \ldots, p - 1. \tag{3.4.14}$$

Proof. Let V be spanned by the columns of $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_m]$ and let it contain the vectors $\mathbf{V}^{(0)} = \mathbf{GB} \in \mathbb{C}^{n \times k}$ and $\mathbf{V}^{(i+1)} = (\mathbf{GE})\,\mathbf{V}^{(i)}$, $i = 0, \ldots, p-1$. For all $i = 0, \ldots, p-1$ define $\mathbf{x}^{(i)} \in \mathbb{R}^{m \times k}$ by $\mathbf{Vx}^{(i)} := \mathbf{V}^{(i)} \in V^k$. Then

$$
\begin{aligned}
\mathbf{x}^{(i)} \; &\underset{(3.4.8),i\times}{=} \; \left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^i \mathbf{x}^{(0)} \Longrightarrow \\
\mathbf{Vx}^{(i)} = \mathbf{V}\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^i \mathbf{x}^{(0)} \; &\underset{(3.4.9)}{=} \; (\mathbf{GE})^i \,\mathbf{Vx}^{(0)} \underset{\mathbf{x}^{(0)}=\hat{\mathbf{G}}\hat{\mathbf{B}},(3.4.5)}{\Longrightarrow} \\
\mathbf{V}\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^i \hat{\mathbf{G}}\hat{\mathbf{B}} \; &= \; (\mathbf{GE})^i \,\mathbf{GB} \Longrightarrow \\
\underbrace{\mathbf{C}^{\mathrm{T}}\mathbf{V}}_{\hat{\mathbf{C}}^{\mathrm{T}}}\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^i \hat{\mathbf{G}}\hat{\mathbf{B}} \; &= \; \mathbf{C}^{\mathrm{T}} (\mathbf{GE})^i \,\mathbf{GB},
\end{aligned}
\tag{3.4.15}
$$

i.e., (3.4.14) holds due to (3.4.13). Matrix $\mathbf{V}$ does not need to have orthogonal columns and $\mathbf{D}$ does not need to be $\mathbf{0}$. By construction $\mathcal{K}_p(\mathbf{GE}, \mathbf{GB}) \subset V$. $\qquad\square$

Let $p \in \mathbb{N}$ and let $m := \dim(V)$ be the dimension of V. Theorem 3.4.3 requires that the linear vector space V at least contains the columns of $\mathbf{V}^{(i)} \in \mathbb{C}^{n \times k}$, $i = 0, \ldots, p-1$. To construct reduced systems we do not work with V directly but instead use matrix $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_m]$ whose column vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ span V. Due to this definition of $\mathbf{V}$ there is little to no direct relation between column $\mathbf{v}_i$ of $\mathbf{V}$ and $\mathbf{V}^{(i)}$: The first one is always a vector, the latter one can consist of multiple column vectors. Also, frequently the vectors $\mathbf{v}_i$ are orthogonal whereas the columns of $\mathbf{V}^{(i)}$ do not need to be. Furthermore, possibly $m < kp$ (if the columns of $\mathbf{V}^{(i)}$, $i = 0, \ldots, p-1$ are not linearly independent and V contains only their span) and $m > kp$ (if in addition to the columns of $\mathbf{V}^{(i)}$, V contains sufficiently many other linearly independent directions – as happens for multi-point moment-matching in Corollary 3.4.1).

Theorem 3.4.4. *Let* $\mathbf{A}, \ldots, \mathbf{H}$ *be as in* (3.1.3) *and* $\hat{\mathbf{A}}, \ldots, \hat{\mathbf{H}}$ *be as in* (3.1.6). *Assume that* W *in* (3.1.6) *is such that* $\mathcal{K}_p(\mathbf{G}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}, \mathbf{G}^{\mathrm{T}}\mathbf{C}) \subset W$. *Then* $\hat{\mathbf{H}}$ *matches moments* $0, \ldots, p-1$ *for all* $s \in \mathbb{C}$

$$
\partial^i \mathbf{H}(s) = \partial^i \hat{\mathbf{H}}(s) \qquad \forall i = 0, \ldots, p-1. \tag{3.4.16}
$$

Proof. Let W be spanned by the columns of $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$ and let it contain the vectors $\mathbf{W}^{(0)} = \mathbf{G}^{\mathrm{T}}\mathbf{C} \in \mathbb{C}^{n \times l}$ and $\mathbf{W}^{(i+1)} = (\mathbf{EG})^{\mathrm{T}}\mathbf{W}^{(i)}$, $i = 0, \ldots, p-1$. For all $i =$

$0, \ldots, p - 1$ define $\mathbf{x}^{(i)} \in \mathbb{R}^{m \times l}$ by $\mathbf{W}\mathbf{x}^{(i)} := \mathbf{W}^{(i)} \in W^l$. Then

$$
\mathbf{x}^{(i)} \underset{(3.4.10),i\times}{=} \left(\left(\hat{\mathbf{E}}\hat{\mathbf{G}}\right)^{\mathrm{T}}\right)^i \mathbf{x}^{(0)} \implies
$$

$$
\mathbf{W}\mathbf{x}^{(i)} = \mathbf{W}\left(\left(\hat{\mathbf{E}}\hat{\mathbf{G}}\right)^{\mathrm{T}}\right)^i \mathbf{x}^{(0)} \underset{(3.4.11)}{=} \left(\left(\mathbf{E}\mathbf{G}\right)^{\mathrm{T}}\right)^i \mathbf{W}\mathbf{x}^{(0)} \underset{\mathbf{x}^{(0)}=\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{C}},(3.4.5)}{\implies}
$$

$$
\mathbf{W}\left(\left(\hat{\mathbf{E}}\hat{\mathbf{G}}\right)^{\mathrm{T}}\right)^i \hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{C}} = \left(\left(\mathbf{E}\mathbf{G}\right)^{\mathrm{T}}\right)^i \mathbf{G}^{\mathrm{T}}\mathbf{C} \underset{(\cdot)^{\mathrm{T}}}{\iff}
$$

$$
\mathbf{C}^{\mathrm{T}}\mathbf{G}\left(\mathbf{E}\mathbf{G}\right)^i = \hat{\mathbf{C}}^{\mathrm{T}}\hat{\mathbf{G}}\left(\hat{\mathbf{E}}\hat{\mathbf{G}}\right)^i \mathbf{W}^{\mathrm{T}} \iff \tag{3.4.17}
$$

$$
\mathbf{C}^{\mathrm{T}}\left(\mathbf{G}\mathbf{E}\right)^i \mathbf{G} = \hat{\mathbf{C}}^{\mathrm{T}}\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^i \hat{\mathbf{G}}\mathbf{W}^{\mathrm{T}} \implies
$$

$$
\mathbf{C}^{\mathrm{T}}\left(\mathbf{G}\mathbf{E}\right)^i \mathbf{G}\mathbf{B} = \hat{\mathbf{C}}^{\mathrm{T}}\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^i \hat{\mathbf{G}}\underbrace{\mathbf{W}^{\mathrm{T}}\mathbf{B}}_{\hat{\mathbf{B}}}
$$

i.e., (3.4.16) holds due to (3.4.13). Matrix $\mathbf{W}$ does not need to have orthogonal columns and $\mathbf{D}$ does not need to be $\mathbf{0}$. □

Theorem 3.4.5 shows that for an appropriate choice of $\mathbf{W}$ one can double the amount of matched moments, and therefore better approximate the original transfer function $\mathbf{H}$.

Theorem 3.4.5. *Let* $\mathbf{A}, \ldots, \mathbf{H}$ *be as in* (3.1.3)*,* $\hat{\mathbf{A}}, \ldots, \hat{\mathbf{H}}$ *be as in* (3.1.6)*. If* V, W *are (generated) as in Theorem 3.4.3 and in Theorem 3.4.4 or if* $\mathbf{A}$ *and* $\mathbf{E}$ *are symmetric,* $\mathbf{B} = \mathbf{C}$*,* V *is generated as in Theorem 3.4.3 and* $W = V$ *then* $\hat{\mathbf{H}}$ *matches moments* $0, \ldots, 2p - 1$ *for all* $s \in \mathbb{C}$

$$
\partial^i \mathbf{H}(s) = \partial^i \hat{\mathbf{H}}(s) \qquad \forall i = 0, \ldots, 2p - 1. \tag{3.4.18}
$$

Proof. First assume that V, W are generated as in Theorem 3.4.3 and in Theorem 3.4.4. From the one but last line of (3.4.15) resp. (3.4.17) (use index j) one obtains

$$
V\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^i \hat{\mathbf{G}}\hat{\mathbf{B}} = \left(\mathbf{G}\mathbf{E}\right)^i \mathbf{G}\mathbf{B}, \qquad \mathbf{C}^{\mathrm{T}}\left(\mathbf{G}\mathbf{E}\right)^j \mathbf{G} = \hat{\mathbf{C}}^{\mathrm{T}}\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^j \hat{\mathbf{G}}\mathbf{W}^{\mathrm{T}} \implies
$$

$$
\mathbf{C}^{\mathrm{T}}\left(\mathbf{G}\mathbf{E}\right)^j \mathbf{G}\mathbf{E}\left(\mathbf{G}\mathbf{E}\right)^i \mathbf{G}\mathbf{B} = \hat{\mathbf{C}}^{\mathrm{T}}\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^j \hat{\mathbf{G}}\underbrace{\mathbf{W}^{\mathrm{T}}\mathbf{E}V}_{\hat{\mathbf{E}}}\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^i \hat{\mathbf{G}}\hat{\mathbf{B}} \iff
$$

$$
\mathbf{C}^{\mathrm{T}}\left(\mathbf{G}\mathbf{E}\right)^{i+j+1} \mathbf{G}\mathbf{B} \underset{(3.4.4)}{=} \hat{\mathbf{C}}^{\mathrm{T}}\left(\hat{\mathbf{G}}\hat{\mathbf{E}}\right)^{i+j+1} \hat{\mathbf{G}}\hat{\mathbf{B}}
$$

for all $i, j \in \{0, \ldots, p - 1\}$. Matrices $\mathbf{V}$ and $\mathbf{W}$ do not need to have orthogonal columns and $\mathbf{D}$ does not need to be $\mathbf{0}$.

Next, assume symmetric $\mathbf{A}$ and $\mathbf{E}$, $\mathbf{B} = \mathbf{C}$, and that V is (generated) as in Theorem 3.4.3. Because $\mathbf{B} = \mathbf{C}$, $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$, $\mathbf{E} = \mathbf{E}^{\mathrm{T}}$ and $\mathbf{X}^{-\mathrm{T}} = \left(\mathbf{X}^{\mathrm{T}}\right)^{-1}$ for every matrix $\mathbf{X}$ imply $\mathbf{G} = \mathbf{G}^{\mathrm{T}}$ W in Theorem 3.4.4 is identical to V in Theorem 3.4.3. □

Theorems 3.4.3 and 3.4.4 can be extended to the case of *multi-point moment-matching* ($s \in \{s_1, s_2, \ldots\}$).

Corollary 3.4.1. Let $s \in \mathbb{C}$ be fixed, write $\mathbf{G}(s)$ instead of $\mathbf{G}$. Theorem 3.4.3 resp. Theorem 3.4.4 only require the inclusions $\mathcal{K}_p(\mathbf{G}(s)\mathbf{E}, \mathbf{G}(s)\mathbf{B}) \subset V$ resp. $\mathcal{K}_p(\mathbf{G}(s)^\mathrm{T}\mathbf{E}^\mathrm{T}, \mathbf{G}(s)^\mathrm{T}\mathbf{C}) \subset W$. Hence to match the required p or $2p$ amount of moments at $s \in \{s_1, s_2, \ldots\}$ one needs to use only V and W such that

$$\bigcup_{i=1,2,\ldots} \mathcal{K}_p(\mathbf{G}(s_i)\mathbf{E}, \mathbf{G}(s_i)\mathbf{B}) \subset V, \qquad \bigcup_{i=1,2,\ldots} \mathcal{K}_p(\mathbf{G}(s_i)^\mathrm{T}\mathbf{E}^\mathrm{T}, \mathbf{G}(s_i)^\mathrm{T}\mathbf{C}) \subset W.$$

Remark 3.4.1. Matching multiple moments of $\mathbf{H}$ at a certain points s_i "implies" that $\hat{\mathbf{H}}$ interpolates $\mathbf{H}$ at s_i (i.e., $\hat{\mathbf{H}}(s_i) = \mathbf{H}(s_i)$) and perhaps approximates $\mathbf{H}$ accurately for s close to s_i.

The *case* $\mathbf{E} = \mathbf{I}$ below is inspired by a remark in [75, p17, -5], we write $\mathbf{GE}$ instead of $\mathbf{G}$, since this is the notation used therein.

Corollary 3.4.2. Let $\mathbf{E} = \mathbf{I}$ and $\mathbf{A}, \ldots, \mathbf{H}$ be as in (3.1.3) and $\hat{\mathbf{A}}, \ldots, \hat{\mathbf{H}}$ be as in (3.1.6). Then substituting the requirement $\mathcal{K}_p(\mathbf{G}^\mathrm{T}\mathbf{E}^\mathrm{T}, \mathbf{C}) \subset W$ for requirement $\mathcal{K}_p(\mathbf{G}^\mathrm{T}\mathbf{E}^\mathrm{T}, \mathbf{G}^\mathrm{T}\mathbf{C}) \subset W$ in Theorem 3.4.4 ensures that $\hat{\mathbf{H}}$ *matches moments* $0, \ldots, 2p - 2$ for all $s \in \mathbb{C}$ in Theorem 3.4.5. In addition substituting the requirement $\mathcal{K}_p(\mathbf{GE}, \mathbf{B}) \subset V$ for requirement $\mathcal{K}_p(\mathbf{GE}, \mathbf{GB}) \subset V$ in Theorem 3.4.3 ensures that $\hat{\mathbf{H}}$ *matches moments* $0, \ldots, 2p - 3$ for all $s \in \mathbb{C}$.

Proof. If $\mathbf{E} = \mathbf{I}$ then

$$\mathcal{K}_i(\mathbf{G}^\mathrm{T}\mathbf{E}^\mathrm{T}, \mathbf{G}^\mathrm{T}\mathbf{C}) = \mathcal{K}_i(\mathbf{G}^\mathrm{T}, \mathbf{G}^\mathrm{T}\mathbf{C}) \subset \mathcal{K}_{i+1}(\mathbf{G}^\mathrm{T}, \mathbf{C}) = \subset \mathcal{K}_{i+1}(\mathbf{G}^\mathrm{T}\mathbf{E}^\mathrm{T}, \mathbf{C})$$

for all i. Hence the requirement

$$\mathcal{K}_i(\mathbf{G}^\mathrm{T}\mathbf{E}^\mathrm{T}, \mathbf{G}^\mathrm{T}\mathbf{C}) \subset W \implies \mathcal{K}_{i-1}(\mathbf{G}^\mathrm{T}\mathbf{E}^\mathrm{T}, \mathbf{G}^\mathrm{T}\mathbf{C}) \subset W$$

which predicts the loss of 1 moment and which can further be corroborated by an example (numerical experiment). $\qquad \square$

Remark 3.4.2. Numerical experiments show for $\mathbf{E} \neq \mathbf{I}$ the proposed alternative Krylov spaces in Corollary 3.4.2 do not need to preserve any moments.

The *Lanczos biorthogonalization* (see [78, Section 7.1.1]) creates the required bases V and W simultaneously.

3.5   Conclusions

In this chapter, we explained the general ideas behind the model reduction methods based on the projections onto Krylov subspaces. These ideas can be also straightforwardly applied to reduce coupled systems. However, this approach may not result in an optimal, with respect to the matrix structure, reduced-order models. In the next chapter, we focus on the reduction methods that are designed for the coupled or interconnected systems.

Chapter 4

Block-structure preserving Model Order Reduction

4.1 Introduction

In the previous chapter we showed how model order reduction can be based on Krylov subspace projections. The starting point for the reduction was a linear time-invariant system, that in the Laplace domain is given by (3.1.2)

$$s\mathbf{E}\mathbf{X}(s) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s)$$
$$\mathbf{Y}(s) = \mathbf{C}^T\mathbf{X}(s).$$

In this chapter we will focus on linear time-invariant systems that describe interconnected respectively coupled systems. We will present a mathematical description of such systems and present an overview of the reduction methods that are especially tailored to be applied to this type of models.

The left side of Figure 4.1 represents a schematic model of an *interconnected system* which consists of four sub-systems and a number of interconnections. These interconnections can be realized in different ways, which will be focused on in Chapter 5. The
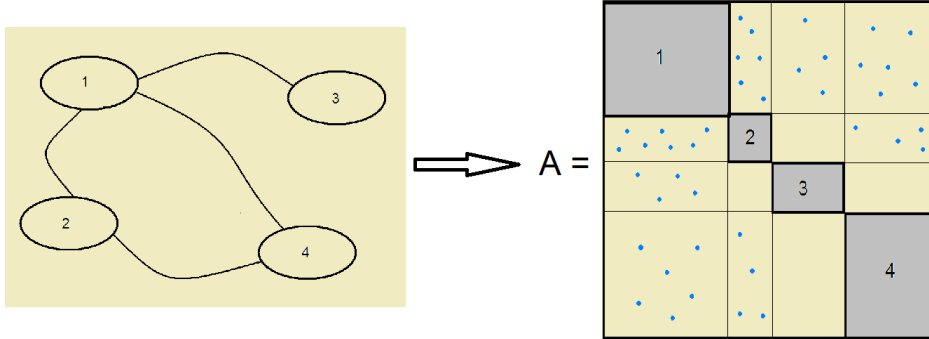
Figure 4.1: Modeling of a coupled system.

right side of Figure 4.1 shows the system matrix $\mathbf{A}$ which corresponds to the graph on the left. The matrix $\mathbf{A}$ has a visible bock-structure. Each of the gray diagonal blocks corresponds to one *sub-system*. The off-diagonal blocks are related to the *interconnections*. The blue dots in the off-diagonal blocks show that the two corresponding sub-systems are *interconnected*. The empty off-diagonal blocks show that there is no coupling between the corresponding two sub-systems.

Generalizing, a system of k components, can be described by a linear system

$$
s \begin{bmatrix} \mathbf{E}_{11} & \cdots & \mathbf{E}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{E}_{k1} & \cdots & \mathbf{E}_{kk} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_k \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{k1} & \cdots & \mathbf{A}_{kk} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_k \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \end{bmatrix} \mathbf{U}
$$

$$
\mathbf{Y} = [\mathbf{C}_1^T, \cdots, \mathbf{C}_k^T] \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_k \end{bmatrix},
$$

(4.1.1)

where the $\mathbf{X}_i \in \mathbb{R}^{N_i}$, $\mathbf{N}_i \in \mathbb{N}$, $i = 1, ..., k$, and the corresponding sub-blocks have compatible dimensions. Naturally, we would like to still be able to recognize this type of *block-structure* in a reduced-order system matrix $\hat{\mathbf{A}}$. Unfortunately, if we apply a Krylov subspace reduction technique from Chapter 3 to the matrix $\mathbf{A}$ we unavoidably loose the block-structure and obtain a non-structured *dense* reduced-order matrix $\hat{\mathbf{A}}$ as shown in Figure 4.2. In the next two sections, we present an overview of Krylov-subspace based block-structure preserving reduction techniques. Such techniques applied to a structured matrix $\mathbf{A}$ result in a reduced-order matrix $\hat{\mathbf{A}}$ like the one shown in Figure 4.3. Although the potential sparse nature of the interconnection off-diagonal blocks is lost, one can still recognize the system's general block-structure. The diagonal blocks still correspond to the reduced-order sub-systems and the zero blocks related to uncoupled sub-systems are preserved. The reduction techniques of this type are called block-structure preserving
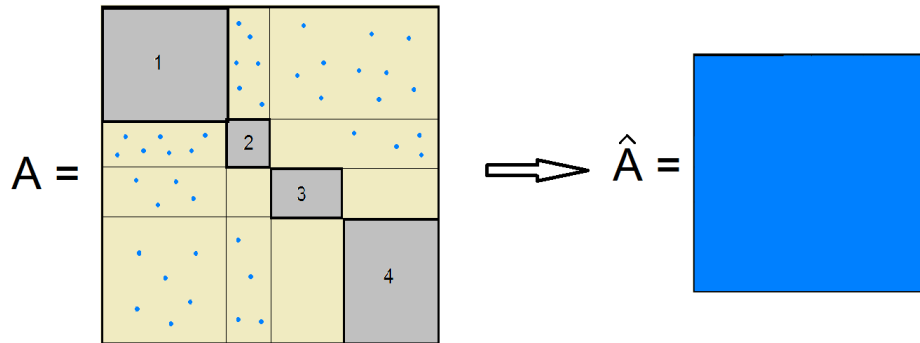
Figure 4.2: Loosing of the structure in the reduced-order matrix $\hat{\mathbf{A}}$.
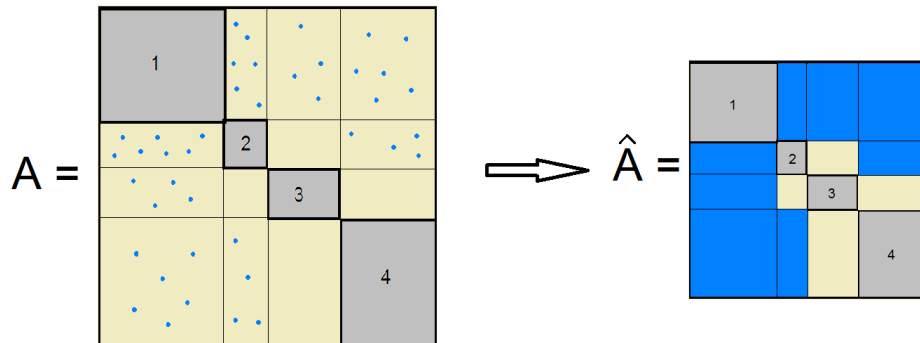
Figure 4.3: Block structure preservation in the reduced-order matrix $\hat{\mathbf{A}}$.

(BSP) methods (see for instance [39]). More information about this type of technique the reader can find in for instance [82].

For the sake of simplicity assume that there are two coupled sub-systems ($k = 2$ in (4.1.1)). Then the system matrix has the block structure

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

We call such a system an *interconnected system* if $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ are explicitly defined by means of their inputs and outputs, i.e., if for instance $\mathbf{A}_{12} = \mathbf{B}_3 \mathbf{C}_4^{\mathsf{T}}$. Elsewise, if $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ are specified in unfactored form, we call the system a *coupled system*. However, it is reasonable to assume that even for the blocks specified in unfactored form there might be defined related input and output operators, i.e., that there can be constructed $\mathbf{B}_3$ and $\mathbf{C}_4$

such that for instance $\mathbf{A}_{12} = \mathbf{B}_3\mathbf{C}_4^{\mathsf{T}}$. Chapter 7 considers possible construction methods for the input and output maps when $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ are specified in unfactored form.

4.2   Moment matching methods for the coupled formulations

We will begin with BSP methods that are directly applicable to coupled systems of the form (4.1.1)

$$
s\begin{bmatrix} \mathbf{E}_{11} & \cdots & \mathbf{E}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{E}_{k1} & \cdots & \mathbf{E}_{kk} \end{bmatrix}\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_k \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{k1} & \cdots & \mathbf{A}_{kk} \end{bmatrix}\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_k \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \end{bmatrix}\mathbf{U}
$$

$$
\mathbf{Y} = [\mathbf{C}_1^T, \cdots, \mathbf{C}_k^T]\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_k \end{bmatrix}.
$$

This type of methods is studied in more detail in for instance [6, 39, 28]. These methods aim at the creation of a reduced-order model whose matrices exhibit the original block-structure and whose transfer function matches a number of moments of the transfer function of the original system (see Section 3.4 for details on moment matching). As for standard Krylov methods, the moment matching property is realized by projecting the original system matrices onto the appropriate input- and/or output-based Krylov subspaces $\tilde{V}$ and $\tilde{W}$ for a chosen expansion point $s \in \mathbb{C}$ (see Theorem 3.4.3 – Theorem 3.4.5). However, to preserve the block structure of the original system, the reduction bases also need to have a special shape. They are created by partitioning the matrices $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{W}}$ into k sub-blocks (with k being the number of sub-systems)

$$
\tilde{\mathbf{V}} = \begin{bmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_k \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_k \end{bmatrix},
$$

where the number of rows in the blocks $\mathbf{V}_i$, $\mathbf{W}_i$, $i = 1, ..., k$ corresponds to the number of rows of the diagonal blocks $\mathbf{A}_{ii}$. Next, the blocks $\mathbf{V}_i$ and $\mathbf{W}_i$ are used to build block-diagonal reduction matrices $\mathbf{V}$ and $\mathbf{W}$

$$
\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & & \\ & \ddots & \\ & & \mathbf{V}_k \end{bmatrix} \quad \text{and} \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & & \\ & \ddots & \\ & & \mathbf{W}_k \end{bmatrix} \tag{4.2.1}
$$

and the reduced-order system is obtained by projecting the original matrices

$$
\hat{\mathbf{A}} = \mathbf{W}^T\mathbf{A}\mathbf{V}, \quad \hat{\mathbf{E}} = \mathbf{W}^T\mathbf{E}\mathbf{V}, \quad \hat{\mathbf{B}} = \mathbf{W}^T\mathbf{B}, \quad \hat{\mathbf{C}} = \mathbf{V}^T\mathbf{C}. \tag{4.2.2}
$$

Note that since the splitting of the matrices $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{W}}$ into sub-blocks may create linearly dependent columns, one needs to apply a *re-orthogonalization* of the matrices $\mathbf{V}$ and $\mathbf{W}$ to remove every possible linear dependence. Moreover, after re-orthogonalization, one has to assure, that the matrices $\mathbf{V}$ and $\mathbf{W}$ have the same number of columns. This can be done by adding the necessary number of random orthogonal columns to the matrix with the smallest amount of columns.

For the reduction bases created in the way described above, the following theorem holds.

Theorem 4.2.1. *Let $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{W}}$ span the input- and output-based Krylov subspaces of the rth order around the expansion point $s \in \mathbb{C}$ for the system* (4.1.1). *If*

$$colspan\tilde{\mathbf{V}} \subseteq colspan\mathbf{V} \qquad and \qquad colspan\tilde{\mathbf{W}} \subseteq colspan\mathbf{W},$$

then a reduced-order system computed as in (4.2.2) *has the transfer function that matches 2p moments of the transfer function of the original system* (4.1.1).

Proof. See Theorem 3.4.5. □

4.2.1 SPRIM

Paper [28] presents *SPRIM*, a structure preserving reduced order method for interconnect macromodeling. Its author focuses on an *RLC circuit* application. Model order reduction methods are of importance to *microchip* manufacturers since complex microchips such as processors contain many interconnected substructures. The relevant equations are (notation as in [28])

$$\mathcal{G}\mathbf{x} + \mathcal{C}\mathbf{x}' = \mathcal{B}u \tag{4.2.3}$$

with

$$\mathcal{G} = \begin{bmatrix} \mathbf{E}_g^{\mathrm{T}}\mathbf{G}\mathbf{E}_g & \mathbf{E}_l^{\mathrm{T}} \\ -\mathbf{E}_l & \mathbf{0} \end{bmatrix}, \quad \mathcal{C} = \begin{bmatrix} \mathbf{E}_c^{\mathrm{T}}\mathbf{C}\mathbf{E}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{L} \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} \mathbf{E}_i^{\mathrm{T}} \\ \mathbf{0} \end{bmatrix},$$

where $\mathbf{G}$, $\mathbf{C}$, and $\mathbf{L}$ are *symmetric positive definite* (square) matrices. The matrices $\mathbf{E}_g$, $\mathbf{E}_c$, $\mathbf{E}_l$ and $\mathbf{E}_i$ are parts of an adjacency matrix $\mathbf{E}$ which describes the connectivity of the electronic circuit, the subscripts g, c, l, i stand for branches containing resistors, capacitors, inductors and current sources. The SPRIM related Laplace domain transfer function $\mathbf{H}_{\mathrm{SPRIM}}$ is

$$\mathbf{H}_{\mathrm{SPRIM}}(s) = \mathcal{B}^{\mathrm{T}} \left(\mathcal{G} + s\mathcal{C} \right)^{-1} \mathcal{B}$$

where $\mathcal{B}, \mathcal{C}$ and $\mathcal{G}$ are re-written

$$\mathcal{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix}, \quad \mathcal{G} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2^{\mathrm{T}} \\ -\mathbf{G}_2 & \mathbf{0} \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0} \end{bmatrix}.$$

The author presents a reduction basis $\mathbf{V}$ of the type (4.2.1) in [28, (21)] and proves in [28, Theorem 3] that it ($\mathbf{W} = \mathbf{V}$) preserves $2p$ moments, double the amount preserved by *PRIMA*. He also addresses the *passivity*.

In this section we present an extension of [28, Theorem 3] based on the notation in Chapter 3. We do not discuss passivity. Related to (4.2.3) the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{E}$ are two by two block matrices with $\mathbf{A}_{22} = \mathbf{0}$ (see [28, Definition of $\mathbf{G}$ in equation 3, p81]). Theorem 4.2.2 below proves [28, Theorem 3, p84], i.e., that all $2p$ moments are matched, even for the case where $\mathbf{A}_{22}$ is symmetric, but not necessarily zero. The preservation of the $2p$ of moments is not due to Theorem 3.4.5 since in the SPRIM case $\mathbf{A}$ is not symmetric and *neither* due to the fact that $\mathbf{V}$ was created by the splitting of a Krylov space.

Theorem 4.2.2. *Let* $\mathbf{A}, \ldots, \mathbf{H}$ *be as in* (3.1.3), $\mathbf{A}_m, \ldots, \mathbf{H}_m$, $V, W, \mathbf{V}, \mathbf{W}$ *be as in* (3.1.6). *In addition let* $n, n_1, n_2 \in \mathbb{N}$, $n_1 + n_2 = n$ *and assume that* $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{E}$ *have block structure* $\mathbf{E}, \mathbf{A}, \mathbf{J} \in \mathbb{R}^{(n_1+n_2) \times (n_1+n_2)}$, $\mathbf{V} \in \mathbb{C}^{(n_1+n_2) \times m}$ *and* $\mathbf{B} \in \mathbb{R}^{(n_1+n_2) \times k}$, *and as in SPRIM satisfy*

$$\mathbf{E} = \begin{bmatrix} \mathbf{Q} & \\ & \mathbf{R} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{M} & \mathbf{N} \\ -\mathbf{N}^{\mathrm{T}} & \mathbf{P} \end{bmatrix}, \quad \mathbf{J} = \begin{bmatrix} \mathbf{I} & \\ & -\mathbf{I} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{U} \\ \mathbf{0} \end{bmatrix},$$

and let $\mathbf{C} = \mathbf{B}$, $\mathbf{M} = \mathbf{M}^{\mathrm{T}}$, $\mathbf{P} = \mathbf{P}^{\mathrm{T}}$, $\mathbf{E} = \mathbf{E}^{\mathrm{T}}$ *(SPRIM assumes* $\mathbf{P} = \mathbf{0}$*). Then*

$$(\mathbf{G}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}})^i \mathbf{G}^{\mathrm{T}}\mathbf{C} = \mathbf{J}(\mathbf{G}\mathbf{E})^i \mathbf{G}\mathbf{B} \implies K_p(\mathbf{G}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}, \mathbf{G}^{\mathrm{T}}\mathbf{C}) = \mathbf{J} K_p(\mathbf{G}\mathbf{E}, \mathbf{G}\mathbf{C}). \tag{4.2.4}$$

If V *is as in Theorem 3.4.3, the columns of* $\mathbf{V}$ *span* V *and* W *is the column span of* $\mathbf{W} = \mathbf{J}\mathbf{V} \in \mathbb{C}^{n \times m}$ *then* $\mathbf{H}_m$ *in* (3.1.6) *matches moments* $0, \ldots, 2p - 1$. *Furthermore, if* V *is as in Theorem 3.4.3 and the columns of the* 2×1 *matrix* $\mathbf{V} = [\mathbf{V}_1; \mathbf{V}_2]$ *span* V *and*

$$\tilde{\mathbf{W}} = \tilde{\mathbf{V}} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix} \in \mathbb{C}^{n \times 2m} \tag{4.2.5}$$

then substituting $\tilde{V}$ *and* $\tilde{W}$ *(the column spans of* $\tilde{\mathbf{V}}$ *respectively* $\tilde{\mathbf{W}}$*) for* V *and* W *in Theorem 3.4.3 respectively in Theorem 3.4.4 also* $\mathbf{H}_m$ *also preserves* $2p$ *moments.*

Proof. First note that $\mathbf{J}^2 = \mathbf{I} \in \mathbb{R}^{n \times n}$, $\mathbf{J}^{-1} = \mathbf{J}$. Using these relations we show for all $\mathbf{X} \in \{\mathbf{A}, \mathbf{E}, \mathbf{F}, \mathbf{G}\}$ that $\mathbf{J}\mathbf{X}\mathbf{J} = \mathbf{X}^{\mathrm{T}}$:

$$\mathbf{J}\mathbf{A}\mathbf{J} = \begin{bmatrix} \mathbf{M} & -\mathbf{N} \\ \mathbf{N}^{\mathrm{T}} & \mathbf{P} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{M}^{\mathrm{T}} & \mathbf{N} \\ -\mathbf{N}^{\mathrm{T}} & \mathbf{P}^{\mathrm{T}} \end{bmatrix} \right)^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}, \qquad \mathbf{J}\mathbf{E}\mathbf{J} = \mathbf{E} \underset{\mathbf{E}=\mathbf{E}^{\mathrm{T}}}{=} \mathbf{E}^{\mathrm{T}}$$

whence in addition

$$\mathbf{J}\mathbf{F}\mathbf{J} = s\mathbf{E}^{\mathrm{T}} - \mathbf{A}^{\mathrm{T}} = (s\mathbf{E} - \mathbf{A})^{\mathrm{T}} = \mathbf{F}^{\mathrm{T}}$$

and

$$\mathbf{J}\mathbf{G}\mathbf{J} = \mathbf{J}\mathbf{F}^{-1}\mathbf{J} = (\mathbf{J}\mathbf{F}\mathbf{J})^{-1} = \mathbf{F}^{-\mathrm{T}} = \mathbf{G}^{\mathrm{T}}.$$

which shows that $K_p(\mathbf{G}^T\mathbf{E}^T, \mathbf{G}^T\mathbf{C}^T)$ contains

$$(\mathbf{G}^T\mathbf{E}^T)^i\mathbf{G}^T\mathbf{C}^T \underset{\mathbf{E}=\mathbf{E}^T,\mathbf{JB}=\mathbf{B},\mathbf{C}=\mathbf{B}}{=} (\mathbf{JGJE})^i\mathbf{JGB} \underset{\mathbf{EJ}=\mathbf{JE}}{=} (\mathbf{JGEJ})^i\mathbf{JGB} = \mathbf{J}(\mathbf{GE})^i\mathbf{GB}$$

which proves (4.2.4).

Secondly, let $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{W}}$ from (4.2.5) and assume that the columns of the unsplitted matrix $\mathbf{V}$ span V which contains $K_p(\mathbf{GE},\mathbf{GC})$. Let $\tilde{V}$ and $\tilde{W}$ be the column spans of $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{W}}$. Because $K_p(\mathbf{GE},\mathbf{GC}) \subset V \subset \tilde{V}$ the first p moments are preserved substituting $\tilde{V}$ for V in Theorem 3.4.3. Because

$$
\begin{aligned}
K_p(\mathbf{G}^T\mathbf{E}^T, \mathbf{G}^T\mathbf{C}) \underset{(4.2.4)}{=} & \ \mathbf{J}K_p(\mathbf{GE},\mathbf{GC}) \\
\subset & \ \mathbf{J}V = \mathbf{J}\text{span}(\mathbf{V}) = \text{span}(\mathbf{JV}) \\
\subset & \ \text{span}(\begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & -\mathbf{V}_2 \end{bmatrix}) = \text{span}(\begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix}) \\
= & \ \text{span}(\tilde{\mathbf{V}}) = \text{span}(\tilde{\mathbf{W}}) = \tilde{W}
\end{aligned}
$$

substituting $\tilde{W}$ for W in Theorem 3.4.4 shows that an additional p moments match. $\qquad\square$

4.2.2 An alternative Krylov method for the coupled formulations

The technique proposed in [22] is motivated by the fact, that for some applications the single-point expansion does not give a sufficient approximation accuracy in the frequency range. On the other hand, using a multi-point expansion can result in excessively large models, especially for systems with many external inputs and outputs. The method proposed in the paper mentioned above, is based on creating a reduction space that consists of a number of sampling matrices $\mathbf{Z}_j$, $j = 1, ..., p$, computed for the system (4.1.1) for p sampling points s_j as follows

$$\mathbf{Z}_j = (s_j\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}.$$

In other words, $\mathbf{Z}_j$, $j = 1, ..., p$ is a vector (or a matrix) that, after projecting the system (4.1.1) onto, will match the 0th moment around the point s_j of the original transfer function, since it consists of the input based starting matrix for the Krylov subspace for s_j as defined in Theorem 3.4.3. After computing p samples, the total sampling matrix $\mathbf{Z}$ is defined as

$$\mathbf{Z} = [\mathbf{Z}_1, ..., \mathbf{Z}_p].$$

Next, following the block-structure presented by the system matrices, matrix $\mathbf{Z}$ is split row-wise into k blocks $\tilde{\mathbf{V}}_i$, $i = 1, ..., k$

$$\mathbf{Z} = \begin{bmatrix} \tilde{\mathbf{V}}_1 \\ \vdots \\ \tilde{\mathbf{V}}_k \end{bmatrix}$$

and a block-diagonal projector is created

$$\tilde{\mathbf{V}} = \begin{bmatrix} \tilde{\mathbf{V}}_1 & & \\ & \ddots & \\ & & \tilde{\mathbf{V}}_k \end{bmatrix}.$$

Finally, the singular value decomposition (SVD) is performed on each of the blocks separately, to produce the orthogonal matrix $\mathbf{V}$

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & & \\ & \ddots & \\ & & \mathbf{V}_k \end{bmatrix}$$

where $\mathbf{V}_i, i = 1, ..., k$ is an orthogonal basis for $\tilde{\mathbf{V}}_i$. At this point, further reduction in size is possible, by removing from the bases $\mathbf{V}_i, i = 1, ..., k$ the columns that correspond to to small singular values. Having the reduction bases $\mathbf{V}$, one can project the original system in the way defined in (4.2.2).

A noticeable advantage of the technique described above is, next to the block-structure preservation, the possibility of reducing different sub-systems with different reduction ratio, determined for each sub-system separately, based on the singular values related to this sub-block as well as the importance of the considered sub-system in the total coupled system.

4.3   Moment matching methods for the uncoupled formulations

Another category of BSP model reduction techniques based on the uncoupled formulation of interconnected sub-systems can be found in for instance [68] or [23]. In this section, we explain the basic ideas based on [89] based on the notation used in this paper (which uses $\mathbf{C}$ instead of $\mathbf{C}^{\mathrm{T}}$).

