A Critical Exposition of Model Order Reduction Techniques: Application to a Slewing Flexible Beam

Stanislao Patalano, Alessandro Mango Furnari, Ferdinando Vitolo, Jean-Luc Dion, Régis Plateaux, Frank Renaud

A Critical Exposition of Model Order Reduction Techniques: Application to a Slewing Flexible Beam

Stanislao Patalano[1] · Alessandro Mango Furnari[1] · Ferdinando Vitolo[1] · Jean-Luc Dion[2] · Regis Plateaux[2] · Frank Renaud[2]

Abstract

Complexity of dynamical systems are increasing more and more as well as their mathematical models. At the same time, simulation of system behaviour assumes a key role to assure the fulfillment of requirements as performances, quality, safety, and robustness. Therefore, due to model complexity, it is often very complex to assess a system behaviour but a reduction of model complexity could enhance the simulation aimed to specific characteristics of the system. Several useful model order reduction (MOR) techniques exist but each of them is often powerful for specific applications. This review paper deals with MOR by critically comparing the most popular MOR techniques from the fields of structural dynamics, numerical mathematics and systems and control. In particular, after different reduction techniques have been presented, a table summarizing their most important features is proposed, for comparison purpose. The motivation for such comparison stems from the fact that the insight obtained by the comparison allows to make a motivated choice for a particular model reduction technique, on the basis of the desired properties retained in the reduced model. Particular attention is paid on reduction techniques from the area of structural dynamics. Finally, the differences among some of the presented reduction techniques are illustrated, on a quantitative level, by means of their application to the case of a slewing flexible beam. In particular, in the application of the different reduction techniques, a consistent-mass finite element model, with only translational degrees of freedom, is employed as beam full model.

1 Introduction

In the last decades, simulation of the dynamical system has become a key activity for improving system characteristics as performances, quality, safety, robustness, etc. Simulation results are affected by the accuracy of mathematical model which tries to give a whole representation of the real system behaviour. The whole representation is becoming an increasingly difficult and time-consuming activity due to a series of technological improvements. Such improvements, in fact, imply real-time big data analysis [1], machine learning [2], high speed communication and their integrations, making systems more and more complex as well as their mathematical models.

In such a context, mathematical model often needs to be reduced (DoF reduction) by using technique as MOR. There are several useful MOR techniques but each of them is generally highly performing for a specific application. Selection of the MOR technique is a crucial activity which affects the reduced model and its results. The first step in all MOR procedures is the selection of the most suitable techniques

✉ Ferdinando Vitolo
ferdinando.vitolo@unina.it

Stanislao Patalano
stanislao.patalano@unina.it

Alessandro Mango Furnari
a.mangofurnari@gmail.com

Jean-Luc Dion
jean-luc.dion@supmeca.fr

Regis Plateaux
regis.plateauxn@supmeca.fr

Frank Renaud
frank.renaud@supmeca.fr

[1] Department of Industrial Engineering, University of Naples Federico II, P.le V. Tecchio, 80, 80125 Naples, Italy

[2] Supméca - Institut Supérieur de Mécanique de Paris, Quartz, EA7393, 3 rue Fernand Hainaut, 93407 Saint-Ouen Cedex, France

for a given application. As there are still no tools to extricate among techniques and allowing univocal choice, the present work aims at providing some guidelines to assist analysts in this step. Here particular attention will be paid on reduction techniques from structural dynamics. Nevertheless, the discussion will be extended to reduction techniques from numerical mathematics and systems and control.

In order to compare different reduction techniques, several features which may be more or less present in each reduction technique and which have been considered the most important in the area of structural dynamics are discussed. The first feature concerns the objective of the approximation. Namely, the reduction procedure could lead to a reduced model which preserves the input–output behaviour or the global behaviour of the full model. Generally, in the area of numerical mathematics and systems and control theory a reduced model which approximates the input–output behaviour of the original system is sought, whereas an approximation of the global dynamics is sought in the structural dynamics community. Nevertheless, in some applications, like active vibration control, an input–output model of a mechanical system could be of interest.

Other important features concern the frequency range in which the accuracy of the reduced model is high and the preservation of the static response of the full model. As matter of fact, depending on the frequency contents of the input signal, the system response could be of interest only in some range of the frequency domain. This is why reduction techniques from structural dynamics and numerical mathematics are generally frequency domain based (Fourier or Laplace domain based) techniques. Nevertheless, it is noted that in balanced truncation reduction (from systems and control) the behaviour in the frequency domain does not form the basis of the 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.

In some fields, like experimental modal analysis, it is very important that the coordinates of the reduced model are measurable coordinates. This is only ensured when the reduced coordinates are physical coordinates. Therefore, the nature of the reduced model coordinates is considered as another important feature of the reduction procedure. In particular, the reduced model coordinates can be: generalized coordinates, physical coordinates, hybrid coordinates or state coordinate.

It is known that, the main goal of model order reduction is to obtain, with the least computational burden, a reduced model which is as accurate as possible in the interested frequency range and whose order is as low as possible. For this reason, the accuracy of the reduced model and the computational burden associated with the reduction procedure are considered as other two important features of the reduction procedure. Here the accuracy of a reduced model is measured by means of the width of the frequency range in which the reduced model approximates the behaviour of the full model. Whereas, the computational burden represents the quantification of the difficulty of the reduced model computation in terms of computer resources required, such as computational time or amount of memory. The importance of the computational burden is evident if we consider that, in the area of structural dynamics, finite element procedures generally lead to models of orders up to $\mathcal{O}(10^7)$. Similarly, in the area of numerical mathematics, large-scale electrical circuits lead to models of order up to $\mathcal{O}(10^7)$. For these applications, it is clear as the need for numerically efficient model reduction procedures arises.

Another important feature is the level of automation of the reduction procedure, once a requirement on the quality of the reduced model is specified. This information is extremely important since, the less the reduction procedure is automatic, the more the accuracy of the reduced model depends on the analyst's experience.

2 Model Order Reduction in Different Fields

In this section, the most significant reduction techniques from three different fields are first presented and then compared. In particular, in Sect. 2.1 will be discussed three families of reduction techniques from structural dynamics; they are: generalized coordinate reduction, physical coordinate reduction and hybrid coordinate reduction. In Sects. 2.2 and 2.3 will be discussed the balanced truncation method, from systems and control, and the Krylov subspace projection method, from numerical mathematics, respectively. It is noted that, these methods are to be considered as the basic algorithms on which many other methods are built. In conclusion, in Sect. 2.4, a qualitative comparison among the aforementioned reduction techniques is carried out.

2.1 Structural Dynamics

In linear structural dynamics, the equations of motion of a mechanical system are written as a set of second-order linear ordinary differential equations, namely,

$$\mathbf{M}\,\ddot{\mathbf{z}}(t) + \mathbf{V}\,\dot{\mathbf{z}}(t) + \mathbf{K}\,\mathbf{z}(t) = \mathbf{f}(t) \tag{1}$$

where, in MOR context, the vectors $\mathbf{z}(t)$ and $\mathbf{f}(t) \in \mathbb{R}^n$ are referred to as *full order coordinates* and full equivalent force vectors; the matrices $\mathbf{M}$, $\mathbf{V}$ and $\mathbf{K} \in \mathbb{R}^{n\times n}$ are, respectively, the mass, damping, and stiffness matrices of the *full order model*.

Model order reduction technique is usually introduced to reduce the size of the full model and leads to a reduced order model. Many model reduction schemes involve the form of coordinate transformation

$$\mathbf{z}(t) = \mathbf{T}\,\mathbf{q}(t) \tag{2}$$

where $\mathbf{T} \in \mathbb{R}^{n\times m}$ is the *coordinate transformation matrix* and $\mathbf{q}(t) \in \mathbb{R}^m$ is the *reduced order coordinates* vector. Since the transformation matrix $\mathbf{T}$ is generally time-invariant, the differentiation of the previous equation with respect to time gives

$$\dot{\mathbf{z}}(t) = \mathbf{T}\,\dot{\mathbf{q}}(t), \quad \ddot{\mathbf{z}}(t) = \mathbf{T}\,\ddot{\mathbf{q}}(t)$$

Introducing these equations into the dynamic equations of equilibrium and premultiplying both sides by the transpose of transformation matrix $\mathbf{T}$, lead to the reduced dynamic equations of equilibrium

$$\mathbf{M}_R\,\ddot{\mathbf{q}}(t) + \mathbf{V}_R\,\dot{\mathbf{q}}(t) + \mathbf{K}_R\,\mathbf{q}(t) = \mathbf{f}_R(t) \tag{3}$$

The matrices $\mathbf{M}_R$, $\mathbf{V}_R$ and $\mathbf{K}_R \in \mathbb{R}^{m\times m}$ are, respectively, the mass, damping, and stiffness matrices of the *reduced order model*; the vector $\mathbf{f}_R(t)$ is the reduced equivalent force vector. Clearly, they are defined as

$$\mathbf{M}_R = \mathbf{T}^T\mathbf{M}\,\mathbf{T}, \quad \mathbf{V}_R = \mathbf{T}^T\mathbf{V}\,\mathbf{T},$$
$$\mathbf{K}_R = \mathbf{T}^T\mathbf{K}\,\mathbf{T}, \quad \mathbf{f}_R(t) = \mathbf{T}^T\mathbf{f}(t)$$

Although the size m of the reduced model is much smaller than the size n of the full model, that is $m \ll n$, the dynamic characteristics of the full model, within the interested frequency range, may be retained in the reduced model.