The system under consideration is an linear system $\mathbf{G}(s)$, composed of k interconnected sub-systems $\mathrm{T}_i(s)$, $i = 1, ...k$. Each sub-system $\mathrm{T}_i(s)$ is assumed to be a linear multiple-input-multiple-output (MIMO) system and has α_i inputs $\mathbf{a}_i$ and β_i outputs $\mathbf{b}_i$ related by

$$\mathbf{b}_i(s) = \mathrm{T}_i(s)\mathbf{a}_i(s). \tag{4.3.1}$$

The input $\mathbf{a}_i$ of each sub-system is a linear combination of the outputs of all sub-systems and of the external input $\mathbf{u}(s) \in \mathbb{R}^m$

$$\mathbf{a}_i(s) = \mathbf{H}_i\mathbf{u}(s) + \sum_{j=1}^{k} \mathbf{K}_{ij}\mathbf{b}_j(s), \tag{4.3.2}$$

where $\mathbf{H}_i \in \mathbb{R}^{\alpha_i \times m}$. The external output $\mathbf{y}(s) \in \mathbb{R}^p$ of $\mathbf{G}(s)$ is a linear combination of the outputs of the sub-systems

$$\mathbf{y}(s) = \sum_{i=1}^{k} \mathbf{F}_i \mathbf{b}_i(s), \tag{4.3.3}$$

with $\mathbf{F}_i \in \mathbb{R}^{p \times \beta_i}$. Let

$$\alpha = \sum_{i=1}^{k} \alpha_i \qquad \text{and} \qquad \beta = \sum_{i=1}^{k} \beta_i.$$

The transfer function of the unconnected sub-systems is given by

$$\mathbf{b}(s) = \mathrm{T}(s)\mathbf{a}(s), \tag{4.3.4}$$

where

$$\mathbf{b}(s) = \begin{bmatrix} \mathbf{b}_1(s) \\ \vdots \\ \mathbf{b}_k(s) \end{bmatrix}, \qquad \mathrm{T}(s) = \begin{bmatrix} \mathrm{T}_1(s) & & \\ & \ddots & \\ & & \mathrm{T}_k(s) \end{bmatrix}, \qquad \mathbf{a}(s) = \begin{bmatrix} \mathbf{a}_1(s) \\ \vdots \\ \mathbf{a}_k(s) \end{bmatrix}$$

are respectively in $\mathbb{R}^\beta$, $\mathbb{R}^{\beta \times \alpha}$, and $\mathbb{R}^\alpha$. Let $\mathbf{F} \in \mathbb{R}^{p \times \beta}$, $\mathbf{K} \in \mathbb{R}^{\alpha \times \beta}$, and $\mathbf{H} \in \mathbb{R}^{\alpha \times m}$

$$\mathbf{F} = [\mathbf{F}_1, ..., \mathbf{F}_k], \qquad \mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \cdots & \mathbf{K}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{K}_{k1} & \cdots & \mathbf{K}_{kk} \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_k \end{bmatrix}. \tag{4.3.5}$$

Using (4.3.5), equations (4.3.2) and (4.3.3) can be rewritten as

$$\mathbf{a}(s) = \mathbf{H}u(s) + \mathbf{K}\mathbf{b}(s) \quad \text{and} \quad \mathbf{y}(s) = \mathbf{F}\mathbf{b}(s). \tag{4.3.6}$$

From (4.3.1) and (4.3.6), it follows that

$$\mathbf{y}(s) = \mathbf{F}[\mathbf{I}_\beta - \mathrm{T}(s)\mathbf{K}]^{-1}\mathrm{T}(s)\mathbf{H}u(s).$$

Let $n_i \in \mathbb{N}$, $i = 1, ..., k$ be the order of $\mathrm{T}_i(s)$ and $(\mathbf{A}_i,\ \mathbf{B}_i,\ \mathbf{C}_i,\ \mathbf{D}_i)$ a minimal state space realization of $\mathrm{T}_i(s)$. If we define

$$n = \sum_{i=1}^{k} n_i,$$

then

$$\mathrm{T}(s) = \mathbf{C}(s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$$

with

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_k \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & & \\ & \ddots & \\ & & \mathbf{B}_k \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & & \\ & \ddots & \\ & & \mathbf{C}_k \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & & \\ & \ddots & \\ & & \mathbf{D}_k \end{bmatrix}.$$

It implies

$$\mathbf{G}(s) = \mathbf{F}[\mathbf{I}_\beta - \mathrm{T}(s)\mathbf{K}]^{-1}\mathrm{T}(s)\mathbf{H} \tag{4.3.7}$$

and a state space realization of $\mathbf{G}(s)$ is given by $(\mathbf{A}_G, \ \mathbf{B}_G, \ \mathbf{C}_G, \ \mathbf{D}_G)$, where

$$
\begin{aligned}
\mathbf{A}_G &= \mathbf{A} + \mathbf{BK}(\mathbf{I}_\beta - \mathbf{DK})^{-1}\mathbf{C}, & \mathbf{B}_G &= \mathbf{B}(\mathbf{I}_\alpha - \mathbf{KD})^{-1}\mathbf{H}, \\
\mathbf{C}_G &= \quad \mathbf{F}(\mathbf{I}_\beta - \mathbf{DK})^{-1}\mathbf{C}, & \mathbf{D}_G &= \mathbf{FD}(\mathbf{I}_\alpha - \mathbf{KD})^{-1}\mathbf{H}.
\end{aligned}
\tag{4.3.8}
$$

If $\mathbf{D} = \mathbf{0}$, the state space realization (4.3.8) of $\mathbf{G}(s)$ reduces to

$$\mathbf{A}_G = \mathbf{A} + \mathbf{BKC}, \quad \mathbf{B}_G = \mathbf{BH}, \quad \mathbf{C}_G = \mathbf{FC} \quad \text{and} \quad \mathbf{D}_G = \mathbf{0}.$$

The following lemma holds:

Lemma 4.3.1. *([89, Lemma 4])*
Let $\lambda \in \mathbb{C}$ be neither a pole of $\mathrm{T}(s)$ nor a pole of $\mathbf{G}(s)$. Define

$$\mathbf{V} := \begin{bmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_k \end{bmatrix} \in \mathbb{C}^{n \times r},$$

such that $\mathbf{V}_i \in \mathbb{C}^{n_i \times r}$. Assume that either

$$\mathcal{K}_k((\lambda\mathbf{I} - \mathbf{A}_G)^{-1}, \ (\lambda\mathbf{I} - \mathbf{A}_G)^{-1}\mathbf{B}_G) \subseteq \text{Im } V$$

or

$$\mathcal{K}_k((\lambda\mathbf{I} - \mathbf{A})^{-1}, \ (\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}) \subseteq \text{Im } V.$$

Construct matrices $\mathbf{Z}_i \in \mathbb{C}^{n_i \times r}$, such that $\mathbf{Z}_i^T \mathbf{V}_i = \mathbf{I}_r$. Project each sub-system as follows:

$$(\hat{\mathbf{A}}_i, \ \hat{\mathbf{B}}_i, \ \hat{\mathbf{C}}_i) := (\mathbf{Z}_i^T \mathbf{A}_i \mathbf{V}_i, \ \mathbf{Z}_i^T \mathbf{B}_i, \ \mathbf{C}_i \mathbf{V}_i).$$

Define

$$\hat{\mathbf{G}}(s) := \hat{\mathbf{C}}_G(s\mathbf{I} - \hat{\mathbf{A}}_G)^{-1}\hat{\mathbf{B}}_G + \mathbf{D}_G.$$

Then, $\hat{\mathbf{G}}(s)$ interpolates $\mathbf{G}(s)$ at λ up to the first k derivatives.

Proof. See the proof of [89, Lemma 4].

$\square$

A comparison of this approach with the model reduction techniques described in Section 4.2, was made for instance in [68]. Some advantages are mentioned, such as the possibility to apply the most suitable reduction method to each sub-system or to speed up the computations using paralellism. Also some drawbacks are mentioned, such as that the separate reduction of the sub-systems usually yields a larger approximated model than a coupled approach, and that sub-system based model reduction is restricted to interconnected systems whose sub-systems have a small number of internal inputs and outputs.

Chapter 5

Separate Bases Reduction Algorithm

5.1 Introduction

Model order reduction techniques, designed especially for coupled or interconnected systems, became a new field of research in recent years. As described in Chapter 4, the common feature of this type of methods is the use of a special block-diagonal form reduction basis $\mathbf{V}$

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & & \\ & \ddots & \\ & & \mathbf{V}_k \end{bmatrix} \tag{5.1.1}$$

that results from the splitting a matrix $\tilde{\mathbf{V}}$ created by a Krylov method applied directly to the coupled system. This approach allows for preservation of the zero-blocks in the coupled system's coefficient matrix. Such blocks appear when two of the sub-systems are not coupled (interconnected) or the coupling holds only in one direction. An example of uni-directional coupling can be a case of a vibrating structure, where the movement of the structure causes acoustic noise, but there is no influence (feedback) of the acoustic behavior of the system on it's dynamics.

Due to the fact that the zero-blocks are preserved in the reduced system, such MOR techniques are called *block structure preserving* (*BSP*) model reduction methods. Their application usually results in a good approximation of the original model. For most of them one can prove the moment matching property. However, this type of methods also has three important drawbacks:

- Though $\mathbf{V}$ in (5.1.1) (possibly) matches the same (amount of) moments as $\tilde{\mathbf{V}}$ it has k times more column vectors and therefore leads to a k times larger reduced system.

- The calculation of $\tilde{\mathbf{V}}$ requires (repeatedly) solving systems with the entire coupled system's coefficient matrix which can be computationally (time- and memory-wise) expensive.

- In practice, the reduction techniques based on an uncoupled formulation of the system (see e.g. [89]) are restricted to the case of interconnected systems with a limited number of interconnections. Otherwise, the reduction procedure is not very efficient, since the dimension of the reduction basis (hence, the reduced-order model) grows very fast. Moreover, such techniques assume that the inputs $\mathbf{B}$ and outputs $\mathbf{C}$ of the sub-systems are both explicitly available. In case of a coupled system these are not explicitly available, only their product $\mathbf{BC}$ is.

In the remainder of this thesis, we will focus on the second and third issue. We present a reduction algorithm suited for systems, coupled through a large number of couplings. In this chapter we introduce a reduction technique based on an uncoupled formulation of a coupled system and extend it to a two-sided method in Chapter 6. The presented algorithm we call a *Separate Bases Reduction* (*SBR*) algorithm. It creates a reduction basis for each sub-system separately, hence is computationally cheaper compared to the reduction techniques that use a coupled formulation such as the BSP methods discussed in Chapter 4. However, the algorithms presented in Chapters 5 and 6 still suffer from the third point of the drawback list presented above. They can be easily applied to interconnected system of a form shown in Figure 5.1, where the sub-systems are not strongly interconnected (i.e. each sub-system exchanges information only with a small number of other sub-systems). In Chapters 7 and 8 we suggest a way to relax this limitation. We also show how to apply the SBR algorithm to strongly coupled systems, i.e. to the systems, where many degrees of freedom of one sub-system are coupled to many degrees of freedom of other sub-systems and where the internal input and output matrices are not explicitly given in the system formulation. Examples of these types of coupled problems are shown in Figure 5.2. Figure 5.2(a) presents a coupled system, that consists of two sub-structures, for instance a solid body and a fluid. The coupling occurs at the interface, where all degrees of freedom of one sub-domain which are sufficiently close to the interface influence similar degrees of freedom of the second sub-domain and vice versa. A different type of *strong coupling* is shown in Figure 5.2(b). This
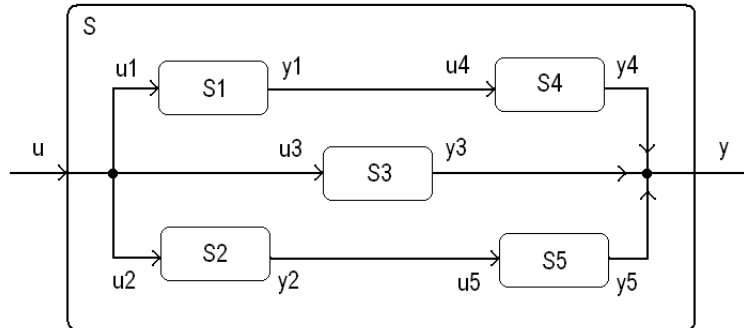
Figure 5.1: Schematic representation of the interconnected system *S*

picture shows a situation, where all degrees of freedom related to both physical quantities *u* and *e* are located inside the same domain. Such situations appear for instance in case of modeling of systems, where the dynamics of the structure is influenced by an electromagnetic field (and vice versa). In the depicted case the change of the velocity of the node $u(x_i)$ influences the electromagnetic field $x \mapsto e(x)$ at the node x_i, and at many nodes in the neighborhood of x_i.

In the next sections of this chapter, we will introduce the linear system describing the situation shown in Figure 5.1 and describe the reduction methodology for this case. Coupled systems of the form as shown in Figure 5.2 will be tackled in Chapters 7 and 8.

5.2 Interconnected system – system definition

In this section we introduce the family of linear *interconnected systems* to which the reduction algorithm is to be applied. For the sake of simplicity we focus on a system of two-subsystems where one sub-system's output is used as a part of the other sub-system's input and vice versa (see for instance Figure 5.3). However, the proposed method can easily be extended to systems composed of an arbitrary number of sub-systems.

5.2.1 The uncoupled formulation

The time domain behavior of each of the sub-systems S_1 and S_2 is modeled by a system of first order differential-algebraic equations after which the frequency domain behavior is obtained via Laplace transformation (see Chapter 3 for more details). For the two

(a) An interface coupling.   (b) A strong coupling between different physical domains.
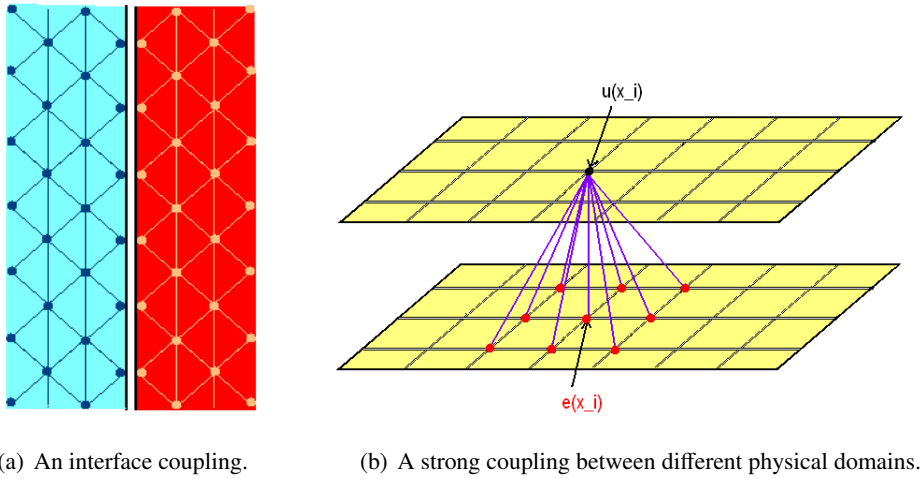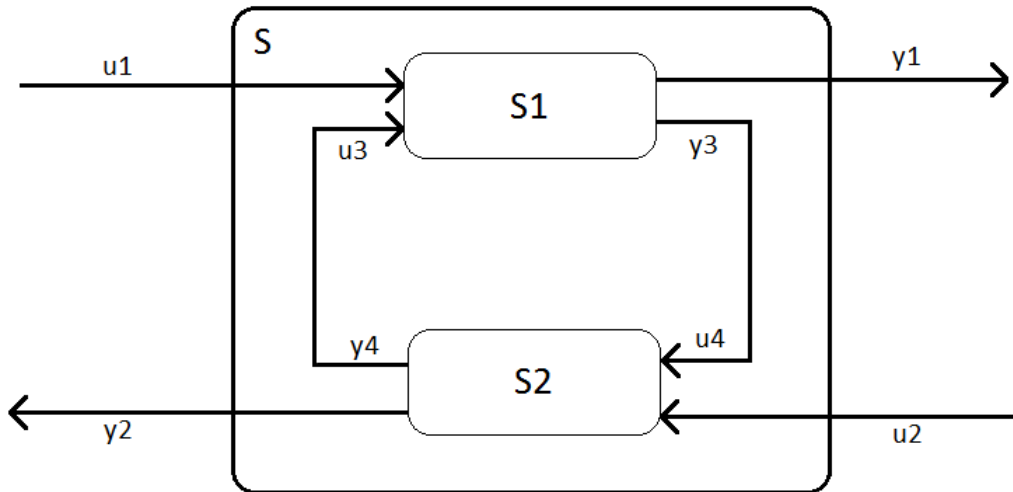
Figure 5.2: Different types of strong coupling

Figure 5.3: Schematic representation of the interconnected system

sub-system example in Figure 5.3, this procedure leads to the Laplace domain systems

$$S_1 \; : \; \begin{cases} s\mathbf{E}_{11}\mathbf{x}_1 = \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{B}_1\mathbf{u}_1 + \mathbf{B}_3\mathbf{u}_3, \\ \quad\; \mathbf{y}_1 = \mathbf{C}_1^T\mathbf{x}_1, \\ \quad\; \mathbf{y}_3 = \mathbf{C}_3^T\mathbf{x}_1, \end{cases}$$

$$S_2 \; : \; \begin{cases} s\mathbf{E}_{22}\mathbf{x}_2 = A_{22}\mathbf{x}_2 + \mathbf{B}_2\mathbf{u}_2 + \mathbf{B}_4\mathbf{u}_4, \\ \quad\; \mathbf{y}_2 = \mathbf{C}_2^T\mathbf{x}_2, \\ \quad\; \mathbf{y}_4 = \mathbf{C}_4^T\mathbf{x}_2. \end{cases}$$

For sub-system i, $i = 1, 2$, let

- $N_i \in \mathbb{N}$ be the number of degrees of freedom of system i;

- $m_i \in \mathbb{N}$ be the number of its *external inputs* of system i;

- $m_{2+i} \in \mathbb{N}$ be the number of its *internal inputs* of system i;

- $p_i \in \mathbb{N}$ be the number of its *external outputs* of system i;

- $p_{2+i} \in \mathbb{N}$ be the number of its *internal outputs* of system i.

Then based thereon

- $\mathbf{A}_{ii}, \mathbf{E}_{ii} \in \mathbb{R}^{N_i \times N_i}$, $i = 1, 2$;

- $\mathbf{B}_i \in \mathbb{R}^{N_1 \times m_i}$, $i = 1, 3$;

- $\mathbf{B}_i \in \mathbb{R}^{N_2 \times m_i}$, $i = 2, 4$;

- $\mathbf{C}_i \in \mathbb{R}^{N_1 \times p_i}$, $i = 1, 3$;

- $\mathbf{C}_i \in \mathbb{R}^{N_2 \times p_i}$, $i = 2, 4$;

- $\mathbf{u}_i \in \mathbb{R}^{m_i}$, $i = 1, \ldots, 4$;

- $\mathbf{y}_i \in \mathbb{R}^{p_i}$, $i = 1, \ldots, 4$.

Using matrix notation, the *system S_1* and *system S_2* can be described as

$$
S_1 : \begin{cases} s\mathbf{E}_{11}\mathbf{x}_1 = \mathbf{A}_{11}\mathbf{x}_1 + [\mathbf{B}_1 \ \mathbf{B}_3] \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_3 \end{bmatrix} \\ \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_3^T \end{bmatrix} \mathbf{x}_1, \end{cases}
\tag{5.2.1}
$$

$$
S_2 : \begin{cases} s\mathbf{E}_{22}\mathbf{x}_2 = \mathbf{A}_{22}\mathbf{x}_2 + [\mathbf{B}_2 \ \mathbf{B}_4] \begin{bmatrix} \mathbf{u}_2 \\ \mathbf{u}_4 \end{bmatrix} \\ \begin{bmatrix} \mathbf{y}_2 \\ \mathbf{y}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_2^T \\ \mathbf{C}_4^T \end{bmatrix} \mathbf{x}_2. \end{cases}
\tag{5.2.2}
$$

5.2.2   The coupled system

When the output of S_1 is used as an input of S_2 and the output of S_2 is used as an input of S_1, equations (5.2.1) and (5.2.2) reduce to an interconnected Laplace domain system. Due to the design of the system depicted in Figure 5.3 one has

$$\begin{cases} \mathbf{u}_3 = \mathbf{y}_4 = \mathbf{C}_4^T \mathbf{x}_2 \\ \mathbf{u}_4 = \mathbf{y}_3 = \mathbf{C}_3^T \mathbf{x}_1, \end{cases} \tag{5.2.3}$$

which in addition implies

$$\begin{cases} m_3 = p_4 \\ m_4 = p_3. \end{cases}$$

Using relation (5.2.3), the interconnected system (5.2.1) can be represented as a single *coupled system* $\mathbf{S}$ of equations

$$S \; : \; \begin{cases} s\mathbf{E}_{11}\mathbf{x}_1 = \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{B}_1\mathbf{u}_1 + \mathbf{B}_3\mathbf{C}_4^T\mathbf{x}_2, \\ s\mathbf{E}_{22}\mathbf{x}_2 = A_{22}\mathbf{x}_2 + \mathbf{B}_2\mathbf{u}_2 + \mathbf{B}_4\mathbf{C}_3^T\mathbf{x}_1, \\ \qquad \mathbf{y}_1 = \mathbf{C}_1^T\mathbf{x}_1, \\ \qquad \mathbf{y}_2 = \mathbf{C}_2^T\mathbf{x}_2 \end{cases}$$

and in matrix form

$$S \; : \; \begin{cases} s\begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \\ \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^T \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}. \end{cases} \tag{5.2.4}$$

Let $N = N_1 + N_2$, $m = m_1 + m_2$, $p = p_1 + p_2$ and define

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix}, \tag{5.2.5}$$

where $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{N \times N}$, $\mathbf{B} \in \mathbb{R}^{N \times m}$, $\mathbf{C} \in \mathbb{R}^{N \times p}$. The matrices defined in (5.2.5) show a special block structure. The sub-systems' matrices $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ form the diagonal blocks of the system matrix $\mathbf{A}$ of S. The off-diagonal blocks are the products $\mathbf{B}_3\mathbf{C}_4^T$ and $\mathbf{B}_4\mathbf{C}_3^T$ of the internal input and output matrices of the sub-system. Similarly, the input and output matrices $\mathbf{B}$ and $\mathbf{C}$ are block structured, as well as in the matrix $\mathbf{E}$.

5.3   Transfer functions of the uncoupled and coupled systems

One of the questions arising at this point is the relation between the transfer functions of the sub-systems 1 and 2, and the transfer function of the coupled system. In this section we will study this issue.

Let us begin with the uncoupled sub-systems. At $s \in \mathbb{C}$ the transfer function of sub-system 1 defined in (5.2.1) is given by

$$
\begin{aligned}
\mathbf{H}(s) &= \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_3^T \end{bmatrix} (s\mathbf{E}_{11} - \mathbf{A}_{11})^{-1} [\mathbf{B}_1 \ \mathbf{B}_3] \\
&= \begin{bmatrix} \mathbf{C}_1^T (s\mathbf{E}_{11} - \mathbf{A}_{11})^{-1} \mathbf{B}_1 & \mathbf{C}_1^T (s\mathbf{E}_{11} - \mathbf{A}_{11})^{-1} \mathbf{B}_3 \\ \mathbf{C}_3^T (s\mathbf{E}_{11} - \mathbf{A}_{11})^{-1} \mathbf{B}_1 & \mathbf{C}_3^T (s\mathbf{E}_{11} - \mathbf{A}_{11})^{-1} \mathbf{B}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{11}(s) & \mathbf{H}_{12}(s) \\ \mathbf{H}_{21}(s) & \mathbf{H}_{22}(s) \end{bmatrix}.
\end{aligned}
$$
(5.3.1)

For the sub-system 2 defined in (5.2.2), similarly

$$
\begin{aligned}
\mathbf{G}(s) &= \begin{bmatrix} \mathbf{C}_2^T \\ \mathbf{C}_4^T \end{bmatrix} (s\mathbf{E}_{22} - \mathbf{A}_{22})^{-1} [\mathbf{B}_2 \ \mathbf{B}_4] \\
&= \begin{bmatrix} \mathbf{C}_2^T (s\mathbf{E}_{22} - \mathbf{A}_{22})^{-1} \mathbf{B}_2 & \mathbf{C}_2^T (s\mathbf{E}_{22} - \mathbf{A}_{22})^{-1} \mathbf{B}_4 \\ \mathbf{C}_4^T (s\mathbf{E}_{22} - \mathbf{A}_{22})^{-1} \mathbf{B}_2 & \mathbf{C}_4^T (s\mathbf{E}_{22} - \mathbf{A}_{22})^{-1} \mathbf{B}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{11}(s) & \mathbf{G}_{12}(s) \\ \mathbf{G}_{21}(s) & \mathbf{G}_{22}(s) \end{bmatrix}.
\end{aligned}
$$
(5.3.2)

At $s \in \mathbb{C}$ the transfer function of the coupled system (5.2.4) is

$$
\begin{aligned}
\mathbf{Z}(s) &= \mathbf{C}^T (s\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} = \begin{bmatrix} \mathbf{C}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^T \end{bmatrix} \left(s \begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{Z}_{11}(s) & \mathbf{Z}_{12}(s) \\ \mathbf{Z}_{21}(s) & \mathbf{Z}_{22}(s) \end{bmatrix}.
\end{aligned}
$$
(5.3.3)

Based on definitions eqs. (5.3.1) to (5.3.3) we wil express the components of the transfer function $\mathbf{Z}(s)$ in terms of the components of the transfer functions $\mathbf{H}(s)$ and $\mathbf{G}(s)$ in two manners. First we follow the typical approach used in the field of systems and control (more details can be found in for instance [84]). Secondly we use the Sherman-Morrison-Woodbury formula.

The systems and control approach

The starting point of this approach are two transfer functions $\mathbf{H}(s)$ and $\mathbf{G}(s)$ of the sub-systems 1 and 2, respectively. For each sub-system, its transfer function relates its inputs to outputs:

$$
\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{11}(s) & \mathbf{H}_{12}(s) \\ \mathbf{H}_{21}(s) & \mathbf{H}_{22}(s) \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_3 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{y}_2 \\ \mathbf{y}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{11}(s) & \mathbf{G}_{12}(s) \\ \mathbf{G}_{21}(s) & \mathbf{G}_{22}(s) \end{bmatrix} \begin{bmatrix} \mathbf{u}_2 \\ \mathbf{u}_4 \end{bmatrix}
$$

and

$$
\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{11}(s) & \mathbf{Z}_{12}(s) \\ \mathbf{Z}_{21}(s) & \mathbf{Z}_{22}(s) \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}.
$$
(5.3.4)

Systems (5.3.1) and (5.3.2) in combination with relation (5.2.3) lead to

$$\mathbf{y}_1 = \mathbf{H}_{11}(s)\mathbf{u}_1 + \mathbf{H}_{12}(s)\mathbf{y}_4 \tag{5.3.5}$$

$$\mathbf{y}_3 = \mathbf{H}_{21}(s)\mathbf{u}_1 + \mathbf{H}_{22}(s)\mathbf{y}_4 \tag{5.3.6}$$

$$\mathbf{y}_2 = \mathbf{G}_{11}(s)\mathbf{u}_2 + \mathbf{G}_{12}(s)\mathbf{y}_3 \tag{5.3.7}$$

$$\mathbf{y}_4 = \mathbf{G}_{21}(s)\mathbf{u}_2 + \mathbf{G}_{22}(s)\mathbf{y}_3. \tag{5.3.8}$$

Substituting $\mathbf{y}_4$ of (5.3.8) for $\mathbf{y}_4$ in (5.3.6) we obtain

$$\mathbf{y}_3 = \mathbf{H}_{21}(s)\mathbf{u}_1 + \mathbf{H}_{22}(s)\mathbf{y}_4 = \mathbf{H}_{21}(s)\mathbf{u}_1 + \mathbf{H}_{22}(s)[\mathbf{G}_{21}(s)\mathbf{u}_2 + \mathbf{G}_{22}(s)\mathbf{y}_3]$$

and hence

$$\mathbf{y}_3 = [\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}[\mathbf{H}_{21}(s)\mathbf{u}_1 + \mathbf{H}_{22}(s)\mathbf{G}_{21}(s)\mathbf{u}_2]. \tag{5.3.9}$$

With this result and (5.3.8), we can also express $\mathbf{y}_4$ in terms of $\mathbf{u}_1$ and $\mathbf{u}_2$

$$\begin{aligned}
\mathbf{y}_4 &= \mathbf{G}_{21}(s)\mathbf{u}_2 + \mathbf{G}_{22}(s)\mathbf{y}_3 = \mathbf{G}_{21}(s)\mathbf{u}_2 \\
&\quad + \mathbf{G}_{22}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}[\mathbf{H}_{21}(s)\mathbf{u}_1 + \mathbf{H}_{22}(s)\mathbf{G}_{21}(s)\mathbf{u}_2].
\end{aligned} \tag{5.3.10}$$

Using (5.3.9) and (5.3.10) in (5.3.5) and (5.3.7), we arrive at

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{H}_{11}(s)\mathbf{u}_1 + \mathbf{H}_{12}(s)\mathbf{y}_4 = \mathbf{H}_{11}(s)\mathbf{u}_1 \\
&\quad + \mathbf{H}_{12}(s)\Big(\mathbf{G}_{21}(s)\mathbf{u}_2 + \mathbf{G}_{22}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}[\mathbf{H}_{21}(s)\mathbf{u}_1 + \mathbf{H}_{22}(s)\mathbf{G}_{21}(s)\mathbf{u}_2]\Big) \\
&= \Big(\mathbf{H}_{11}(s) + \mathbf{H}_{12}(s)\mathbf{G}_{22}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}\mathbf{H}_{21}(s)\Big)\mathbf{u}_1 \\
&\quad + \Big(\mathbf{H}_{12}(s)\mathbf{G}_{21}(s) + \mathbf{H}_{12}(s)\mathbf{G}_{22}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}\mathbf{H}_{22}(s)\mathbf{G}_{21}\Big)\mathbf{u}_2
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{y}_2 &= \mathbf{G}_{11}(s)\mathbf{u}_2 + \mathbf{G}_{12}(s)\mathbf{y}_3 = \mathbf{G}_{11}(s)\mathbf{u}_2 + \mathbf{G}_{12}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}[\mathbf{H}_{21}(s)\mathbf{u}_1 \\
&\quad + \mathbf{H}_{22}(s)\mathbf{G}_{21}(s)\mathbf{u}_2] = \mathbf{G}_{12}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}\mathbf{H}_{21}(s)\mathbf{u}_1 \\
&\quad + \Big(\mathbf{G}_{11}(s) + \mathbf{G}_{12}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}\mathbf{H}_{22}(s)\mathbf{G}_{21}(s)\Big)\mathbf{u}_2.
\end{aligned}$$

This shows that the components of $\mathbf{Z}(s)$, as defined in (5.3.4), are

$$\mathbf{Z}_{11}(s) = \mathbf{H}_{11}(s) + \mathbf{H}_{12}(s)\mathbf{G}_{22}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}\mathbf{H}_{12}(s) \tag{5.3.11}$$

$$\mathbf{Z}_{12}(s) = \mathbf{H}_{12}(s)\mathbf{G}_{21}(s) + \mathbf{H}_{12}(s)\mathbf{G}_{22}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}\mathbf{H}_{22}(s)\mathbf{G}_{21}(s) \tag{5.3.12}$$

$$\mathbf{Z}_{21}(s) = \mathbf{G}_{12}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}\mathbf{H}_{21}(s) \tag{5.3.13}$$

$$\mathbf{Z}_{22}(s) = \mathbf{G}_{11}(s) + \mathbf{G}_{12}(s)[\mathbf{I} - \mathbf{H}_{22}(s)\mathbf{G}_{22}(s)]^{-1}\mathbf{H}_{22}(s)\mathbf{G}_{21}(s). \tag{5.3.14}$$

Computing the transfer function of the coupled system using the Sherman-Morrison-Woodbury formula

The evaluation of the transfer function of the coupled system, as defined in (5.3.3), requires a computation of an inverse of a block matrix. For a system consisting of an arbitrary number of sub-systems, a suitable tool towards this end is the Sherman-Morrison-Woodbury formula (see for instance [40] and references therein). This formula allows for a computationally cheap matrix inversion, as long as the considered matrix can be easily expressed as a sum of a matrix for which an inverse is known (or easy to compute) and a (low rank) correction. Let $\mathbf{L}$ be non-singular and let matrices $\mathbf{J}, \mathbf{M}, \mathbf{N}$ be of compatible size. Then the formula of $\mathbf{K} = \mathbf{L} + \mathbf{MJN}^T$ is (after [40])

$$\mathbf{K}^{-1} = (\mathbf{L} + \mathbf{MJN}^T)^{-1} = \mathbf{L}^{-1} - \mathbf{L}^{-1}\mathbf{M}(\mathbf{J}^{-1} + \mathbf{N}^T\mathbf{L}^{-1}\mathbf{M})^{-1}\mathbf{N}^T\mathbf{L}^{-1}, \qquad (5.3.15)$$

In our case, the matrix to be inverted can be decomposed into

$$
\begin{aligned}
\mathbf{K} &= s\begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} s\mathbf{E}_{11} - \mathbf{A}_{11} & -\mathbf{B}_3\mathbf{C}_4^T \\ -\mathbf{B}_3\mathbf{C}_3^T & s\mathbf{E}_{22} - \mathbf{A}_{22} \end{bmatrix} \\
&= \begin{bmatrix} s\mathbf{E}_{11} - \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & s\mathbf{E}_{22} - \mathbf{A}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{0} \end{bmatrix} \\
&= \mathbf{L} - \begin{bmatrix} \mathbf{0} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{0} \end{bmatrix}
\end{aligned}
$$

where $\mathbf{L}$ is a block-diagonal matrix, whose inverse can be calculated by computing the inverses of each sub-block separately and the correction matrix can be factored

$$
\begin{bmatrix} \mathbf{0} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{B}_3 \\ \mathbf{B}_4 & \mathbf{0} \end{bmatrix} \mathbf{I} \begin{bmatrix} \mathbf{C}_3^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T \end{bmatrix} = \mathbf{MJN}^T.
$$

Abbreviate $\mathbf{G}_i(s) = (s\mathbf{E}_{ii} - \mathbf{A}_{ii})^{-1}$, $\mathbf{P}_i(s) = \mathbf{G}_i(s)\mathbf{E}_{ii}$ and $\mathbf{R}_i(s) = \mathbf{G}_i(s)[\mathbf{B}_i\ \mathbf{B}_{2+i}]$, $i = 1, 2$ and omit the argument s when possible. Note that $\mathbf{R}_i = [\mathbf{R}_{i1}, \mathbf{R}_{i2}] = [\mathbf{G}_i\mathbf{B}_i, \mathbf{G}_i\mathbf{B}_{2+i}]$ consists of two blocks. Substituting the formulas for $\mathbf{L}, \mathbf{M}, \mathbf{J}$ and $\mathbf{N}$ into the the Sherman-

Morrison-Woodbury formula, we get

$$
\left(s \begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3 \mathbf{C}_4^T \\ \mathbf{B}_4 \mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix} \right)^{-1} = \left(\begin{bmatrix} s\mathbf{E}_{11} - \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & s\mathbf{E}_{22} - \mathbf{A}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{B}_3 \\ \mathbf{B}_4 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C}_3^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T \end{bmatrix} \right)^{-1}
$$

$$
= \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{B}_3 \\ \mathbf{B}_4 & \mathbf{0} \end{bmatrix} \circ
$$

$$
\left(\mathbf{I} - \begin{bmatrix} \mathbf{C}_3^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T \end{bmatrix} \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{B}_3 \\ \mathbf{B}_4 & \mathbf{0} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{C}_3^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T \end{bmatrix} \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \circ
$$

$$
\left(\mathbf{I} - \begin{bmatrix} \mathbf{0} & \mathbf{C}_3^T \mathbf{R}_{12} \\ \mathbf{C}_4^T \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{C}_3^T \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T \mathbf{G}_2 \end{bmatrix},
$$

(5.3.16)

where the entries of $\mathbf{G}$ and $\mathbf{R}$ depend on s. Using this result and eqs. (5.3.1) to (5.3.3), one can find the formula for the transfer function of the coupled system

$$
\mathbf{Z}(s) = \begin{bmatrix} \mathbf{C}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^T \end{bmatrix} \left(\begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \circ \right.
$$

$$
\left. \left(\mathbf{I} - \begin{bmatrix} \mathbf{0} & \mathbf{C}_3^T \mathbf{R}_{12} \\ \mathbf{C}_4^T \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{C}_3^T \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T \mathbf{G}_2 \end{bmatrix} \right) \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathbf{C}_1^T \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^T \mathbf{R}_{21} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{C}_1^T \mathbf{R}_{12} \\ \mathbf{C}_2^T \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \circ
$$

$$
\left(\mathbf{I} - \begin{bmatrix} \mathbf{0} & \mathbf{C}_3^T \mathbf{R}_{12} \\ \mathbf{C}_4^T \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{C}_3^T \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T \mathbf{R}_{21} \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathbf{H}_{11}(s) & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{11}(s) \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{H}_{12}(s) \\ \mathbf{G}_{12}(s) & \mathbf{0} \end{bmatrix} \left(\mathbf{I} - \begin{bmatrix} \mathbf{0} & \mathbf{H}_{22}(s) \\ \mathbf{G}_{22}(s) & \mathbf{0} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{H}_{21}(s) & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{21}(s) \end{bmatrix}.
$$

(5.3.17)

It is easy to show, that the formulation (5.3.17) is equivalent to the formulation given by eqs. (5.3.11) to (5.3.14). Moreover, (5.3.17) provides an elegant relation between the components of the transfer functions of the sub-systems and the coupled system, that reveals the symmetry and the structure of the coupled system. In addition it shows that the relation between the transfer functions is not straightforward. Since several sub-expressions such as $(s\mathbf{E}_{ii} - A_{ii})^{-1}$ reoccur frequently, we will introduce abbreviations in the upcoming sections.

Formula (5.3.17) reveals a structure which is more difficult to find in (5.3.11) – (5.3.14) and can be used to calculate the transfer function of the coupled system if the transfer functions of the individual sub-systems are available. The involved inverse is of a small matrix which means that calculation of the transfer function of the coupled system is relatively cheap.

5.4   Standard block structure preserving reduction

In this section we will recall the general ideas of the standard block-structure preserving methods.

As explained in Chapter 3, a typical block structure preserving (*BSP*) model reduction method applied to the system (5.2.4) consists of the following three steps:

1. Create the matrix $\tilde{\mathbf{V}}$ whose columns span the nth Krylov subspace around $s_0 \in \mathbb{C}$

$$\tilde{\mathbf{V}} = \mathcal{K}_n(\mathbf{P}(s_0), \mathbf{R}(s_0)),$$

   where $\mathbf{P}(s_0)$ and $\mathbf{R}(s_0)$ are

$$\mathbf{P}(s_0) = (s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{E} \in \mathbb{R}^{N \times N} \quad \text{and} \quad \mathbf{R}(s_0) = (s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} \in \mathbb{R}^N.$$

2. Build a the block-diagonal reduction matrix $\mathbf{V}$ with $N_1 + N_2 = N$ rows

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix},$$

   where $\mathbf{V}_1$ and $\mathbf{V}_2$ contain the first N_1 respectively last N_2 rows of the matrix $\tilde{\mathbf{V}}$.

3. Project the original system onto a lower-dimensional space

$$\hat{\mathbf{E}} = \mathbf{V}^T\mathbf{E}\mathbf{V}, \quad \hat{\mathbf{A}} = \mathbf{V}^T\mathbf{A}\mathbf{V}, \quad \hat{\mathbf{B}} = \mathbf{V}^T\mathbf{B}, \quad \hat{\mathbf{C}} = \mathbf{V}^T\mathbf{C}.$$

When possible we write $\mathbf{P}$ and $\mathbf{R}$ rather than $\mathbf{P}(s_0)$ respectively $\mathbf{R}(s_0)$. The model reduction methods based on this idea are widely applied and popular due to a good accuracy of the reduced-order systems that they deliver. However, as already mentioned in Section 5.1, they have a few drawbacks, one of them being the high cost of the construction of the reduction basis. The main computational cost of this type of methods is related to evaluation of $\mathbf{x} \mapsto (s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{x}$, which involves solving a system of equations with a large coefficient matrix. In the next section we introduce an alternative structure preserving method which for some cases can significantly reduce the computational costs.

5.5   Separate Bases Reduction algorithm

In Section 5.4, the reduction basis is built using the coupled formulation of the system (5.2.4). The construction of this basis requires repeated evaluations of $\mathbf{x} \mapsto (s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{x}$ where $s_0\mathbf{E} - \mathbf{A}$ is an $N \times N$ matrix. For large N this procedure can be computationally very expensive or even unfeasible. In such cases one can try to make use of a natural

block structure of the coupled system and for instance replace the evaluations involving $(s_0\mathbf{E} - \mathbf{A})^{-1}$ by evaluations involving $(s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}$ and $(s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-1}$, i.e., by evaluations involving only the coefficient matrices of both sub-systems. If N is large and for instance $N_1 = N_2 = N/2$ then the serial computation of $(s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}$ and $(s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-1}$ may be much faster that of $(s_0\mathbf{E} - \mathbf{A})^{-1}$. Further acceleration can be achieved through parallelism.

Following this idea, we introduce a new model reduction algorithm, called Separate Bases Reduction (*SBR*) algorithm. Here the Krylov subspaces that create the reduction bases correspond to the uncoupled sub-systems (as defined in (5.2.1) and (5.2.2)) rather than to the coupled system (5.2.4). The procedure is as follows:

1. Create two matrices $\mathbf{V}_1$ and $\mathbf{V}_2$, one for each sub-system:

   - For the sub-system S_1, build a matrix $\mathbf{V}_1$, whose columns span the n_1th Krylov subspace around $s_0 \in \mathbb{C}$

   $$\mathbf{V}_1 = \mathcal{K}_{n_1}(\mathbf{P}_1(s_0), \mathbf{R}_1(s_0)),$$

   where $\mathbf{P}_1(s_0)$ and $\mathbf{R}_1(s_0)$ are

   $$\mathbf{P}_1(s_0) = (s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}\mathbf{E}_{11} \quad \text{and} \quad \mathbf{R}_1(s_0) = (s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}[\mathbf{B}_1 \ \mathbf{B}_3].$$

   Matrix $\mathbf{V}_1$ has N_1 rows.

   - For the sub-system S_2, build a matrix $\mathbf{V}_2$, whose columns span the n_2th Krylov subspace around $s_0 \in \mathbb{C}$

   $$\mathbf{V}_2 = \mathcal{K}_{n_2}(\mathbf{P}_2(s_0), \mathbf{R}_2(s_0)),$$

   where $\mathbf{P}_2(s_0)$ and $\mathbf{R}_2(s_0)$ are

   $$\mathbf{P}_2(s_0) = (s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-1}\mathbf{E}_{22} \quad \text{and} \quad \mathbf{R}_2(s_0) = (s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-1}[\mathbf{B}_2 \ \mathbf{B}_4].$$

   Matrix $\mathbf{V}_2$ has N_2 rows.

2. Build the block-diagonal reduction matrix $\mathbf{V}$ with $N_1 + N_2 = N$ rows

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix}.$$

3. Project the original system onto a lower-dimensional space

$$\hat{\mathbf{E}} = \mathbf{V}^T\mathbf{E}\mathbf{V}, \quad \hat{\mathbf{A}} = \mathbf{V}^T\mathbf{A}\mathbf{V}, \quad \hat{\mathbf{B}} = \mathbf{V}^T\mathbf{B}, \quad \hat{\mathbf{C}} = \mathbf{V}^T\mathbf{C}.$$

In the sequel, when possible without causing confusion, we omit the argument s_0 of $\mathbf{P}_i$ and $\mathbf{R}_i$, $i = 1, 2$. In the next section, we will compare the SBR algorithm with a standard BSP reduction method, by examining their most important properties.

5.6   Separate Bases Reduction algorithm – properties

In this section we will discuss the differences and similarities between Separate Bases Reduction algorithm and standard block structure preserving model reduction methods (see Chapter 3).

Block-structure preservation

As described in Section 5.5, the SBR algorithm uses reduction matrices of the block-diagonal form

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix}.$$

Therefore, its application preserves the block structure of the coupled system matrices.

Rank and orthogonality

The sub-blocks $\mathbf{V}_1$ and $\mathbf{V}_2$ of the projector $\mathbf{V}$ are constructed separately, using one of the Krylov basis building algorithms. Hence, both of them have a full column rank and, as a result, the matrix $\mathbf{V}$ also has a full column rank. If the sub-blocks $\mathbf{V}_1$ and $\mathbf{V}_2$ have orthogonal columns then also matrix $\mathbf{V}$ has (automatically) orthogonal columns, i.e., no explicit orthogonalization has to be applied.

Computational cost

The difference between the computational costs for a standard block structure preserving method and the Separate Bases Reduction algorithm comes from the fact, that the SBR algorithm computes the reduction bases for the set of uncoupled systems instead of using the coupled formulation of the system. This approach can significantly reduce the computational time and storage requirements needed during the model reduction process.
The main cost of the Krylov basis construction lies in the evaluation of the matrix pencil inverse function $\mathbf{x} \mapsto (s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{x}$. For coupled models with many degrees n of freedom this evaluation may be unfeasible. But for sub-problems of smaller size evaluation may be possible. The amount of computational work required for the solution of $(s_0\mathbf{E} - \mathbf{A})\mathbf{x} = \mathbf{d}$ depends on the employed solution method which at its turn relies on specific properties of the matrix $s_0\mathbf{E} - \mathbf{A}$ (symmetry, monotone, positive definite, etc.). Different methods lead to different amounts of computational work: The minimal amount of work of $O(n)$ operations is usually achieved by multigrid methods (see [91]), other methods such as GMRES, PCG, CGS and BiCGstab(l) (see [79], [85], and [33]) are more expensive. Classical fixed point methods such as Jacobi, Gauss-Seidel and matrix-splitting based methods are usually the slowest.

Size of the reduction space

Another difference with respect to the standard BSP reduction methods is the size of the reduction matrix $\mathbf{V}$ and, as a result, dimension of the reduced order model.
Let us consider the coupled system (5.2.4) and assume, for simplicity, that there is no need for deflation (all columns turn out to be linearly independent) while building the matrix $\mathbf{V}$. We will apply a typical reduction procedure like described in Section 5.4 and the SBR algorithm. In both cases, we will build a Krylov subspace of order n and estimate the size of the reduction space and reduced order model.

We begin with the analysis of the standard structure preserving algorithm. The nth Krylov subspace built for the coupled system for the starting matrices as defined in Section 5.4 will be of the form

$$\tilde{\mathbf{V}} = \mathcal{K}_n\big(\mathbf{P}, \mathbf{R}\big) = \text{colspan}\{\mathbf{R}, \ldots, \ \mathbf{P}^{n-1}\mathbf{R}\}$$

where $\mathbf{P} = (s_0\mathbf{E}-\mathbf{A})^{-1}\mathbf{E}$ and $\mathbf{R} = (s_0\mathbf{E}-\mathbf{A})^{-1}\mathbf{B}$. Since $\mathbf{B} \in \mathbb{R}^{N\times m}$, each of the components $\mathbf{P}^j\mathbf{R}$ of the matrix $\tilde{\mathbf{V}}$ has m columns. Thus, for a degree n Krylov space, assuming no deflation, the size of $\tilde{\mathbf{V}}$ is $N\times(nm)$. Next, the block-diagonal reduction matrix $\mathbf{V}$ is created by splitting the rows of $\tilde{\mathbf{V}}$ according to the dimensions of the sub-problems. In our case, the coupled system consists of two sub-systems, so the final size of the reduction matrix $\mathbf{V}$ is $N \times (2nm)$. This leads to a reduced model of order $2nm$.

Next, we will focus on the *SBR* algorithm. In this case two matrices $\mathbf{V}_1$ and $\mathbf{V}_2$, are built separately and we assume that each of them corresponds to an nth degree Krylov subspace based on the appropriate matrices (for $i = 1, 2$ define $\mathbf{G}_i(s_0) = (s_0\mathbf{E}_{ii} - \mathbf{A}_{ii})^{-1}$, $\mathbf{P}_i(s_0) = \mathbf{G}_i\mathbf{E}_{ii}$ and $\mathbf{R}_i(s_0) = \mathbf{G}_i[\mathbf{B}_i \ \mathbf{B}_{2+i}]$ and observe that $\mathbf{R}_i = [\mathbf{R}_{i1}, \mathbf{R}_{i2}]$ where $\mathbf{R}_{i1}$ and $\mathbf{R}_{i2}$ are $\mathbf{G}_i\mathbf{B}_i$, respectively $\mathbf{G}_i\mathbf{B}_{2+i}$). For the sub-system S_1, we create the matrix $\mathbf{V}_1$

$$\mathbf{V}_1 = \mathcal{K}_n\big(\mathbf{P}_1, \ \mathbf{R}_1\big).$$

Here, $\mathbf{R}_1, [\mathbf{B}_1 \ \mathbf{B}_3] \in \mathbb{R}^{N_1\times(m_1+m_3)}$, so each component $\mathbf{P}_1^j\mathbf{R}_1$ of the matrix $\mathbf{V}_1$ has (m_1+m_3) columns whence $\mathbf{V}_1$ has $n \times (m_1 + m_3)$ columns.
For the sub-system S_2, we create

$$\mathbf{V}_2 = \mathcal{K}_n\big(\mathbf{P}_2, \ \mathbf{R}_2\big).$$

Similarly, since $\mathbf{R}_2, [\mathbf{B}_2 \ \mathbf{B}_4] \in \mathbb{R}^{N_2\times(m_2+m_4)}$, every component $\mathbf{P}_2^j\mathbf{R}_2$ of the matrix $\mathbf{V}_2$ has $(m_2 + m_4)$ columns, and matrix $\mathbf{V}_2$ has $n \times (m_2 + m_4)$ columns.

Next, matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ are used as diagonal blocks of the reduction matrix $\mathbf{V}$, resulting in a reduced model of order

$$n \times (m_1 + m_3) + n \times (m_2 + m_4) = n \times (m + m_3 + m_4).$$

This result shows that the SBR algorithm creates a smaller reduced order model than standard BSP methods if $(m_3 + m_4) < m$. This is for instance the case for coupled systems for which the number of internal inputs is not larger than the number of external inputs. If there are many more internal inputs than external ones, the size of the SBR algorithm based reduction matrix will grow very fast compared to the size of the BSP reduction matrix. However, this problem can be avoided for the category of systems for which the internal input matrices $\mathbf{B}_2$ and $\mathbf{B}_4$ can be approximated by only a small number of dominant components. This approach will be explained in more detail in Chapters 7 and 8.

The moment matching property

In order to assess the *SBR moment matching* properties we compare the column-spaces of the BPS and SBR reduction matrices. For simplicity, without loss of generalization, we focus at the *SISO* case (the coupled system is SISO) where in addition $\mathbf{B}_i$, $\mathbf{C}_i$, $i = 1, \ldots, 4$ related to the sub-systems are column-vectors which implies that all products $\mathbf{C}_i^T(\ldots)\mathbf{B}_j$, $i, j = 1, \ldots, 4$, are scalars. A similar analysis is possible for the *MIMO* case (a MIMO coupled system with sub-system matrices $\mathbf{B}_i$, $\mathbf{C}_i$).

Theorem 5.6.1. *Let the coupled system be as in Figure 5.3, described by* (5.2.1) *and* (5.2.2). *Assume that all inputs and outputs are column-vectors, i.e., $m_i = p_i = 1$, $i = 1, 2, 3, 4$. Then the SBR reduced-order model transfer function matches at least the same (number of) moments as the BSP reduced-order model transfer function.*

Proof. First, we examine the reduction space built by a standard BSP method. To match the first k moments at $s \in \mathbb{C}$, of the coupled system of the form (5.2.4), one has to construct the Krylov space

$$\tilde{\mathbf{V}} = \mathcal{K}_k(\mathbf{P}, \mathbf{R}),$$

where

$$\mathbf{P} = (s\mathbf{E} - \mathbf{A})^{-1}\mathbf{E} \quad \text{and} \quad \mathbf{R} = (s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}.$$

The ith Krylov step for the BSP method adds to the reduction basis the column span of the following matrix $\mathbf{V}_{\text{BSP}}^{(i)}$

$$\mathbf{V}_{\text{BSP}}^{(i)} = \begin{bmatrix} \mathbf{V}_{11}^{(i)} & \mathbf{V}_{12}^{(i)} & 0 & 0 \\ 0 & 0 & \mathbf{V}_{21}^{(i)} & \mathbf{V}_{22}^{(i)} \end{bmatrix} \tag{5.6.1}$$

with blocks of the form

$$\begin{aligned} \tilde{\mathbf{V}}_{\text{BSP}}^{(i)} &= \begin{bmatrix} \mathbf{V}_{11}^{(i)} & \mathbf{V}_{12}^{(i)} \\ \mathbf{V}_{21}^{(i)} & \mathbf{V}_{22}^{(i)} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}_1^{i-1}\mathbf{R}_{11} + \sum_{j=0}^{i-1} \alpha_j \mathbf{P}_1^j \mathbf{R}_{12} & \sum_{j=0}^{i-1} \gamma_j \mathbf{P}_1^j \mathbf{R}_{12} \\ \sum_{j=0}^{i-1} \beta_j \mathbf{P}_2^j \mathbf{R}_{22} & \mathbf{P}_2^{i-1}\mathbf{R}_{21} + \sum_{j=0}^{i-1} \delta_j \mathbf{P}_2^j \mathbf{R}_{22} \end{bmatrix} \end{aligned} \tag{5.6.2}$$

By (5.3.16) there exist scalars a, b, c, d and by construction (induction) there exist coefficient vectors $\alpha = [\alpha_1, \ldots, \alpha_j] \in \mathbb{R}^{i-2}, \beta, \gamma, \delta \in \mathbb{R}^{i-2}$ such that

$$
\begin{aligned}
\tilde{\mathbf{V}}_{\text{BSP}}^{(i)} \quad &= \quad \mathbf{P}\tilde{\mathbf{V}}_{\text{BSP}}^{(i-1)} \\[4pt]
&= \quad \left(s \begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix} \tilde{\mathbf{V}}_{\text{BSP}}^{(i-1)} \\[4pt]
&\overset{\underset{(5.3.16)}{\text{induction}}}{=} \left(\begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{C}_3^T\mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T\mathbf{G}_2 \end{bmatrix} \right) \begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix} \tilde{\mathbf{V}}_{\text{BSP}}^{(i-1)} \\[4pt]
&= \quad \left(\begin{bmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{C}_3^T\mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T\mathbf{P}_2 \end{bmatrix} \right) \\[4pt]
&\qquad \begin{bmatrix} \mathbf{P}_1^{i-2}\mathbf{R}_{11} + \sum_{j=0}^{i-2} \alpha_j \mathbf{P}_1^j \mathbf{R}_{12} & \sum_{j=0}^{i-2} \gamma_j \mathbf{P}_1^j \mathbf{R}_{12} \\ \sum_{j=0}^{i-2} \beta_j \mathbf{P}_2^j \mathbf{R}_{22} & \mathbf{P}_2^{i-2}\mathbf{R}_{21} + \sum_{j=0}^{i-2} \delta_j \mathbf{P}_2^j \mathbf{R}_{22} \end{bmatrix} \\[4pt]
&= \quad \begin{bmatrix} \mathbf{P}_1^{i-1}\mathbf{R}_{11} + \sum_{j=1}^{i-1} \alpha_j \mathbf{P}_1^j \mathbf{R}_{12} & \sum_{j=1}^{i-1} \gamma_j \mathbf{P}_1^j \mathbf{R}_{12} \\ \sum_{j=1}^{i-1} \beta_j \mathbf{P}_2^j \mathbf{R}_{22} & \mathbf{P}_2^{i-1}\mathbf{R}_{21} + \sum_{j=1}^{i-1} \delta_j \mathbf{P}_2^j \mathbf{R}_{22} \end{bmatrix} + \\[4pt]
&\qquad \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{C}_3^T\mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_4^T\mathbf{P}_2 \end{bmatrix} \\[4pt]
&\qquad \begin{bmatrix} \mathbf{P}_1^{i-2}\mathbf{R}_{11} + \sum_{j=0}^{i-2} \alpha_j \mathbf{P}_1^j \mathbf{R}_{12} & \sum_{j=0}^{i-2} \gamma_j \mathbf{P}_1^j \mathbf{R}_{12} \\ \sum_{j=0}^{i-2} \beta_j \mathbf{P}_2^j \mathbf{R}_{22} & \mathbf{P}_2^{i-2}\mathbf{R}_{21} + \sum_{j=0}^{i-2} \delta_j \mathbf{P}_2^j \mathbf{R}_{22} \end{bmatrix} \\[4pt]
&\overset{\mathbf{C}_i^T(\ldots)\mathbf{B}_j \in \mathbb{C}}{=} \begin{bmatrix} \mathbf{P}_1^{i-1}\mathbf{R}_{11} + \sum_{j=1}^{i-1} \alpha_j \mathbf{P}_1^j \mathbf{R}_{12} & \sum_{j=1}^{i-1} \gamma_j \mathbf{P}_1^j \mathbf{R}_{12} \\ \sum_{j=1}^{i-1} \beta_j \mathbf{P}_2^j \mathbf{R}_{22} & \mathbf{P}_2^{i-1}\mathbf{R}_{21} + \sum_{j=1}^{i-1} \delta_j \mathbf{P}_2^j \mathbf{R}_{22} \end{bmatrix} + \\[4pt]
&\qquad \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} * & * \\ * & * \end{bmatrix} \\[4pt]
&= \quad \begin{bmatrix} \mathbf{P}_1^{i-1}\mathbf{R}_{11} + \sum_{j=1}^{i-1} \alpha_j \mathbf{P}_1^j \mathbf{R}_{12} & \sum_{j=1}^{i-1} \gamma_j \mathbf{P}_1^j \mathbf{R}_{12} \\ \sum_{j=1}^{i-1} \beta_j \mathbf{P}_2^j \mathbf{R}_{22} & \mathbf{P}_2^{i-1}\mathbf{R}_{21} + \sum_{j=1}^{i-1} \delta_j \mathbf{P}_2^j \mathbf{R}_{22} \end{bmatrix} + \\[4pt]
&\qquad \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{22} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mu_1 & \mu_2 \\ \mu_3 & \mu_4 \end{bmatrix} \\[4pt]
&= \quad \begin{bmatrix} \mathbf{P}_1^{i-1}\mathbf{R}_{11} + \sum_{j=1}^{i-1} \alpha_j \mathbf{P}_1^j \mathbf{R}_{12} & \sum_{j=1}^{i-1} \gamma_j \mathbf{P}_1^j \mathbf{R}_{12} \\ \sum_{j=1}^{i-1} \beta_j \mathbf{P}_2^j \mathbf{R}_{22} & \mathbf{P}_2^{i-1}\mathbf{R}_{21} + \sum_{j=1}^{i-1} \delta_j \mathbf{P}_2^j \mathbf{R}_{22} \end{bmatrix} + \\[4pt]
&\qquad \begin{bmatrix} \mu_1 \mathbf{R}_{12} & \mu_2 \mathbf{R}_{12} \\ \mu_3 \mathbf{R}_{22} & \mu_4 \mathbf{R}_{22} \end{bmatrix} \\[4pt]
&= \quad \begin{bmatrix} \mathbf{P}_1^{i-1}\mathbf{R}_{11} + \sum_{j=0}^{i-1} \hat{\alpha}_j \mathbf{P}_1^j \mathbf{R}_{12} & \sum_{j=0}^{i-1} \hat{\gamma}_j \mathbf{P}_1^j \mathbf{R}_{12} \\ \sum_{j=0}^{i-1} \hat{\beta}_j \mathbf{P}_2^j \mathbf{R}_{22} & \mathbf{P}_2^{i-1}\mathbf{R}_{21} + \sum_{j=0}^{i-1} \hat{\delta}_j \mathbf{P}_2^j \mathbf{R}_{22} \end{bmatrix}
\end{aligned}
$$
(5.6.3)

where $\hat{\alpha} = [\mu_1, \alpha]$, $\hat{\beta} = [\mu_3, \beta]$, $\hat{\gamma} = [\mu_2, \gamma]$ and $\hat{\delta} = [\mu_4, \delta]$. It is easy to see that the column span of the matrix constructed from the matrix $\tilde{\mathbf{V}}_{\text{BSP}}^{(i)}$ by splitting its rows, has the same column span as the matrix defined in (5.6.1). Finally, the reduction basis $\mathbf{V}_{\text{BSP}}$ after k steps of the BSP algorithm has the following form

$$
\mathbf{V}_{\text{BSP}} = [\mathbf{V}_{\text{BSP}}^{(1)}, \ldots, \mathbf{V}_{\text{BSP}}^{(k)}].
$$
(5.6.4)

Now we will examine the SBR reduction space algorithm. Let $\mathbf{P}_i, \mathbf{R}_i = [\mathbf{R}_{i1}, \mathbf{R}_{i2}]$, $i =$

1, 2 be as defined before. For $s \in \mathbb{C}$ SBR builds two Krylov subspaces

$$\mathbf{V}_1 = \mathcal{K}_k(\mathbf{P}_1, \mathbf{R}_1), \quad \text{and} \quad \mathbf{V}_2 = \mathcal{K}_k(\mathbf{P}_2, \mathbf{R}_2).$$

One can easily prove, that the ith step of the Krylov iteration within the SBR algorithm adds to the reduction basis the column span of the following matrix $\mathbf{V}_{SBR}^{(i)}$

$$\mathbf{V}_{SBR}^{(i)} = \begin{bmatrix} \mathbf{V}_1^{(i)} & 0 \\ 0 & \mathbf{V}_2^{(i)} \end{bmatrix}, \tag{5.6.5}$$

where

$$\mathbf{V}_1^{(i)} = \begin{bmatrix} \mathbf{P}_1^{i-1}\mathbf{R}_{11}, \mathbf{P}_1^{i-1}\mathbf{R}_{12} \end{bmatrix}$$

and

$$\mathbf{V}_2^{(i)} = \begin{bmatrix} \mathbf{P}_2^{i-1}\mathbf{R}_{21}, \mathbf{P}_2^{i-1}\mathbf{R}_{22} \end{bmatrix}.$$

Finally, the reduction basis $\mathbf{V}_{SBR}$ after k steps of the SBR algorithm has the following form

$$\mathbf{V}_{SBR} = [\mathbf{V}_{SBR}^{(1)}, \dots, \mathbf{V}_{SBR}^{(k)}]. \tag{5.6.6}$$

Comparing (5.6.2) and (5.6.5), we observe that

$$\text{colspan}\mathbf{V}_{BSP} \subset \text{colspan}\mathbf{V}_{SBR}.$$

Because the dimensions of the spaces are equal for our case (SISO external and column-vectors $\mathbf{B}_i, \mathbf{C}_i$ for the sub-systems) one finds that in addition

$$\text{colspan}\mathbf{V}_{BSP} = \text{colspan}\mathbf{V}_{SBR}. \tag{5.6.7}$$

Because $\text{colspan}\mathbf{V}_{BSP} \subset \text{colspan}\mathbf{V}_{SBR}$ the SBR reduced-order model transfer function matches (at least) the same (number of) moments as the BSP reduced-order model transfer function which at its turn (Theorem 2, [25]) matches the same (number of) moments as the original coupled system's transfer function. For the more general case where $\mathbf{B}_i$, $\mathbf{C}_i$, $i = 1, \dots, 4$ are matrices one should also obtain

$$\text{colspan}\mathbf{V}_{BSP} \subseteq \text{colspan}\mathbf{V}_{SBR} \tag{5.6.8}$$

which is sufficient to prove the moment matching property of the SBR reduced-order system.   □

5.7   Numerical examples

In this section we will compare the performance of the SBR algorithm and a standard BSP method using two variants of the same test problem.
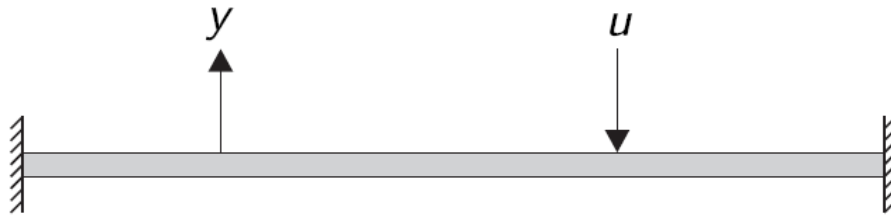
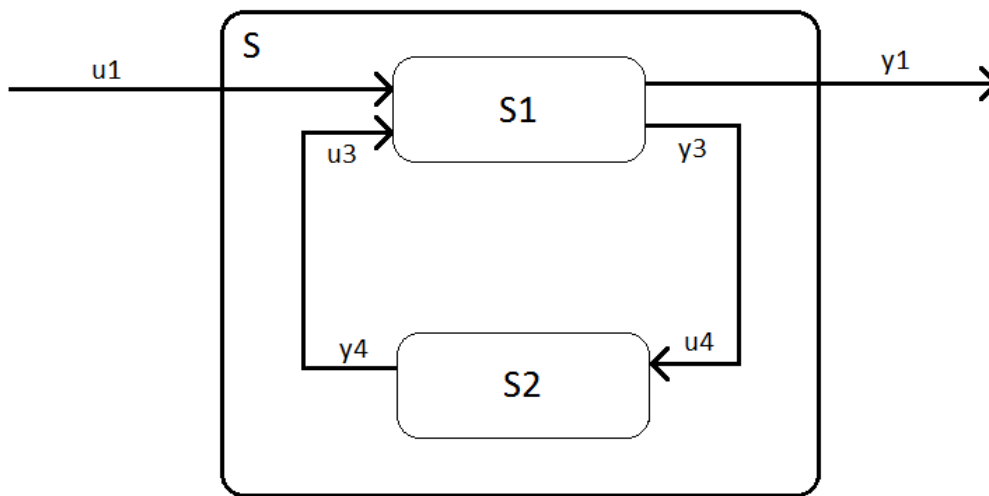Figure 5.4: A linear beam subject to mechanical vibrations

Figure 5.5: Schematic representation of the linear beam-controller system

Our test problem consists of an interconnected system which consists of two sub-systems. One sub-system is a linear *beam* subject to mechanical vibrations shown in Figure 5.4 and the other one is a controller. The system's schematic representation is shown in Figure 5.5. The controller sub-system is a part of a closed loop because its only inputs and outputs are connected to the beam's sub-system. Hence, the sub-systems are described

by the following equations: The beam's *system S_1* is given by equations:

$$S_1 \; : \; \begin{cases} s\mathbf{I}_{11}\mathbf{x}_1 \;=\; \mathbf{A}_{11}\mathbf{x}_1 + [\mathbf{B}_1 \; \mathbf{B}_3]\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_3 \end{bmatrix} \\[2mm] \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_3 \end{bmatrix} \;=\; \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_3^T \end{bmatrix}\mathbf{x}_1. \end{cases} \tag{5.7.1}$$

The controller's *system S_2* is described by equations:

$$S_2 \; : \; \begin{cases} s\mathbf{I}_{22}\mathbf{x}_2 \;=\; \mathbf{A}_{22}\mathbf{x}_2 + \mathbf{B}_4\mathbf{u}_4 \\[1mm] \mathbf{y}_4 \;=\; \mathbf{C}_4^T\mathbf{x}_2. \end{cases} \tag{5.7.2}$$

Combined, the interconnected system is:

$$S \; : \; \begin{cases} s\begin{bmatrix} \mathbf{I}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0} \end{bmatrix}\mathbf{u}_1 \\[3mm] \mathbf{y}_1 = [\mathbf{C}_1^T \; \mathbf{0}]\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}. \end{cases} \tag{5.7.3}$$

Using the notation introduced in (5.2.5) we have

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{22} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{0} \end{bmatrix}. \tag{5.7.4}$$

The examples which follow differ in both number of degrees of freedom and number of internal input/output connections.

Example 5.7.1. The interconnected system has 120 degrees of freedom, 60 for each sub-system. There is one external input and one external output, and each of the sub-systems has one internal input and one internal output. Hence, $\mathbf{A}_{ii}, \mathbf{I}_{ii} \in \mathbb{R}^{60\times60}$, $i = 1,2$ and $\mathbf{B}_i, \mathbf{C}_i \in \mathbb{R}^{60\times1}$, $i = 1,3,4$. For both BSP and SBR, multi-point expansion was used, i.e., the Krylov spaces were accumulated for the same set of two frequencies $s_1, s_2 \in \mathbb{C}$.

Let $\mathbf{P}$ and $\mathbf{R}$ be defined as in Section 5.4, $\mathbf{P}_i$ and $\mathbf{R}_i$ be defined as in Section 5.5, all related to the interconnected system (5.2.4) shown in Figure 5.5. The difference between the latter system and the interconnected system of the general form studied in this section (see Figure 5.3), leads to small differences in the required Krylov spaces. For instance, in stead of SBR space $\mathcal{K}_4(\mathbf{P}_2(s_1), \mathbf{R}_2(s_1))$ we only need and therefore create $\mathcal{K}_4(\mathbf{P}_2(s_1), \mathbf{R}_{22}(s_1))$.

For the BSP method, Krylov bases $\mathcal{K}_6(\mathbf{P}(s_1), \mathbf{R}(s_1))$ and $\mathcal{K}_6(\mathbf{P}(s_2), \mathbf{R}(s_2))$ are constructed. The columns are combined into the column span $[\mathcal{K}_6(\mathbf{P}(s_1), \mathbf{R}(s_1)), \mathcal{K}_6(\mathbf{P}(s_2), \mathbf{R}(s_2))]$ (which turns out to contain 12 linearly independent vectors) which is split into the BSP block-diagonal projector (which turns out to consist of 24 linearly independent vectors).

Note that the combination of Krylov spaces as well as the row-based splitting of the combined span can lead to linearly dependent columns.

For the SBR method, $\mathcal{K}_4(\mathbf{P}_1(s_1), [\mathbf{R}_{11}(s_1)), \mathbf{R}_{12}(s_1)]$ and $\mathcal{K}_4(\mathbf{P}_1(s_2), [\mathbf{R}_{11}(s_2), \mathbf{R}_{12}(s_2)])$, respectively $\mathcal{K}_4(\mathbf{P}_2(s_1), \mathbf{R}_{22}(s_1))$ and $\mathcal{K}_4(\mathbf{P}_2(s_2), \mathbf{R}_{22}(s_2))$ are constructed which leads to $2 \cdot 8 + 2 \cdot 4 = 24$ linearly independent columns.

Figure 5.6 shows the magnitude plots with respect to the frequency of the frequency response functions of the original and the reduced-order systems. The dashed vertical lines mark the expansion points for which the reduction bases were computed. It can be
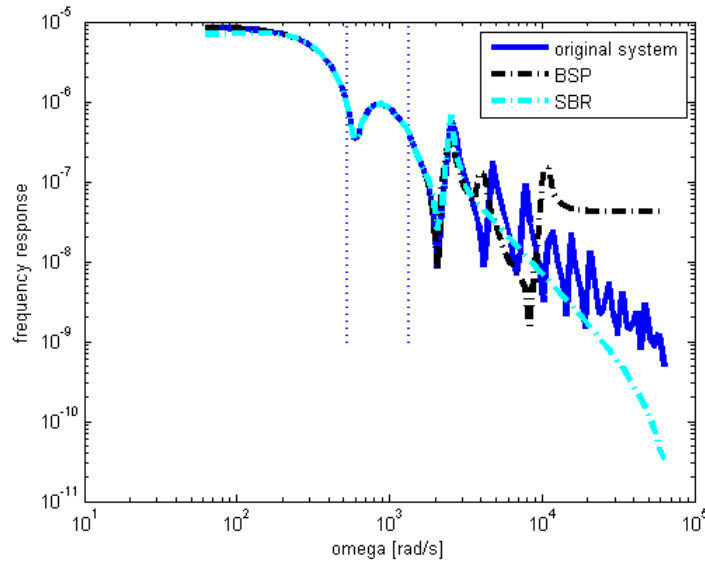
Figure 5.6: Magnitude plots of the frequency response functions of the original and reduced-order systems

observed, that both of the reduction algorithms give reduced-order models with related reduced-order frequency response functions that exactly match the original frequency response function at the expansion points and approximate the original one well in the expansion points' neighborhood. For the frequencies away from this neighborhood, the accuracy of the reduced-order models deteriorates. This is an expected result, since Krylov subspace based methods only match locally good based on the underlying Taylor series approximation. This fact can be also noticed in Figure 5.7, that shows the magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original frequency response function as a function of frequency. Observe that SBR performs almost as well as the standard BSP method, its relative error is at
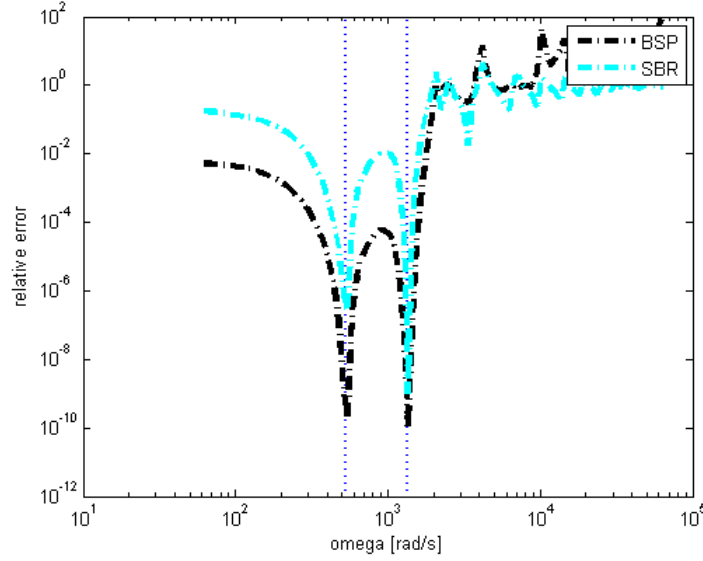
Figure 5.7: Magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original frequency response function

most 2% around the expansion points. That SBR performs a bit less than BSP is probably due to the fact that BSP in addition matches the moments 5 and 6 – SBR did not have to match these moments due to Theorem 5.6.1 and indeed does not.

Let $\mathbf{V}_{\text{BSP}}$ denote the block-diagonal reduction basis built by the standard BSP method and $\mathbf{V}_{\text{SBR}}$ denote the block-diagonal reduction basis built by the SBR algorithm. With the use of these, define the reduced system's components:

$$\hat{\mathbf{E}}_k = \mathbf{V}_k^T \mathbf{V}_k, \quad \hat{\mathbf{A}}_k = \mathbf{V}_k^T \mathbf{A} \mathbf{V}_k, \quad \hat{\mathbf{B}}_k = \mathbf{V}_k^T \mathbf{B}, \quad \hat{\mathbf{C}}_k = \mathbf{V}_k^T \mathbf{C}, \quad k = \text{BSP}, \text{SBR}. \quad (5.7.5)$$

Based on (3.2.1), the formula for the ith derivative at point s of the transfer function $H(s)$ of the system defined by (5.7.3) and (5.7.4) is

$$\partial^i \mathbf{H}(s) = (-1)^i i! \mathbf{C}^T \left((s\mathbf{E} - \mathbf{A})^{-1} \mathbf{E} \right)^i (s\mathbf{E} - \mathbf{A})^{-1} \mathbf{B}, \quad (5.7.6)$$

and those for the reduced-order systems (5.7.5) are

$$\partial^i \mathbf{H}_i(s) = (-1)^i i! \hat{\mathbf{C}}_k^T \left((s\hat{\mathbf{E}}_k - \hat{\mathbf{A}}_k)^{-1} \hat{\mathbf{E}}_k \right)^i (s\hat{\mathbf{E}}_k - \hat{\mathbf{A}}_k)^{-1} \hat{\mathbf{B}}_k, \quad k = \text{BSP}, \text{SBR}. \quad (5.7.7)$$

Tables 5.1 and 5.2 tabulate the values of the first 10 derivatives ($i = 0, \ldots, 9$) at s_1 respectively s_2 of the transfer functions $\mathbf{H}$, $\mathbf{H}_{\text{BSP}}$ and $\mathbf{H}_{\text{SBR}}$ in eqs. (5.7.6) to (5.7.7).