Based on the type of coordinates retained as the reduced order coordinates, the existing MOR techniques, in the area of structural dynamics, fall into three basic categories: generalized coordinate reduction, physical coordinate reduction and hybrid coordinate reduction. Each category may include several subcategories [3].

2.1.1 Generalized Coordinate Reduction

In the context of this work, all the coordinates which are measurable coordinates, like position and angular coordinates for example, are referred to as *physical coordinates*. All those coordinates which are not physical coordinates are generally referred to as *generalized coordinates*. The modal coordinates and the Ritz coordinates are two types of frequently used generalized coordinates.

Modal coordinate reduction is one of the most widely used model reduction techniques for linear and weakly nonlinear structural dynamic systems. It is actually a combination of mode superposition and modal truncation methods. Mode superposition method transforms the full, large size of the finite element model from physical coordinates in physical space to modal coordinates in modal space using the eigenvector matrix of this system. The modal truncation scheme removes those modal coordinates or modes that have unimportant contributions to the system responses. Generally, only a few modes have a significant effect on the system dynamic characteristics within the frequency range of interest. The number of modal coordinates retained is, thus, much smaller than that of physical coordinates. The coordinates transformation takes the form

$$\mathbf{z}(t) = \mathbf{\Phi}_m\,\boldsymbol{\eta}_m(t) \tag{4}$$

where the eigenvector matrix $\mathbf{\Phi}_m \in \mathbb{R}^{n\times m}$ is defined by the m eigenvectors of the full model corresponding to the frequencies of interest, whereas the vector $\boldsymbol{\eta}_m(t)$ contains m modal coordinates. Generally, we are interested in the dynamic characteristics of the model at its low-frequency range. Thus, if the m lower frequencies are of interest, the eigenvector matrix takes the form

$$\mathbf{\Phi}_m = \left[\boldsymbol{\varphi}_1\,\boldsymbol{\varphi}_2\,\ldots\,\boldsymbol{\varphi}_m\right] \tag{5}$$

where $\boldsymbol{\varphi}_1\,\boldsymbol{\varphi}_2\,\ldots\,\boldsymbol{\varphi}_m$ are the first m eigenvectors of the full model. If the eigenvector matrix $\mathbf{\Phi}_m$ is normalized with respect to the mass matrix, substituting Eq. (4) in Eq. (1) and premultiplying the latter by the transpose of the eigenvector matrix, lead to the following reduced model

$$\ddot{\boldsymbol{\eta}}_m(t) + \mathbf{V}_m\,\dot{\boldsymbol{\eta}}_m(t) + \mathbf{\Lambda}_m\,\boldsymbol{\eta}_m(t) = \mathbf{\Phi}_m^T\,\mathbf{f}(t) \tag{6}$$

where $\mathbf{V}_m$ is, for proportionally damped systems, a diagonal matrix, and $\mathbf{\Lambda}_m = \mathrm{diag}(\omega_1^2\,\omega_2^2\,\ldots\,\omega_m^2)$ is a diagonal matrix containing the first m eigenvalues of the full model. In the literature, this procedure is generally referred to as the *mode superposition method* or *mode displacement method*.

Mode superposition methods share a common property consisting on using a small number of free vibration modes to represent the dynamics of the structure with some reduced number of generalized degrees of freedom. This operation therefore reduces the size of the system to be solved and could result in important computational gains. However, when mode displacement method is employed to reduce the size of a model, there are some important drawback on the expansion procedures used in practice. First, to ensure that the dynamic responses computed have enough accuracy, all the modes whose frequencies are up to at least two or three times the highest exciting frequency should be retained in the summation as specified by the Rubin's rule [4]. Therefore, the number of modes required to be included in the mode superposition is usually high, which leads to a very expensive computational effort in the eigenvalue analysis. Second 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 [5]. Moreover, these methods generally do not propose the computation of an error bound for the response. Consequently, the success of the methods is established on the basis of a posteriori error comparisons. Typically, either the errors on the eigenfrequencies and eigenvectors or the

errors on the input–output representation are used to show the success of the applied method [6].

To accelerate the convergence of the low-frequency reduction the *mode acceleration* method was proposed [7–9]. This technique improves the original mode superposition method by adding different vectors to the expansion procedure [10, 11]. The system response is represented as in Eq. (4) but with a correction term, namely

$$\mathbf{z}(t) = \mathbf{\Phi}_m \boldsymbol{\eta}_m(t) + \left[\mathbf{K}^{-1} - \sum_{i=1}^{m} \frac{\boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^T}{\omega_i^2}\right] \mathbf{f}(t) \tag{7}$$

The first term at the right-hand side of Eq. (7) is the same as that in mode superposition. The second term represents the static correction of truncated modes. With the introduction of the correction term, the number of normal modes required in the mode superposition is significantly reduced if the same accuracy of responses is required. The convergent condition of the mode acceleration is that all exciting frequencies are contained in the frequency band of the kept modes [9, 12].

Actually, it is not always necessary to employ the eigenvectors of the full model in the transformation from physical coordinates to generalized coordinates. If the loads are known, for example, the approximately chosen Ritz vectors can be used as a good representation of the system response. They are an attractive alternative of mode superposition method when a model is subjected to fixed spatial distribution of dynamic loads and the eigenvectors of the model are not the best choice of basis, e.g., eigenvectors that are orthogonal to the loading are not excited even if their frequencies are contained in the loading frequency bandwidth [11]. Therefore, Ritz vector methods can be regarded as a generalized mode superposition approach in which the exact eigenvectors are replaced by more generally defined Ritz vectors [8].

The Ritz vector method has the similar transformation as modal reduction in Eq. (4). The generalized coordinates in the Ritz vector method are referred to as *Ritz coordinates*. The construction of the Ritz vectors is generally more computationally efficient than the exact eigenvectors. However, the dynamic equations of motion of the reduced model obtained from Ritz vector methods are generally coupled while they are usually uncoupled in the modal coordinate reduction [3].

Load-dependent Ritz vectors (LDRVs) are a particular and efficient class of Ritz vectors in which loading information on the structure is used to generate the vectors. In the LDRV method, the first Ritz vector is the static deformation of a structure due to a particular applied load pattern. Additional orthogonal vectors can be computed using inverse iteration and Gram–Schmidt orthogonalization. The Ritz vectors generated by this scheme automatically include static correction. Because the Ritz vectors are created based on a specific load pattern, few Ritz vectors are typically needed to achieve the same level of accuracy in response analysis under that excitation. However, the Ritz vectors have to be regenerated if the load pattern changes. Therefore, the reduced model obtained from the Ritz vector method is load-dependent. Different load-patterns have different reduced models. This is a weakness of the LDRV method.