Table 5.1: Derivatives of the original and reduced-order transfer functions for the first expansion point s_1 for the first example, multiplied by 10^6

i	$\partial^i \mathbf{H}(s_1)$	$\partial^i \mathbf{H}_{\text{BSP}}(s_1)$	$\partial^i \mathbf{H}_{\text{SBR}}(s_1)$
0	-0.839024819103714	-0.839024819105314	-0.839024819033797
1	0.024749951224057	0.024749951223572	0.024749951222262
2	-0.000439271234399	-0.000439271234405	-0.000439271234414
3	0.000006748103698	0.000006748103700	0.000006748103699
4	-0.000000102481958	-0.000000102481959	-0.000000102061233
5	0.000000001588268	0.000000001588268	0.000000001565032
6	-0.000000000025086	-0.000000000025076	-0.000000000024532
7	0.000000000000401	0.000000000000379	0.000000000000406
8	-0.000000000000008	-0.000000000000077	-0.000000000000005
9	0.000000000000004	0.000000000000071	0.000000000000005

Table 5.2: Derivatives of the original and reduced-order transfer functions for the second expansion point s_2 for the first example, multiplied by 10^6

i	$\partial^i \mathbf{H}(s_2)$	$\partial^i \mathbf{H}_{\text{BSP}}(s_2)$	$\partial^i \mathbf{H}_{\text{SBR}}(s_2)$
0	-0.384563353001982	-0.384563353003402	-0.384563353367612
1	0.002782774780907	0.002782774781066	0.002782774785200
2	-0.000023735573604	-0.000023735573635	-0.000023735573609
3	0.000000317217673	0.000000317217674	0.000000317217672
4	-0.000000006002628	-0.000000006002628	-0.000000006004274
5	0.000000000140000	0.000000000140000	0.000000000138146
6	-0.000000000003777	-0.000000000003777	-0.000000000003671
7	0.000000000000118	0.000000000000118	0.000000000000117
8	-0.000000000000004	-0.000000000000004	-0.000000000000005
9	0.000000000000000	0.000000000000000	0.000000000000000

Recall that for the BSP approximation, Krylov spaces of degree 6 were constructed. Tables 5.1 and 5.2 show that the six first derivatives ($i = 0, \ldots, 5$) indeed match the first six derivatives of the original transfer function, modulo some round-off errors. The tables show that the higher derivatives ($i = 6, \ldots, 9$) also happen to be well approximated.

For the SBR approximation, Krylov spaces of degree 4 were constructed. Tables 5.1 and 5.2 show that the first four derivatives ($i = 0, \ldots, 3$) match the first four derivatives of the original transfer function, modulo some round-off errors. Also for the SBR approxi-

mations, the tables show that the higher order derivatives ($i = 4, \ldots, 9$) are quite well approximated.

Finally, consider the computational times involved with the construction of $\mathbf{V}_{\text{BSP}}$ respectively $\mathbf{V}_{\text{SBR}}$. Because of the relatively small size of the considered system, mentioned/measured times are for 1000 repeated runs: The constructing $\mathbf{V}_{\text{BSP}}$ and $\mathbf{V}_{\text{SBR}}$ takes 17 respectively 7 seconds. For systems of higher dimensions, the differences in computational times will be significantly larger.

Example 5.7.2. However, competitive or better performance of the SBR algorithm relative to the BSP algorithm can not always be assured. By construction, the standard BSP method will give much better results than SBR when the number of internal inputs is much greater than the number of external inputs and relatively large with respect to the total number of degrees of freedom of the coupled system.

Figure 5.8 shows the magnitude plots with respect to the frequency of the original and reduced-order frequency response functions of the system, for which the application of the SBR algorithm does not give satisfactory results. The dashed vertical line marks the expansion point for which the reduction bases were computed. Here, the fully coupled
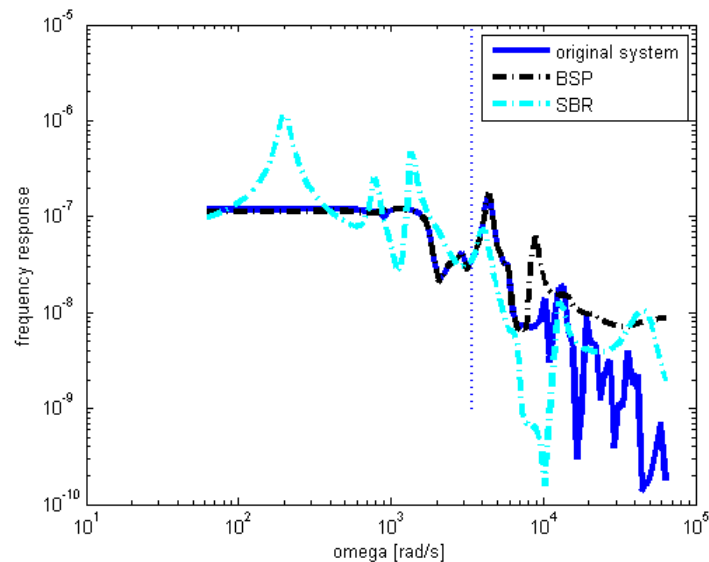
Figure 5.8: Magnitude plots of the frequency response functions of the original and reduced-order systems

system has 80 degrees of freedom, 40 for each sub-system. There is one external input

and one external output, and each of the sub-systems has ten internal inputs and ten internal outputs. Hence, $\mathbf{A}_{ii}, \mathbf{I}_{ii} \in \mathbb{R}^{40 \times 40}, i = 1, 2$, $\mathbf{B}_1, \mathbf{C}_1 \in \mathbb{R}^{40 \times 1}$, and $\mathbf{B}_j, \mathbf{C}_j \in \mathbb{R}^{40 \times 10}, j = 3, 4$. The reduction bases are based on a single expansion point $s \in \mathbb{C}$.

For the BPS method, we constructed Krylov basis $\mathcal{K}_{21}(\mathbf{P}(s), \mathbf{R}(s))$ (42 columns) and related $\mathbf{V}_{\text{BSP}}$ (as in Section 5.4) of rank 42 (no linearly dependent columns after splitting).

For the SBR method, based on matrices $\mathbf{P}_i$ and $\mathbf{R}_i$ defined earlier, we constructed Krylov basis $\mathcal{K}_2(\mathbf{P}_1(s), [\mathbf{R}_{11}(s), \mathbf{R}_{12}(s)])$ (22 columns) and Krylov basis $\mathcal{K}_2(\mathbf{P}_2(s), \mathbf{R}_{22}(s))$ (20 columns) and related $\mathbf{V}_{\text{SBR}}$ (as in Section 5.5) of rank 42 (also here no linearly dependent columns as was to be expected).

Figure 5.9 shows the magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original frequency response function as a function of frequency. In this case, the standard BSP method performs much better than the SBR method, although both reduced-order systems have the same dimension. The reason for it is that the coupled system has only 1 external input but 10 internal inputs. Hence, every Krylov iteration step adds 1 vector to the basis (which results in adding 2 columns to the split block-diagonal reduction matrix) in case of the standard BSP method, while in case of the SBR algorithm, every Krylov step for the sub-system 1 adds 11 new vectors and for the sub-system 2 there are 10 new vectors created – so every step
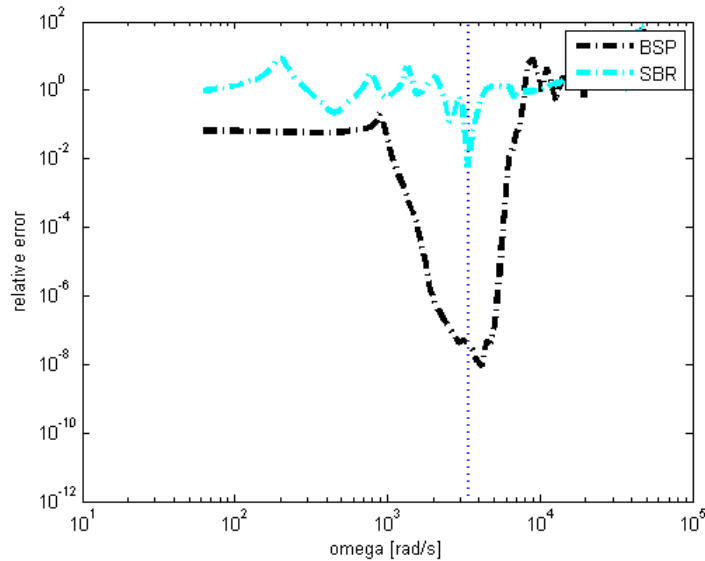
Figure 5.9: Magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original frequency response function

Table 5.3: Derivatives of the original and reduced-order transfer functions for the expansion point s for the second example, multiplied by 10^7

i	$\partial^i \mathbf{H}(s)$	$\partial^i \mathbf{H}_{\mathrm{BSP}}(s)$	$\partial^i \mathbf{H}_{\mathrm{SBR}}(s)$
0	-0.349984611544531	-0.349984598976983	-0.349984603422285
1	0.000580754070987	0.000580754084250	0.000580754105948
2	-0.000001928114532	-0.000001928114520	-0.000005770673504
3	0.000000012770698	0.000000012770698	0.000000013683621
4	-0.000000000067912	-0.000000000067912	-0.000000000098227
5	0.000000000000859	0.000000000000859	0.000000000001220
6	-0.000000000000014	-0.000000000000014	-0.000000000000011
7	0.000000000000000	0.000000000000000	0.000000000000000
8	-0.000000000000000	-0.000000000000000	-0.000000000000000
9	0.000000000000000	0.000000000000000	0.000000000000000

adds in total 21 columns to the block-diagonal reduction basis).

Again, compare the first 10 derivatives of the original and approximated transfer functions. Table 5.3 shows the values of the derivatives at the expansion point. Also in this case the reduced-order transfer functions match at least as many moments as the degree of the respective Krylov spaces. However, in this example, the standard BSP method created a Krylov space of degree 21, whereas the SBR method could afford a Krylov space of degree only 2, which explains the difference in accuracy.

5.8   Conclusions

In this chapter, we introduced a new reduction method called Separate Bases Reduction algorithm. It is a block-structure preserving algorithm, since it uses a reduction basis of a block-diagonal form. The difference between our method and other common BSP methods is that the SBR algorithm exploits the uncoupled formulation of the coupled/interconnected system. This makes our method potentially computationally faster than BSP methods. It also allows for different order of reduction for different subsystems. The SBR algorithm is the most beneficial if the number of the internal inputs and outputs is relatively small compared to the number of external inputs and outputs, as well as compared to the dimensions of the system, since the growth of the reduction bases is proportional to the number of internal inputs/outputs. In Chapters 7 and 8 we will show how this disadvantage can be reduced. We have shown that SBR matches at least the (amount of) moments that BSP does. It is easy to see that the application of the SBR method requires a special structure of the coupled system matrix $\mathbf{A}$, namely, the coupling blocks $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ need to be of the form $\mathbf{BC}^T$. However, in industrial

applications the components **B** and **C** of the product may not be given explicitly. The way to overcome this problem will be presented in Chapter 8.

In the next chapter we present an improved version of the SBR algorithm that uses a two-sided projection, i.e. reduces the system by projecting it onto two Krylov subspaces, one based on the input information, and another based on the sub-systems' output space.

Chapter 6

Two-sided Separate Bases
Reduction Algorithm

6.1 Introduction

For the Separate Bases Reduction algorithm, this chapter shows that one can double the number of matched moments by using two-sided projection as shown in Theorem 3.4.5. Afterwards numerical results are presented.

As explained in Section 5.4 a standard BSP method constructs a basis for the Krylov subspace

$$\mathcal{K}_n\left((s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}, \ (s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\right),$$

for $s_0 \in \mathbb{C}$ where (using Definition(5.2.5)) where

$$(s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} = \left(s_0\begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix}\right)^{-1}\begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix} \tag{6.1.1}$$

depends on the external inputs but not on the external outputs $\mathbf{C}_1, \mathbf{C}_2$. Similarly, as in Section 5.5, the SBR algorithm constructs bases for the Krylov spaces

$$\mathcal{K}_{n_1}\left((s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}\mathbf{E}_{11}, \; (s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}[\mathbf{B}_1 \; \mathbf{B}_3]\right)$$

and

$$\mathcal{K}_{n_2}\left((s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-1}\mathbf{E}_{22}, \; (s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-1}[\mathbf{B}_2 \; \mathbf{B}_4]\right),$$

for $s_0 \in \mathbb{C}$ which neither involve the external outputs $\mathbf{C}_1$ and $\mathbf{C}_2$ (nor internal outputs $\mathbf{C}_3$ and $\mathbf{C}_4$). In this chapter we propose a two-sided SBR algorithm which exploits the external and internal outputs.

6.2   Two-sided structure preserving methods

In this section we will explain how the two-sided projection idea can be implemented in case of the block-structure preserving methods. A detailed explanation of the two-sided methods one can find for instance in [37]. Generally speaking, the use of a two-sided reduction method means, that the system is projected onto two subspaces, $\mathbf{V}$ and $\mathbf{W}$, based on input and output matrices, respectively. In case of the coupled system (5.2.4), the reduction matrices $\mathbf{V}$ and $\mathbf{W}$, for an expansion point $s_0 \in \mathbb{C}$, are built according to the following algorithm:

1. Create matrix $\tilde{\mathbf{V}}$, whose columns span the nth Krylov subspace around $s_0 \in \mathbb{C}$

$$\tilde{\mathbf{V}} = \mathcal{K}_n(\mathbf{P}(s_0), \mathbf{R}(s_0)),$$

   where $\mathbf{P}(s_0)$ and $\mathbf{R}(s_0)$ are

$$\mathbf{P}(s_0) = (s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{E} \quad \text{and} \quad \mathbf{R}(s_0) = (s_0\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}.$$

2. Create matrix $\tilde{\mathbf{W}}$, whose columns span the nth Krylov subspace around $s_0 \in \mathbb{C}$

$$\tilde{\mathbf{W}} = \mathcal{K}_n(\mathbf{S}(s_0), \mathrm{T}(s_0)),$$

   where $\mathbf{S}(s_0)$ and $\mathrm{T}(s_0)$ are

$$\mathbf{S}(s_0) = (s_0\mathbf{E} - \mathbf{A})^{-T}\mathbf{E}^T \quad \text{and} \quad \mathrm{T}(s_0) = (s_0\mathbf{E} - \mathbf{A})^{-T}\mathbf{C}.$$

3. Build the block-diagonal reduction matrix $\mathbf{V}$ with $N_1 + N_2 = N$ rows

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix},$$

   where $\mathbf{V}_1$ and $\mathbf{V}_2$ contain the first N_1 respectively last N_2 rows of the matrix $\tilde{\mathbf{V}}$.

4. Build the block-diagonal reduction matrix $\mathbf{W}$ with $N_1 + N_2 = N$ rows

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 \end{bmatrix},$$

   where $\mathbf{W}_1$ and $\mathbf{W}_2$ contain the first N_1 respectively last N_2 rows of the matrix $\tilde{\mathbf{W}}$.

5. Project the original system onto the lower-dimensional space

$$\hat{\mathbf{E}}_{\text{BSP}} = \mathbf{W}^T \mathbf{E} \mathbf{V}, \quad \hat{\mathbf{A}}_{\text{BSP}} = \mathbf{W}^T \mathbf{A} \mathbf{V}, \quad \hat{\mathbf{B}}_{\text{BSP}} = \mathbf{W}^T \mathbf{B}, \quad \hat{\mathbf{C}}_{\text{BSP}} = \mathbf{V}^T \mathbf{C}.$$

Different algorithms lead to $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{W}}$ (and hence $\mathbf{V}$ and $\mathbf{W}$) with different specific properties (such as orthogonality or bi-orthogonality). Some properties and their advantages and disadvantages are discussed in [80].

The described BSP algorithm results in a block-structured reduced order system and uses all of the inputs *and* outputs. Consequently, the BSP-based reduced order system's transfer function matches twice as many moments of the original system's transfer function as the only inputs based one in Section 5.4.

6.3 Two-sided Separate Bases Reduction algorithm

The two-sided projection technique introduced in the previous section can be adapted to similarly improve the moment matching properties of the SBR algorithm. With the uncoupled formulation (5.2.1) and (5.2.2) in mind we define the reduction algorithm as follows.

1. For the sub-system S_1, create two matrices:
   - Matrix $\mathbf{V}_1$, whose columns span the n_1th Krylov subspace around $s_0 \in \mathbb{C}$

$$\mathbf{V}_1 = \mathcal{K}_{n_1}(\mathbf{P}_1(s_0), \mathbf{R}_1(s_0)),$$

   where $\mathbf{P}_1(s_0)$ and $\mathbf{R}_1(s_0)$ are

$$\mathbf{P}_1(s_0) = (s_0 \mathbf{E}_{11} - \mathbf{A}_{11})^{-1} \mathbf{E}_{11} \quad \text{and} \quad \mathbf{R}_1(s_0) = (s_0 \mathbf{E}_{11} - \mathbf{A}_{11})^{-1}[\mathbf{B}_1 \ \mathbf{B}_3].$$

   Matrix $\mathbf{V}_1$ has N_1 rows.
   - Matrix $\mathbf{W}_1$, whose columns span the n_1th Krylov subspace around $s_0 \in \mathbb{C}$

$$\mathbf{W}_1 = \mathcal{K}_{n_1}(\mathbf{S}_1(s_0), \mathrm{T}_1(s_0)),$$

   where $\mathbf{S}_1(s_0)$ and $\mathrm{T}_1(s_0)$ are

$$\mathbf{S}_1(s_0) = (s_0 \mathbf{E}_{11} - \mathbf{A}_{11})^{-T} \mathbf{E}_{11}^T \quad \text{and} \quad \mathrm{T}_1(s_0) = (s_0 \mathbf{E}_{11} - \mathbf{A}_{11})^{-T}[\mathbf{C}_1 \ \mathbf{C}_3].$$

   Matrix $\mathbf{W}_1$ has N_1 rows.

2. For the sub-system S_2, create two matrices:

   - Matrix $\mathbf{V}_2$, whose columns span the n_2th Krylov subspace around $s_0 \in \mathbb{C}$

$$\mathbf{V}_2 = \mathcal{K}_{n_2}(\mathbf{P}_2(s_0), \mathbf{R}_2(s_0)),$$

   where $\mathbf{P}_2(s_0)$ and $\mathbf{R}_2(s_0)$ are

$$\mathbf{P}_2(s_0) = (s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-1}\mathbf{E}_{22} \quad \text{and} \quad \mathbf{R}_2(s_0) = (s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-1}[\mathbf{B}_2\ \mathbf{B}_4].$$

   Matrix $\mathbf{V}_2$ has N_2 rows.

   - Matrix $\mathbf{W}_2$, whose columns span the n_2th Krylov subspace around $s_0 \in \mathbb{C}$

$$\mathbf{W}_2 = \mathcal{K}_{n_2}(\mathbf{S}_2(s_0), \mathbf{T}_2(s_0)),$$

   where $\mathbf{S}_2(s_0)$ and $\mathbf{T}_2(s_0)$ are

$$\mathbf{S}_2(s_0) = (s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-T}\mathbf{E}_{22}^{T} \quad \text{and} \quad \mathbf{T}_2(s_0) = (s_0\mathbf{E}_{22} - \mathbf{A}_{22})^{-T}[\mathbf{C}_2\ \mathbf{C}_4].$$

   Matrix $\mathbf{W}_2$ has N_2 rows.

3. Build two block-diagonal reduction matrices $\mathbf{V}$ and $\mathbf{W}$ with $N_1 + N_2 = N$ rows

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix} \qquad \text{and} \qquad \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 \end{bmatrix}.$$

4. Project the original system onto the lower-dimensional space

$$\hat{\mathbf{E}}_{\text{SBR}} = \mathbf{W}^T\mathbf{E}\mathbf{V}, \quad \hat{\mathbf{A}}_{\text{SBR}} = \mathbf{W}^T\mathbf{A}\mathbf{V}, \quad \hat{\mathbf{B}}_{\text{SBR}} = \mathbf{W}^T\mathbf{B}, \quad \hat{\mathbf{C}}_{\text{SBR}} = \mathbf{V}^T\mathbf{C}.$$

Again, different algorithms lead to $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{W}_1$, $\mathbf{W}_2$ with different properties. Also the above SBR algorithm results in a block-structured reduced order system and uses all of the inputs *and* outputs. Consequently, also the above SBR-based reduced order system's transfer function matches twice as many moments of the original system's transfer function as the only inputs based one in Section 5.4.

6.4 Numerical examples

In this section we compare the two-sided SBR reduction algorithm from Section 6.3 to the two-sided BSP method from Section 6.2. The test cases are the two examples introduced in Section 5.7. For better comparison of the performance of the two-sided algorithms the one-sided numerical results are added to the pictures.

Example 6.4.1. The interconnected system is as in Example 5.7.1. It has 120 degrees of freedom, 60 for each sub-system. There is one external input and one external output, and each of the sub-systems has one internal input and one internal output. Hence, $\mathbf{A}_{ii}, \mathbf{I}_{ii} \in \mathbb{R}^{60 \times 60}$, $i = 1, 2$ and $\mathbf{B}_i, \mathbf{C}_i \in \mathbb{R}^{60 \times 1}$, $i = 1, 3, 4$. The reduced-order models computed for this example have 24 degrees of freedom, obtained after using expansion around two points $s_i, i = 1, 2$. Let $\mathbf{P}$, $\mathbf{R}$, $\mathbf{S}$ and T be defined as in Section 6.2, $\mathbf{P}_i$, $\mathbf{R}_i$, $\mathbf{S}_i$ and T_i be defined as in Section 6.3, all related to the interconnected system (5.2.4) shown in Figure 5.5.

To obtain 24 degrees of freedom for BSP we create bases for the four linear spans $\mathcal{K}_6(\mathbf{P}(s_i), \mathbf{R}(s_i))$ and $\mathcal{K}_6(\mathbf{S}(s_i), \mathrm{T}(s_i))$, $i = 1, 2$, after which we construct the block-diagonal reduction matrices $\mathbf{V}_{\mathrm{BSP}}$ and $\mathbf{W}_{\mathrm{BSP}}$ as described in Section 6.2. The latter matrices turn out to consist of 24 linearly independent columns each.

For the SBR method we create bases for the linear spans $\mathcal{K}_4(\mathbf{P}_1(s_i), \mathbf{R}_1(s_i))$, $\mathcal{K}_4(\mathbf{P}_2(s_i), \mathbf{R}_{22}(s_i))$ and $\mathcal{K}_4(\mathbf{S}_1(s_i), \mathrm{T}_1(s_i))$, $\mathcal{K}_4(\mathbf{S}_2(s_i), \mathrm{T}_{22}(s_i))$, $i = 1, 2$, and create the block-diagonal reduction matrices $\mathbf{V}_{\mathrm{SBR}}$, $\mathbf{W}_{\mathrm{SBR}}$, which turn out to consist of 24 linearly independent columns each.

Figure 6.1 shows the magnitude plots with respect to the frequency of the frequency response functions of the original and the reduced-order systems. The dashed vertical lines mark the expansion points for which the reduction bases were computed. Figure 6.2 shows the magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original frequency response function as a function of frequency. Note that the two-sided version of the SBR algorithm produces considerably better results than the one-sided SBR algorithm. In this example the two-sided SBR is still less accurate than the two-sided standard BSP method around the expansion points but it approximates the original system better for higher frequencies. It also performs better than the one-sided standard BSP method.

Next we consider the two-sided SBR moment matching property. Tables 6.1 and 6.2 show the values of the first 10 derivatives $i = 0, ..., 9$ of the transfer functions of (1) the original system ($\mathbf{H}$); (2) the system reduced using two-sided BSP ($\mathbf{H}_{\mathrm{BSP}}$); and (3) the system reduced using the two-sided SBR method ($\mathbf{H}_{\mathrm{SBR}}$) – all based on the two expansion points s_1 and s_2.

For the two-sided BSP method, the values in the table agree with the expectations based on theory – the number of derivatives matched is equal to the double number of columns contained by one of the reduction bases. In this case, the number of columns was 6, so the reduced-order model should match $2 \cdot 6$ derivatives. In the table we limited the number of compared derivatives to 10 and it is clear, that the first 10 values match the
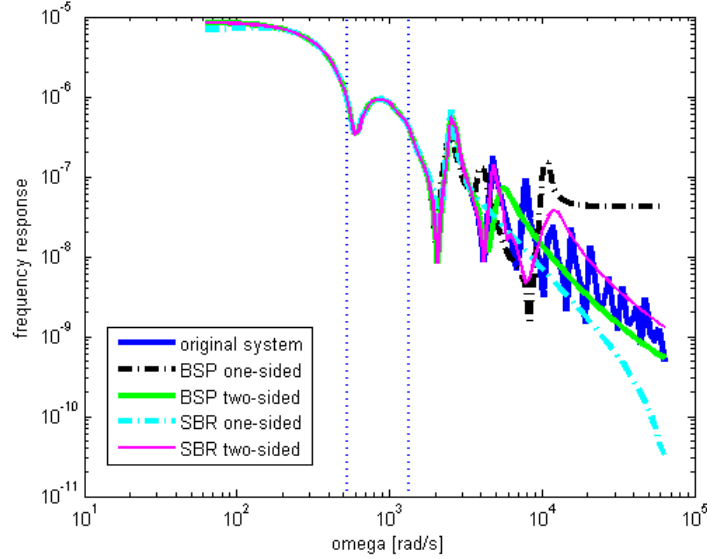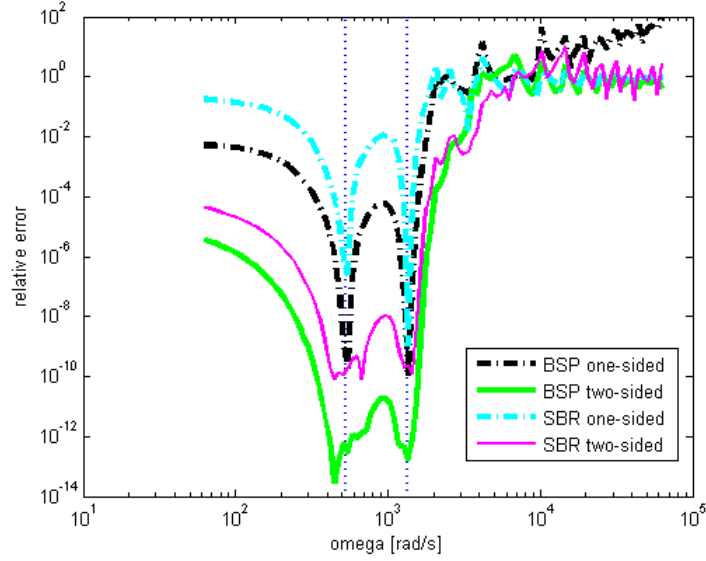
Figure 6.1: Magnitude plots of the frequency response functions of the original and reduced-order systems

Table 6.1: Derivatives of the original and reduced-order transfer functions using the two-sided method for the first expansion point s_1 for the first example, multiplied by 10^6

i	$\partial^i \mathbf{H}(s_1)$	$\partial^i \mathbf{H}_{\mathrm{BSP}}(s_1)$	$\partial^i \mathbf{H}_{\mathrm{SBR}}(s_1)$
0	-0.839024819103714	-0.839024819103787	-0.839024819123461
1	0.024749951224057	0.024749951224058	0.024749951225162
2	-0.000439271234399	-0.000439271234399	-0.000439271249222
3	0.000006748103698	0.000006748103698	0.000006748084130
4	-0.000000102481958	-0.000000102481958	-0.000000102470683
5	0.000000001588268	0.000000001588268	0.000000001573684
6	-0.000000000025086	-0.000000000025086	-0.000000000028534
7	0.000000000000401	0.000000000000401	0.000000000003063
8	-0.000000000000008	-0.000000000000008	-0.000000000001525
9	0.000000000000004	0.000000000000004	0.000000000001877

values of the original system. In case of the SBR algorithm, one can also see that for the expansion point s_2 the number of matched derivatives equal double number of columns contained by the basis $(2 \cdot 4)$. Moreover, the last two derivatives are also equal to the last two derivatives of the original transfer function. On the other hand, for the expansion

Figure 6.2: Magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original frequency response function

Table 6.2: Derivatives of the original and reduced-order transfer functions using two-sided method for the second expansion point s_2 for the first example, multiplied by 10^6

i	$\partial^i \mathbf{H}(s_2)$	$\partial^i \mathbf{H}_{\mathrm{BSP}}(s_2)$	$\partial^i \mathbf{H}_{\mathrm{SBR}}(s_2)$
0	-0.384563353001982	-0.384563353001980	-0.384563352994440
1	0.002782774780907	0.002782774780906	0.002782774780961
2	-0.000023735573604	-0.000023735573604	-0.000023735573638
3	0.000000317217673	0.000000317217673	0.000000317217674
4	-0.000000006002628	-0.000000006002628	-0.000000006002628
5	0.000000000140000	0.000000000140000	0.000000000140000
6	-0.000000000003777	-0.000000000003777	-0.000000000003777
7	0.000000000000118	0.000000000000118	0.000000000000118
8	-0.000000000000004	-0.000000000000004	-0.000000000000004
9	0.000000000000000	0.000000000000000	0.000000000000000

point (s_1) only the first 7 derivatives can be considered as matched, probably due to the exploited software (since theoretically 8 moments should match). We have not been able to track down the cause (yet) but have observed that this "mismatch" rarely occurs. In

this example, for most (other) expansion points s the 8 moments match.

Example 6.4.2. Here, the fully coupled system is as in Example 5.7.2. It has 80 degrees of freedom, 40 for each sub-system. There is one external input and one external output, and each of the sub-systems has ten internal inputs and ten internal outputs. Hence, $\mathbf{A}_{ii}, \mathbf{I}_{ii} \in \mathbb{R}^{40 \times 40}, i = 1, 2,\ \mathbf{B}_1, \mathbf{C}_1 \in \mathbb{R}^{40 \times 1},\ \text{and } \mathbf{B}_j, \mathbf{C}_j \in \mathbb{R}^{40 \times 10}, j = 3, 4$. The reduction matrices are based on a single expansion point $s \in \mathbb{C}$ and the reduced-order model has 42 degrees of freedom.

To obtain 42 degrees of freedom for BSP method we create bases for the two linear spans $\mathcal{K}_{21}(\mathbf{P}(s), \mathbf{R}(s))$ and $\mathcal{K}_{21}(\mathbf{S}(s), \mathbf{T}(s))$ after which we construct the block-diagonal reduction matrices $\mathbf{V}_{\mathrm{BSP}}$ and $\mathbf{W}_{\mathrm{BSP}}$ as described in Section 6.2. The latter matrices turn out to consist of 42 linearly independent columns each.

For the SBR method we create bases for the linear spans $\mathcal{K}_2(\mathbf{P}_1(s), \mathbf{R}_1(s))$, $\mathcal{K}_2(\mathbf{P}_2(s), \mathbf{R}_{22}(s))$ and $\mathcal{K}_2(\mathbf{S}_1(s), \mathbf{T}_1(s))$, $\mathcal{K}_2(\mathbf{S}_2(s), \mathbf{T}_{22}(s))$, $i = 1, 2$, and create the block-diagonal reduction matrices $\mathbf{V}_{\mathrm{SBR}}$, $\mathbf{W}_{\mathrm{SBR}}$, which turn out to consist of 42 linearly independent columns each.

Figure 6.3 shows the magnitude plots with respect to the frequency of the frequency response functions of the original and the reduced-order systems. The dashed vertical line marks the expansion point for which the reduction bases were computed. Figure 6.4 shows the magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original frequency response function as a function of frequency. In this case, the use of the two-sided version of the SBR algorithm neither gives satisfactory results. However, the accuracy of its approximation around the expansion point is noticeably higher for the two-sided version than for the one-sided. This can be explained if we again look at the first 10 derivatives of the original transfer function and the transfer functions of the reduced-order models computed using two-sided BSP method ($\mathbf{H}_{\mathrm{BSP}}$) and using the two-sided SBR method ($\mathbf{H}_{\mathrm{SBR}}$), for the expansion point s, presented in table 6.3. As expected, $\mathbf{H}_{\mathrm{BSP}}$ matches all first 10 derivatives. Also for the two-sided SBR algorithm the double number of derivatives $2 \cdot 2$. Moreover, also several higher derivatives are accurately approximated.

6.5   Conclusions

In this chapter, we proposed a two-sided variant of the SBR algorithm introduced in Chapter 5. As shown by numerical experiments, this variant increases the amount of matched moments. The numerical tests corroborated that $\mathbf{H}_{\mathrm{SBR}}$ matches twice as many derivatives of the original transfer function $\mathbf{H}$ at the expansion points as the dimension
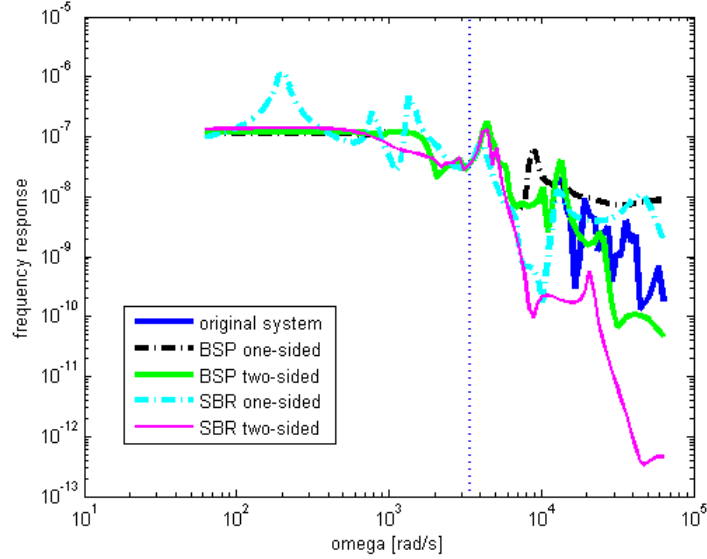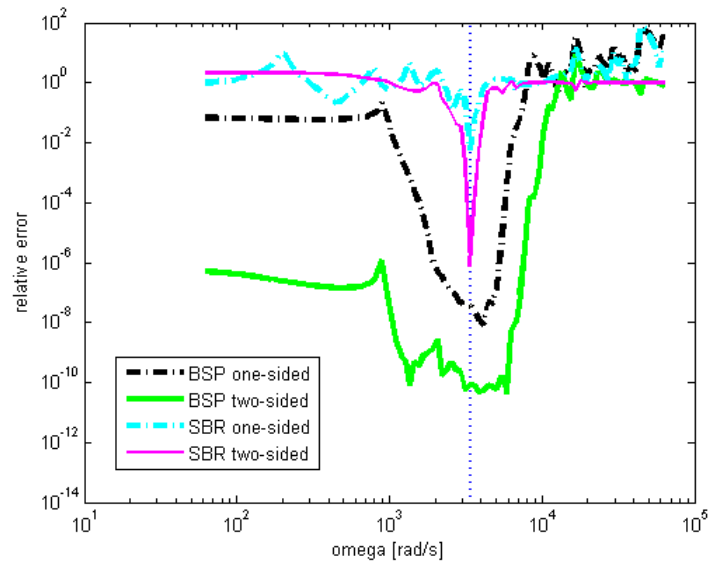
Figure 6.3: Magnitude plots of the frequency response functions of the original and reduced-order systems

Table 6.3: Derivatives of the original and reduced-order transfer functions using two-sided method for the expansion point s for the second example, multiplied by 10^7

i	$\partial^i \mathbf{H}(s)$	$\partial^i \mathbf{H}_{\text{BSP}}(s)$	$\partial^i \mathbf{H}_{\text{SBR}}(s)$
0	-0.349984611544531	-0.349984611562807	-0.349984611962056
1	0.000580754070987	0.000580754070958	0.000580754070360
2	-0.000001928114532	-0.000001928114532	-0.000001928114560
3	0.000000012770698	0.000000012770698	0.000000012770698
4	-0.000000000067912	-0.000000000067912	-0.000000000071351
5	0.000000000000859	0.000000000000859	0.000000000000934
6	-0.000000000000014	-0.000000000000014	-0.000000000000015
7	0.000000000000000	0.000000000000000	0.000000000000000
8	-0.000000000000000	-0.000000000000000	-0.000000000000000
9	0.000000000000000	0.000000000000000	0.000000000000000

of smallest reduced-order sub-system. The results obtained by application of the two-sided variant of the SBR technique seem to be very competitive with (and for some cases even considerably better than) the results obtained by standard one-sided BSP methods. However, the standard two-sided BSP techniques still perform better, since for a good

Figure 6.4: Magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original frequency response function

match with the original transfer function not only moment matching is important, but the whole space spanned by the reduction bases. On the other hand, it should be stressed that usually the SBR algorithm is considerably cheaper computationally than standard BSP reduction techniques, which implies that one has a choice between larger but faster to compute or smaller but slower to compute reduced-order models.

In case of the two-sided SBR algorithm the same limitation holds as for the one-sided variant. The algorithm is most beneficial for a system with a small number of internal inputs and outputs, compared to the number of the external inputs and outputs as well as to the dimension of the system. For other systems the two-sided SBR algorithm is not recommended.

Chapter 7

Low-rank Approximations of Couplings

7.1 Introduction

In Chapters 5 and 6 we introduced the structure preserving separate bases reduction (SBR) model order reduction method which is suited for the reduction of coupled linear systems. Although the SBR method can conceptually be applied to an arbitrary coupled system it turns out to be efficient mainly if the sub-systems are weakly coupled. In the case of many interconnected internal inputs and outputs the SBR reduction bases will quickly reach very large dimensions while only matching a minimal (not sufficient) number of moments.

In this chapter, we study the couplings in more details. We introduce a singular value decomposition (SVD) and a generalized singular value decomposition (GSVD) as tools to determine whether all of the information about the coupling that is contained in the system description is relevant, or whether it is possible to neglect part of it and still have a good accuracy of the coupled problem's solution. For more information on the use of GSVD see also [53].

In this chapter we focus on a linear system of equations of the form

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}, \tag{7.1.1}$$

where $\mathbf{A}_{ii} \in \mathbb{R}^{N_i \times N_i}$, $i = 1, 2$, $\mathbf{A}_{12} \in \mathbb{R}^{N_1 \times N_2}$, $\mathbf{A}_{21} \in \mathbb{R}^{N_2 \times N_1}$, $\mathbf{x}_i, \mathbf{b}_i \in \mathbb{R}^{N_i}$, $N_1 + N_2 = N$ and $N_i \in \mathbb{N}$, $i = 1, 2$.

Here, the blocks $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ describe the coupling between the two sub-domains in which the degrees of freedom $\mathbf{x}_1$ and $\mathbf{x}_2$ are defined. We are interested in finding low-rank approximations $\tilde{\mathbf{A}}_{12}$ and $\tilde{\mathbf{A}}_{21}$ of the coupling matrices $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$, respectively, that solve the system

$$\begin{bmatrix} \mathbf{A}_{11} & \tilde{\mathbf{A}}_{12} \\ \tilde{\mathbf{A}}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}, \tag{7.1.2}$$

where $[\tilde{\mathbf{x}}_1^T, \tilde{\mathbf{x}}_2^T]^T$ is a good approximation of the solution of the system (7.1.1).