The external load vector $\mathbf{f}(t)$ is often represented by a superposition of the spatial matrix $\mathbf{G}$ (loading patterns) and the time-dependent vector $\mathbf{h}(t)$, namely,

$$\mathbf{f}(t) = \mathbf{G}\,\mathbf{h}(t) = \sum_{i=1}^{r} \mathbf{g}_i\, h_i(t) \tag{8}$$

where

$$\mathbf{G} = \left[\mathbf{g}_1\, \mathbf{g}_2 \ldots \mathbf{g}_r\right], \quad \mathbf{h}(t) = \left\{h_1(t)\, h_2(t) \ldots h_r(t)\right\}^T \tag{9}$$

and r denotes the total number of loading patterns. For many types of loading, the number of loading patterns is small. Assume that m general Ritz vectors are obtained and denoted by $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_m$, and be $\mathbf{\Psi}_m = \mathbf{K}^{-1}\mathbf{G} \in \mathbb{R}^{n\times m}$ the matrix containing them. The physical coordinates $\mathbf{z}(t) \in \mathbb{R}^n$ may be represented by the Ritz coordinates $\boldsymbol{\rho}_m(t) \in \mathbb{R}^m$, as

$$\mathbf{z}(t) = \mathbf{\Psi}_m\, \boldsymbol{\rho}_m(t) \tag{10}$$

Introducing this coordinate transformation into Eq. (1) and premultiplying the resultant equation by the transpose of $\mathbf{\Psi}_m$ give

$$\mathbf{M}_R\, \ddot{\boldsymbol{\rho}}_m(t) + \mathbf{V}_R\, \dot{\boldsymbol{\rho}}_m(t) + \mathbf{K}_R\, \boldsymbol{\rho}_m(t) = \mathbf{G}_R\, \mathbf{h}(t) \tag{11}$$

in which the system matrices of the reduced model are defined as

$$\mathbf{M}_R = \mathbf{\Psi}_m^T \mathbf{M}\, \mathbf{\Psi}_m, \quad \mathbf{V}_R = \mathbf{\Psi}_m^T \mathbf{V}\, \mathbf{\Psi}_m,$$
$$\mathbf{K}_R = \mathbf{\Psi}_m^T \mathbf{K}\, \mathbf{\Psi}_m, \quad \mathbf{G}_R = \mathbf{\Psi}_m^T \mathbf{G}$$

Clearly, the reduced model dynamic equations of equilibrium in Eq. (11) are dependent on the particular choice of Ritz vectors $\mathbf{\Psi}_m$. For ease of exposition, we discuss only the single loading pattern case remanding to other work (like [13]) for a more detailed discussion of Ritz vector reduction techniques. Therefore, in the following $\mathbf{G}$ is a n-dimensional vector, denoted by $\mathbf{g}$, and $\mathbf{h}(t)$ is a scalar function, denoted by $h(t)$.

One of the most popular LDRV method is the WYD algorithm. It was first developed by Wilson et al. [11], in which the LDRVs were derived by a Krylov sequence. The method, further developed by Nour-Omid and Clough [14, 15] and Leger and Clough [16], employs a special Krylov sequence,

see also Balmès [17]. It starts with the static solution of Eq. (1) for a given load pattern:

$$\bar{\boldsymbol{\psi}}_1 = \mathbf{K}^{-1}\mathbf{g} \tag{12}$$

This vector is mass normalized by

$$\boldsymbol{\psi}_1 = \frac{\bar{\boldsymbol{\psi}}_1}{\left(\bar{\boldsymbol{\psi}}_1^T \mathbf{M}\, \bar{\boldsymbol{\psi}}_1\right)^{1/2}} \tag{13}$$

The inertia term neglected is considered in the successive steps to generate new Ritz vectors:

$$\hat{\boldsymbol{\psi}}_i = \mathbf{K}^{-1}\mathbf{M}\, \boldsymbol{\psi}_{i-1} \quad i = 2, 3, \ldots, m \tag{14}$$

The Gram–Schmidt mass orthogonalization and normalization are employed for these vectors:

$$\bar{\boldsymbol{\psi}}_i = \hat{\boldsymbol{\psi}}_i - \sum_{j=1}^{i-1} \left(\boldsymbol{\psi}_j^T \mathbf{M}\, \hat{\boldsymbol{\psi}}_i\right) \boldsymbol{\psi}_j \tag{15}$$

$$\boldsymbol{\psi}_i = \frac{\bar{\boldsymbol{\psi}}_i}{\left(\bar{\boldsymbol{\psi}}_i^T \mathbf{M}\, \bar{\boldsymbol{\psi}}_i\right)^{1/2}} \tag{16}$$

This process is repeated until enough Ritz vectors have been generated or no more independent Ritz vector can be generated by the process. The LDRVs generated using the above procedure are referred to as WYD vectors. The static completeness condition for the generated Ritz vector set is satisfied automatically.

To let the Ritz vectors span the configuration space at a desired frequency and to efficiently capture possible dynamic deformations for the desired frequency range, the *quasi-static Ritz vector method* is introduced. It extends the above LDRV methods by employing a quasi-static procedure. In details, the first QSRV is chosen as a quasi-static mode corresponding to the loading pattern **g** by solving the following quasi-static equilibrium equation [13]

$$\bar{\boldsymbol{\psi}}_1 = \left(\mathbf{K} - \omega_c^2 \mathbf{M}\right)^{-1} \mathbf{g} \tag{17}$$

where ω_c is referred to as the *centering frequency* [18], since it is usually chosen at the midpoint of the frequency range of interest or excitation frequency. Applying normalization, the first QSRV is given by

$$\boldsymbol{\psi}_1 = \frac{\bar{\boldsymbol{\psi}}_1}{\left(\bar{\boldsymbol{\psi}}_1^T \mathbf{M}\, \bar{\boldsymbol{\psi}}_1\right)^{1/2}} \tag{18}$$

For $i = 2, 3, \ldots, m$, the quasi-static recurrence procedure is given by

$$\hat{\boldsymbol{\psi}}_i = \left(\mathbf{K} - \omega_c^2 \mathbf{M}\right)^{-1} \mathbf{M}\, \boldsymbol{\psi}_{i-1} \tag{19}$$

$$\bar{\boldsymbol{\psi}}_i = \hat{\boldsymbol{\psi}}_i - \sum_{j=1}^{i-1} \left(\boldsymbol{\psi}_j^T \mathbf{M}\, \hat{\boldsymbol{\psi}}_i\right) \boldsymbol{\psi}_j \tag{20}$$

$$\boldsymbol{\psi}_i = \frac{\bar{\boldsymbol{\psi}}_i}{\left(\bar{\boldsymbol{\psi}}_i^T \mathbf{M}\, \bar{\boldsymbol{\psi}}_i\right)^{1/2}} \tag{21}$$

This quasi-static recurrence is continued until a given criterion is satisfied. Note that if $\omega_c = 0$, then the QSRV algorithm reduces to the WYD algorithm. Thus, the tuning parameter, ω_c, extends the flexibility and generality of the LDRV methods, which allows the Ritz vectors to best represent both the deformation shape and the frequency content of the dynamic response. As a result, improved computational efficiency and response accuracy are expected.

In the above procedure, only one centering frequency is employed. However, more than one centering frequency can be employed in the QSRV algorithm. Using multiple centering frequencies makes the technique more robust for general cases. This procedure is discussed in [13].

To determine how many LDRVs are necessary for a given problem, a participation factor, p_i, was defined by Wilson et al. [11] and Nour-Omid and Clough [14] to measure the significance of one particular Ritz vector, $\boldsymbol{\psi}_i$, in the total response

$$p_i = \boldsymbol{\psi}_i^T \mathbf{g} \tag{22}$$

This participation factor is computed for each Ritz vector and is used to terminate the vector generation process when its value drops below some threshold. However, since it does not take dynamic effects into account, this is a purely static measure and only suitable for low-frequency problems. In [13], the modal assurance criterion (MAC) is used as a measure for the participation factor, having the form

$$p_i = \frac{\left|\boldsymbol{\psi}_i^T \mathbf{s}\right|}{\left[\left(\boldsymbol{\psi}_i^T \boldsymbol{\psi}_i\right)\left(\mathbf{s}^T \mathbf{s}\right)\right]^{1/2}} \tag{23}$$

where

$$\mathbf{s} = \left(\mathbf{K} - \Omega^2 \mathbf{M}\right)^{-1} \mathbf{g} \tag{24}$$

is the frequency response due to the loading pattern, **g**, and Ω is a specified frequency. At the maximum value of one, the Ritz vector exactly matches the frequency response deformation shape, while the minimum value of zero denotes that the Ritz vector is orthogonal to the frequency response deformation and hence not likely to contribute significantly to the reduced model response. The frequency parameter, Ω, is chosen to represent a dominant frequency of the loading pattern, **g**.