In this chapter we use methods that approximate the coupling blocks $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ by matrices of lower rank. The main two reasons for this choice are:

- If the coupling blocks $\tilde{\mathbf{A}}_{12}$ and $\tilde{\mathbf{A}}_{21}$ of the system (7.1.2) are of sufficiently low rank, then the system (7.1.2) can be solved efficiently by applying the *Sherman-Morrison-Woodbury* formula (5.3.15) introduced in Section 5.3

$$\mathbf{K}^{-1} = (\mathbf{L} + \mathbf{MJN}^T)^{-1} = \mathbf{L}^{-1} - \mathbf{L}^{-1}\mathbf{M}(\mathbf{J}^{-1} + \mathbf{N}^T\mathbf{L}^{-1}\mathbf{M})^{-1}\mathbf{N}^T\mathbf{L}^{-1};$$

- If the matrix

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

  is the system matrix $\mathbf{A}$ of the coupled system (5.2.4), then the dimensions of the internal input and output matrices $\mathbf{B}_3$, $\mathbf{B}_4$, $\mathbf{C}_3$ and $\mathbf{C}_4$ are directly related to the rank of the coupling blocks $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$. If it is possible to find a good low-rank approximation of the coupling blocks, then the approximated internal input and output matrices $\tilde{\mathbf{B}}_3$, $\tilde{\mathbf{B}}_4$, $\tilde{\mathbf{C}}_3$ and $\tilde{\mathbf{C}}_4$ with

$$\tilde{\mathbf{A}}_{12} = \tilde{\mathbf{B}}_3 \tilde{\mathbf{C}}_4^T \qquad \text{and} \qquad \tilde{\mathbf{A}}_{21} = \tilde{\mathbf{B}}_4 \tilde{\mathbf{C}}_3^T$$

  will have smaller sizes than the original ones.

In the next sections, we introduce the singular value decomposition and its generalized version as the tools to find the low-rank approximations of the coupling blocks. We are especially interested in low-rank approximations of the form $\mathbf{BC}^T$ or $\mathbf{BDC}^T$ where $\mathbf{B}, \mathbf{C}$ are of low rank (only have a few linearly independent columns) and $\mathbf{D}$ is a diagonal matrix (with a non-negative diagonal) since such approximations are suited for the application of the Sherman-Morrison-Woodbury formula (5.3.15). Both the SVD and the GSVD provide factorizations of the form $\mathbf{BDC}^T$. The difference is that the SVD can be applied to only one block such as $\mathbf{A}_{12}$ and the GSVD must be applied to two blocks such as $\mathbf{A}_{11}$ and $\mathbf{A}_{12}$ simultaneously. The latter method turns out to better take into account the couplings.

7.2   SVD and GSVD

This section presents the standard SVD and in more detail the GSVD . Standard defacto GSVD-documentation such as [55] states that each matrix can be factored into $\mathbf{UCX}^T$ where $\mathbf{U}$ is orthogonal (unitary), $\mathbf{C}$ is diagonal with entries in decreasing or increasing order and $\mathbf{X}$ is non-singular. Here, we stick to the notation used in the literature. Note that this matrix $\mathbf{C}$ differs from our output matrix $\mathbf{C}$. Wherever the difference between output matrix and the GSVD-component matrix may be not clear, we make an explicit remark which of them is meant. Reference works such as [35, Theorem 2.5.2] provide a bit more information but not all possibly needed details. We will present the GSVD from [2] and describe it in full detail. Let us start with the SVD.

Theorem 7.2.1. *(The SVD of $\mathbf{A} \in \mathbb{R}^{n \times m}$)*
There exists orthogonal matrices $\mathbf{U} \in \mathbb{R}^{n \times n}$, $\mathbf{V} \in \mathbb{R}^{m \times m}$ and a diagonal matrix $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p, 0, \ldots) \in \mathbb{R}^{n \times m}$, $0 \le p = \mathrm{rank}(\mathbf{A}) \le \min(n, m)$ such that

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T \tag{7.2.1}$$

and $\sigma_1 \ge \ldots \ge \sigma_p > 0$.

Proof. See for instance [35, Theorem 2.5.2]. □

Below we provide our definition of explicit rank-revealing and show its relation to a property called rank-revealing in the QR literature (see for instance [10]).

Definition 7.2.1. *A factorization $\mathbf{A} = \mathbf{BC}^T$ is called* explicitly rank-revealing *if both $\mathbf{B}$ and $\mathbf{C}$ have full column rank.*

Our definition of explicit rank-revealing is more strict than the one of rank revealing in the QR literature, since we in addition require both $\mathbf{B}$ and $\mathbf{C}$ to be of full column rank (our factorization reveals the rank of $\mathbf{A}$ as is shown by Theorem 8.3.1 in Chapter 8). In the QR literature a QR method is called *rank-revealing* if it 'reveals' the rank of the matrix $\mathbf{A}$ – and not necessarily if $\mathbf{Q}$ and $\mathbf{R}^T$ have full column rank. For our applications later on we expect $\mathbf{A}$ to be singular. Note that implementations of factorizations of such a matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{C}^T}$$

can lead to factors $\mathbf{B}$, $\mathbf{C}^T$ which have zero columns, respectively zero rows. If one deletes each zero column k in $\mathbf{B}$ (here column $k = 3$) and simultaneously row k in $\mathbf{C}^T$ then

$$
\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\left(\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}^T \right)}_{\mathbf{C}}
$$

where both factors $\mathbf{B}$, $\mathbf{C}$ have full column rank. Another, mainly QR-bound, rank-revealing definition exists in the literature: One can use a permutation $\mathbf{P}$ to (swap) the third and fourth column:

$$
\mathbf{AP} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{P} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{Q}} \underbrace{\left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{P} \right)}_{\mathbf{R}}
$$

to obtain a matrix $\mathbf{Q}$ of full column rank (preferably orthogonal or unitary) and a block matrix $\mathbf{R}$ of the form

$$
\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}
$$

with non-singular square $\mathbf{R}_1$ and with or without the zero blocks. For our example, $\mathbf{R} = \mathbf{C}^T$ has no zero-blocks. But for square singular $\mathbf{A}$

$$
\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}
$$

would produce such blocks and lead to $\mathbf{B}$ and $\mathbf{C}$ which are not of full column rank. In such case, one should consider an alternative factorization

$$
\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C} \\ \mathbf{0} \end{bmatrix}^T
$$

where $\mathbf{B}$ and $\mathbf{C}$ are taken to be of full column rank. The proof that one can construct such factors depends on the factorization algorithm under consideration. Below we will show that the SVD and the GSVD (just as standard Householder reflections and other algorithms) can produce explicitly rank-revealing factorizations.

Theorem 7.2.2. *(The explicitly rank-revealing SVD variant rSVD of $\mathbf{A} \in \mathbb{R}^{n \times m}$)*
There exist, possibly non-square, matrices $\mathbf{U} \in \mathbb{R}^{n \times p}$, $\mathbf{V} \in \mathbb{R}^{m \times p}$ with orthogonal columns and a diagonal matrix $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{p \times p}$, $0 \le p = \mathrm{rank}(\mathbf{A}) \le \min(n, m)$ such that

$$
\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T \tag{7.2.2}
$$

and $\sigma_1 \geq \ldots \geq \sigma_p > 0$, i.e., the diagonal matrix Σ does not contain rows or columns which are the zero vector. Furthermore there exist matrices $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{Y} \in \mathbb{R}^{m \times p}$ of full column rank such that

$$\mathbf{A} = \mathbf{XY}^T. \tag{7.2.3}$$

Proof. Observe that the diagonal of Σ in Theorem 7.2.1 contains p non-zero and $\min(n, m) - p$ zeros, and that therefore Σ can be written as a block matrix:

$$\Sigma = \begin{matrix} \\ p \\ n-p \end{matrix} \begin{pmatrix} \overset{p}{\Sigma_{11}} & \overset{m-p}{\mathbf{0}} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

where $\Sigma_{11} = \text{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{p \times p}$. This partition into blocks induces a similar partition of $\mathbf{U}$ and $\mathbf{V}$ into 1×2 block-matrices:

$$\mathbf{U} = n \begin{pmatrix} \overset{p}{\mathbf{U}_1} & \overset{n-p}{\mathbf{U}_2} \end{pmatrix}, \qquad \mathbf{V} = m \begin{pmatrix} \overset{p}{\mathbf{V}_1} & \overset{m-p}{\mathbf{V}_2} \end{pmatrix}$$

whence

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} = \mathbf{U}_1 \Sigma_{11} \mathbf{V}_1^T.$$

The last claim follows straightforward: Because the singular values are non-negative

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \mathbf{U}\Sigma^{1/2}\Sigma^{1/2}\mathbf{V}^T = \underbrace{\mathbf{U}\Sigma^{1/2}}_{\mathbf{X}} \underbrace{\left(\mathbf{V}\Sigma^{1/2}\right)^T}_{\mathbf{Y}^T}.$$

The matrices $\mathbf{X} = \mathbf{U}\Sigma^{1/2}$ and $\mathbf{Y} = \mathbf{V}\Sigma^{1/2}$ are of full column rank because $\Sigma^{1/2}$ is of full column rank. This implies colspan $\mathbf{U}\Sigma^{1/2}$ = colspan $\mathbf{U}$ and because $\mathbf{U}$ is of full column rank (is orthogonal) therefore also $\mathbf{X}$ is. We call the columns of $\mathbf{U}$ and $\mathbf{V}$ the *principal components* or *dominant components*. $\square$

Similarly we need to show that there exists and use an explicitly rank-revealing variant *rGSVD* of the GSVD. The factorization [35, Theorem 8.7.4] and the one employed by `MATLAB` can return zeros on the diagonal of $\mathbf{C}$ and $\mathbf{S}$ and are not straightforwardly suited. The reason that there can be zeros on the diagonals is that `MATLAB` does not return the ranks l and k as computed for instance by `LAPACK` but defines $\mathbf{C}$ and $\mathbf{S}$ in terms of the dimensions of the input matrices $\mathbf{A}$ and $\mathbf{B}$. First, consider the standard GSVD as presented in the reference work [35].

Theorem 7.2.3. *(The standard GSVD as presented in [35, Theorem 8.7.4])*
Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\mathbf{B} \in \mathbb{R}^{p \times n}$ *and let* $q = \min(p, n)$. *Then there exist orthogonal matrices*
$\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{p \times p}$, *and a square non-singular matrix* $\mathbf{X} \in \mathbb{R}^{n \times n}$ *and diagonal matrices*
$\mathbf{C} = \mathrm{diag}(c_1, \ldots, c_n)$, $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_q)$ *such that*

$$\begin{aligned}
\mathbf{A} &= \mathbf{U}\mathbf{C}\mathbf{X}^T \\
\mathbf{B} &= \mathbf{V}\mathbf{S}\mathbf{X}^T,
\end{aligned} \tag{7.2.4}$$

such that $\mathbf{C}^2 + \mathbf{S}^2 = \mathbf{I}$.

Proof. See [35, Theorem 8.7.4]. This theorem and its proof do not mention that if
$\mathrm{rank}([\mathbf{A}; \mathbf{B}]) > m$ then $\mathbf{C}$ is not necessarily a diagonal matrix, but has a non-zero upper
diagonal $p = \mathrm{rank}([\mathbf{A}; \mathbf{B}]) - m$, i.e., there exists a natural number p such that matrix $\mathbf{C}$
has only possible non-zero entries c_{ij} when $j - i = p$. Online GSVD-documentation [55]
mentions that it is possible that $\mathbf{C}$ is a matrix with only one non-zero upper diagonal, but
does not specify which diagonal it would be.   □

In Theorem 7.2.4 we present a GSVD based on `LAPACK` version 3.3.1 which creates $\mathbf{C}, \mathbf{S}$
with positive diagonals. An alternative would be the GSVD presented in [90] based
on a CS factorization as in [86]. The dimensions m, n, p below are as defined in [35,
Theorem 8.7.4] and [2], and the ranks k, l are as defined in [2].

Because `MATLAB` and [35, Theorem 8.7.4] use the input-matrices' dimensions instead of
the ranks k and l in some cases zeros are added to $\mathbf{S}$, even for the "*economy-sized SVD*"
version, which increases the amount of columns of $\mathbf{V}$ and or $\mathbf{U}$. For large problems
where as little as possible computer memory should be used to calculate a GSVD, we
developed the explicitly rank-revealing version for $\mathbf{B}$ in Theorem 7.2.5. It uses the least
amount of memory possible and is based on [2].

Theorem 7.2.4. *(A GSVD based on [2])*
Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\mathbf{B} \in \mathbb{R}^{p \times n}$ *and let* $0 < l := \mathrm{rank}(B)$, $q := \mathrm{rank}([A; B])$, $k = q - l$.
Then there exist orthogonal matrices $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{p \times p}$, *and* $\mathbf{Q} \in \mathbb{R}^{n \times n}$, *non-singular*
upper-triangular matrix $\mathbf{R} \in \mathbb{R}^{n \times q}$, *diagonal matrix* $\mathbf{D}_2 = \mathrm{diag}(s_1, s_2, \ldots, s_{\min\{p, q\}}) \in$
$\mathbb{R}^{p \times q}$ *and matrix with one non-zero diagonal* $\mathbf{D}_1 = \mathrm{diag}(c_1, c_2, \ldots) \in \mathbb{R}^{m \times q}$ *such that*

$$\begin{aligned}
\mathbf{A} &= \mathbf{U}\mathbf{D}_1 \left(\mathbf{Q}\mathbf{R}\right)^T, \\
\mathbf{B} &= \mathbf{V}\mathbf{D}_2 \left(\mathbf{Q}\mathbf{R}\right)^T.
\end{aligned} \tag{7.2.5}$$

Proof. For the proof of the existence of this factorization see [35, Theorem 8.7.4]. The
definition of the dimensions l and k can be found in `LAPACK` version 3.3.1, source code
file `dggsvd.f`. The related `LAPACK` documentation [2] states: There exist (1) orthogonal

matrices $\mathbf{U}, \mathbf{V}, \mathbf{Q}$; (2) matrices $\mathbf{D}_1, \mathbf{D}_2$ with one non-zero diagonal; and (3) an upper block triangular matrix $\hat{\mathbf{R}}$ such that

$$\mathbf{A} = \mathbf{U}\mathbf{D}_1\mathbf{X}^T, \qquad \mathbf{B} = \mathbf{V}\mathbf{D}_2\mathbf{X}^T,$$

where $\mathbf{X} = (\mathbf{Q}\mathbf{R})^T$ can be non-square and is of full (column) rank. In full detail, if $m - q \geq 0$ then

$$
\underbrace{m\ \begin{pmatrix} \overset{n}{\mathbf{A}} \end{pmatrix}}_{\mathbf{A}} = \underbrace{m\ \begin{pmatrix} \overset{q}{\mathbf{U}_1} & \overset{m-q}{\mathbf{U}_2} \end{pmatrix}}_{\mathbf{U}} \circ \underbrace{\begin{matrix} q \\ m-q \end{matrix}\begin{pmatrix} \overset{q}{\hat{\mathbf{C}}} \\ \mathbf{0} \end{pmatrix}}_{\mathbf{D}_1} \circ \underbrace{q\ \begin{pmatrix} \overset{n-q}{\mathbf{0}} & \overset{q}{\mathbf{R}} \end{pmatrix}}_{\hat{\mathbf{R}}} \circ \underbrace{\begin{matrix} n-q \\ q \end{matrix}\begin{pmatrix} \overset{n}{\mathbf{Q}_1^T} \\ \mathbf{Q}_2^T \end{pmatrix}}_{\mathbf{Q}^T}
$$

$$
\underbrace{p\ \begin{pmatrix} \overset{n}{\mathbf{B}} \end{pmatrix}}_{\mathbf{B}} = \underbrace{p\ \begin{pmatrix} \overset{l}{\mathbf{V}_1} & \overset{p-l}{\mathbf{V}_2} \end{pmatrix}}_{\mathbf{V}} \circ \underbrace{\begin{matrix} l \\ p-l \end{matrix}\begin{pmatrix} \overset{k}{\mathbf{0}} & \overset{l}{\hat{\mathbf{S}}} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}}_{\mathbf{D}_2} \circ \underbrace{\begin{matrix} k \\ l \end{matrix}\begin{pmatrix} \overset{n-q}{\mathbf{0}} & \overset{q}{[\mathbf{R}_{11}\ \ \mathbf{R}_{12}]} \\ \mathbf{0} & [\mathbf{0}\ \ \mathbf{R}_{22}] \end{pmatrix}}_{\hat{\mathbf{R}}} \circ \underbrace{\begin{matrix} n-q \\ q \end{matrix}\begin{pmatrix} \overset{n}{\mathbf{Q}_1^T} \\ \mathbf{Q}_2^T \end{pmatrix}}_{\mathbf{Q}^T}
$$

where the upper-triangular $\mathbf{R}$ is written as a 2×2 block matrix and

$$\hat{\mathbf{C}} = \begin{matrix} k \\ l \end{matrix}\begin{pmatrix} \overset{k}{\mathbf{I}} & \overset{l}{\mathbf{0}} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}, \quad \hat{\mathbf{S}} = \mathbf{S},$$

with

$$\mathbf{C} = \mathrm{diag}(c_1, \ldots, c_l), 1 > c_1 \geq c_2 \geq \ldots \geq c_l, \quad \mathbf{S} = \mathrm{diag}(s_1, \ldots, s_l), 0 < s_1 \leq s_2 \leq \ldots \leq s_l,$$

and

$$\mathbf{C}^2 + \mathbf{S}^2 = \mathbf{I}.$$

If $m - q < 0$ then

$$
\underbrace{m\ \begin{pmatrix} \overset{n}{\mathbf{A}} \end{pmatrix}}_{\mathbf{A}} = \underbrace{\mathbf{U}}_{\substack{\mathbf{A} \\ \mathbb{R}^{m \times m}}} \circ \underbrace{m\ \begin{pmatrix} \overset{m}{\hat{\mathbf{C}}} & \overset{q-m}{\mathbf{0}} \end{pmatrix}}_{\mathbf{D}_1} \circ \underbrace{\begin{matrix} m \\ q-m \end{matrix}\begin{pmatrix} \overset{n-q}{\mathbf{0}} & \overset{q}{\begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ \mathbf{0} & \mathbf{R}_{22} & \mathbf{R}_{23} \end{bmatrix}} \\ \mathbf{0} & [\mathbf{0}\ \mathbf{0}\ \mathbf{R}_{33}] \end{pmatrix}}_{\hat{\mathbf{R}}} \circ \underbrace{\begin{matrix} n-q \\ q \end{matrix}\begin{pmatrix} \overset{n}{\mathbf{Q}_1^T} \\ \mathbf{Q}_2^T \end{pmatrix}}_{\mathbf{Q}^T}
$$

$$
\underbrace{p\ \begin{pmatrix} \overset{n}{\mathbf{B}} \end{pmatrix}}_{\mathbf{B}} = \underbrace{p\ \begin{pmatrix} \overset{l}{\mathbf{V}_1} & \overset{p-l}{\mathbf{V}_2} \end{pmatrix}}_{\mathbf{V}} \circ \underbrace{\begin{matrix} l \\ p-l \end{matrix}\begin{pmatrix} \overset{k}{\mathbf{0}} & \overset{l}{\hat{\mathbf{S}}} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}}_{\mathbf{D}_2} \circ \underbrace{\begin{matrix} k \\ l \end{matrix}\begin{pmatrix} \overset{n-q}{\mathbf{0}} & \overset{q}{\begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \end{bmatrix}} \\ \mathbf{0} & \begin{bmatrix} \mathbf{0} & \mathbf{R}_{22} & \mathbf{R}_{23} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{33} \end{bmatrix} \end{pmatrix}}_{\hat{\mathbf{R}}} \circ \underbrace{\begin{matrix} n-q \\ q \end{matrix}\begin{pmatrix} \overset{n}{\mathbf{Q}_1^T} \\ \mathbf{Q}_2^T \end{pmatrix}}_{\mathbf{Q}^T}
$$

where the upper-triangular $\mathbf{R}$ is written as a 3×3 block matrix and

$$
\hat{\mathbf{C}} = \begin{matrix} k \\ m-k \end{matrix} \begin{pmatrix} \overset{k}{\mathbf{I}} & \overset{m-k}{\mathbf{0}} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}, \qquad \hat{\mathbf{S}} = \begin{matrix} m-k \\ q-m \end{matrix} \begin{pmatrix} \overset{m-k}{\mathbf{S}} & \overset{q-m}{\mathbf{0}} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}
$$

with

$$
\mathbf{C} = \mathrm{diag}(c_1, \ldots, c_{m-k}), 1 > c_1 \geq c_2 \geq \ldots \geq c_{m-k},
$$
$$
\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_{m-k}), 0 < s_1 \leq s_2 \leq \ldots \leq s_{m-k},
$$

and

$$
\mathbf{C}^2 + \mathbf{S}^2 = \mathbf{I}.
$$

To obtain the `MATLAB`-style output where the entries in $\mathbf{C}$ are ascending and the entries in $\mathbf{S}$ are descending, the order of the columns and rows in $\mathbf{S}$ (and likewise in $\mathbf{C}$) needs to be reversed: Let $k > 0$ and

$$
\mathbf{R}_k = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{k \times k}
$$

and observe that $\mathbf{R}_k^{-1} = \mathbf{R}_k^T = \mathbf{R}_k$. Hence, if $m - q \geq 0$ then

$$
\mathbf{A} = \left(\mathbf{U} \begin{bmatrix} \mathbf{R}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-q} \end{bmatrix} \right) \circ \left(\begin{bmatrix} \mathbf{R}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-q} \end{bmatrix} \mathbf{D}_1 \mathbf{R}_q \right) \circ \left(\mathbf{R}_q \hat{\mathbf{R}} \mathbf{Q}^T \right)
$$
$$
= \left(\mathbf{U} \begin{bmatrix} \mathbf{R}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-q} \end{bmatrix} \right) \circ \begin{bmatrix} \mathbf{R}_q \hat{\mathbf{C}} \mathbf{R}_q \\ \mathbf{0} \end{bmatrix} \circ \left(\mathbf{R}_q \hat{\mathbf{R}} \mathbf{Q}^T \right)
$$
$$
\mathbf{B} = \left(\mathbf{V} \begin{bmatrix} \mathbf{R}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-l} \end{bmatrix} \right) \begin{bmatrix} \mathbf{R}_l \hat{\mathbf{S}} \mathbf{R}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \left(\mathbf{R}_q \hat{\mathbf{R}} \mathbf{Q}^T \right).
$$

If $m - q < 0$

$$
\mathbf{A} = (\mathbf{U} \mathbf{R}_m) \circ \begin{bmatrix} \mathbf{0} & \mathbf{R}_m \hat{\mathbf{C}} \mathbf{R}_m \end{bmatrix} \circ \left(\mathbf{R}_q \hat{\mathbf{R}} \mathbf{Q}^T \right)
$$
$$
\mathbf{B} = \left(\mathbf{V} \begin{bmatrix} \mathbf{R}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-l} \end{bmatrix} \right) \begin{bmatrix} \mathbf{R}_l \hat{\mathbf{S}} \mathbf{R}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \left(\mathbf{R}_q \hat{\mathbf{R}} \mathbf{Q}^T \right).
$$

can be worked out similarly. In this case $[\mathbf{0}, \mathbf{R}_m \hat{\mathbf{C}} \mathbf{R}_m]$ contains only non-zero entries at its diagonal $q - m$ and $\mathbf{R}_l \hat{\mathbf{S}} \mathbf{R}_l$ only on its main diagonal. $\qquad \square$

Theorem 7.2.5. *(An explicitly rank-revealing GSVD variant based on Theorem 7.2.4, denoted rGSVD)*

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{p \times n}$ and let $0 < l := \mathrm{rank}(B)$, $q := \mathrm{rank}([A; B])$, $k = q - l$. Then there exist orthogonal matrices $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{Q} \in \mathbb{R}^{n \times n}$, matrix $\mathbf{V} \in \mathbb{R}^{p \times l}$ with orthogonal columns, non-singular upper-triangular matrix $\mathbf{R} \in \mathbb{R}^{n \times q}$, diagonal matrix $\mathbf{S} = \mathrm{diag}(s_1, s_2, \ldots, s_l) \in \mathbb{R}^{l \times l}$ and matrix with one non-zero diagonal $\mathbf{D}_1 = \mathrm{diag}(c_1, c_2, \ldots) \in \mathbb{R}^{m \times q}$ such that

$$
\mathbf{A} = \mathbf{U} \mathbf{D}_1 (\mathbf{Q} \mathbf{R})^T,
$$
$$
\mathbf{B} = \mathbf{V} \mathbf{S} \left(\hat{\mathbf{Q}} \hat{\mathbf{R}} \right)^T \tag{7.2.6}
$$

where $\mathbf{S}$ *does not contain zeros,* $\hat{\mathbf{Q}}$ *consists of columns of* $\mathbf{Q}$ *and* $\hat{\mathbf{R}}$ *is a diagonal block of* $\mathbf{R}$ *which is non-singular.*

Proof. Note that due to Theorem 7.2.4

$$\mathbf{R}_q \hat{\mathbf{R}} \mathbf{Q}^T = \begin{bmatrix} \mathbf{0} & [\mathbf{0} \quad \mathbf{R}_l \mathbf{R}_{22}] \\ \mathbf{0} & [\mathbf{R}_k \mathbf{R}_{11} \quad \mathbf{R}_k \mathbf{R}_{12}] \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} = \begin{bmatrix} [\mathbf{0} \quad \mathbf{R}_l \mathbf{R}_{22}] \mathbf{Q}_2^T \\ [\mathbf{R}_k \mathbf{R}_{11} \quad \mathbf{R}_k \mathbf{R}_{12}] \mathbf{Q}_2^T \end{bmatrix}.$$

This shows that there exists a *explicitly rank-revealing GSVD (rGSVD)* $\mathbf{B} = \mathbf{X} \tilde{\mathbf{S}} \mathbf{Y}^T$

$$
\begin{aligned}
\mathbf{B} &= \left(\mathbf{V} \begin{bmatrix} \mathbf{R}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-l} \end{bmatrix} \right) \begin{bmatrix} \mathbf{R}_l \hat{\mathbf{S}} \mathbf{R}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} [\mathbf{0} \quad \mathbf{R}_l \mathbf{R}_{22}] \mathbf{Q}_2^T \\ [\mathbf{R}_k \mathbf{R}_{11} \quad \mathbf{R}_k \mathbf{R}_{12}] \mathbf{Q}_2^T \end{bmatrix} \\
&= \begin{bmatrix} (\mathbf{V} \mathbf{R}_l)(\mathbf{R}_l \hat{\mathbf{S}} \mathbf{R}_l) & \mathbf{0} \end{bmatrix} \begin{bmatrix} [\mathbf{0} \quad \mathbf{R}_l \mathbf{R}_{22}] \mathbf{Q}_2^T \\ [\mathbf{R}_k \mathbf{R}_{11} \quad \mathbf{R}_k \mathbf{R}_{12}] \mathbf{Q}_2^T \end{bmatrix} \\
&= (\mathbf{V} \mathbf{R}_l)(\mathbf{R}_l \hat{\mathbf{S}} \mathbf{R}_l)([\mathbf{0} \mathbf{R}_l \mathbf{R}_{22}] \begin{bmatrix} \mathbf{Q}_{21}^T \\ \mathbf{Q}_{22}^T \end{bmatrix}) \\
&= (\mathbf{V} \mathbf{R}_l)(\mathbf{R}_l \hat{\mathbf{S}} \mathbf{R}_l)(\mathbf{R}_l \mathbf{R}_{22} \mathbf{Q}_{22}^T) \\
&= \mathbf{X} \tilde{\mathbf{S}} \mathbf{Y}^T
\end{aligned}
$$

with $\mathbf{X}, \mathbf{Y}$ of full column rank and $\tilde{\mathbf{S}}$ a diagonal matrix with positive diagonal entries. Note that by construction $\mathbf{Q}_2^T = [\mathbf{Q}_{21}^T; \mathbf{Q}_{22}^T]$, the columns of $\mathbf{Q}_{22}$ are columns of the orthogonal matrix $\mathbf{Q}$, i.e., $\mathbf{Q}_{22}$ is orthogonal. The columns of $\mathbf{X}$ and $\mathbf{Y}$ we call the (GSVD-based) *principal components* or *dominant components*. Note that both $\mathbf{R}_q \hat{\mathbf{C}} \mathbf{R}_q$ and $\mathbf{R}_l \hat{\mathbf{S}} \mathbf{R}_l$ only contain entries on their main diagonal. Also, rank($\mathbf{X}$) = rank($\mathbf{B}$) and obviously, multiplications with $\mathbf{R}_l$ are only to ensure that the entries of $\mathbf{S}$ are in descending order. □

Now we show that if $\mathbf{B}$ has all its non-zero entries in a sub-block then also $\mathbf{V}$ has:

Corollary 7.2.1. (*The rGSVD preserves the block structure* (zero blocks) of matrix $\mathbf{B}$) Assume that $\mathbf{B} \in \mathbb{R}^{n \times m}$ is a sparse matrix which only contains non-zero entries in a *window* $(i_1, \ldots, i_2) \times (j_1, \ldots, j_2)$ (i.e., in a sub-block of size $n_1 \times m_1$). Then there is an explicitly rank-revealing factor $\mathbf{V}$ of $\mathbf{B} = \mathbf{V} \mathbf{S} \mathbf{X}^T$ which also only contains non-zero entries in window of size $n_1 \times m_1$.

Proof. Since $\mathbf{B} = \mathbf{V} \mathbf{S} \mathbf{X}^T$ there is a relation between the column spans of $\mathbf{B}$ and $\mathbf{V}$: The latter column span must contain the former. By construction of the explicitly rank-revealing variant $\mathbf{V}$ at the end of the proof of Theorem 7.2.4 its rank is the rank of $\mathbf{B}$. This implies that also $\mathbf{V}$ can only have non-zeros in a similarly sized window, i.e., it has a similar (non-zero) block structure as $\mathbf{B}$. □

Corollary 7.2.2. (The relation between the GSVD of $\mathbf{A}$, $\mathbf{B}$ and *GSVD of* $\mathbf{AM}$, $\mathbf{BM}$)
If $\mathbf{A} = \mathbf{USX}^T$, $\mathbf{B} = \mathbf{VCX}^T$ is the GSVD of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{M}$ is a (compatible) non-singular
matrix then $\mathbf{AM} = \mathbf{USX}^T\mathbf{M}$, $\mathbf{B} = \mathbf{VCX}^T\mathbf{M}$ is the GSVD of $\mathbf{AM}$, $\mathbf{BM}$.

Proof. The result is straightforward since the proposed GSVD fits all requirements of a
GSVD. □

Corollary 7.2.3. (*GSVD(*$\mathbf{A}$, $\mathbf{B}$*) applied to a non-singular matrix* $\mathbf{A}$ *results in diagonal* $\mathbf{S}$
and diagonal $\mathbf{C}$)
In the case of a block matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \in \mathbb{R}^{(m+p)\times(m+p)}$$

where $\mathbf{A}_{11}$ is of full column rank the GSVD leads to a diagonal matrix $\mathbf{C}$ *and* $\mathbf{S}$.

Proof. Follows from $q = \mathrm{rank}([\mathbf{A}_{11}; \mathbf{A}_{21}]) = m$. □

7.3   The GSVD of the coupling blocks

In this section we propose the application of the GSVD to the sub-blocks of the system
matrices that describe the coupling between two sub-systems or physical sub-domains.
Based on this we are able to determine how much of the coupling information contained
in the system description is relevant and necessary to obtain an accurate solution of the
approximated coupled system. We propose the GSVD instead of the SVD since with the
GSVD we obtain dominant modes of the couplings relative to the individual systems.

First we describe how to apply the GSVD from Section 7.2 to the coupled system (7.1.1)

$$\begin{bmatrix} \mathbf{A}_{11} & \tilde{\mathbf{A}}_{12} \\ \tilde{\mathbf{A}}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}.$$

This system consists of two sets of equations

$$\mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{x}_2 = \mathbf{b}_1 \qquad\qquad (7.3.1)$$
$$\mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}\mathbf{x}_2 = \mathbf{b}_2. \qquad\qquad (7.3.2)$$

Equation (7.3.1) describes the behavior of the variable $\mathbf{x}_1$ (related to one of the sub-
systems, sub-structures or physical sub-domains) and the way it is coupled to the variable
$\mathbf{x}_2$. Similarly (7.3.2) describes the behavior of the variable $\mathbf{x}_2$ and the way it depends on
variable $\mathbf{x}_1$. As stated in the GSVD Theorem 7.2.3 one of the necessary conditions for
applying the GSVD to a pair of the matrices ($\mathbf{A}$, $\mathbf{B}$) is that the matrices $\mathbf{A}$ and $\mathbf{B}$ have
the same number of columns. In case of the system (7.1.1), the obvious pairs would be

$(\mathbf{A}_{11}, \mathbf{A}_{21})$ and $(\mathbf{A}_{22}, \mathbf{A}_{12})$. However, this is not the most optimal choice from a practical point of view. In industrial applications, it is often the case, that the software suited to generate or reduce the coupled system (7.1.1) is not available. However, there may exist software able to deal with the sub-problem (7.3.1) and software that can deal with the sub-problem (7.3.2). Figure 7.1 shows the schematic representation of the coupled
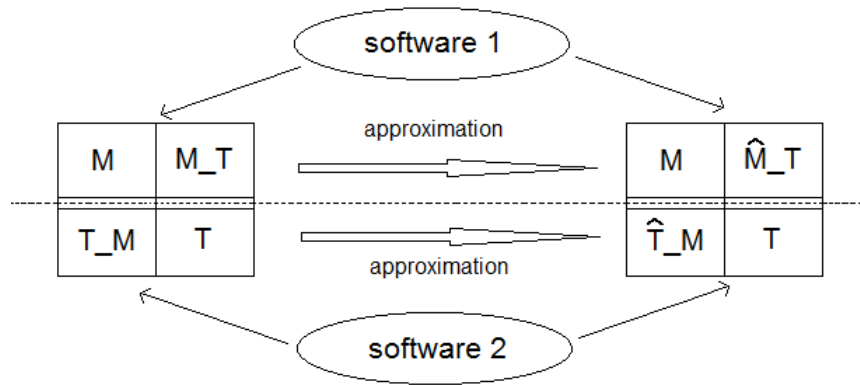
Figure 7.1: Schematic representation of the potential solution (reduction) strategy for the coupled system consisting of thermal and mechanical sub-systems

thermo-mechanical system, where the system matrices of the mechanical sub-problem M and thermal sub-problem T are generated by two different types of software. Frequently, the matrix M_T that determines how the mechanical sub-problem M is influenced by the thermal sub-problem T, is delivered by the same tool that generates the mechanical system matrix M, which allows for further analysis (solution, reduction) of the system consisting of both of the matrices. Similarly for the thermal sub-problem T. Hence, the preferred way of choosing the pairs of matrices for the GSVD is the equation-wise one. Unfortunately, in general the matrix pairs $(\mathbf{A}_{11}, \mathbf{A}_{12})$ and $(\mathbf{A}_{22}, \mathbf{A}_{21})$ involve matrices with different number of columns whence GSVD can not be applied to them. On the other hand, since the matrix pairs are chosen row-wise, they have the same number of rows and creating pairs

$$(\mathbf{A}_{11}^T, \mathbf{A}_{12}^T) \qquad \text{and} \qquad (\mathbf{A}_{22}^T, \mathbf{A}_{21}^T) \tag{7.3.3}$$

assures that the components have the same number of columns.

Observe that the GSVD in Theorem 7.2.3 ensures a factorization $\mathbf{X}\mathbf{S}^T\mathbf{V}^T$ where $\mathbf{S}$ is a diagonal matrix only for matrix $\mathbf{B}$, and not for $\mathbf{A}$. Because later on we assume $\mathbf{S}$ to be diagonal, we apply GSVD to the ordered pair $(\mathbf{A}_{11}^T, \mathbf{A}_{12}^T)$ and not to $(\mathbf{A}_{12}^T, \mathbf{A}_{11}^T)$.

Some systems are of the form

$$\left(s \begin{bmatrix} \mathbf{E}_{11} & \\ & \mathbf{E}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \right) \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}, \tag{7.3.4}$$

rather than (7.1.1). In such cases one needs to consider not the GSVD of blocks of $\mathbf{A}$ but for instance the GSVD related to blocks of $\mathbf{E}^{-1}\mathbf{A}$. The latter matrix has block structure

$$\begin{bmatrix} \mathbf{E}_{11}^{-1}\mathbf{A}_{11} & \mathbf{E}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{E}_{22}^{-1}\mathbf{A}_{21} & \mathbf{E}_{22}^{-1}\mathbf{A}_{22} \end{bmatrix}.$$

It turns out that the GSVD of $(\mathbf{E}_{11}^{-1}\mathbf{A}_{11})^T$ and $(\mathbf{E}_{11}^{-1}\mathbf{A}_{12})^T$ is straightforwardly related to the one of GSVD $(\mathbf{A}_{11}^T$ and $\mathbf{A}_{12}^T)$:

$$\begin{matrix} \mathbf{A}_{11}^T = \mathbf{U}\mathbf{S}\mathbf{X}^T \\ \mathbf{A}_{12}^T = \mathbf{V}\mathbf{C}\mathbf{X}^T \end{matrix} \implies \begin{matrix} (\mathbf{E}_{11}^{-1}\mathbf{A}_{11})^T = \mathbf{A}_{11}^T\mathbf{E}_{11}^{-T} = \mathbf{U}\mathbf{S}\mathbf{X}^T\mathbf{E}_{11}^{-T}, \\ (\mathbf{E}_{11}^{-1}\mathbf{A}_{12})^T = \mathbf{A}_{12}^T\mathbf{E}_{11}^{-T} = \mathbf{V}\mathbf{C}\mathbf{X}^T\mathbf{E}_{11}^{-T} \end{matrix}$$

based on Corollary 7.2.2.

In the remainder of this thesis, we will always apply the GSVD to the pairs as defined in (7.3.3). In the next section, we show how the components obtained after application of the GSVD to the pairs given by (7.3.3) can be used to find low-rank approximations of the coupling blocks.

7.4   Low-rank approximations of the couplings

In the previous section we suggested and motivated a manner of the selection of the pairs of the system sub-matrices for the GSVD application. Here we will explain how to obtain low-rank approximations of the coupling-blocks of the system matrix.