2.1.2 Physical Coordinate Reduction

In the physical coordinate reduction, the reduced model is obtained by removing part of the physical coordinates of the full model. Thus, the coordinates of the reduced model actually belong to a subset of the full model. Each coordinate has its physical meaning. This is the most straightforward model reduction among the three categories. The condensation technique for the deletion of unwanted degrees of freedom was first proposed by Guyan and Irons [19, 20]. Because the dynamic effect is ignored in the condensation, this method is usually referred to as static condensation. Since the late 1960s, this technique has been widely used in many static and dynamic problems. In this technique, the total degrees of freedom are first divided into the master and slave degrees of freedom. Then, the relationship, called dynamic condensation matrix, of the responses or mode shapes between these two sets of degrees of freedom is defined by dynamic condensation schemes. Using the dynamic condensation matrix, the system matrices of a full model can be condensed to the size spanned only by the master degrees of freedom. Also, the measured data from a modal test can be expanded to the size of the full finite element model. This approach has been included in many textbooks in the areas of structural dynamics and mechanical vibrations like [21–24]. Today, it is still one of the most popular condensation methods.

The Guyan condensation matrix is derived from the static response problem. There exist two main modifications to improve the accuracy of Guyan condensation. One is the generalized Guyan condensation, which is Guyan condensation combined with the generalized inverse of stiffness matrix. The other is the quasi-static condensation, which is a combination of Guyan condensation and the frequency shifted technique. A combination of the latter two methods is also possible.

The static equation of equilibrium, that is,

$$\mathbf{K}\,\mathbf{z} = \mathbf{f} \tag{25}$$

was used in the derivation of Guyan condensation. Assume that the total degrees of freedom of the full model are categorized as the *master degrees of freedom* and the *slave degrees of freedom*, which are respectively, the kept degrees of freedom and the deleted degrees of freedom. They are simply referred to as the masters and slaves and indicated by m and s, respectively. With this arrangement, the static Eq. (25) may be partitioned as

$$\begin{bmatrix} \mathbf{K}_{mm} & \mathbf{K}_{ms} \\ \mathbf{K}_{sm} & \mathbf{K}_{ss} \end{bmatrix} \begin{Bmatrix} \mathbf{z}_m \\ \mathbf{z}_s \end{Bmatrix} = \begin{Bmatrix} \mathbf{f}_m \\ \mathbf{f}_s \end{Bmatrix} \tag{26}$$

where $\mathbf{z}_s \in \mathbb{R}^s$ is the displacement vector corresponding to the slaves, which are to be condensed, and $\mathbf{z}_m \in \mathbb{R}^m$ is the vector corresponding to the masters, which are to be retained. From Eq. (26), the vector $\mathbf{z}_s$ may be expressed in terms of $\mathbf{z}_m$ as

$$\mathbf{z}_s = -\mathbf{K}_{ss}^{-1}\mathbf{K}_{sm}\,\mathbf{z}_m + \mathbf{K}_{ss}^{-1}\mathbf{f}_s \tag{27}$$

Clearly, the displacements at the slaves consist of two parts due to the linearity of this model. One part results from the displacements at the masters and is called attached displacements. Another part results from the external forces acting on the slaves and is called relative displacements.

The external forces at the slaves were assumed to be zero by Guyan [19], that is, $\mathbf{f}_s = \mathbf{0}$. Actually, this assumption is only necessary for deriving the relation of the displacements between the masters and slaves. Letting $\mathbf{f}_s = \mathbf{0}$ on the right-hand side of Eq. (27) leads to

$$\mathbf{z}_s = \mathbf{R}_G\,\mathbf{z}_m \tag{28}$$

$\mathbf{R}_G \in \mathbb{R}^{s\times m}$ is referred to as the *Guyan condensation matrix* and is defined as

$$\mathbf{R}_G = -\mathbf{K}_{ss}^{-1}\mathbf{K}_{sm} \tag{29}$$

Equation (28) is the relation of displacements between the masters and slaves. The corresponding condensation matrix is a load-independent matrix because the external forces at the slaves were ignored in the derivation. Using the condensation matrix in Eq. (29), the displacement vector $\mathbf{z}$ may be expressed as

$$\mathbf{z} = \mathbf{T}_G\,\mathbf{z}_m \tag{30}$$

where the *coordinate transformation matrix* $\mathbf{T}_G \in \mathbb{R}^{n\times m}$ relates the responses at all the degrees of freedom to those at the masters and is defined as

$$\mathbf{T}_G = \begin{bmatrix} \mathbf{I} \\ \mathbf{R}_G \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{K}_{ss}^{-1}\mathbf{K}_{sm} \end{bmatrix} \tag{31}$$

where $\mathbf{I}$ is an $m \times m$ identity matrix.

The condensation matrix provides the means to reduce the stiffness matrix. This condensation is usually used in the static problem to eliminate unwanted degrees of freedom such as the internal degrees of freedom in finite element models. Nevertheless, the static condensation technique is also used to reduce the size of dynamic problems. In particular, the condensation matrix may be obtained from the dynamic equations when the dynamic effects are ignored. Hence, error is introduced when the dynamic problem is considered. The magnitude of these errors depends on the natural properties of the full model and how many and what degrees of freedom are selected as the masters.

In Guyan condensation, only the under part of the stiffness matrix, that is $\mathbf{K}_{sm}$ and $\mathbf{K}_{ss}$, is used in the condensation matrix. This means that the condensation matrix is independent of the stiffness concerning the masters. The generalized Guyan condensation implements all elements of the stiffness matrix into the condensation matrix. Letting

$$\mathbf{K}_m = \begin{bmatrix} \mathbf{K}_{mm} \\ \mathbf{K}_{sm} \end{bmatrix}, \quad \mathbf{K}_s = \begin{bmatrix} \mathbf{K}_{ms} \\ \mathbf{K}_{ss} \end{bmatrix} \tag{32}$$

in Eq. (26), we have

$$\mathbf{K}_m \mathbf{z}_m + \mathbf{K}_s \mathbf{z}_s = \mathbf{f} \tag{33}$$

Letting $\mathbf{f} = \mathbf{0}$ in Eq. (33), we obtain

$$\mathbf{z}_s = \mathbf{R}_{Gg} \mathbf{z}_m \tag{34}$$

where

$$\mathbf{R}_{Gg} = -\mathbf{K}_s^+ \mathbf{K}_m \tag{35}$$

is the *generalized Guyan condensation matrix*. The superscript + denotes the generalized inverse of the matrix, namely

$$\mathbf{K}_s^+ = \left(\mathbf{K}_s^T \mathbf{K}_s\right)^{-1} \mathbf{K}_s^T \tag{36}$$

The *generalized Guyan transformation matrix* is given by

$$\mathbf{T}_{Gg} = \begin{bmatrix} \mathbf{I} \\ \mathbf{R}_{Gg} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{K}_s^+ \mathbf{K}_m \end{bmatrix}. \tag{37}$$

The generalized Guyan condensation usually has much higher accuracy than Guyan condensation. However, the computational work of the former is higher than the latter.

To make the reduced model close to the full model in any given frequency range the *frequency shift technique* has been applied to many dynamic problems. Usually, the general eigenproblem of a full undamped model, referred to as the full eigenproblem, is given by

$$(\mathbf{K} - \lambda \mathbf{M})\boldsymbol{\varphi} = \mathbf{0} \tag{38}$$

where λ and $\boldsymbol{\varphi}$ are the eigenvalue (square of natural frequency) and the corresponding eigenvector (mode shape) of the full model, respectively. If an eigenvalue shift is applied to the general eigenproblem (38), we have

$$\left(\overline{\mathbf{K}} - \bar{\lambda} \mathbf{M}\right)\boldsymbol{\varphi} = \mathbf{0} \tag{39}$$

where the *dynamic stiffness matrix* $\overline{\mathbf{K}}$ and eigenvalue with shift $\bar{\lambda}$ are given by

$$\overline{\mathbf{K}} = (\mathbf{K} - q\,\mathbf{M}), \quad \bar{\lambda} = \lambda - q \tag{40}$$

where q is the value of the eigenvalue shift. Equation (39) may be rewritten in a partitioned form as

$$\left(\begin{bmatrix} \overline{\mathbf{K}}_{mm} & \overline{\mathbf{K}}_{ms} \\ \overline{\mathbf{K}}_{sm} & \overline{\mathbf{K}}_{ss} \end{bmatrix} - \bar{\lambda} \begin{bmatrix} \mathbf{M}_{mm} & \mathbf{M}_{ms} \\ \mathbf{M}_{sm} & \mathbf{M}_{ss} \end{bmatrix} \right) \begin{Bmatrix} \boldsymbol{\varphi}_m \\ \boldsymbol{\varphi}_s \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{0} \end{Bmatrix} \tag{41}$$

The second equation of Eq. (41) leads to

$$\boldsymbol{\varphi}_s = -(\overline{\mathbf{K}}_{ss} - \bar{\lambda}\,\mathbf{M}_{ss})^{-1}(\overline{\mathbf{K}}_{sm} - \bar{\lambda}\,\mathbf{M}_{sm})\boldsymbol{\varphi}_m \tag{42}$$

Letting $\bar{\lambda} = 0$ in Eq. (42), the following relation of the eigenvector between the masters and slaves is obtained

$$\boldsymbol{\varphi}_s = \overline{\mathbf{R}}_G \boldsymbol{\varphi}_m \tag{43}$$

where the condensation matrix $\overline{\mathbf{R}}_G \in \mathbb{R}^{s \times m}$ is defined as

$$\overline{\mathbf{R}}_G = -\overline{\mathbf{K}}_{ss}^{-1}\overline{\mathbf{K}}_{sm} = -(\mathbf{K}_{ss} - q\,\mathbf{M}_{ss})^{-1}(\mathbf{K}_{sm} - q\,\mathbf{M}_{sm}) \tag{44}$$

The corresponding coordinate transformation matrix $\overline{\mathbf{T}}_G$ is given by

$$\overline{\mathbf{T}}_G = \begin{bmatrix} \mathbf{I} \\ \overline{\mathbf{R}}_G \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ -(\mathbf{K}_{ss} - q\,\mathbf{M}_{ss})^{-1}(\mathbf{K}_{sm} - q\,\mathbf{M}_{sm}) \end{bmatrix} \tag{45}$$

The condensation described above is referred to as *quasi-static condensation*, and the matrix defined in Eq. (44) is called the *quasi-static condensation matrix* [25]. Clearly, the dynamic effect at the frequency $\omega = \sqrt{q}$ is included exactly in the condensation matrix. Moreover, if we set the eigenvalue shifting value $q = 0$ in Eq. (44), the classical Guyan condensation will be obtained. Because the dynamic influence is considered in the condensation matrix, this condensation is actually a dynamic condensation.

The frequency shifting technique may also be implemented into the generalized Guyan condensation matrix defined in Eq. (35). With this application, the condensation matrix is given by

$$\overline{\mathbf{R}}_{Gg} = -(\mathbf{K}_s - q\,\mathbf{M}_s)^+(\mathbf{K}_m - q\,\mathbf{M}_m) \tag{46}$$

and the corresponding coordinate transformation matrix $\overline{\mathbf{T}}_G$ is

$$\overline{\mathbf{T}}_{Gg} = \begin{bmatrix} \mathbf{I} \\ \overline{\mathbf{R}}_{Gg} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ -(\mathbf{K}_s - q\,\mathbf{M}_s)^+(\mathbf{K}_m - q\,\mathbf{M}_m) \end{bmatrix} \tag{47}$$

For convenience, Eq. (46) is called the *generalized quasi-static static condensation matrix*, and the method is called *generalized quasi-static Guyan condensation*.

2.1.3 Hybrid Coordinate Reduction: Fixed Interface Method

It is known that (see for example [3]) the accuracy of Guyan condensation may be improved by increasing the lowest natural frequency of the slave model.[1] Some ways to pursue this goal are: optimal selection of master degrees of freedom, increase of the number of master degrees of freedom and use of frequency shift technique. When the hybrid coordinate reduction is employed, the effects of the modes at the lowest frequency range of the slave model are

[1] The slave model is the full model with all its masters grounded.

included in the condensation and the cut frequency[2] change from the lowest natural frequency to the higher-order natural frequency. This is another way to increase the accuracy of Guyan condensation.

Fixed interface component mode synthesis (see Bampton and Craig [26]) is a typical *hybrid coordinate reduction* because the coordinates of the reduced model consist of some physical coordinates of the full model and part of the modal coordinates of the model with interface degrees of freedom fixed. One feature of this method is that it employs a "mixed" function base that consists of two different classes of shape functions, namely, (1) static response functions and (2) admissible functions that are linearly independent of the functions of class one [27]. Of course, the mode shape of the slave model satisfies the requirement and is a good selection. The Guyan condensation matrix may be used for the function base of class one. In order to take into account the contribution of the slaves, a second class of admissible shape functions associated with a finite number p ($p \ll s$) of additional coordinates can be introduced. In the particular case that the modal admissible function $\tilde{\boldsymbol{\varphi}}$ is defined by the following eigenproblem

$$\mathbf{K}_{ss}\widetilde{\boldsymbol{\Phi}} = \mathbf{M}_{ss}\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Lambda}} \tag{48}$$

the transformation of slave coordinates $\boldsymbol{\varphi}_s$ to the modal coordinate takes the following form

$$\boldsymbol{\varphi}_s = \mathbf{R}_G\,\boldsymbol{\varphi}_m + \widetilde{\boldsymbol{\Phi}}_p\,\mathbf{a} \tag{49}$$

where the matrix $\widetilde{\boldsymbol{\Phi}}_p \in \mathbb{R}^{s\times p}$ consists of the lowest p mode shapes of the slave model and the vector $\mathbf{a} \in \mathbb{R}^p$ contains the modal coordinates. The eigenvector of the full model may thus be expressed as

$$\begin{Bmatrix} \boldsymbol{\varphi}_m \\ \boldsymbol{\varphi}_s \end{Bmatrix} = \mathbf{T}_a \begin{Bmatrix} \boldsymbol{\varphi}_m \\ \mathbf{a} \end{Bmatrix} \tag{50}$$

where the transformation matrix $\mathbf{T}_a \in \mathbb{R}^{n\times(m+p)}$ is given by

$$\mathbf{T}_a = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_G & \widetilde{\boldsymbol{\Phi}}_p \end{bmatrix} \tag{51}$$

Many other hybrid coordinate reduction schemes are available in the literature.

2.2 Systems and Control: Balanced Truncation

In system analysis and control engineering context, MOR aims at obtaining a reduced model whose input–output behaviour is preserved as much as possible. In this context, balanced truncation is the most popular method, since it not only leads to a reduced model with a well approximated input–output behaviour, but it also preserves stability and provides an error bound, which gives a direct measure of the reduced model quality.

The balanced realization method was proposed by Moore and it is based on the observability and controllability of a system. Later, the stability preservation property was found by Pernebo and Silverman, whereas the error bound was derived by Enns and Glover [28–31].

The concepts of observability and controllability originated in control engineering. In simple terms, a system is observable if the states of the system may be deduced from the output. Similarly, a system is controllable if an input exists that enables the states of the system to attain any arbitrary value. The observability and controllability measures provide a numerical indication of how much each state is observable and controllable. The states that are least observable and controllable may be deleted from the system, since they give a small contribute to the desirable system behaviour. The resulting balanced model is in a form that is convenient for model order reduction in system analysis and control engineering. Recently, this method has been successfully implemented to finite element models with local nonlinearities [32].

Control design and system analysis typically work with the first-order differential equations in state-space, namely,

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\,\mathbf{x}(t) + \mathbf{B}\,\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\,\mathbf{x}(t) + \mathbf{D}\,\mathbf{u}(t) \end{aligned} \tag{52}$$

where the order of the model is given by the dimension of the state vector $\mathbf{x}(t)$, which will be considered to be n.

In balanced truncation, a reduced 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 behaviour. Second, a reduced model is obtained on the basis of this balanced realization by discarding the states with the smallest influence.

For linear, asymptotically stable have negative real part), time-invariant systems [33], the controllability and observability Gramians are defined as

$$\begin{aligned} \mathbf{W}_c &= \int_0^\infty e^{\mathbf{A}t}\,\mathbf{B}\,\mathbf{B}^T e^{\mathbf{A}^T t}\,dt \\ \mathbf{W}_o &= \int_0^\infty e^{\mathbf{A}^T t}\,\mathbf{C}^T\mathbf{C}\,e^{\mathbf{A}t}\,dt \end{aligned} \tag{53}$$

These two Gramians are conveniently calculated from the algebraic Lyapunov equations:

[2] The cut frequency is the upper bound of the valid frequency range of the reduced model.

$$\begin{aligned}\mathbf{A}\,\mathbf{W}_c + \mathbf{W}_c\,\mathbf{A}^T + \mathbf{B}\,\mathbf{B}^T &= \mathbf{0}\\ \mathbf{A}^T\mathbf{W}_o + \mathbf{W}_o\,\mathbf{A} + \mathbf{C}^T\mathbf{C} &= \mathbf{0}\end{aligned} \tag{54}$$

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 $\mathcal{O}(10^3)$ [6, 28, 34].