Definition 7.4.1. *Let k be an integer and define $1 : k$ by $1 : k = [1, 2, \ldots, k]$ be a positive integer vector. Assume that index vectors $\mathbf{I} \in \mathbb{N}^m$ and $\mathbf{J} \in \mathbb{N}^m$ only contain positive integer entries. Let $\mathbf{A}$ be a matrix with columns $\mathbf{a}_i$ and define $\mathbf{A}(:, \mathbf{J})$ to be the column matrix $[\mathbf{a}_{j_1}, \mathbf{a}_{j_2}, \ldots, \mathbf{a}_{j_m}]$. Similarly let $\mathbf{A}$ be a matrix with rows $\mathbf{a}_i$ and define $\mathbf{A}(\mathbf{I}, :)$ to be the row matrix $[\mathbf{a}_{i_1}; \mathbf{a}_{i_2}; \ldots; \mathbf{a}_{i_m}]$. Finally, let $\mathbf{A}(\mathbf{I}, \mathbf{J})$ be $\mathbf{B}(\mathbf{I}, :)$ where $\mathbf{B} = \mathbf{A}(:, \mathbf{J})$. For the remainder of this thesis define $\mathbf{X}^{(k)}$, $\mathbf{S}^{(k)}$ and $\mathbf{V}^{(k)}$ by $\mathbf{X}(:, 1 : k)$, $(\mathbf{S}(1 : k, 1 : k))^T$ and $(\mathbf{V}(:, 1 : k))^T$, and similarly $\mathbf{X}_i^{(k)}$, $\mathbf{S}_i^{(k)}$ and $\mathbf{V}_i^{(k)}$, $i = 1, 2$.*

Following the notation of GSVD Theorem 7.2.3, decomposing the pairs given by (7.3.3), one obtains

$$\begin{cases} \mathbf{A}_{11}^T = \mathbf{U}_1\mathbf{C}_1\mathbf{X}_1^T, \\ \mathbf{A}_{12}^T = \mathbf{V}_1\mathbf{S}_1\mathbf{X}_1^T \end{cases}, \qquad \begin{cases} \mathbf{A}_{22}^T = \mathbf{U}_2\mathbf{C}_2\mathbf{X}_2^T, \\ \mathbf{A}_{21}^T = \mathbf{V}_2\mathbf{S}_2\mathbf{X}_2^T, \end{cases}$$

which results in coupling blocks factorizations

$$\mathbf{A}_{12} = \mathbf{X}_1 \mathbf{S}_1^T \mathbf{V}_1^T \tag{7.4.1}$$

$$\mathbf{A}_{21} = \mathbf{X}_2 \mathbf{S}_2^T \mathbf{V}_2^T. \tag{7.4.2}$$

Assuming that we applied an explicitly rank-revealing version of the GSVD algorithm 7.2.3, based on (7.4.1) and (7.4.2) we define coupling block $\mathbf{A}_{12}$ *low-rank approximation* $\tilde{\mathbf{A}}_{12}$ of rank k_1 and coupling block $\mathbf{A}_{21}$ *low-rank approximation* $\tilde{\mathbf{A}}_{21}$ of rank k_2 by

$$\tilde{\mathbf{A}}_{12} = \mathbf{X}_1^{(k_1)} \mathbf{S}_1^{(k_1)} \mathbf{V}_1^{(k_1)} \tag{7.4.3}$$

$$\tilde{\mathbf{A}}_{21} = \mathbf{X}_2^{(k_2)} \mathbf{S}_2^{(k_2)} \mathbf{V}_2^{(k_2)}. \tag{7.4.4}$$

Finally we use $\tilde{\mathbf{A}}_{12}$ and $\tilde{\mathbf{A}}_{21}$ for the approximation (7.1.2) of the coupled system (7.1.1). In the next section we will show the results of applying this approximation strategy to a model of a coupled thermo-mechanical system.

7.5   Numerical example

The test example consists of a linear system of equations which describes a stationary coupled *thermo-mechanical problem*: A(n offline calculated) current heats up a three-dimensional two-layered bar which deforms due to its layers' different expansion coefficients. The beam's movement is constrained – one of its ends is fixed. The imposed boundary conditions are:

1. Temperature: prescribed at the fixed end – adiabatic heating on surfaces;

2. Displacements: zero (x, y, z) displacements at the bar's fixed end.

The dimensions of the beam are $0.1\text{m} \times 1\text{m} \times 0.02\text{m}$.

The linear system

The temperature and displacements are discretized with finite elements on a hexahedral mesh. After elimination of the degrees of freedom on the Dirichlet boundary (prescribed temperature and zero displacements) one obtains a block-matrix $A \in \mathbb{R}^{1200 \times 1200}$. The temperature vector $x_1 \in \mathbb{R}^{300}$ and nodal displacements vector $x_2 \in \mathbb{R}^{900}$ solve

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \tag{7.5.1}$$
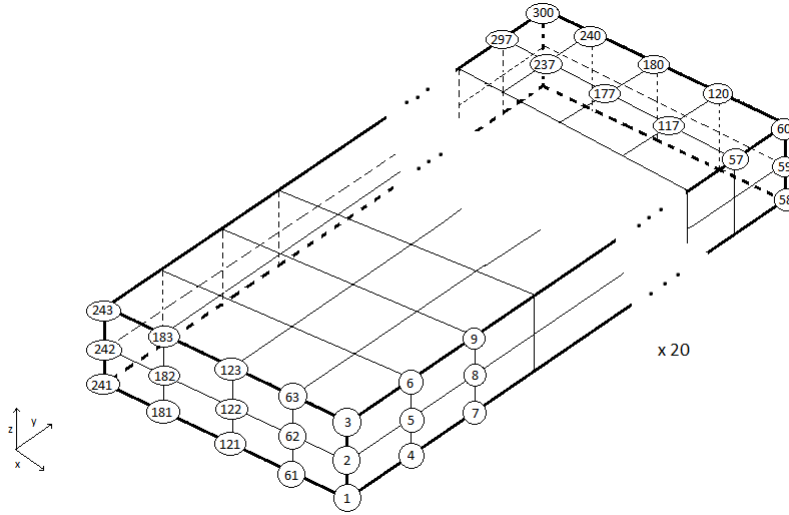
Figure 7.2: Numbering of the nodes of the discretized beam model

where $\mathbf{A}_{11} \in \mathbb{R}^{300 \times 300}$, $\mathbf{A}_{21} \in \mathbb{R}^{900 \times 300}$, and $\mathbf{A}_{11} \in \mathbb{R}^{900 \times 900}$. Let us define

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & 0 \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}. \qquad (7.5.2)$$

The numbering of the nodes is presented in Figure 7.2. Figure 7.3 and Figure 7.4 show the solution's temperature and displacement parts, respectively. Observe that the displacement is more dominant in the z than in the x, y directions.

The unscaled linear system

Because we intend to construct an approximation based on available "industrial" software, we work with the system of equations as it is given. It turns out that the matrix $\mathbf{A}$ is badly scaled, most likely due to not scaling/non-dimensionalizing the physics equations by the software. The spectral condition number $\kappa_2(\mathbf{A})$ is $\kappa_2(\mathbf{A}) \approx 3.1566 \times 10^{14}$ which implies that for double precision IEEE arithmetic $\mathbf{A}$ is *numerically singular* – the matlab `rank(A)` command returns $1195 \neq 1200$. The same command shows that $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are of full column rank.

Scaled version of the linear system

Next, we scale the system of equations based on the values of the diagonal entries of $\mathbf{A}$ such that the scaled system has entries 1 at the diagonal, i.e., we apply a *diagonal*
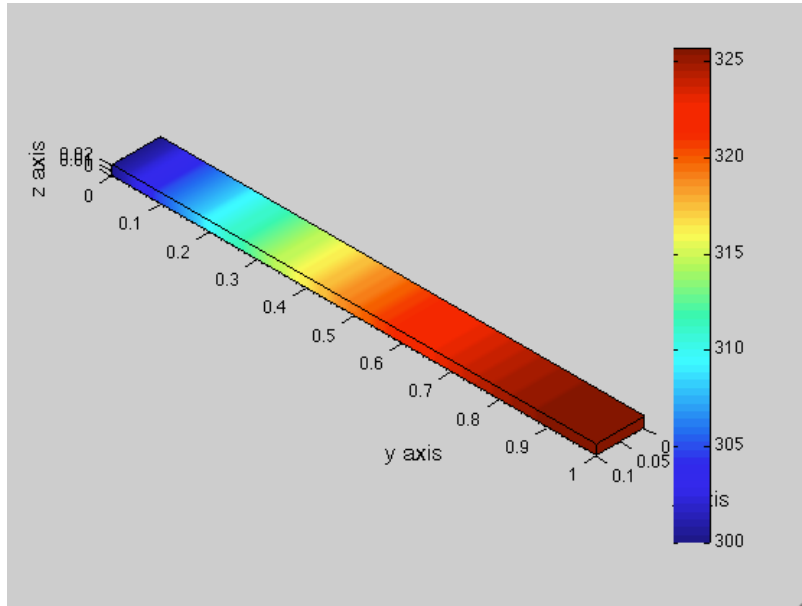
Figure 7.3: Temperatures (in kelvin) along the beam (courtesy of Fabio Freschi, Torino University)

scaling. Both row and column scaling is performed (matrix $\mathbf{A}$ is multiplied from the left and right side) which implies that we both scale the equations *and* "non-dimensionalize" the degrees of freedom: Using

$$\mathbf{D} = \text{diag}(\sqrt{a_{ii}^{-1}}) \tag{7.5.3}$$

we solve

$$\mathbf{D}\begin{bmatrix} \mathbf{A}_{11} & 0 \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\mathbf{D}\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} = \mathbf{D}\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}, \tag{7.5.4}$$

where $\mathbf{v} = \mathbf{D}^{-1}\mathbf{x}$. Let $\mathbf{E} = \mathbf{DAD}$. Then by construction

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_{11} & 0 \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{bmatrix} \tag{7.5.5}$$

and has diagonal entries equal to 1. After the scaling, numerical test shows that $\mathbf{E}$ has full column rank and that $\kappa_2(\mathbf{E}) \approx 1.4456 \times 10^8$. Figure 7.5 and Figure 7.6 show plots of the solution vectors v_1 and v_2, respectively. If the matrix had been even worse conditioned, one could have scaled the x, y and z components of the solution independently, since the displacement in z direction is several orders of magnitudes larger than those in x and y direction.
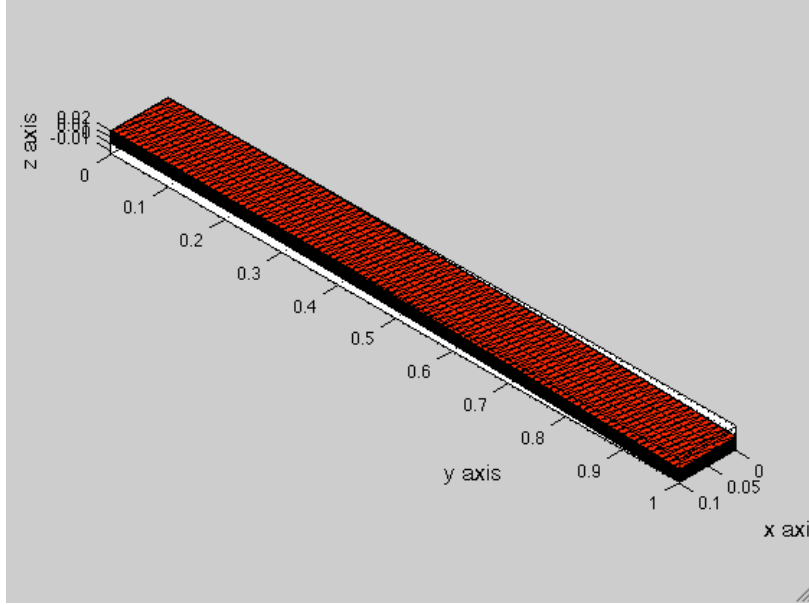
Figure 7.4: Displacements (in meters) along the beam (courtesy of Fabio Freschi, Torino University)

Low-rank approximation of the scaled system

Finally, the original system was approximated by a system obtained by replacing the exact coupling block by a GSVD-based one. In this case, the thermo-mechanical coupling is defined by the $\mathbf{E}_{21}$ block of the scaled matrix. To compute the GSVD, this block was used together with the $\mathbf{E}_{22}$ block. To assure correct dimensions required in the GSVD algorithm, transposed versions of these matrices were employed. The decomposed system has the following form

$$\begin{cases} \mathbf{E}_{22}^T = \mathbf{U}\mathbf{C}\mathbf{X}^T, \\ \mathbf{E}_{21}^T = \mathbf{V}\mathbf{S}\mathbf{X}^T. \end{cases} \tag{7.5.6}$$

A plot of the generalized singular values of the matrix $\mathbf{E}_{21}$, diagonal values of $\mathbf{S}$, is shown in figure 7.7. The approximation of the coupling sub-block $\tilde{\mathbf{E}}_{21} \approx \mathbf{E}_{21}$ is obtained by taking into account only a number of dominant components computed by the GSVD

$$\tilde{\mathbf{E}}_{21} = \mathbf{X}^{(k)}\mathbf{S}^{(k)}\mathbf{V}^{(k)}, \tag{7.5.7}$$

where k is the rank of the approximation. Next the approximated system

$$\begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \tilde{\mathbf{E}}_{21} & \mathbf{E}_{22} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{v}}_1 \\ \tilde{\mathbf{v}}_2 \end{bmatrix} = \mathbf{D} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \tag{7.5.8}$$
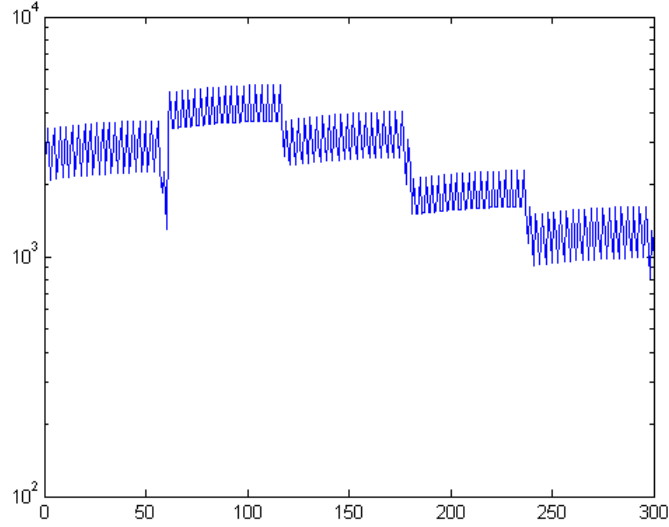
Figure 7.5: The scaled system's solution temperatures (in scaled kelvin) vector $\mathbf{v}_1$ with respect to the temperature degrees of freedom

is solved. In case of the thermal part of the problem, the approximated solution $\tilde{\mathbf{v}}_1$ is equal to the exact one $\mathbf{v}_1$. This is due to the fact, that in the considered problem the coupling is uni-directional (the temperature affects the shape of the beam but not the other way round) and the values of the thermal variables are not influenced by the values of the displacements. Let us examine the accuracy of the displacement variable $\tilde{\mathbf{v}}_2$. Since $\tilde{\mathbf{v}}_2$ describes the displacement of the nodes in three different spatial directions and the magnitudes of the displacements are different for each of the directions, the good way of judging the final result seems to be by comparing the approximated total displacement vector of the node with the exact one. Hence, we define the exact and approximated displacement vectors for a node i, $i = 1, ..., 300$, $\mathbf{r}$ and $\tilde{\mathbf{r}}$ respectively, as follows

$$\mathbf{r}_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \qquad \text{and} \qquad \tilde{\mathbf{r}}_i = \sqrt{\tilde{x}_i^2 + \tilde{y}_i^2 + \tilde{z}_i^2}, \qquad (7.5.9)$$

where x_i, y_i, z_i, $\tilde{x}_i$, $\tilde{y}_i$, $\tilde{z}_i$ are the entries of the vectors $\mathbf{v}_2$ and $\tilde{\mathbf{v}}_2$ corresponding to the displacements of the node i in the x, y, and z direction. The relative approximation errors are estimated for each node according to the formula

$$\epsilon_i = \frac{|\mathbf{r}_i - \tilde{\mathbf{r}}_i|}{|\mathbf{r}_i|}. \qquad (7.5.10)$$

Table 7.1 shows ϵ_i for the solution of the approximated system, where the coupling block $\mathbf{E}_{21}$ was approximated using $k = 15$ dominant components. Each column of the table
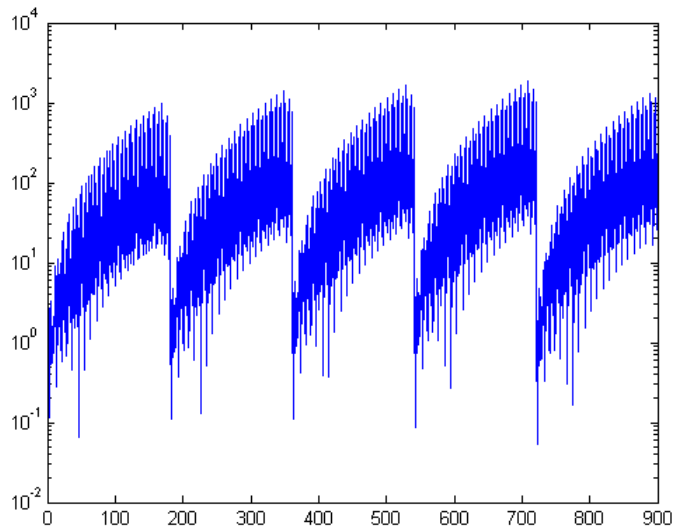
Figure 7.6: The scaled system's solution nodal displacements (in scaled meters) vector $\mathbf{v}_2$ with respect to the displacement degrees of freedom
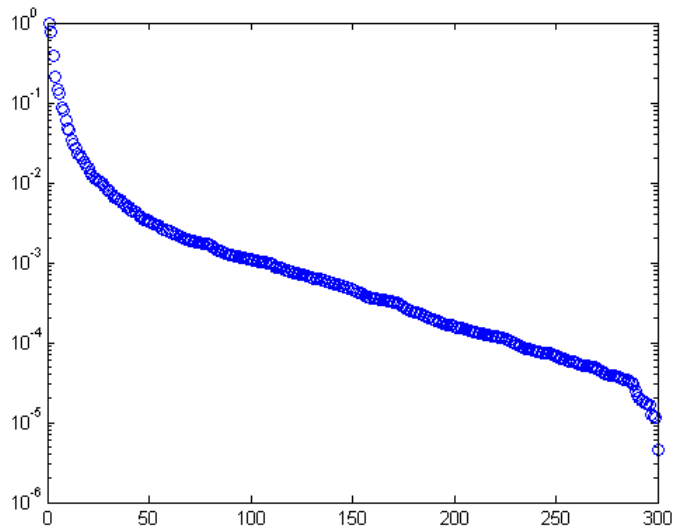
Figure 7.7: Generalized singular values $\mathbf{C}$ of the matrix $\mathbf{E}_{21}{}^T$

shows the error for one of the 15 discretization lines along the beam. Every line consists of 20 nodes, numbered from 1 at the top (node at the fixed end of the beam) to 20 at the bottom (node at the free end of the beam). Figure 7.8 illustrates how the lines are defined. As shown in the table 7.1, the highest relative error for the displacements is
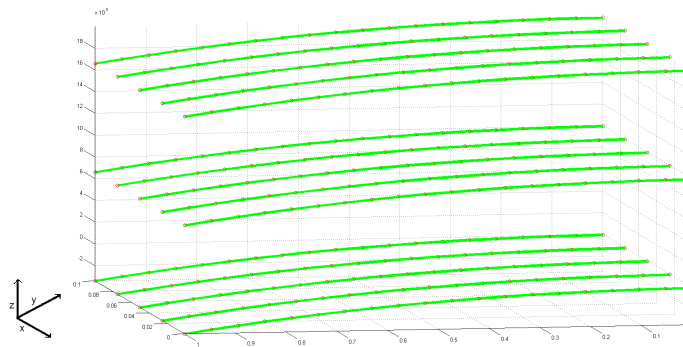
Figure 7.8: Discretization lines along the beam

computed for the nodes that are close to the fixed end of the beam. This can be explained by the fact, that the displacements of the nodes, that are located close to the fixed end, are small (very close to zero). On the other hand, looking at the displacements close to the free end of the beam, one can see, that already a rank 15 approximation results in a relative error around 2%. This makes the approximation procedure valuable, since the evaluation of the unknown displacements for nodes at the free side of the beam is of interest. The displacements of nodes at the fixed end of the beam are pre-described and do not need to be computed.

7.6   Conclusions

In this chapter, we focused on the coupling blocks of a linear system of equations. We introduced the singular value decomposition (SVD), the generalized singular value decomposition (GSVD), and their explicitly rank-revealing variants as tools to evaluate the importance of the couplings. We showed that the explicitly rank-revealing GSVD of a pair of matrices (**A**, **B**) preserves a block-structure of the matrix **B**. Finally, we presented numerical results of application of the GSVD to the coupling block of a thermo-mechanical problem.

In the next chapter, we propose a way of combining the GSVD idea with the SBR algorithm, to make the latter one more efficient for strongly coupled systems.

Table 7.1: Relative error ϵ_i (%) for all nodes of the beam, $k = 15$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	41,44	49,71	97,83	61,02	25,24	58,71	76,69	35,47	37,22	135,87	67,15	27,70	47,55	45,97	23,18
2	25,88	13,15	34,16	34,09	9,80	21,97	40,60	16,49	14,57	21,16	17,46	12,65	26,29	22,59	7,82
3	19,95	5,58	15,34	23,93	7,97	9,62	25,70	12,23	7,28	1,46	7,82	8,96	19,04	8,94	12,40
4	12,17	6,85	13,58	13,92	5,40	9,91	14,50	6,25	7,68	5,93	8,34	9,32	9,16	6,87	15,73
5	10,63	6,00	9,49	11,61	5,68	7,61	11,97	6,24	6,83	7,33	6,58	7,60	12,46	6,20	10,80
6	8,41	4,23	6,72	8,97	4,06	5,52	9,08	4,36	5,13	5,68	4,01	5,50	10,99	3,89	7,38
7	6,07	3,67	6,05	6,29	3,07	5,04	6,26	2,86	4,55	4,14	3,39	4,81	6,57	2,91	6,94
8	4,81	3,04	4,92	4,93	2,52	4,23	4,97	2,29	3,93	3,83	2,75	4,01	5,35	2,51	5,47
9	3,91	1,88	3,41	4,02	1,49	2,98	4,10	1,31	2,89	3,36	1,42	2,93	5,35	1,28	3,62
10	3,20	1,78	2,85	3,26	1,40	2,56	3,32	1,19	2,58	2,79	1,47	2,73	4,17	1,27	3,49
11	2,67	1,68	2,32	2,67	1,29	2,13	2,69	1,02	2,23	2,33	1,42	2,43	3,00	1,34	3,38
12	2,28	1,07	1,54	2,33	0,84	1,59	2,39	0,75	1,87	2,13	0,82	1,98	3,24	0,61	2,51
13	1,94	0,77	1,19	2,01	0,73	1,43	2,10	0,84	1,81	1,92	0,74	1,76	3,33	0,67	2,05
14	1,39	0,53	0,69	1,39	0,25	0,99	1,43	0,28	1,43	1,37	0,75	1,33	2,09	0,72	2,02
15	1,06	0,68	0,98	1,06	0,58	1,18	1,10	0,65	1,50	1,12	0,91	1,36	1,57	0,99	1,74
16	0,84	0,75	1,33	0,87	0,78	1,46	0,91	0,89	1,66	0,94	0,89	1,56	1,96	0,96	1,18
17	0,49	0,69	0,99	0,43	0,60	0,99	0,44	0,59	1,05	0,44	0,67	1,04	1,11	0,65	0,52
18	0,87	1,18	1,04	0,79	1,07	0,93	0,79	1,01	0,83	0,83	0,96	0,82	0,65	1,18	0,62
19	0,77	1,08	1,12	0,70	1,02	1,08	0,64	0,98	1,05	0,44	0,84	1,27	0,41	0,83	1,10
20	1,86	1,89	1,94	1,86	1,89	1,93	1,93	1,95	1,98	2,07	2,10	2,35	2,16	2,22	2,53

Chapter **8**

Low-rank approximations based SBR algorithm

8.1 Introduction

In Chapters 5 and 6 we presented the Separate Bases Reduction algorithm – a block-structure preserving model reduction method for coupled systems. As discussed in these chapters, one of the SBR method's disadvantages is that the sizes of the its Krylov subspaces increase very fast for systems with a large number of internal inputs and outputs. Hence, the use of the SBR algorithm was recommended for the cases, in which the number of internal inputs and outputs was considerably smaller than the dimension of the system or comparable to the number of the external inputs and outputs. In this chapter, we approximate the internal inputs (outputs) by their GSVD-based dominant parts. This improves the efficiency of the SBR method. In addition we will prove that both the SBR algorithm and its low-rank based variant can be applied to coupled systems for which the internal input and output operators $\mathbf{B}$ and $\mathbf{C}$ are not explicitly availabe.

8.2   Implicitly defined couplings

In Chapter 5, we introduced the interconnected system of the form

$$
S \ : \ \begin{cases} s\begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \\ \qquad \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^T \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}. \end{cases}
\tag{8.2.1}
$$

as a result of the coupling of the two sub-systems, (5.2.1) and (5.2.2). Here, the coupling blocks are given by the explicit products of the internal inputs and outputs of the two sub-systems, namely $\mathbf{B}_3\mathbf{C}_4^T$ and $\mathbf{B}_4\mathbf{C}_3^T$. Having such a formulation at our disposal, we can apply the SBR algorithm in a straightforward way. However, for some applications it may be impossible to obtain matrices $\mathbf{B}_3$, $\mathbf{B}_4$, $\mathbf{C}_3$ and $\mathbf{C}_4$. In the following sections we propose a way of transforming an interconnected system with *implicitly defined couplings* of a form

$$
S \ : \ \begin{cases} s\begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \\ \qquad \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^T \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \end{cases},
\tag{8.2.2}
$$

with

$$
\mathbf{E} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix}
\tag{8.2.3}
$$

into a form that can be reduced using the SBR algorithm. Our goal is to find decompositions (factorizations) of the two coupling blocks

$$
\mathbf{A}_{12} = \tilde{\mathbf{B}}_3\tilde{\mathbf{C}}_4^T \qquad \text{and} \qquad \mathbf{A}_{21} = \tilde{\mathbf{B}}_4\tilde{\mathbf{C}}_3^T
\tag{8.2.4}
$$

that provide a good (with respect to the corresponding Krylov subspaces) approximation of the original internal inputs and outputs of the coupled system (8.2.2). A factorization of the type $\mathbf{A} = \mathbf{BC}$ is not be unique. The next section shows how to deal with this.

8.3   Decomposition theorem

In this section, related to (8.2.2), first we show that a factorization $\mathbf{A} = \mathbf{BC}$ is not unique and next we prove that if $\mathbf{A}_{12} = \mathbf{B}_1\mathbf{C}_1$ and simultaneously $\mathbf{A}_{12} = \mathbf{B}_2\mathbf{C}_2$ then $\mathcal{K}_p(\mathbf{A}_{11}, \mathbf{B}_1) = \mathcal{K}_p(\mathbf{A}_{11}, \mathbf{B}_2)$ if $\mathbf{C}_1$ and $\mathbf{C}_2$ are of *full column rank*. The proofs will be for the input-based Krylov subspaces. Similar theory applies to the output-based Krylov subspaces.

First, a factorization of the type $\mathbf{A} = \mathbf{BC}$ is not unique since $\mathbf{A} = \mathbf{IA}$ and $\mathbf{A} = \mathbf{AI}$ are two different factorizations. Even a QR factorization $\mathbf{A} = \mathbf{QR}$ is not unique since if $\mathbf{A} = \mathbf{QR}$ then $\mathbf{A} = (\mathbf{Q\overline{S}})(\mathbf{SR})$ for all complex valued diagonal matrices $\mathbf{S}$ with unit-length diagonal elements ($\mathbf{\overline{S}}$ denotes the complex conjugate of $\mathbf{S}$). Also other factorizations such as Gaussian-elimination based $\mathbf{A} = \mathbf{LU}$ exist.

Since we aim at the use of $\mathbf{B}$ for the generation of a Krylov subspace $\mathcal{K}_p(\mathbf{A}_{11}, \mathbf{B}_1)$ we will next show that the non-uniqueness does not need to be an issue. To this end we prove the following Lemma 8.3.1 and Theorem 8.3.1.

Lemma 8.3.1. *Let* $\mathbf{B} \in \mathbb{R}^{n \times p}$, $\mathbf{C} \in \mathbb{R}^{p \times m}$, $m, n, p \in \mathbb{N}$. *Then*

$$\mathrm{rank}(\mathbf{C}) = p \implies colspan\ \mathbf{BC} = colspan\ \mathbf{B}.$$

Proof. Matrix $\mathbf{C}$ has rank p which implies $p \leq m$ and that $\mathbf{C}$ has p linearly independent columns of length p. Thus based on

$$\mathrm{colspan}\ \mathbf{C} = \{\mathbf{Cx} \colon \mathbf{x} \in \mathbb{R}^m\} \tag{8.3.1}$$

one finds

$$\mathrm{colspan}\ \mathbf{C} \underset{(8.3.1)}{=} \{\mathbf{Cx} \colon \mathbf{x} \in \mathbb{R}^m\} \underset{p \leq m}{=} \mathbb{R}^p \tag{8.3.2}$$

whence

$$\mathrm{colspan}\ \mathbf{BC} \underset{(8.3.1)}{=} \{\mathbf{BCx} \colon \mathbf{x} \in \mathbb{R}^m\} \underset{(8.3.2)}{=} \{\mathbf{By} \colon \mathbf{y} \in \mathbb{R}^p\} \underset{(8.3.1)}{=} \mathrm{colspan}\ \mathbf{B}.$$

Note: The condition that $\mathbf{C}$ has full column rank is sufficient but not necessary. It can be relaxed: If for instance $\mathbf{B}$ has only $2 \leq p$ linearly independent columns, e.g. the ith and the jth column, then a sufficient condition is colspan $\mathbf{C} = \mathrm{colspan}\ \{\mathbf{e}_i, \mathbf{e}_j\} \subset \mathbb{R}^p$. $\qquad\square$

Theorem 8.3.1. *Let* $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{R}^{n \times p}$, $\mathbf{C}_1, \mathbf{C}_2 \in \mathbb{R}^{p \times m}$ *and* $m, n, p \in \mathbb{N}$.
If

$$\mathrm{rank}(\mathbf{C}_1) = \mathrm{rank}(\mathbf{C}_2) = p \quad and \quad \mathbf{B}_1\mathbf{C}_1 = \mathbf{B}_2\mathbf{C}_2$$

then

$$colspan\ \mathbf{B}_1 = colspan\ \mathbf{B}_2.$$

Proof. Observe that

$$\mathrm{colspan}\ \mathbf{B}_1 \underset{\mathrm{Lem.8.3.1}}{=} \mathrm{colspan}\ \mathbf{B}_1\mathbf{C}_1 = \mathrm{colspan}\ \mathbf{B}_2\mathbf{C}_2 \underset{\mathrm{Lem.8.3.1}}{=} \mathrm{colspan}\ \mathbf{B}_2.$$

$$\square$$

Next we prove that certain Krylov subspaces are identical.

Theorem 8.3.2. *Let* $\mathbf{A} \in \mathbb{R}^{n \times n}$ *is non-singular and* $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{R}^{n \times m}$, $n, m \in \mathbb{N}$. *Then*

$$colspan\ \mathbf{B}_1 = colspan\ \mathbf{B}_2 \implies \mathcal{K}_p(\mathbf{A}, \mathbf{B}_1) = \mathcal{K}_p(\mathbf{A}, \mathbf{B}_2).$$

Proof. Note that

$$\begin{aligned}
colspan\ \mathbf{B}_1 &= colspan\ \mathbf{B}_2 &\iff \\
\{\mathbf{B}_1\mathbf{x} \colon \mathbf{x} \in \mathbb{R}^m\} &= \{\mathbf{B}_2\mathbf{x} \colon \mathbf{x} \in \mathbb{R}^m\} &\iff \\
\{\mathbf{A}\mathbf{B}_1\mathbf{x} \colon \mathbf{x} \in \mathbb{R}^m\} &= \{\mathbf{A}\mathbf{B}_2\mathbf{x} \colon \mathbf{x} \in \mathbb{R}^m\} &\iff \\
colspan\ \mathbf{A}\mathbf{B}_1 &= colspan\ \mathbf{A}\mathbf{B}_2,
\end{aligned}$$

which, repeatedly applied, shows that $colspan\ \mathbf{A}^k\mathbf{B}_1 = colspan\ \mathbf{A}^k\mathbf{B}_2$ for all $k \geq 0$ whence $\mathcal{K}_p(\mathbf{A}, \mathbf{B}_1) = \mathcal{K}_p(\mathbf{A}, \mathbf{B}_2)$. $\qquad\qquad \square$

Theorem 8.3.1 in combination with Theorem 8.3.2 show that every factorization of an off-diagonal block of the form $\mathbf{A}_{12} = \mathbf{B}\mathbf{C}^{\mathrm{T}}$ with $\mathbf{C}$ of *full column rank* leads to the same krylov space $\mathcal{K}_p(\mathbf{A}_{11}, \mathbf{B})$. The following sections show how to use this property for the application of the SBR method to an arbitrary coupled system (8.2.2).

8.4   Decomposition theorem – numerical example

In Section 8.3 we showed that the Krylov space does not depend on the factors of the decomposition $\mathbf{A}_{12} = \mathbf{B}\mathbf{C}^{\mathrm{T}}$ when these factors are of maximal column rank. To illustrate this numerically, we calculate these factors of $\mathbf{A}_{12}$ with different factorization techniques, based on a *QR factoration* and *LU factorization*. For simplicity, we use a one-sided variant of the SBR method. The system used for the test is a linear beam coupled to a controller, like 5.7.3 in Section 5.7. Only the beam system has an external input and external output. Hence, the considered system is of a form

$$S \ : \ \begin{cases} s\begin{bmatrix} \mathbf{I}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{B}_3\mathbf{C}_4^T \\ \mathbf{B}_4\mathbf{C}_3^T & \mathbf{A}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0} \end{bmatrix}\mathbf{u}_1 \\ \qquad\qquad \mathbf{y}_1 = [\mathbf{C}_1^T\ \mathbf{0}]\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}. \end{cases} \qquad (8.4.1)$$

Let $\mathbf{A}_{12} = \mathbf{B}_3\mathbf{C}_4^T$ and $\mathbf{A}_{21} = \mathbf{B}_4\mathbf{C}_3^T$. Here, the full coupled system has 80 degrees of freedom, 40 for each sub-system. Both of the sub-systems have 5 internal inputs and 5 internal outputs. It means, that the coupling blocks $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ are of rank 5. For all cases, the same number of Krylov iterations is performed and the reduced-order systems

are of the order 55 (originally 80). The first sub-system was reduced from order 40 down to 30 and the second from order 40 down to 25.

To reduce the original system, we will build three reduction matrices involving an nth-order Krylov sub-space as follows:

- **Reduction matrix based on the original internal input blocks**
  The diagonal sub-blocks of the reduction matrix span the Krylov subspaces

$$\mathbf{V}_1 = \mathcal{K}_n(\mathbf{P}_1, \mathbf{R}_1),$$

  where

$$\mathbf{P}_1 = (s\mathbf{I}_{11} - \mathbf{A}_{11})^{-1} \quad \text{and} \quad \mathbf{R}_1 = (s\mathbf{I}_{11} - \mathbf{A}_{11})^{-1}[\mathbf{B}_1 \ \mathbf{B}_3]$$

  and

$$\mathbf{V}_2 = \mathcal{K}_n(\mathbf{P}_2, \mathbf{R}_2),$$

  where

$$\mathbf{P}_2 = (s\mathbf{I}_{22} - \mathbf{A}_{22})^{-1} \quad \text{and} \quad \mathbf{R}_2 = (s\mathbf{I}_{22} - \mathbf{A}_{22})^{-1}\mathbf{B}_4.$$

  The block-diagonal reduction matrix $\mathbf{V}$ is of the form

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix}.$$

- **Reduction matrix based on a QR decomposition of the coupling blocks**
  Based on a QR decomposition of the coupling matrices $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$, we get

$$\mathbf{A}_{12} = Q_1\mathcal{R}_1 \quad \text{and} \quad \mathbf{A}_{21} = Q_2\mathcal{R}_2.$$

  We use an *rank-revealing* version of the QR algorithm, i.e., Q_1, Q_2, $\mathcal{R}_1^{\mathrm{T}}$, $\mathcal{R}_2^{\mathrm{T}}$ are of full column rank. Hence, the matrices Q_1 and Q_2 used to build the Krylov subspaces have the same rank (and most likely amount of columns) as $\mathbf{B}_3$ and $\mathbf{B}_4$. Next, the reduction sub-blocks are created

$$\mathbf{V}_1^{QR} = \mathcal{K}_n(\mathbf{P}_1^{QR}, \mathbf{R}_1^{QR}),$$

  where

$$\mathbf{P}_1^{QR} = (s\mathbf{I}_{11} - \mathbf{A}_{11})^{-1} \quad \text{and} \quad \mathbf{R}_1^{QR} = (s\mathbf{I}_{11} - \mathbf{A}_{11})^{-1}[\mathbf{B}_1 \ Q_1]$$

  and

$$\mathbf{V}_2^{QR} = \mathcal{K}_n(\mathbf{P}_2^{QR}, \mathbf{R}_2^{QR}),$$

where

$$\mathbf{P}_2^{QR} = (s\mathbf{I}_{22} - \mathbf{A}_{22})^{-1} \qquad \text{and} \qquad \mathbf{R}_2^{QR} = (s\mathbf{I}_{22} - \mathbf{A}_{22})^{-1}Q_2.$$

The block-diagonal reduction matrix $\mathbf{V}^{QR}$ is of the form

$$\mathbf{V}^{QR} = \begin{bmatrix} \mathbf{V}_1^{QR} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2^{QR} \end{bmatrix}.$$

- **Reduction matrix based on the LU decomposition of the coupling blocks**
  Based on the LU decomposition of the coupling matrices $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$, we get

$$\mathbf{A}_{12} = \mathcal{L}_1\mathcal{U}_1 \qquad \text{and} \qquad \mathbf{A}_{21} = \mathcal{L}_2\mathcal{U}_2.$$

We use a *rank-revealing* version of the LU algorithm, i.e., $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{U}_1^{\mathrm{T}}$ and $\mathcal{U}_2^{\mathrm{T}}$ are of full column rank. Hence, the matrices $\mathcal{L}_1$ and $\mathcal{L}_2$ used to build the Krylov subspaces have the same rank (and most likely amount of columns) as $\mathbf{B}_3$ and $\mathbf{B}_4$. Next, the reduction sub-blocks are created

$$\mathbf{V}_1^{LU} = \mathcal{K}_n(\mathbf{P}_1^{LU}, \mathbf{R}_1^{LU}),$$

where

$$\mathbf{P}_1^{LU} = (s\mathbf{I}_{11} - \mathbf{A}_{11})^{-1} \qquad \text{and} \qquad \mathbf{R}_1^{LU} = (s\mathbf{I}_{11} - \mathbf{A}_{11})^{-1}[\mathbf{B}_1 \ \mathcal{L}_1]$$

and

$$\mathbf{V}_2^{LU} = \mathcal{K}_n(\mathbf{P}_2^{LU}, \mathbf{R}_2^{LU}),$$

where

$$\mathbf{P}_2^{LU} = (s\mathbf{I}_{22} - \mathbf{A}_{22})^{-1} \qquad \text{and} \qquad \mathbf{R}_2^{LU} = (s\mathbf{I}_{22} - \mathbf{A}_{22})^{-1}\mathcal{L}_2.$$

The block-diagonal reduction matrix $\mathbf{V}^{LU}$ is of the form

$$\mathbf{V}^{LU} = \begin{bmatrix} \mathbf{V}_1^{LU} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2^{LU} \end{bmatrix}.$$

Figure 8.1 shows the magnitude plots with respect to the frequency of the frequency response functions of the three reduced-order systems, created using original, QR-, and LU-decomposition based input matrices. The plots are almost identical, which is confirmed in Figure 8.2, that shows the relative errors between the reduced-order frequency response function of the original system and the frequency response functions computed based on both decompositions. The small differences between the three frequency response functions should be caused by round-off errors.

The next section shows how Theorems 8.3.1 and 8.3.2 in combination with GSVD can be used to improve the performance of the SBR algorithm applied to coupled systems with a high number of couplings (or interconnections).
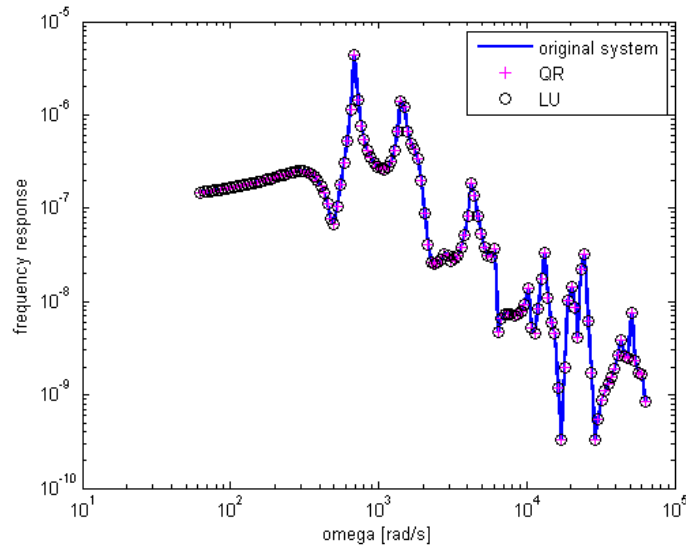
Figure 8.1: Magnitude plots of the frequency response functions of the reduced-order systems based on different decompositions of the coupling blocks
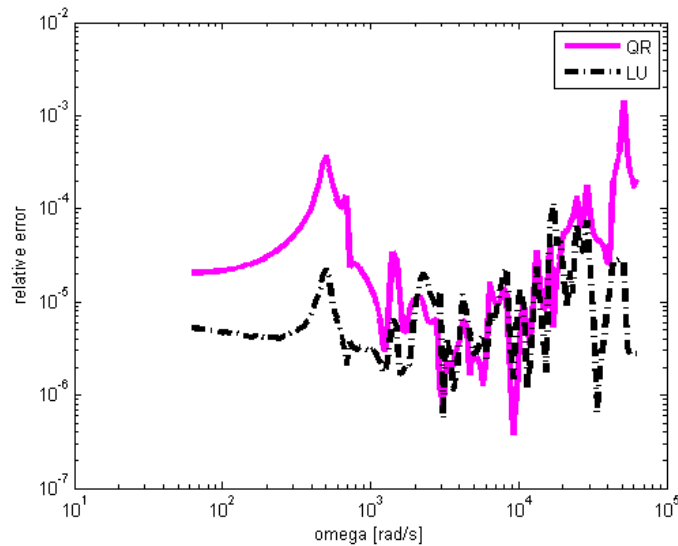
Figure 8.2: Magnitude plots of the relative errors of the reduced-order frequency response functions based on different decompositions of the coupling blocks with respect to the reduced-order frequency response function based on the original input and output matrices

8.5   Low-rank approximations based SBR algorithm

As shown in Chapter 7 in Sections 7.4 and 7.5 for some coupled systems it is not always necessary to take into account all of the coupling components. Sometimes only a small number of them determines the behavior of the system and the rest can be neglected without much loss of accuracy. This section extends the application of the SBR algorithm to coupled (or interconnected) systems characterized by a high number of couplings of which only a small percentage is relevant to obtain an accurate solution.

Chapters 5 and 6 pointed out that the standard SBR method should be applied only for the systems with a relatively small number of internal inputs and outputs. That is, only for coupled systems where few degrees of freedom of one sub-system (related to one physical domain or to a physcial quantity) are coupled/connected to the other subsystem, which implies that the coupling blocks $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ of the system (8.2.2) are of low rank. Otherwise, the SBR method produces reduction bases which increase in size too fast with respect to the number of Krylov iterations. However, if only a part of the components of the high rank coupling blocks is relevant, we can decrease the growth speed of the reduction bases. To do so, we first need to determine, which components of the coupling are important and should be kept, and which ones can be neglected. One of the ways to make this decision, is to apply the generalized singular value decomposition (GSVD) of Theorem 7.2.3 to the coupling matrices $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$. As explained in Section 7.3, for (8.2.2), the GSVD should be applied to the pairs $(\mathbf{A}_{11}^T, \mathbf{A}_{12}^T)$ and $(\mathbf{A}_{22}^T, \mathbf{A}_{21}^T)$.

Following the notation of Theorem 7.2.3 one has

$$\begin{array}{ccc} \mathbf{A}_{11}^T = \mathbf{U}_1 \mathbf{C}_1 \mathbf{X}_1^T & & \mathbf{A}_{22}^T = \mathbf{U}_2 \mathbf{C}_2 \mathbf{X}_2^T \\ \mathbf{A}_{12}^T = \mathbf{V}_1 \mathbf{S}_1 \mathbf{X}_1^T & \text{and} & \mathbf{A}_{21}^T = \mathbf{V}_2 \mathbf{S}_2 \mathbf{X}_2^T \end{array}$$

which results in the expressions for the coupling blocks

$$\mathbf{A}_{12} = \mathbf{X}_1 \mathbf{S}_1^T \mathbf{V}_1^T \tag{8.5.1}$$

$$\mathbf{A}_{21} = \mathbf{X}_2 \mathbf{S}_2^T \mathbf{V}_2^T. \tag{8.5.2}$$

Note, that here the matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ are not used to denote external output matrices, but components of the GSVD. Assuming that the coupling blocks are of the form (8.2.4), since $\mathbf{S}_1$ and $\mathbf{S}_2$ are real-valued non-negative diagonal, we can define the input and output matrices as following products

$$\tilde{\mathbf{B}}_3 = \mathbf{X}_1 \mathbf{S}_1^{1/2}, \qquad \tilde{\mathbf{C}}_4 = \mathbf{V}_1 \mathbf{S}_1^{1/2} \tag{8.5.3}$$

$$\tilde{\mathbf{B}}_4 = \mathbf{X}_2 \mathbf{S}_2^{1/2}, \qquad \tilde{\mathbf{C}}_3 = \mathbf{V}_2 \mathbf{S}_2^{1/2}. \tag{8.5.4}$$

Since $\mathbf{S}_1$ and $\mathbf{S}_2$ are diagonal matrices with non-negative entries their square roots are diagonal matrices with entries $\sqrt{[\mathbf{S}_1]_{ii}}$ and $\sqrt{[\mathbf{S}_2]_{ii}}$. Constructing the inputs and outputs as in (8.5.3) and (8.5.4), all of $\mathbf{B}_i$ and $\mathbf{C}_i$, $i = 3, 4$ are scaled by $\sqrt{\mathbf{S}_1}$ or $\sqrt{\mathbf{S}_2}$.

According to the Theorems 8.3.1 and 8.3.2, $\mathcal{K}_p(\mathbf{A}_{11}, \mathbf{B}_3\mathbf{C}_4^T) = \mathcal{K}_p(\mathbf{A}_{11}, \tilde{\mathbf{B}}_3\tilde{\mathbf{C}}_4^T)$ and $\mathcal{K}_p(\mathbf{A}_{22}, \mathbf{B}_4\mathbf{C}_3^T) = \mathcal{K}_p(\mathbf{A}_{11}, \tilde{\mathbf{B}}_4\tilde{\mathbf{C}}_3^T)$. Moreover, using a type of the decomposition that orders the components with respect to their importance has an additional benefit. It makes it possible to approximate the inputs and outputs leaving only the most relevant components and, as a result, reduces the dimensions of the blocks. In some cases, this reduction is sufficient to allow for an efficient application of the SBR algorithm.

Let us now compare the procedures of building the standard and GSVD-based Krylov subspaces. Here, we will limit the discussion to the case of creation of a Krylov space based on inputs of the sub-system (5.2.1), but a similar analysis applies to all the other cases, i.e. input-based Krylov subspace for (5.2.2) and output-based Krylov subspaces for both sub-systems, (5.2.1) and (5.2.2). As defined in Chapter 5, matrices $\mathbf{A}_{11} \in \mathbb{R}^{N_1 \times N_1}$, $\mathbf{B}_1 \in \mathbb{R}^{N_1 \times m_1}$ and $\mathbf{B}_3 \in \mathbb{R}^{N_1 \times m_3}$. Assume, that $\mathbf{B}_3$ has full column rank m_3 and that application of GSVD to the pair $(\mathbf{A}_{11}^T, \mathbf{A}_{12}^T)$ leads to

$$\tilde{\mathbf{B}}_3 = \mathbf{X}_1\mathbf{S}_1^{1/2} = [\tilde{\mathbf{b}}_1, ..., \tilde{\mathbf{b}}_{m_3}] \in \mathbb{R}^{N_1 \times m_3}.$$

where both $\mathbf{X}_1$ and $\mathbf{S}_1$ in (8.5.3) are of full column rank. Next, let $\hat{\mathbf{B}}_3 = \mathbf{X}_1^{(k)}(\mathbf{S}_1^{(k)})^{1/2}$ approximate $\tilde{\mathbf{B}}_3$ with the use of k dominant components. Then

$$\hat{\mathbf{B}}_3 = [\tilde{\mathbf{b}}_1, ..., \tilde{\mathbf{b}}_k] \in \mathbb{R}^{N_1 \times k}. \tag{8.5.5}$$

For simplicity, we assume that $m_1 + m_3$ is a multiple of $m_1 + k$, so there exists $\lambda \in \mathbb{N}$ such that

$$m_1 + m_3 = \lambda(m_1 + k). \tag{8.5.6}$$

The pth Krylov subspace created by the SBR algorithm for the sub-system (5.2.1) for $s_0 \in \mathbb{C}$ is

$$\mathcal{K}_p(\mathbf{P}_1, \mathbf{R}_1) = \text{colspan}\{\mathbf{R}_1, \mathbf{P}_1\mathbf{R}_1, ..., \mathbf{P}_1^{p-1}\mathbf{R}_1\},$$

where

$$\mathbf{P}_1 = (s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}\mathbf{E}_{11} \quad \text{and} \quad \mathbf{R}_1 = (s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}[\mathbf{B}_1 \ \mathbf{B}_3]$$

and consists of $p(m_1 + m_3)$ columns (assuming that no linear dependence occurs). Likewise,

$$\mathcal{K}_{\lambda p}(\mathbf{P}_1, \hat{\mathbf{R}}_1) = \text{colspan}\{\hat{\mathbf{R}}_1, \mathbf{P}_1\hat{\mathbf{R}}_1, ..., \mathbf{P}_1^{\lambda p-1}\hat{\mathbf{R}}_1\}, \tag{8.5.7}$$

where

$$\mathbf{P}_1 = (s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}\mathbf{E}_{11} \quad \text{and} \quad \hat{\mathbf{R}}_1 = (s_0\mathbf{E}_{11} - \mathbf{A}_{11})^{-1}[\mathbf{B}_1 \ \hat{\mathbf{B}}_3]$$

consists also of $p(m_1 + m_3)$ columns, but *approximately* matches λ as many moments of the original transfer function.

Projecting system (5.2.1) onto a subspace $\mathcal{K}_{\lambda p}(\mathbf{P}_1, \hat{\mathbf{R}}_1)$ in (8.5.7) does not preserve the moments of the transfer function of this sub-system. However, if the column span of the matrix $\hat{\mathbf{B}}_3$ gives a good approximation of the column span of the matrix $\mathbf{B}_3$ we can

expect that the reduced-order system obtained by projection onto the space (8.5.7) will give an accurate approximation of the appropriate number of moments of the transfer function of the original system. Moreover, if the matrix $\mathbf{B}_3$ can be approximated by $\hat{\mathbf{B}}_3$ with a significantly smaller number of columns, λ times more steps may be used during the Krylov procedure (to approximate a higher number of moments) or one can use more expansion points, keeping the reduced-order model still relatively small.

In the next section, we present the results of a numerical test that show the advantage of using the low-rank approximation based SBR algorithm for a system with high order of coupling.

8.6   The low-rank approximation based SBR algorithm – numerical example

In this section, we consider Example 5.7.2 which was already discussed in Section 5.7 and is identical to Example 6.4.2 in Section 6.4. The difficulty of this test case is that here the coupling blocks of the system (5.7.3) are of rank 10 (the coupled system has 10 internal inputs and 10 internal outputs), while each of the sub-systems contains only 40 degrees of freedom (80 degrees of freedom in total). As shown in Example 5.7.2 and Example 6.4.2 the standard SBR algorithm generates to many columns to be competitive. However, the use of low-rank approximations makes the SBR algorithm more competitive. Figure 8.3 shows the magnitude plots with respect to the frequency of the original and reduced-order frequency response functions. In case of the two-sided BSP method and the two-sided SBR algorithm based reduced-order systems, the reduction bases were created in the same manner as described in Example 6.4.2 in Chapter 6 and the original system was reduced to 42 degrees of freedom. The low-rank approximation based two-sided SBR algorithm created the reduction bases for rank 3 approximations of the coupling blocks, i.e. the internal input and output matrices $\mathbf{B}_i, \mathbf{C}_i \in \mathbb{R}^{40 \times 10}$, $i = 3, 4$ were approximated by $\hat{\mathbf{B}}_i, \hat{\mathbf{C}}_i \in \mathbb{R}^{40 \times 3}$, $i = 3, 4$. Hence, every Krylov step was adding 4 new columns to the reduction basis (3 corresponding to $\hat{\mathbf{B}}_3$ or $\hat{\mathbf{C}}_4$ and 1 corresponding to $\mathbf{B}_1$ or $\mathbf{C}_1$) in case of the sub-system S_1 and 3 new columns (corresponding to $\hat{\mathbf{B}}_4$ or $\hat{\mathbf{C}}_3$) in case of the sub-system S_2. To construct the reduced-order system of dimension 42, the low-rank approximation based SBR algorithm performed 6 iterations for each sub-system (for both, input and output related bases). Figure 8.4 shows the magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original one. Note that the two-sided SBR algorithm based on low-rank approximations of the internal inputs and outputs leads to much better results than the SBR algorithm applied to the original coupling blocks. The two-sided low-rank based reduced-order transfer function $\mathbf{H}_{\mathrm{low-rank-SBR}}$ approximates $\mathbf{H}$ less accurate than the standard two-sided BSP transfer function but in the neighborhood of the expansion point
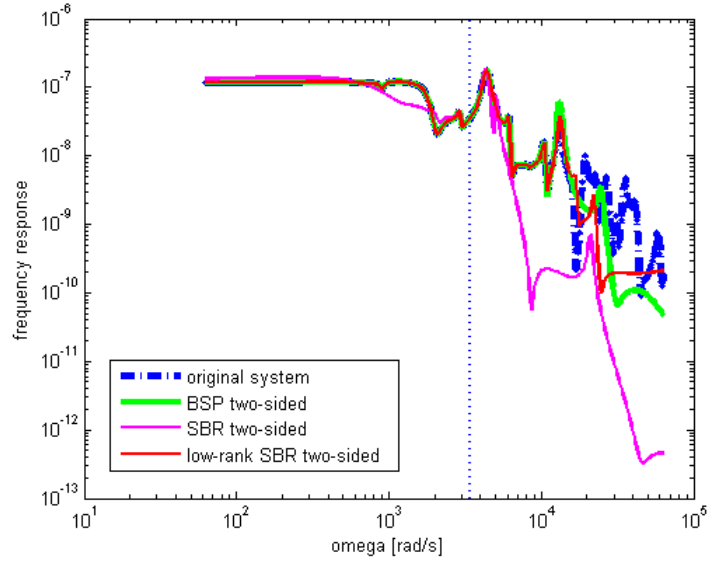
Figure 8.3: Magnitude plots of the frequency response functions of the original and reduced-order systems

s the relative error is still below 2%. Table 8.1 shows that not only the first 6 derivatives are matched but also the 7th one is well approximated.

Table 8.1: Derivatives of the original and low-rank approximation based reduced-order transfer functions for the expansion point (s) for the second example, multiplied by 10^7

i	$\partial^i \mathbf{H}(s)$	$\partial^i \mathbf{H}_{\mathrm{low-rank-SBR}}(s)$
0	-0.349984611544531	-0.349975323605725
1	0.000580754070987	0.000580770193275
2	-0.000001928114532	-0.000001928787960
3	0.000000012770698	0.000000012766510
4	-0.000000000067912	-0.000000000068062
5	0.000000000000859	0.000000000000859
6	-0.000000000000014	-0.000000000000014
7	0.000000000000000	0.000000000000000
8	-0.000000000000000	-0.000000000000000
9	0.000000000000000	0.000000000000000
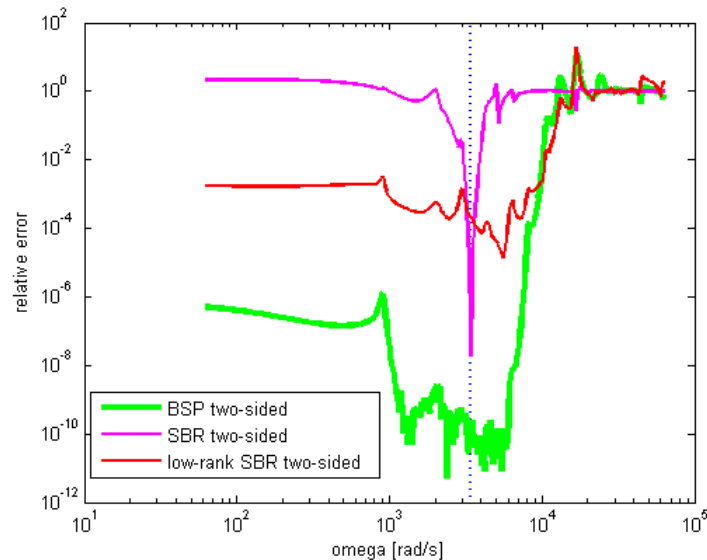
Figure 8.4: Magnitude plots of the relative errors of the reduced-order frequency response functions with respect to the original frequency response function

8.7   Conclusions

In this chapter we proposed a solution to the limitation of the SBR algorithm which is that it is mainly applicable to slightly coupled (or interconnected) systems. Using GSVD introduced in Chapter 7 we showed how to find and use the most relevant coupling components with respect to each sub-system. Just using the relevant couplings (principal components) we were able to efficiently apply the SBR algorithm to an example for which the standard SBR algorithm (due to the high number of interconnections) in Chapters 5 and 6 was not competitive. In addition we showed how the SBR method can be efficiently used for a small coupled system since we need to add (many) fewer columns per Krylov space iteration. This approach gives us the freedom to apply the SBR algorithm to any coupled system, as long as the couplings can be well approximated by a relatively low number of their dominant components. The advantage is in the possibility of using the uncoupled formulation of the coupled system, which, in some cases, can be computationally very beneficial. However, this type of reduction technique does not anymore ensure the exact matching of the moments of the original system's transfer function. Instead, we can expect an approximation of the moments which depends on the quality of our approximations of the inputs and outputs, see also [52].

Chapter 9

The OCE benchmark

9.1 Introduction

The benchmark system treated in this chapter is a model of a printhead delivered by Océ Technologies B.V. in the Netherlands. It is a MEMS (micro-electro-mechanical-system) based design, containing a large number of individual channels integrated into a single chip. A schematic overview of a single channel (a side and bottom view) is shown in Figure 9.1. The dotted line depicts the ink flow; the ink, coming from the reservoir, enters through a restriction (1), from which it flows into the actuation chamber (2). Below the actuation chamber, a 300 μm long feed-through is placed (3), after which the nozzle plate is reached. The nozzle plate is 75 μm thick and consists of a pyramid shaped funnel (4) and a nozzle (5) with a radius of 11 μm.

The main goal is to suppress acoustic pressure waves, which can be generated in a number of ways, such as the non-continuous ink supply by many thousands of ink channels, residual vibrations at the inlet of the ink channels, fast movement of the printhead, resonance of the whole structure, etc.

The models of such devices used for simulations can reach large dimensions, hence application of the model order reduction techniques is often required, to decrease the simulation time. In this chapter, we study the application of the GSVD based approximations for the coupling blocks in the model of the printhead.
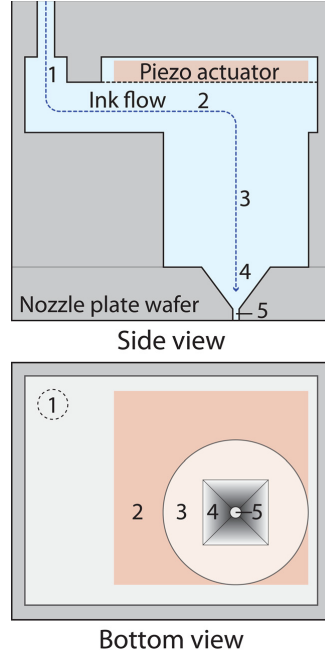
Figure 9.1: A schematic overview of a single channel (courtesy of Herman Wijshoff)

9.2   The second and first order system

The related system of equations is a second order system. Let $n_1, n_2 \in \mathbb{N}$ and $n = n_1 + n_2$. The *second order system* of interest is

$$\begin{cases} \mathbf{Mx}'' + \mathbf{Kx} = \mathbf{b} \\ \qquad\quad \mathbf{y} = \mathbf{cx} \end{cases} \tag{9.2.1}$$

with $(n_1 + n_2) \times (n_1 + n_2)$, 2×2 block-matrices

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{0} \\ \mathbf{M}_{22} & \mathbf{M}_{22} \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{0} & \mathbf{K}_{22} \end{bmatrix} \tag{9.2.2}$$

and $\mathbf{M}_{21} = -\rho \mathbf{K}_{12}^{\mathrm{T}}$. The first sub-system corresponds to the displacement of the structure and the second sub-system describes the pressure of the fluid. The related *Laplace transformation*

$$\begin{cases} \tilde{w}^2 \mathbf{Mx} + \mathbf{Kx} = \mathbf{b} \\ \qquad\qquad\quad \mathbf{y} = \mathbf{cx} \end{cases}$$

leads to transfer function

$$H(w) = \mathbf{c}(\mathbf{K} + \tilde{w}^2 \mathbf{M})^{-1}\mathbf{b}, \qquad \tilde{w} \in \mathbb{C}.$$

Searching for purely *oscillatory modes* implies that the related $\tilde{w}$ is purely imaginary, i.e., that one is interested in positive real values w of:

$$H(w) = \mathbf{c}(\mathbf{K} - w^2\mathbf{M})^{-1}\mathbf{b}, \qquad w \in \mathbb{R}. \tag{9.2.3}$$

Let $\mathbf{x}_2 = \mathbf{x}_1'$. Then the *first order system* reformulation of (9.2.1) is

$$\begin{cases} \mathbf{x}_2 = \mathbf{x}_1' \\ \mathbf{M}\mathbf{x}_2' + \mathbf{K}\mathbf{x}_1 = \mathbf{b} \end{cases} \implies \begin{cases} \mathbf{x}_1' - \mathbf{x}_2 = \mathbf{0} \\ \mathbf{M}\mathbf{x}_2' + \mathbf{K}\mathbf{x}_1 = \mathbf{b} \end{cases}$$

which implies

$$\begin{cases} \underbrace{\begin{bmatrix} \mathbf{I} & \\ & \mathbf{M} \end{bmatrix}}_{\mathbf{E}} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}' = \underbrace{\begin{bmatrix} & \mathbf{I} \\ -\mathbf{K} & \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix}}_{\mathbf{B}} \\ \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{c} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}. \end{cases}$$

Its related transfer function is

$$H(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}. \tag{9.2.4}$$

Solution of $\mathbf{FX} = \mathbf{B}$:

$$\begin{cases} s\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0} \\ s\mathbf{M}\mathbf{x}_2 + \mathbf{K}\mathbf{x}_1 = \mathbf{b} \end{cases} \implies \begin{cases} s\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0} \\ s^2\mathbf{M}\mathbf{x}_1 + \mathbf{K}\mathbf{x}_1 = \mathbf{b} \end{cases} \implies \begin{cases} \mathbf{x}_1 = (s^2\mathbf{M} + \mathbf{K})^{-1}\mathbf{b} \\ \mathbf{x}_2 = s\mathbf{x}_1. \end{cases}$$

This implies that

$$\mathbf{y}_1 = \mathbf{c}(s^2\mathbf{M} + \mathbf{K})^{-1}\mathbf{b}$$

is identical to $\mathbf{y}$ if and only if $s = iw, w \in \mathbb{R}$.

In the sequel we will examine the second order system.

9.3   Sparsity patterns and magnitudes of the blocks of M, K

There are three available discretizations for the *OCE application*: *coarse*: `1188_1050`, *medium*: `4752_5304` and *fine*: `20748_35775`. The numbers relate to the amount of degrees of freedom as follows: Case `4752_5304` implies $n_1 = 4752$ and $n_2 = 5304$. Extracted from ANSYS, the blocks $\mathbf{M}_{11}, \mathbf{M}_{21}, \mathbf{M}_{22}, \mathbf{K}_{11}, \mathbf{K}_{12}, \mathbf{K}_{22}$ in (9.2.2) are very differently scaled: For instance, for the medium case their absolute value greatest resp. smallest entries (*magnitude*) are of the order

$$\begin{aligned} \mathbf{M} &= O\left(\begin{bmatrix} 10^{-10} & \mathbf{0} \\ 10^{-5} & 10^{-18} \end{bmatrix}\right), \quad \mathbf{K} = O\left(\begin{bmatrix} 10^{+8} & 10^{-8} \\ \mathbf{0} & 10^{-4} \end{bmatrix}\right), \\ \mathbf{M} &= O\left(\begin{bmatrix} 10^{-12} & \mathbf{0} \\ 10^{-6} & 10^{-20} \end{bmatrix}\right), \quad \mathbf{K} = O\left(\begin{bmatrix} 10^{-12} & 10^{-9} \\ \mathbf{0} & 10^{-6} \end{bmatrix}\right). \end{aligned} \tag{9.3.1}$$

For the calculation of the transfer function furthermore note that $w \in [0, 2\pi 1500]$. Thus approximately, $w^2 \in [0, 10^8]$. The use of the standard MATLAB '\' operations to solve $(\mathbf{K} - w^2\mathbf{M})\mathbf{x} = \mathbf{b}$ leads to error messages and abortions, not to solutions. An alternative, the use of the MATLAB package Factorize, alleviates this problem, but (too) severe round-off remains. Furthermore, the '\' operation turns out to be very slow for this poorly scaled problem. Investigation shows that that $\mathbf{K}_{11}$ contains entries in $[10^{-12}, 10^{+8}]$. The use of standard double precision floating point *IEEE arithmetic* involved in matrix operations such as matrix multiplication is bound to round-away contributions of the smaller entries.

Further investigation shows that all diagonal blocks but $\mathbf{K}_{11}$ are symmetric. For the results shown in this thesis the slightly non-symmetric ANSYS block $\mathbf{K}_{11}$ has been used as is. The results would be the same if one had instead used its symmetric part $(\mathbf{K}_{11} + \mathbf{K}_{11}^{\mathrm{T}})/2$ (tested). It has also been shown that indeed $\mathbf{M}_{21} = -\rho\mathbf{K}_{12}^{\mathrm{T}}$ for all three examples, where $\rho = 1090$.

Observe that the determination of the smallest absolute value positive entry of a sparse MATLAB matrix with MATLAB is not trivial: The smallest entry of a sparse matrix usually is zero (since the default entry has value zero), MATHWORKS and other sources do not provide an on-the-shelf solution. To obtain the smallest non-zero entry we have written a MATLAB function vfilter which for a full or sparse matrix $\mathbf{X}$ writes all entries $\mathbf{X}_{ij}$ such that $|\mathbf{X}_{ij}| > \epsilon \geq 0$ column-wise into a full vector. The use of this function applied to matrix X and $\epsilon = 0$ in combination with min provides the smallest absolute value entry of X.

Naturally, small entries should only be discarded if they are not relevant to the system of interest, i.e., if the the system is properly scaled, which is the topic of discussion of the next section.

9.4   Scaling the second order system

We need to scale the matrices $\mathbf{K}$ and $\mathbf{M}$ ($\mathbf{E}$ and $\mathbf{A}$) to obtain a *numerically robust* solution of the system

$$\mathbf{F}(w)\mathbf{x} = \mathbf{b} \iff (\mathbf{K} - w^2\mathbf{M})\mathbf{x} = \mathbf{b} \iff \begin{bmatrix} \mathbf{K}_{11} - w^2\mathbf{M}_{11} & \mathbf{K}_{12} \\ \rho w^2\mathbf{K}_{12}^{\mathrm{T}} & \mathbf{K}_{22} - w^2\mathbf{M}_{22} \end{bmatrix} \mathbf{x} = \mathbf{b},$$

which depends on w. For the problem of interest we expect symmetric blocks $\mathbf{M}_{11}, \mathbf{M}_{22}, \mathbf{K}_{11}$ and $\mathbf{K}_{22}$, and $\mathbf{M}_{21} = -\rho\mathbf{K}_{12}^{\mathrm{T}}$. This implies that this system could be scaled (preconditioned) into a symmetric one (*symmetry scaling*), for which efficient linear solvers exist.

(a) M_{11}

(b) M_{22}

(c) K_{11}

(d) K_{22}

(e) K_{12}
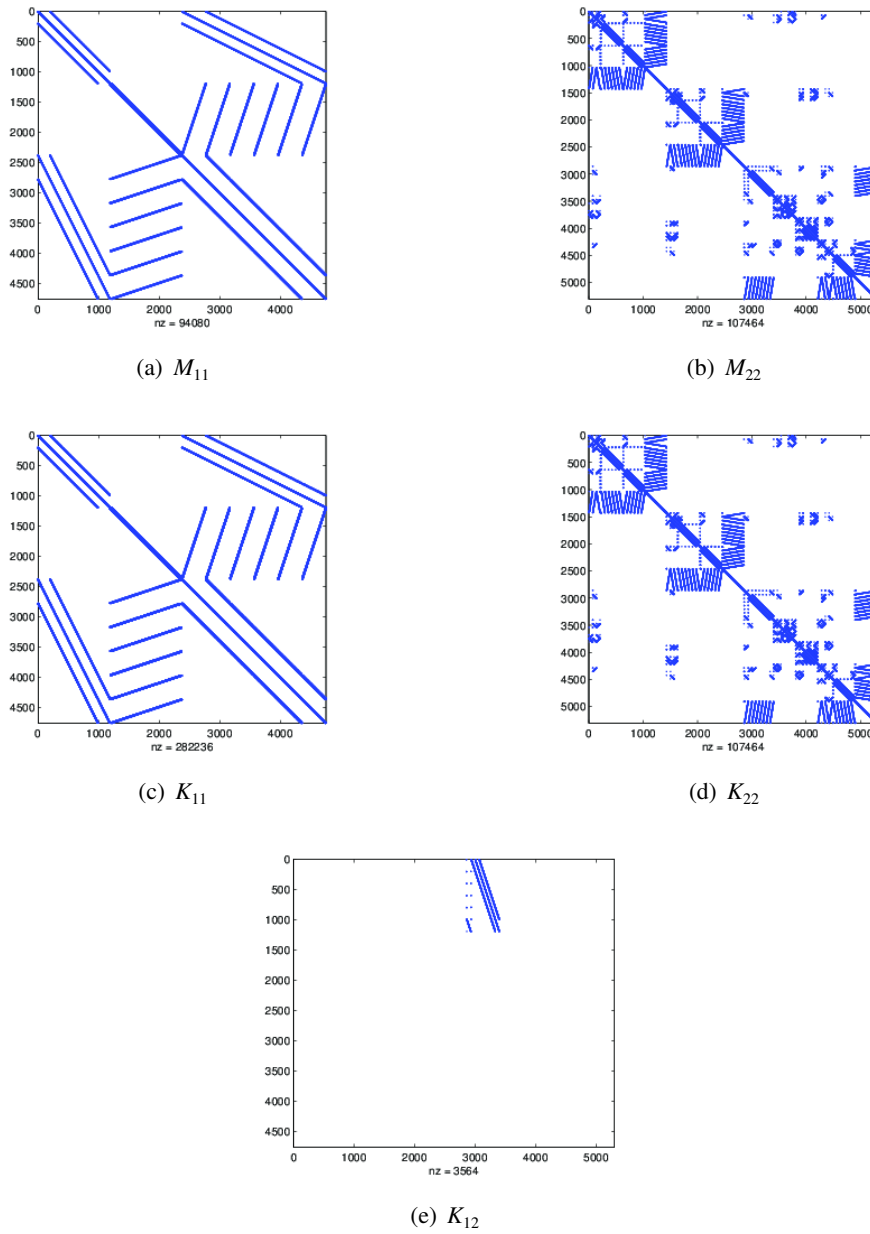
Figure 9.2: Sparsity pattern of matrix blocks.

This can be done as follows: Observe that for a two by two matrix

$$\mathbf{A} = \begin{bmatrix} a & d \\ c & b \end{bmatrix}, \quad \mathbf{D}_1 = \begin{bmatrix} 1 & \\ & \sqrt{c/d} \end{bmatrix} \implies \mathbf{D}_1^{-1}\mathbf{A}\mathbf{D}_1 = \begin{bmatrix} a & \sqrt{cd} \\ \sqrt{cd} & b \end{bmatrix}$$

can be scaled to a symmetric one. Hence, based on $c = \rho w^2$ and $d = 1$, define

$$\mathbf{D}_1 = \begin{bmatrix} \mathbf{I}_{n_1} & \\ & \sqrt{\rho w^2}\mathbf{I}_{n_2} \end{bmatrix}.$$

Furthermore, to better scale the entries inside and between blocks (create diagonal elements of magnitude 1), define

$$\mathbf{D}_2 = \mathrm{diag}(1/\sqrt{[\mathbf{D}_1^{-1}\mathbf{F}\mathbf{D}_1]_{11}}, \ldots, 1/\sqrt{[\mathbf{D}_1^{-1}\mathbf{F}\mathbf{D}_1]_{nn}}).$$

We now scale with a *diagonal scaling*:

$$\hat{\mathbf{M}} := \underbrace{\mathbf{D}_2\mathbf{D}_1^{-1}}_{\mathbf{Q}}\mathbf{M}\underbrace{\mathbf{D}_1\mathbf{D}_2}_{\mathbf{P}},$$
$$\hat{\mathbf{K}} := \mathbf{Q}\mathbf{K}\mathbf{P}, \hat{\mathbf{b}} := \mathbf{Q}\mathbf{b}, \hat{\mathbf{c}} := \mathbf{c}\mathbf{P},$$

which, by invariance under inputs and outputs transformations means that

$$\hat{\mathbf{H}}(w) := \mathbf{c}(\hat{\mathbf{K}} - w^2\hat{\mathbf{M}})^{-1}\hat{\mathbf{b}}$$

is identical to $\mathbf{H}$ in (9.2.3) for all w. Obviously $\mathbf{D}_1$ is non-singular except for $w = 0$ and $\mathbf{D}_2$ exists and is non-singular when all diagonal entries of $\mathbf{D}_1^{-1}\mathbf{F}\mathbf{D}_1$ are non-zero.

The factors $\mathbf{P} = \mathbf{P}(w)$ and $\mathbf{Q} = \mathbf{Q}(w)$ depend on w. This is fine for the construction of Krylov spaces to match moments. However, to plot the transfer function $\mathbf{H}$ one needs to evaluate $\mathbf{c}(\mathbf{K} - w_k^2\mathbf{M})$ for many $w_k \in [0, 10^8]$. Repeated calculation of $\mathbf{P}(w_k)$ and $\mathbf{Q}(w_k)$ would be (too) costly, so we decided to use the w-independent factors $\mathbf{P} := \mathbf{P}(\hat{w})$ $\mathbf{Q} := \mathbf{Q}(\hat{w})$ for all w where $\hat{w}$ is the average of all w_k. For the OCE example, to plot the transfer functions, we sample the provided region of interest: $w_k = 5\pi \cdot k, k = 0, \ldots, 600$. The value of $\hat{w}$ turns out to be w_{301} which is close to but not too close to a *pole* of $\mathbf{H}$ and such that all diagonal entries of $\mathbf{D}_1^{-1}\mathbf{F}\mathbf{D}_1$ are non-zero.

9.5   The structure and the GSVD of $\mathbf{K}_{12}$

Here we briefly comment on the GSVD of the scaled $\mathbf{K}_{12}^{\mathrm{T}}$. Figure 9.2 and numerical investigation show that $\mathbf{K}_{12} \in \mathbb{R}^{1188 \times 1050}$ is a *sparse matrix* which contains a small *non-zero sub-block* of size 295×175 (window $(3, \ldots, 297) \times (561, \ldots, 735)$). This is

typical for applications where the different physical quantities are defined in bordering sub-domains and are coupled via the mutual boundary – if one numbers the degrees of freedom on the mutual boundary consecutively. Since $\mathbf{K}_{12}^{\mathrm{T}}$ has this structure it is of the type as required for Corollary 7.2.1. This means that also $\mathbf{V}$ has all its non-zero entries in the same sub-block, i.e., it only has possible non-zero entries from row 561 to 735. This information is of importance, because the standard GSVD implementations such as `MATLAB`'s do not use this information and generate $\mathbf{V}$ which contains round-off (non-zero) entries outside the window, as can be seen in Figure 9.3. For the medium test case the results are worse, as to be expected: For $p = 5$ and $\epsilon = 0$, $\mathbf{K}_{12}^{(p)}$ (see below) is a full matrix.

To work around this problem we have written a `MATLAB` function `spfilter` which for a full or sparse matrix $\mathbf{X}$ copies all entries $\mathbf{X}_{ij}$ such that $|\mathbf{X}_{ij}| > M(\mathbf{X}) \cdot \epsilon$ into a sparse matrix $\mathbf{Y}$, where $M(\mathbf{X}) := \max\{|X|_{ij}\}_{i,j}$. This way, using $\epsilon = 10^{-11}$, both $\mathbf{K}_{12} = \mathbf{X}\mathbf{S}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}$ and all of its dominant parts $\mathbf{K}_{12}^{(p)} := \mathbf{X}^{(p)}\mathbf{S}^{(p)}\mathbf{V}^{(p)}$ (for some $p \leq n$) (based on Definition 7.4.1) have similar sparsity patterns.