If a system is controllable, observable, asymptotically stable, and

$$\mathbf{W}_c = \mathbf{W}_o = \mathbf{\Sigma} = \mathrm{diag}(\sigma_1\,\sigma_2\,\ldots\,\sigma_n) \tag{55}$$

then it is called a *balanced realization*. The Hankel singular values, calculated as the square roots of the eigenvalues of $(\mathbf{W}_c\mathbf{W}_o)$, namely,

$$\sigma_i = \sqrt{\lambda_i(\mathbf{W}_c\mathbf{W}_o)} \quad i = 1, 2, \ldots, n$$

are ordered as $\sigma_i > \sigma_j$ for $i < j$. Since the controllable and observable Gramians represent how an individual state contributes to the system controllability and observability, if results $\sigma_i = 0$, the corresponding state is uncontrollable or unobservable. Moreover, form Eq. (55) it is clear that the realization is balanced in the sense that states that are easy to control are also easy to observe.

Generally, the requirements in Eq. (55) are very difficult to satisfy. A balanced transformation was introduced by Moore to make the controllability and observability Gramians diagonal and equal, as shown in Eq. (55). The balanced transformation takes the form

$$\mathbf{x}(t) = \mathbf{T}_b\,\bar{\mathbf{x}}(t) \tag{56}$$

in which $\mathbf{T}_b$ is a $n \times n$ square matrix and the elements of $\bar{\mathbf{x}}(t)$ are referred to as *balanced coordinates*. Using the transformation (56), the system (52) become

$$\begin{aligned}\dot{\bar{\mathbf{x}}}(t) &= \bar{\mathbf{A}}\,\bar{\mathbf{x}}(t) + \bar{\mathbf{B}}\,\mathbf{u}(t)\\ \mathbf{y}(t) &= \bar{\mathbf{C}}\,\bar{\mathbf{x}}(t) + \bar{\mathbf{D}}\,\mathbf{u}(t)\end{aligned} \tag{57}$$

in which the matrices are given by

$$\bar{\mathbf{A}} = \mathbf{T}_b^{-1}\mathbf{A}\,\mathbf{T}_b, \quad \bar{\mathbf{B}} = \mathbf{T}_b^{-1}\mathbf{B}, \quad \bar{\mathbf{C}} = \mathbf{C}\,\mathbf{T}_b, \quad \bar{\mathbf{D}} = \mathbf{D}.$$

The matrix $\mathbf{\Sigma}$ can be partitioned as

$$\mathbf{\Sigma} = \begin{bmatrix}\mathbf{\Sigma}_m & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_s\end{bmatrix} \tag{58}$$

where $\mathbf{\Sigma}_m$ and $\mathbf{\Sigma}_s$ are respectively the largest and smallest Hankel singular values. Rearranging the balanced coordinates as two groups $\bar{\mathbf{x}}_m(t) \in \mathbb{R}^m$ and $\bar{\mathbf{x}}_s(t) \in \mathbb{R}^s$, according to the partition (58), the system (57) may be rewritten as

$$\begin{aligned}\begin{Bmatrix}\dot{\bar{\mathbf{x}}}_m(t)\\ \dot{\bar{\mathbf{x}}}_s(t)\end{Bmatrix} &= \begin{bmatrix}\bar{\mathbf{A}}_{mm} & \bar{\mathbf{A}}_{ms}\\ \bar{\mathbf{A}}_{sm} & \bar{\mathbf{A}}_{ss}\end{bmatrix}\begin{Bmatrix}\bar{\mathbf{x}}_m(t)\\ \bar{\mathbf{x}}_s(t)\end{Bmatrix} + \begin{bmatrix}\bar{\mathbf{B}}_m\\ \bar{\mathbf{B}}_s\end{bmatrix}\mathbf{u}(t)\\ \mathbf{y}(t) &= \begin{bmatrix}\bar{\mathbf{C}}_m & \bar{\mathbf{C}}_s\end{bmatrix}\begin{Bmatrix}\bar{\mathbf{x}}_m(t)\\ \bar{\mathbf{x}}_s(t)\end{Bmatrix} + \bar{\mathbf{D}}\,\mathbf{u}(t).\end{aligned} \tag{59}$$

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 behaviour. The ignorance of the uncontrollable or unobservable states will not affect the input–output properties. Therefore, after the omission of the states corresponding to the smaller Henkel singular values in the balanced coordinates, the input–output properties of the reduced model is approximately equal to those of the original system. Based on this, the reduced model in balanced coordinates is given by [35]

$$\begin{aligned}\dot{\bar{\mathbf{x}}}_m(t) &= \bar{\mathbf{A}}_{mm}\,\bar{\mathbf{x}}_m(t) + \bar{\mathbf{B}}_m\,\mathbf{u}(t)\\ \tilde{\mathbf{y}}(t) &= \bar{\mathbf{C}}_m\,\bar{\mathbf{x}}_m(t) + \bar{\mathbf{D}}\,\mathbf{u}(t)\end{aligned} \tag{60}$$

where the output $\tilde{\mathbf{y}}(t)$ is an approximation of the output $\mathbf{y}(t)$ of the full model, and the quality of this approximation can be assessed by means the definition of an error bound. Moreover, it is proven that when $\mathbf{\Sigma}_m$ and $\mathbf{\Sigma}_s$ have no diagonal entries in common, namely when results $\sigma_m > \sigma_{m+1}$, the reduced model is asymptotically stable [6, 29]. A detailed analysis of balanced truncation and error bound definition lies outside the scope of this work, but more details can be found in [30, 31].

2.3 Numerical Mathematics: Krylov Subspace Projection

Typical applications of Krylov subspace projection for model order reduction are large electronic circuits with large linear subnetworks of components (see e.g. [36, 37]) and micro electro-mechanical systems (MEMS). For an application in structural vibrations, see e.g. [38]. The main goal of Krylov subspace methods is to find an approximation, in some range of the frequency domain, of the transfer function (TF) of the original system. In particular, a so-called moment expansion of the transfer function is considered and reduction focuses on matching the first coefficients (moments) of this expansion.

One of the basic and earliest reduction methods involving the usage of Krylov subspace is Asymptotic Waveform Evaluation (AWE), proposed by Pillage and Rohrer in 1990 [39, 40]. However, the main focus of AWE is on finding a Padé approximation of the transfer function, rather than on the construction of a Krylov subspace. Nevertheless, there is a close relation between the Padé approximations and Krylov subspace methods. Later, in 1995, in [41] a method called Padé Via Lanczos (PVL) was proposed and the relation between the Padé approximation and Krylov subspace was shown. In 1998, a new reduction technique, PRIMA, was introduced in [37], that 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.

For the sake of simplicity, a strictly proper[3] SISO system is considered (although all methods discussed here can be extended to MIMO systems),

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\,\mathbf{x}(t) + \mathbf{b}\,u(t) \\ y(t) &= \mathbf{c}\,\mathbf{x}(t) \end{aligned} \tag{61}$$

where $\mathbf{b} \in \mathbb{R}^{n\times 1}$ and $\mathbf{c} \in \mathbb{R}^{1\times n}$ are respectively column and row vectors.

The problem with AWE is that practical experience indicates that applicability of the method stops once eight or more moments are matched. Being the order of the reduced model equal to the number of moments matched, it cannot be higher than eight or so.

Iterative projection methods have long been used in linear system solutions and have recently become popular for model order reduction [42–45]. Methods based on this concept truncate the solution of the original system in an appropriate basis. To illustrate the concept, consider a basis transformation $\mathbf{T} \in \mathbb{R}^{n\times n}$ that maps the original state vector $\mathbf{x}(t) \in \mathbb{R}^n$ as follows

$$\begin{Bmatrix} \mathbf{x}_m(t) \\ \mathbf{x}_s(t) \end{Bmatrix} = \mathbf{T}\,\mathbf{x}(t) \tag{62}$$

where results, $\mathbf{x}_m \in \mathbb{R}^m$ and $\mathbf{x}_s \in \mathbb{R}^s$. The basis transformation $\mathbf{T}$ can then be written as

$$\mathbf{T} = \begin{bmatrix} \mathbf{W}_m^T \\ \mathbf{W}_s^T \end{bmatrix} \tag{63}$$

and its inverse as

$$\mathbf{T}^{-1} = [\mathbf{V}_m\,\mathbf{V}_s] \tag{64}$$

Since results

$$\mathbf{W}_m^T\,\mathbf{V}_m = \mathbf{I} \tag{65}$$

we conclude that

$$\mathbf{\Pi} = \mathbf{V}_m\,\mathbf{W}_m^T \tag{66}$$

is an oblique projection along the kernel of $\mathbf{W}_m$ onto the m-dimensional subspace that is spanned by the columns of the matrix $\mathbf{V}_m$ [46].