In `MATLAB` there are different but equivalent manners for the filtering of entries from a matrix. However, most of them do not terminate or lead to out of memory errors even for the small case. Functions `vfilter` and `spfilter` contain information on manners which somehow do not lead to the desired result.

Explicit multiplication with factors $\mathbf{X}^{(p)}$, $\mathbf{S}^{(p)}$ and $\mathbf{V}^{(p)}$ for the multiplication with $\mathbf{x} \mapsto \mathbf{K}\mathbf{x}$ is likely to be the more efficient then the use of multiplication with $\mathbf{K}_{12}^{(p)}$.

9.6 A GSVD-based approximation of $\mathbf{K}_{12}$

In this section we analyse how the GSVD based approximation of $\mathbf{K}_{12}$ influences the solution of the static problem

$$\begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{0} & \mathbf{K}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \tag{9.6.1}$$

Based on the definition of $\mathbf{K}_{12}^{(p)}$ the approximation leads to system

$$\begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12}^{(p)} \\ \mathbf{0} & \mathbf{K}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1^{(p)} \\ \mathbf{y}_2^{(p)} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}. \tag{9.6.2}$$

We intend to estimate

$$\left\| \left(|x_i - y_i^{(p)}| / |x_i| \right)_i \right\|_\infty \tag{9.6.3}$$
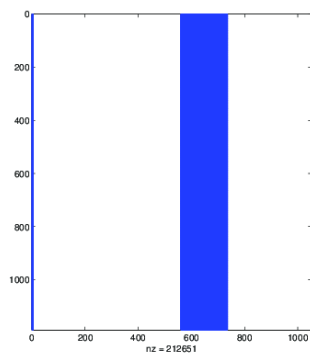
(a) $p = 5, \epsilon = 0$
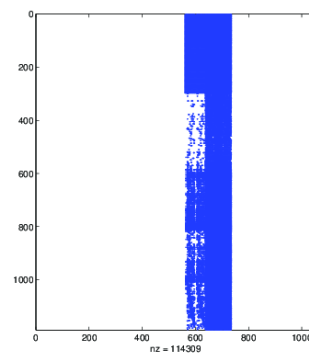
(b) $p = 5, \epsilon = 10^{-}14$
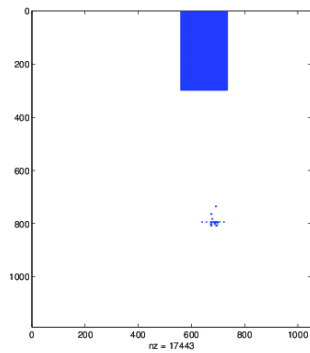
(c) $p = 5, \epsilon = 10^{-}12$
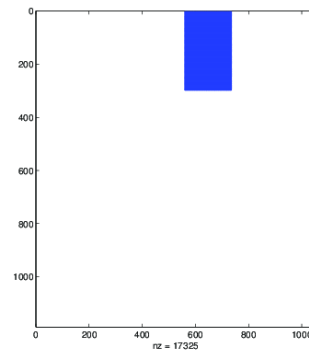
(d) $p = 5, \epsilon = 10^{-}10$

Figure 9.3: Entries of $\mathbf{K}_{12}^{(p)}$ greater than $\epsilon \cdot M(\mathbf{K}_{12}^{(p)})$, small case

Table 9.1: Relative errors due to use of the GSVD approximation

p	$\|\|x(i) - y^{(p)}(i)\|/\|x(i)\|\|_\infty$
1	3.750080647e-008
2	1.427208120e-007
3	1.119493657e-007
4	1.468582269e-007
5	1.500944068e-007

over the set of indices i for which x_i is non-zero (outside round-off region). To determine this set, we first solved (9.6.1) and made a log-plot of its sorted entries, shown in Figure 9.4. Based on this plot we decided to omit all entries smaller than 10^{-7} and ob-
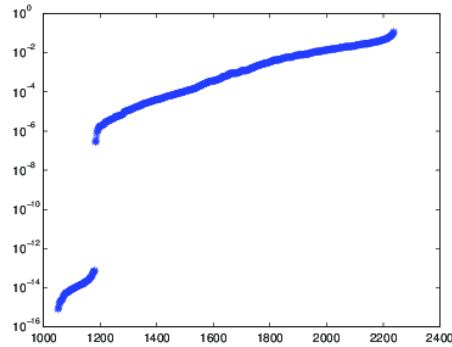
Figure 9.4: Entries of $\mathbf{x}$ in (9.6.1), sorted

tained the results in Table 9.1. The accuracy does not seem to be (very) sensitive to the amount of principal components used, which is due to the fact that the scaled $\mathbf{K}_{12}$ block is still of magnitude 10^5 smaller than the scaled diagonal blocks $\mathbf{K}_{11}$ and $\mathbf{K}_{22}$. However, Section 9.7 shows that different amounts of principal components do have a remarkable effect on the related transfer function.

9.7 The $K_{12}^{(p)}$ GSVD-approximation based transfer function

The aim is to determine a principal component analysis (*PCA*) based *rank-revealing* factorization $\mathbf{K}_{12} \doteq \mathbf{B}\mathbf{C}^\mathrm{T}$ where $\mathbf{B}$ and $\mathbf{C}$ are constructed with the use of the first $p \leq n$ *principal components*, based on the scaled versions of $\mathbf{K}$ (and if needed $\mathbf{M}$) as constructed above.

To the scaled matrix $\mathbf{K}$ (which depends on $\hat{w}$) we apply a GSVD to $\mathbf{K}_{11}^\mathrm{T}$ and $\mathbf{K}_{12}^\mathrm{T}$ such

that $\mathbf{K}_{11}^{\mathrm{T}} = \mathbf{UCX}^{\mathrm{T}}$ and $\mathbf{K}_{12}^{\mathrm{T}} = \mathbf{VSX}^{\mathrm{T}}$. Hence,

$$\mathbf{K}_{12} = \mathbf{XS}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}} = \underbrace{\mathbf{X}\sqrt{\mathbf{S}^{\mathrm{T}}}}_{\mathbf{B}}\underbrace{\sqrt{\mathbf{S}^{\mathrm{T}}}\mathbf{V}^{\mathrm{T}}}_{\mathbf{C}^{\mathrm{T}}}.$$

Figure 9.5 shows all of the diagonal values of the matrix S and Figure 9.6 shows the first 1000 of them. Next, for $p = 1, \ldots, 5$ we approximate $\mathbf{K}_{12}$ by the contribution of its p
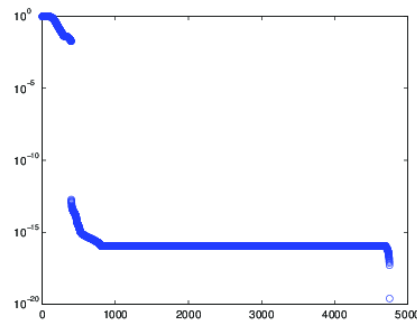
Figure 9.5: Diagonal elements of S.
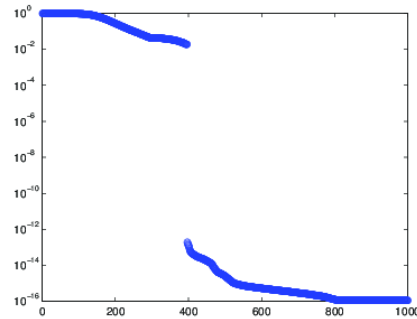
Figure 9.6: First 1000 diagonal elements of S.

most dominant modes

$$\mathbf{K}^{(p)} = (\mathbf{X}^{(p)}\sqrt{\mathbf{S}^{(p)}})(\sqrt{\mathbf{S}^{(p)}}\mathbf{V}^{(p)})$$

and plot the related transfer functions, together with the transfer function related to $\mathbf{K}_{12}$ (blue) in Figure 9.7 One can observe that the transfer function related to $\mathbf{K}^{(p)}$ closely approximates p peaks of the original transfer function (the one for $\mathbf{K}_{12}$).
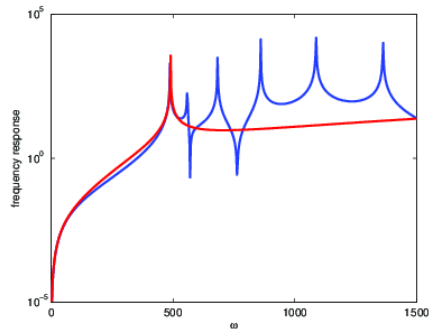
(a) rank 1 approx. of K_{12}
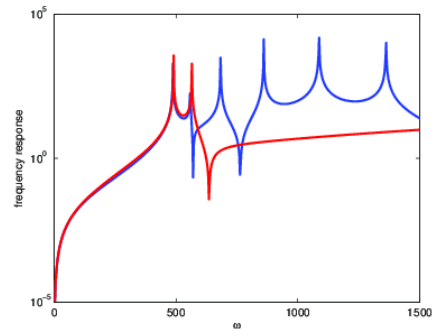
(b) rank 2 approx. of K_{12}
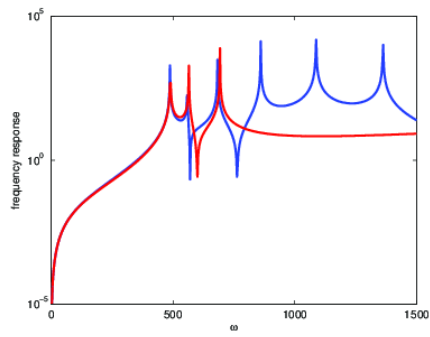
(c) rank 3 approx. of K_{12}
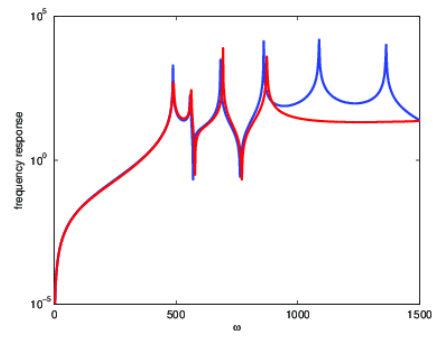
(d) rank 4 approx. of K_{12}
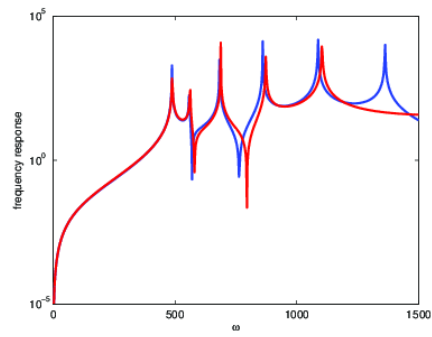
(e) rank 5 approx. of K_{12}

Figure 9.7: Low-rank approximations of block K_{12}.

9.8   The GSVD approximation of $M_{11}^{-1}K_{12}$

In fact, we need to apply the GSVD to $M_{11}^{-1}K_{12}$ rather than K_{12}. Fortunately, there is a straightforward relation between the GSVD of (K_{11}, K_{12}) and $(M_{11}^{-1}K_{11}, M_{11}^{-1}K_{12})$, as explained by Corollary 7.2.2. To see this, abbreviate $K := K_{12}$ and $M := M_{11}$ and observe that

$$
\begin{aligned}
K^T &= VSX^T \implies \\
K &= XS^TV^T \implies \\
M^{-1}K &= M^{-1}XS^TV^T \implies \\
M^{-1}K &= \underbrace{(M^{-1}X)\sqrt{S^T}}_{Y}\ \underbrace{\sqrt{S^T}V^T}_{Z}
\end{aligned}
$$

which leads to the principal component based approximation:

$$
M^{-1}K \doteq M^{-1}X^{(p)}S^{(p)}V^{(p)}.
$$

One first rewrites (9.2.4) to produce the term sI, for instance as follows:

$$
\begin{aligned}
H(w) &= c(K - w^2M)^{-1}b \implies \\
H(w) &= c(M^{-1}K - w^2I)M^{-1}b \implies \\
H(w) &= -c(w^2I - M^{-1}K)M^{-1}b.
\end{aligned}
\tag{9.8.1}
$$

Observe that the inverse of block-matrix M in (9.2.2) is

$$
M^{-1} = \begin{bmatrix} M_{11}^{-1} & 0 \\ -M_{22}^{-1}M_{21}M_{11}^{-1} & M_{22}^{-1} \end{bmatrix}
$$

whence

$$
M^{-1}K = \begin{bmatrix} M_{11}^{-1}K_{11} & M_{11}^{-1}K_{12} \\ -M_{22}^{-1}M_{21}M_{11}^{-1}K_{11} & -M_{22}^{-1}M_{21}M_{11}^{-1}K_{12} + M_{22}^{-1}K_{22} \end{bmatrix}
$$

Now, SBR applied to the first row of this system leads to the approximation

$$
M^{-1}K \doteq \begin{bmatrix} M_{11}^{-1}K_{11} & M_{11}^{-1}X^{(p)}S^{(p)}V^{(p)} \\ -M_{22}^{-1}M_{21}M_{11}^{-1}K_{11} & -M_{22}^{-1}M_{21}M_{11}^{-1}K_{12} + M_{22}^{-1}K_{22} \end{bmatrix}
$$

which shows that one can use the GSVD-based approximation

$$
H(w) \doteq c(K^{(p)} - w^2M)^{-1}b
$$

where

$$
K = \begin{bmatrix} K_{11} & X^{(p)}S^{(p)}V^{(p)} \\ 0 & K_{22} \end{bmatrix}.
$$

9.9 Conclusions

In this chapter, we applied the GSVD based low-rank approximation idea to the industrial case of an inkjet printhead. We proposed a way of scaling of a badly conditioned dynamical system. We explained the relations between the first and the second order formulation of the model. The recovery of the sparsity patterns in the components created by the GSVD was treated in more detail. Moreover, the stiffness matrix of the dynamical system, with the low-rank approximation of the coupling block, was studied in the context of solving a linear system and computing a transfer function, giving promising results. Finally, we investigated the GSVD of the product of two system sub-matrices, namely $\mathbf{M}_{11}^{-1}\mathbf{K}_{12}$.

Chapter 10

Conclusions and Recommendations

We proposed a new model order reduction technique for coupled systems. Our method, called the Separate Bases Reduction (SBR) algorithm, belongs to the family of block-structure preserving (BSP) reduction techniques based on the uncoupled formulation of the coupled problem. However, unlike other reduction approaches dealing with the separate sub-system representation, the SBR algorithm can be applied to a wide category of coupled systems, including strongly coupled systems and interconnected systems with many interconnections. This is due to the fact that for such cases we avoid a too fast growth of the reduction bases and related reduced-order model, as long as the coupling can be well approximated by a relatively small number of GSVD principal components. Examples of such strongly coupled systems are systems with an interface coupling, for instance systems describing interactions between a fluid and a solid wall, or systems which for instance describe an electromagnetic-structural coupling in an electronic device. Another advantage of the proposed technique is that it is computationally cheaper than the more common BSP reduction methods which deal with the coupled formulation of the system.

For the initial version of the SBR algorithm (without low-rank approximations of the couplings), we proved the moment matching property. The GSVD based approximation of the couplings only approximates the moments, but numerical experiments show that taking a sufficient number of dominant components still results in accurately approximated moments. What makes the SBR algorithm universal, is the fact, that it can be

applied even if the internal input and output matrices are not known explicitly. We show, that having at our disposal only the coupled system's matrices, external inputs and outputs, and the dimensions of the sub-systems, we are able to create appropriate Krylov subspaces for each sub-system. This property of the reduction method is desirable when dealing with industrial problems for which the separate sub-systems' information may not be available.

The SBR method has been designed keeping in mind the practical use in an industrial environment. It is fairly straightforward to adapt existing software modules and make them suitable for application of SBR. This is certainly not the case for the BSP type methods. Although the reduced-order models obtained by application of the BSP methods frequently show a bit better approximation accuracy, the SBR algorithm is much more beneficial from the point of view of the computational time. This property is especially valuable in case of large industrial applications.

Recommendations for future work

The application of the low-rank approximation based SBR algorithm to several test cases resulted in reduced-order models that accurately approximated the original systems. To make the SBR algorithm more efficient and reliable, some topics should be studied further.

First of all, our GVSD-based reduction version only approximately matches the moments of the reduced system. An a priori approximation error estimate would be helpful to a priori determine the amount of GSVD based components.

Furthermore, it should be examined how to apply the SBR method when the system matrix $\mathbf{E}$ is not block-diagonal.

Badly scaled coupled systems strongly influence the moments' numerical accuracy. The relation between the GSVD of the scaled and unscaled system should be examined.

Bibliography

[1] M.A. Akgün. A new family of mode-superposition methods for response calculations. *Journal of Sound & Vibration*, 167(2):289–302, 1993.

[2] E. Anderson, Z. Bai, and C. Bischof et al. *LAPACK user's guide, Third Edition*. SIAM, Philadelphia, 1999.

[3] A.C. Antoulas. *Approximation of large-scale dynamical systems*. SIAM, Philadelphia, USA, 2005.

[4] A.C. Antoulas. An overview of approximation methods for large-scale dynamical systems. *Annual Reviews in Control*, 29(2):181–190, 2005.

[5] Z. Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43(1-2):9–44, 2002.

[6] Z. Bai, R. Li, and Y. Su. A unified Krylov projection framework for structure-preserving model reduction. In W.H.A. Schilders, H.A. van der Vorst, and J. Rommes, editors, *Model Order Reduction: Theory, Research Aspects and Applications*, volume 13 of *Mathematics in Industry*, pages 75–93. Springer-Verlag Berlin Heidelberg, 2008.

[7] G. Berkooz, P. Holmes, and J.L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25:539–575, 1993.

[8] B. Besselink, A. Lutowska, U. Tabak, N. van de Wouw, M.E. Hochstenbach, H. Nijmeijer, D.J. Rixen, and W.H.A. Schilders. A comparison of model reduction techniques from structural dynamics, numerical mathematics and systems and control. *to be submitted*.

[9] R.H. Bisseling. *Parallel Scientific Computation, A Structured Approach using BSP and MPI*. Oxford Scholarship Online, 2007.

[10] S. Chandrasekaran and I. C. F. Ipsen. On rank-revealing factorisations. *SIAM Journal on Matrix Analysis and Applications*, 15:592–622, 1994.

[11] R.R. Craig and A. Kurdila. *Fundamentals of structural dynamics*. John Wiley & Sons, Hoboken, USA, 2006.

[12] R.R. Craig Jr. Coupling of substructures for dynamic analysis: an overview. In *Proceedings of the 41st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Atlanta, USA*, 2000.

[13] R.R. Craig Jr. and M.C.C. Bampton. Coupling of substructures for dynamic analyses. *AIAA Journal*, 6(7):1313–1319, 1968.

[14] D. de Klerk, D.J. Rixen, and S.N. Voormeeren. General framework for dynamics substructuring: history, review, and classification of techniques. *AIAA Journal*, 46(5):1169–1181, 2008.

[15] U. Desai and D. Pal. A transformation approach to stochastic model reduction. *IEEE Transactions on Automatic Control*, AC-29(12):1097–1100, 1984.

[16] J.M. Dickens, J.M. Nakagawa, and M.J. Wittbrodt. A critique of mode acceleration and modal truncation augmentation methods for modal response analysis. *Computers & Structures*, 62(6):985–998, 1997.

[17] J.M. Dickens and A. Stroeve. Modal truncation vectors for reduced dynamic substructure models. In *Proceedings of the 41st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Atlanta, USA*, 2000.

[18] J.M. Dickens and E.L. Wilson. Numerical methods for dynamic substructure analysis. Technical Report UBC/EERC-80/120, Earthquake Eng. Res. Center, University of California, Berkeley, 1980.

[19] A. Doris, N. van de Wouw, W.P.M.H. Heemels, and H. Nijmeijer. A disturbance attenuation approach for continuous piecewise affine systems: Control design and experiments. *J. Dynamic Systems, Measurement and Control*, 132(4):044502–1 to 044502–7, 2010.

[20] D.F. Enns. Model reduction with balanced realizations: an error bound and a frequency weighted generalization. In *Proceedings of the 23rd IEEE Conference on Decision and Control, Las Vegas, USA*, pages 127–132, 1984.

[21] P. Feldmann and R.W. Freund. Efficient linear circuit analysis by Pade approximation via the Lanczos process. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 14(5):639–649, 1995.

[22] J. Fernández Villena, W.H.A. Schilders, and L. Miguel Silveira. Order reduction techniques for coupled multi-domain electromagnetic based models. *CASA report*, 2008.

[23] J. Fernández Villena, W.H.A. Schilders, and L. Miguel Silveira. Block oriented model order reduction of interconnected systems. *CASA report*, 2009.

[24] K. Fernando and H. Nicholson. Singular perturbational model reduction of balanced systems. *IEEE Transactions on Automatic Control*, AC-27(2):466–468, 1982.

[25] R. W. Freund. Sprim: Structure-preserving reduced-order interconnect macromodelling. In *Proceedings of the 2004 IEEE/ACM International conference on Computer-aided design*, pages 80–87. IEEE Computer Society, Washington, DC, USA, 2004.

[26] R.W. Freund. Krylov-subspace methods for reduced-order modeling in circuit simulation. *Journal of Computational and Applied Mathematics*, 123(1-2):395–421, 2000.

[27] R.W. Freund. Model reduction methods based on Krylov subspaces. *Acta Numerica*, 12:267–319, 2003.

[28] R.W. Freund. SPRIM: structure-preserving reduced-order interconnect macromodeling. In *Proceedings of the IEEE/ACM International Conference on Computer Aided Design*, pages 80–87, 2004.

[29] R.W. Freund and P. Feldmann. Reduced-order modeling of large passive linear circuits by means of the SyPVL algorithm. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, San Jose, USA*, pages 280–287, 1996.

[30] K.A. Gallivan, E. Grimme, and P.M. Van Dooren. Model reduction of large-scale systems: rational Krylov versus balancing techniques. In H. Bulgak and C. Zenger, editors, *Error Control and Adaptivity in Scientific Computing*, pages 177–190. Kluwer Academic, 1999.

[31] M. Géradin and D. Rixen. *Mechanical vibrations: theory and application to structural dynamics*. John Wiley & Sons, 1997.

[32] M. Géradin and D.J. Rixen. *Mechanical Vibrations and Structural Dynamics, 2nd edition*. John Wiley and Sons, 1997.

[33] H.A. van der Vorst G.L.G. Sleijpen and D.R. Fokkema. Bicgstab(l) and other hybrid bi-cg methods. *Numerical Algorithms*, 7:75–109, 1994.

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

[35] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. third edition, Johns Hopkins University Press, Baltimore, 1996.

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

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

[38] S. Gugercin and A.C. Antoulas. A survey of model reduction by balanced truncation and some new results. *International Journal of Control*, 77(8):748–766, 2004.

[39] L. He H. Yu and S.X.D. Tan. Block structure preserving model order reduction. *BMAS – IEEE Behavioral Modeling and Simulation Workshop*, 2005.

[40] H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *SIAM Review*, 23(1):53–60, 1981.

[41] P.J. Heres. *Robust and efficient Krylov subspace methods for model order reduction*. PhD thesis, Eindhoven University of Technology, the Netherlands, 2005.

[42] P.J. Heres, D. Deschrijver, W.H.A. Schilders, and T. Dhaene. Combining krylov subspace methods and identification-based methods for model order reduction. *International Journal of Numerical Modelling : Electronic Networks, Devices and Fields*, 20(6):271–282, 2007.

[43] W.C. Hurty. Vibrations of structural systems by component mode synthesis. *ASCE Journal of the Engineering Dynamics Division*, 86(EM4):51–69, 1960.

[44] W.C. Hurty. Dynamic analysis of structural systems using component modes. *AIAA Journal*, 3(4):678–685, 1965.

[45] R. Ionutiu. *Model Order Reduction for Multi-terminals Systems with Applications to Circuit Simulation*. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2011.

[46] K.Gallivan and E.Grimme. A rational lanczos algorithm for model reduction. *Numerical Algorithms*, 12:33–63, 1996.

[47] S.W. Kim, B.D.O. Anderson, and A.G. Madievski. Error bound for transfer function order reduction using frequency weighted balanced truncation. *Systems & Control Letters*, 24(3):183–192, 1995.

[48] Y.T. Leung. Fast response method for undamped structures. *Engineering Structures*, 5:141–149, 1983.

[49] Y. Liu and B.D.O. Anderson. Singular perturbation approximation of balanced systems. *International Journal of Control*, 50(4):1379–1405, 1989.

[50] Z.-S. Liu, S.-H. Chen, and Y.-Q. Zhao. An accurate modal method for computing response to harmonic excitation. *International Journal of Analytical and Experimental Modal Analysis*, 9(1):1–13, 1994.

[51] B. Lohmann and B. Salimbahrami. Introduction to Krylov subspace methods in model order reduction. In B. Lohmann and A. Gräser, editors, *Methods and Applications in Automation*, pages 1–13. Shaker Verlag, Aachen, 2003.

[52] A. Lutowska, M.E. Hochstenbach, and W.H.A. Schilders. A low-rank gsvd-based model order reduction method for coupled systems. *in preparation*.

[53] A. Lutowska, M.E. Hochstenbach, and W.H.A. Schilders. Model order reduction for complex high-tech systems. In *Scientific Computing in Electrical Engineering, SCEE 2010, Mathematics in Industry*, to appear.

[54] R.H. MacNeal. A hybrid method of component mode synthesis. *Computers & Structures*, 1(4):581–601, 1971.

[55] Matlab. *Matlab online documentation*. `www.mathworks.com`, The Math Works Inc., 2011.

[56] D.G. Meyer. A fractional approach to model reduction. In *Proceedings of the American Control Conference, Atlanta, USA*, pages 1041–1047, 1988.

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

[58] C.T. Mullis and R.A. Roberts. Synthesis of minimum roundoff noise fixed point digital filters. *IEEE Transactions on Circuits and Systems*, CAS-23(9):551–562, 1976.

[59] R. Ober. Balanced parametrization of classes of linear systems. *SIAM Journal on Control and Optimization*, 29(6):1251–1287, 1991.

[60] R. Ober and D. McFarlane. Balanced canonical forms for minimal systems: A normalized coprime factor approach. *Linear Algebra and its Applications*, 122-124:23–64, 1989.

[61] A. Odabasioglu, M. Celik, and L.T. Pileggi. PRIMA: passive reduced-order interconnect macromodeling algorithm. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 17(8):645–654, 1998.

[62] M. Papadrakakis. *Solving large-scale problems in mechanics*. John Wiley & Sons, New York, USA, 1993.

[63] A.V. Pavlov, N. van de Wouw, and H. Nijmeijer. *Uniform Output Regulation of Nonlinear Systems*. Birkhauser, 2005.

[64] L. Pernebo and L.M. Silverman. Model reduction via balanced state space representations. *IEEE Transactions on Automatic Control*, AC-27(2):382–387, 1982.

[65] M. Petyt. *Introduction to finite element vibration analysis*. Cambridge University Press, Cambridge, United Kingdom, 1990.

[66] L.T. Pillage and R.A. Rohrer. Asymptotic waveform evaluation for timing analysis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 9(4):352–366, 1990.

[67] J.W.S. Rayleigh. *The theory of sound*. Dover Publications, New York, 1945.

[68] T. Reis and T. Stykel. A survey on model reduction of coupled systems. In W.H.A. Schilders, H.A. van der Vorst, and J. Rommes, editors, *Model Order Reduction: Theory, Research Aspects and Applications*, volume 13 of *Mathematics in Industry*, pages 133–155. Springer-Verlag Berlin Heidelberg, 2008.

[69] 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.

[70] D. Rixen. High order static correction modes for component mode synthesis. In *Fifth World Congress on Computational Mechanics*, Vienna, Austria, 2002.

[71] D. Rixen. A Lanczos procedure for efficient mode superposition in dynamic analysis. In *Proceedings of the 43st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Denver, USA*, 2002.

[72] D.J. Rixen. Generalized mode acceleration methods and modal truncation augmentation. In *Proceedings of the 42rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Seattle, USA*, 2001.

[73] D.J. Rixen. Dual Craig-Bampton with enrichment to avoid spurious modes. In *IMAC-XXVII: International Modal Analysis Conference, Orlando, FL*, Bethel, CT, 2009. Society for Experimental Mechanics.

[74] V. Rochus, D.J. Rixen, and J.-C. Golinval. Electrostatic coupling of mems structures: transient simulations and dynamic pull-in. *Nonlinear Analysis*, 63(5-7):1619–1633, 2005.

[75] J. Rommes. *Methods for eigenvalue problems with applications in model order reduction*. PhD thesis, 2007.

[76] S. Rubin. Improved component-mode representation for structural dynamic analysis. *AIAA Journal*, 13(8):995–1006, 1975.

[77] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, Philadelphia, USA, 2003.

[78] Y. Saad. *Iterative methods for sparse linear systems – Second Edition*. SIAM, 2003.

[79] Y. Saad and M. H. Schulz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7:856–869, 1986.

[80] B. Salimbahrami and B. Lohmann. Krylov subspace methods in linear model order reduction: Introduction and invariance properties. *Sci. Rep.. Inst. of Automation*, 2002.

[81] J.M.A. Scherpen. Balancing for nonlinear systems. *Systems & Control Letters*, 21:143–153, 1993.

[82] W.H.A. Schilders, H.A. van der Vorst, and J. Rommes. *Model Order Reduction: Theory, Research Aspects and Applications*, volume 13 of *Mathematics in Industry*. Springer-Verlag Berlin Heidelberg, 2008.

[83] L. Sirovich. Turbulence and the dynamics of coherent structures, part I: Coherent structures. *Quarterly of Applied Mathematics*, 45(3):561–571, 1987.

[84] S. Skogestad and I. Postlethwaite. *Multivariable feedback control, Analysis and Design*. John Wiley and Sons, Philadelphia, 2005.

[85] P. Sonneveld. CGS, a fast Lanczos-type solver for nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 10:36–52, 1989.

[86] B. D. Sutton. Computing the complete sc decomposition. 50:33–65, 2009.

[87] M.V. Ugryumova. *Model Order Reduction for Multi-terminals Systems with Applications to Circuit Simulation*. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2000.

[88] N. van de Wouw, A. de Kraker, D.H. van Campen, and H. Nijmeijer. Non-linear dynamics of a stochastically excited beam system with impact. *Int. J. Non-Linear Mech.*, 38(5):767–779, 2003.

[89] A. Vandendorpe and P. Van Dooren. Model reduction of interconnected systems. In W.H.A. Schilders, H.A. van der Vorst, and J. Rommes, editors, *Model Order Reduction: Theory, Research Aspects and Applications*, volume 13 of *Mathematics in Industry*, pages 305–321. Springer-Verlag Berlin Heidelberg, 2008.

[90] J. Wang. Computing the csd and gsvd, 2004.

[91] P. Wesseling. *An Introduction to Multigrid Methods*. John Wiley & Sons, 1992.

[92] E.L. Wilson, M.-W. Yuan, and J.M. Dickens. Dynamic analysis by direct superposition of Ritz vectors. *Earthquake Engineering & Structural Dynamics*, 10(6):813–821, 1982.

[93] Y.Q. Zhao, S.-H. Chen, S. Chai, , and Q.-W. Qu. An improved modal truncation method for responses to harmonic excitation. *Computers & Structures*, 80(1):99–103, 2002.

[94] K. Zhou, J.C. Doyle, and K. Glover. *Robust and optimal control*. Prentice Hall, Upper Saddle River, USA, 1996.

Index

Summary

Model Order Reduction for Coupled Systems using Low-rank Approximations

This thesis presents the result of a study on model order reduction (MOR) methods, that can be applied to coupled systems. The goal of the research was to develop reduction techniques, that preserve special properties of coupled or interconnected system, e.g. block-structure of the underlying matrices. On the other hand, the new techniques should also be able to benefit from the knowledge, that the system they are applied to, consists of two (or more) sub-systems or describes some phenomena in different physical domains. As a result of this study, two main approaches are proposed. Their general description is given in the following paragraphs.

First, the Separate Bases Reduction (SBR) algorithm is developed, which is a projection based MOR technique that uses Krylov subspaces as reduction bases. The novelty of this method is that SBR algorithm, unlike standard reduction methods designed for coupled problems, uses an uncoupled formulation of the system. In other words, an appropriate Krylov subspace is built for each of the sub-system constituting the interconnected system. As a result, the computational costs of application of the SBR algorithm, with respect to time and memory storage needed for calculations, are lower than in case of MOR methods that use the coupled formulation of the system. Moreover, the block-diagonal form of the reduction matrices allows for preservation of the block-structure of the system matrices and keeps the sub-systems (or different physical domains) still recognizable in the reduced-order model. The SBR algorithm was successfully applied to a few test cases, resulting in the reduced systems that approximate the original ones with accuracy comparable to the accuracy of systems reduced by means of other block-structure preserving MOR methods.

The second topic of the research focuses on the couplings between the sub-systems. Here, the off-diagonal blocks of the system matrices that correspond to the couplings, are approximated by matrices of lower rank. As a main tool, a generalized singular value decomposition (GSVD) is used, which is used to determine the most important components of a coupling block with respect to one of the sub-systems. Although this method

does not reduce the dimension of the considered problem, it gives benefits if used before application of a MOR technique. First of all, the use of low-rank approximations of the coupling blocks can decrease the computational costs of the the the Krylov subspaces construction needed for reduction. If the couplings can be approximated by sufficiently low-rank blocks, the necessary matrix inverse calculation can be performed cheaper, by application of the Sherman-Morrison-Woodbury formula. Moreover, the undesired growth of the reduction bases, in case of use of the SBR algorithm to sub-systems with many inputs (outputs), can be lowered by using only dominant components of the input (output) space. The conducted experiments showed, that for some cases, the number of the components used to define the couplings can be significantly reduced.

Samenvatting

Model orde reductiemethoden voor gekoppelde systemen met behulp van lage-rang benaderingen

Dit proefschrift presenteert de resultaten van een onderzoek naar model orde reductie-(MOR) methoden die kunnen worden toegepast op gekoppelde of met elkaar verbonden systemen. Het onderzoek richtte zich op technieken die speciale eigenschappen van zulke systemen behouden, zoals de blokstructuur van de onderliggende matrices. Anderzijds dienen de nieuwe technieken te profiteren van de kennis dat ze worden toegepast op een systeem waarvan de subsystemen verschijnselen in verschillende fysische domeinen beschrijven. Het proefschrift suggereert twee benaderingen die worden beschreven in de volgende alinea's.

Ten eerste is het gescheiden bases reductie (SBR) algoritme ontwikkeld, een op projectie gebaseerde MOR techniek die Krylov deelruimten gebruikt als reductiebasis. Het nieuwe van deze methode is dat het SBR algoritme, in tegenstelling tot standaard reductiemethoden voor gekoppelde systemen, gebruikt maakt van de ontkoppelde subsystemen. Met andere woorden, SBR bouwt een Krylov deelruimte voor elk van de subsystemen onafhankelijk, dusdanig dat men het gekoppelde probleem kan oplossen. Hierdoor behoeft SBR minder rekentijd en computergeheugen dan standaard MOR methoden voor gekoppelde systemen. Bovendien zorgt de blokdiagonale vorm van de reductiematrix voor het behoud van de blokstructuur van het systeem. Daardoor blijven de subsystemen (gerelateerd aan de verschillende fysische domeinen en groetheden) herkenbaar in het gereduceerde model. Het SBR algoritme is met succes toegepast op enkele test-cases. De resulterende gereduceerde systemen hebben een nauwkeurigheid die vergelijkbaar is met die van andere blokstructuurbehoudende MOR methoden.

Het tweede aandachtspunt van het onderzoek was de koppelingen tussen de subsystemen, gerepresenteerd door de matrixblokken buiten de hoofddiagonaal. Deze matrixblokken worden benaderd met lage-rang matrices afkomstig uit een gegeneraliseerde singuliere waarde decompositie (GSVD). Deze decompositie bepaalt de belangrijkste componenten van een koppelingsblok ten opzichte van de subsystemen die het koppelt. Deze methode reduceert de afmeting van het beschouwde systeem niet maar leidt tot

voordelen wanneer gebruikt vóór het toepassen van een MOR methode. In de eerste plaats kan het gebruik van de lage-rang benadering leiden tot een vermindering van de rekentijd en geheugenopslag. Als de koppelingen kunnen worden benaderd met blokken van voldoende lage rang dan kan de benodigde matrix inverse door toepassing van de Sherman-Morrison-Woodbury formule goedkoper worden uitgerekend. Bovendien kan de ongewenste groei van de reductiebases die voorkomt bij het toepassen van het SBR algoritme op subsystemen met veel ingangen of uitgangen worden geremd. Experimenten hebben aangetoond dat in sommige gevallen de koppelingsblokken kunnen worden benaderd met blokken van zeer lage rang.

Acknowledgements

Please, let me acknowledge some of the people who made my period as a PhD student at Eindhoven University of Technology so special and enjoyable.

First of all, I would like to express my gratitude to my promotor, Prof. Wil Schilders, for giving me the opportunity to do my PhD within the CASA group and for his guidance during the past years. I would also like to thank him for making it possible that I could participate in many scientific events, where I met many new people from my field of research, and in addition spent memorable time with them, my friends and colleagues. I am also thankful to my copromotor, dr. Michiel Hochstenbach, for sharing his knowledge.

My PhD project was part of a joint STW project, in which, next to CASA group, also two other groups participated. I would like to mention the group of Prof. Henk Nijmeijer, Nathan van de Wouw, and Bart Besselink (TU Eindhoven) as well as the group of Prof. Daniel Rixen and Umut Tabak (TU Delft). I thank you all for your enthusiasm, interest and inspiring discussions. Our meetings were always fruitful and pleasant. My special thanks go to Bart, for his patience and engagement while introducing me to the engineering world and for the whole set of the "Bart's beam" examples that were extremely helpful for trying out my ideas. I also would like to thank Umut for all his effort to create a "real-life" coupled problem for my research.

I would also like to thank my committee members, who, next to Prof. Henk Nijmeijer and Prof. Daniel Rixen, are Prof. Rob Bisseling, Prof. Michael Günther and Dr. Joost Rommes.

Further, I would like to express my gratitude to Fabio Freschi from the University of Torino who contributed the thermo-mechanical system that served as a numerical example in Chapter 7 of this thesis. Also thanks to Herman Wijshoff for the explanation of the OCE print-head example described in Chapter 9.

Curriculum Vitae

Agnieszka Lutowska was born on March 20, 1983 in Zielona Góra, Poland. In 2002, she started to study Industrial Mathematics at University of Zielona Góra, from which she obtained her Master degree in 2007. Years 2006-2007 she spent as a Socrates/Erasmus student and a guest student at the Department of Mathematics at University of Kaiserslautern, Germany. During this period, she also held a student practice at the Fraunhofer ITWM in Kaiserslautern. The research conducted by her at ITWM was described in her Master thesis.

In October 2007, she started her PhD project at the Department of Mathematics and Computer Science at the Eindhoven University of Technology, the Netherlands. The results of her PhD study are presented in this thesis.