Equations (62) through (64), lead to

$$\mathbf{x}_m(t) = \mathbf{W}_m^T\,\mathbf{x}(t) \tag{67}$$

$$\mathbf{x}(t) = \mathbf{V}_m\,\mathbf{x}_m(t) + \mathbf{V}_s\,\mathbf{x}_s(t) \tag{68}$$

If we substitute Eq. (68) into the Eq. (61), we obtain

$$\begin{aligned} \dot{\mathbf{x}}_m(t) &= \mathbf{W}_m^T\,\mathbf{A}\,\mathbf{V}_m\,\mathbf{x}_m(t) + \mathbf{W}_m^T\,\mathbf{A}\,\mathbf{V}_s\,\mathbf{x}_s(t) \\ &\quad + \mathbf{W}_m^T\,\mathbf{b}\,u(t) \\ y(t) &= \mathbf{c}\,\mathbf{V}_m\,\mathbf{x}_m(t) + \mathbf{c}\,\mathbf{V}_s\,\mathbf{x}_s(t) \end{aligned} \tag{69}$$

that is still an exact expression. The approximation occurs when we would delete the terms involving $\mathbf{x}_s(t)$, namely

$$\mathbf{x}(t) \approx \mathbf{V}_m\,\mathbf{x}_m(t) \tag{70}$$

in which case we obtain a projection of the original system

$$\begin{aligned} \dot{\mathbf{x}}_m(t) &= \mathbf{W}_m^T\,\mathbf{A}\,\mathbf{V}_m\,\mathbf{x}_m(t) + \mathbf{W}_m^T\,\mathbf{b}\,u(t) \\ y(t) &= \mathbf{c}\,\mathbf{V}_m\,\mathbf{x}_m(t) \end{aligned} \tag{71}$$

To produce a good approximation to the original system, the neglected term $\mathbf{V}_s\,\mathbf{x}_s(t)$ must be sufficiently small. This has implications for the choice of the subspaces colsp $(\mathbf{V}_m)$[4] and colsp $(\mathbf{W}_m)$, which depends on the goal of the reduction procedure. In case of Krylov subspace based methods, the aim is to approximate the input–output behaviour of the full model. This is done by matching the moments of the original transfer function. This means that the reduced-order transfer function corresponding to system (71) has the moment matching property. To ensure the satisfaction of the moment matching property, one can choose $\mathbf{V}_m$ and $\mathbf{W}_m$ such that the columns of these matrices span the so-called Krylov subspace [6].

Given a matrix $\hat{\mathbf{A}} \in \mathbb{R}^{n\times n}$ and a vector $\boldsymbol{\psi} \in \mathbb{R}^n$, the k-dimensional *Krylov subspace* $\mathcal{K}_k(\hat{\mathbf{A}}, \mathbf{r})$ is defined as

$$\mathcal{K}_k(\hat{\mathbf{A}}, \mathbf{r}) = \text{span}(\mathbf{r}, \hat{\mathbf{A}}\,\mathbf{r}, \hat{\mathbf{A}}^2\mathbf{r}, \ldots, \hat{\mathbf{A}}^{k-1}\mathbf{r}) \tag{72}$$

If an approximation of the system transfer function is to be found, the matrices $\mathbf{V}_m$ and $\mathbf{W}_m$ are chosen as follows:

$$\text{colsp}\,(\mathbf{V}_m) = \mathcal{K}_m((s_0\,\mathbf{I} - \mathbf{A})^{-1}, (s_0\,\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}) \tag{73}$$

$$\text{colsp}\,(\mathbf{W}_m) = \mathcal{K}_m((s_0\,\mathbf{I} - \mathbf{A})^{-T}, (s_0\,\mathbf{I} - \mathbf{A})^{-T}\mathbf{c}^T) \tag{74}$$

If $\mathbf{V}_m$ and $\mathbf{W}_m$ are built in the way defined in (73) and (74), the model reduction method is called a two-sided method. If only one of the projection matrices ($\mathbf{V}_m$ or $\mathbf{W}_m$) 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 $2m$ moments of the original transfer function. In case of one-sided methods, m moments are matched. Besides the difference in the number of moments matched, the choice to use one- or two-sided methods influences also some other properties of the reduced system [6].

The process of constructing the reduction matrices, $\mathbf{V}_m$ and $\mathbf{W}_m$, is not straightforward and requires the use of

[3] A strictly proper system is a system whose output does not explicitly depend on the input, that is, $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), t)$.

[4] colsp $(\hat{\mathbf{A}})$ denotes the vector space spanned by the columns of the matrix $\hat{\mathbf{A}}$.

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. The classical Arnoldi algorithm generates a set $\mathbf{V}_m$ of orthonormal vectors, i.e.

$$\mathbf{V}_m^T \mathbf{V}_m = \mathbf{I} \tag{75}$$

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

$$\mathbf{W}_m^T \mathbf{V}_m = \mathbf{I} \tag{76}$$

Two sets of basis vectors $\mathbf{V}_m$ and $\mathbf{W}_m$ for Krylov subspaces may also be computed using a two-sided Arnoldi algorithm (see [47]). In this case, both $\mathbf{V}_m$ and $\mathbf{W}_m$ are orthonormal,

$$\mathbf{V}_m^T \mathbf{V}_m = \mathbf{I}, \quad \mathbf{W}_m^T \mathbf{W}_m = \mathbf{I} \tag{77}$$

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. We conclude observing that, so far, there have been no proven a priori error-bounds for the Krylov based reduction techniques [48].

2.4 Qualitative Comparison and Summary Table

In this section a qualitative comparison of the reduction techniques discussed in this work is carried out. A first important difference exists between reduction methods developed in the area of structural dynamics and those developed in the other fields, due to the form used to represent the original system. In particular, 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 taken into account or when proportional damping is considered. This important difference makes all reduction methods developed in the area of structural dynamics generally not suitable to be applied in other fields. 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.

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 Lyapunov's two equations [see Eq. (54)], which are of the size of the original high-order model. This computational complexity seriously hinders the applicability of balanced truncation to systems of very high order. Moreover, a full coordinate transformation has to be computed, before reduction can be performed by means of truncation. As a result, the total computational cost associated to balanced truncation is very high [6].

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 m basis vectors. Furthermore, the computations are less costly since the required matrix operations 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 natural frequencies in this range [6]. It is also noted that, among the reduction techniques from the field of structural dynamics, those that do not require the solution of any eigenproblem have a more less computational burden.

The Krylov subspace based moment matching techniques from numerical mathematics also have a small numerical cost. 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. Here, it is noted that the cost of two-sided moment matching methods are twice as high as the cost of one-sided methods, as two sets of basis vectors need to be obtained in the former [6].

A final general difference can be found in the level of automation of the model reduction techniques from the different fields. It is known that, only balanced truncation method is fully automatic when a requirement on the quality of the reduced model is given. Namely, the existence of an a priori error bound allows for the automatic choice of the reduction order. On the other hand, 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. Nevertheless, the level of automation differs among the reduction techniques discussed in this work. For example, modal truncation techniques are dependent on the frequency range of interest, which needs to be specified a priori. Similarly, the reduction procedure in moment matching techniques from numerical mathematics is dependent on the choice of expansion points. However, the computational procedure in modal truncation and moment matching is fully automatic as soon as a choice is specified for the frequency range of interest and the expansion points, respectively. It is known that all methods which involve DOF condensation need the user's choice of master DOF. Therefore, for these methods the level of automation

Table 1 Synthesis of several reduction techniques

		Input-Output Behaviour	Global Behaviour	Static Response	Low-Frequency Response	Middle-Frequency Response	Type of Coordinate	Accuracy	Computational Burden	Level of Automation
Modal Truncation	Mode Displacement [33,17,2];[93,99]		●		●	●	G	◉	◉	◉
	Mode Acceleration [33,17,2];[98,97,99]		●	●	●		G	●	◉	◉
Ritz Vector Methods	Static Ritz Vector [50,120];[73,89,90]		●	●	●		G	●	⊙	●
	Quasistatic Ritz Vector [50]		●	●	●	●	G	●	⊙	●
DOF Condensation	Static Condensation [52,33,99];[60,95,20]		●	●	●		P	⊙	⊙	○
	Quasistatic Condensation [21,99];[20]		●	●	●	●	P	⊙	⊙	○
	Generalized Static Condensation [99]		●	●	●		P	◉	⊙	○
	Generalized Quasistatic Condensation [99]		●	●	●	●	P	◉	⊙	○
Component Mode Synthesis	Fixed Interface Method [31,33];[21,30,99]		●	●	●		H	●	◉	⊙
	Craig-Bampton TAM [31,18];[63,99]		●	●	●		P	●	◉	⊙
	Free Interface Method [31,33];[32,55,99]		●		●	●	G	◉	◉	⊙
Other Methods	Balanced Truncation [94,17];[85,38,3,109]	●		Time Domain Based Reduction			S	●	●	●
	Krylov Subspace Projection (AWE) [24,17];[109,26,96,39,78]	●		●	●		S	⊙	⊙	◉

○	Very Low	G	The coordinates of the reduced model are generalized coordinates
⊙	Low	P	The coordinates of the reduced model are physical coordinates
◉	Medium	H	The coordinates of the reduced model are hybrid coordinates
●	High	S	The original model is defined in the state space
●	Very High	[Main References];[Other References]	

is very low. Furthermore, thanks to the definition of a participation factor measuring the significance of one particular Ritz vector in the total system response, the level of automation of Ritz vector methods may be considered high, even though this parameter does not form an error bound for the reduced model.

The first step in all model order reduction procedures is the selection of the most suitable technique for a given application. To provide some guidelines to assist analysts in this step, in Table 1 a comparison of the reduction techniques discussed in this work is presented. More generally, this table should provide a view about the strengths and weaknesses of each reduction technique.

Along the rows of the table are listed the reduction techniques discussed in this work, whereas along the columns are listed several features which may be more or less present in each reduction technique and which have been considered the most important in the area of structural dynamics.

Balanced truncation and Krylov subspace projection have been grouped in other methods, although each of them form a family of reduction techniques. These two methods belong to the fields of systems and control and numerical mathematics, respectively. Therefore, a more detailed discussion of such methods lies outside the scope of this work. It is also noted that, among the Krylov subspace projection methods the synthesis shown in the table is referred to the asymptotic waveform evaluation method.

As mentioned in the introduction of this work, the accuracy of a reduced model is measured by means of the width of the frequency range in which the reduced model approximates the behaviour of the full model. Here, the higher accuracy has been assigned to balanced truncation reduction since, as seen in the previous section, it deletes only the states that give a small contribute to the desirable system behaviour. Whereas, the lower accuracy has been assigned to static and quasi-static condensation techniques since they lead to a reduced model whose valid frequency range could be very narrow if the master DOF are not properly chosen. As will be shown at the end of this work, asymptotic waveform evaluation also has a low level of accuracy.

The discussion above, concerning the accuracy of the reduced model, the computational burden and the level of automation of the reduction procedure, is summarized in the last three columns of the table, respectively. These features are measured on a scale from one to five, represented by means of a dot inside a circle. In particular, the blank circle indicates a very low level of that feature and the solid black circle indicates a very high one. It is noted that, the comparison of the accuracy and computational burden of each reduction technique is performed considering the same order of the reduced model. Moreover, it is noted that the letters G, P and H in the previous column suggest that the coordinates of the reduced model are generalized coordinates, physical coordinates and hybrid coordinates, respectively, whereas the letter S indicates that the original model is represented in the state space, and therefore the input and output variables are not involved in the model reduction process (reduction only involves the internal description of the system).

In order to explain how the information contained in this table is read, the quasi-static Ritz vector method is considered as example. This method preserves the global system behaviour and the static response. It is able to approximate the full model behaviour in any interested frequency range. The coordinates of the reduced model are generalized coordinates. Moreover, the accuracy of the reduced model is high, the computational burden is low and the level of automation is high.

In conclusion, this table does not claim to be an exhaustive synthesis of all reduction techniques discussed in this work. However, it aims to represent an incentive, for researches working in the area of model order reduction, in summarizing the results of the great amount of reduction techniques (from different fields) existing in the literature. This could provide a useful tool for all those researchers who are new in this area and want to take advantage of the benefits of model order reduction.

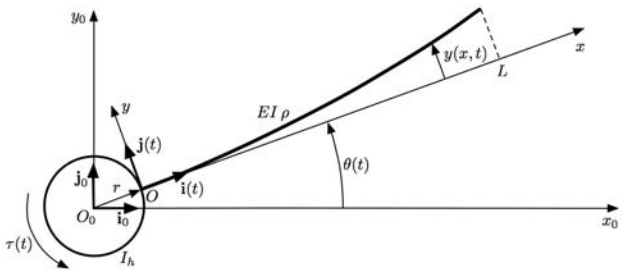

Fig. 1 Slewing flexible beam

3 Quantitative Comparison of Techniques: Application to a Slewing Flexible Beam

In this section some of the reduction techniques reported in Table 1 are applied to a FE model of the slewing flexible beam depicted in Fig. 1. In particular, mode displacement reduction will be employed in Sect. 3.2.1, Ritz vector methods in Sect. 3.2.2, Guyan condensation methods in Sect. 3.2.3, hybrid coordinate reduction in Sect. 3.2.4 and asymptotic waveform evaluation in Sect. 3.2.5. The purpose of this section is to verify the information provided in Table 1. The system frequency response will be employed to measure the accuracy of different reduced models.

3.1 A Simple FE Model of a Slewing Flexible Beam

Since lightweight and flexible components have been widely used in many engineering structures, their dynamic analysis have drawn the attention of many researchers in the last decades. The rotating beam represents the simpler case of such structures and, therefore, many mathematical models have been developed in the literature.

Here a simple finite element consistent-mass model with only translational DOF is considered to model the dynamical behaviour of the slewing flexible beam depicted in Fig. 1. This figure shows a system made up of a flexible beam clamped to a rotating rigid hub.

The following hypothesis have been employed in order to get a simple linear model:

- The beam significant vibration develops into the plane of its rigid motion.
- The amplitude of the applied torque $\tau(t)$ is small enough and its frequencies are away from the beam natural frequencies, so that the small displacement hypothesis can be used to model the beam transversal displacements $y(x, t)$ from its undeformed configuration.
- The angular velocity of the rigid body system is small enough compared to the lowest natural frequency of the flexible beam, so that the dynamic stiffening effect could not be taken into account as well as the non-linear terms

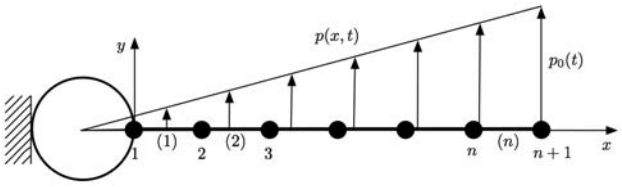

Fig. 2 Finite element model of the rotating beam in Fig. 1

into the translational (along the beam transversal direction) and rotational equilibrium statements of the beam could be neglected [49].
- The angular velocity of the rigid body system is small enough, so that any aerodynamic force can be neglected.
- The damping due to the beam vibration is a Rayleigh damping.
- The beam is an Euler–Bernoulli beam.
- The hub moment of inertia is large enough, so that the rigid body system dynamics is independent from the beam transversal elastic dynamics.

Under the above hypothesis, a non-rotating cantilever beam undergoing a triangular load distribution, representing the effect of the beam rigid body motion (inertial force), could be used in place of the slewing flexible beam in Fig. 1 to model its dynamics. A finite element model of such system is shown in Fig. 2.

Its dynamic equation is

$$\mathbf{M}\,\ddot{\mathbf{z}}(t) + \mathbf{V}\,\dot{\mathbf{z}}(t) + \mathbf{K}\,\mathbf{z}(t) = \mathbf{f}(t) \tag{78}$$

where we consider a Rayleigh damping and the equivalent load vector $\mathbf{f}(t)$ due to the trapezoidal load distribution has the form

$$\mathbf{f}(t) = \mathbf{g}\,\tau(t) \tag{79}$$

The system transfer function $W(s) \in \mathbb{C}$ between the applied torque $\tau(t)$ and the beam tip displacement $y(L, t)$ from the undeformed configuration [that is given by the last element of the nodal displacement vector $\mathbf{z}$(t)] is given by

$$W(s) = \mathbf{h}\left(s^2\,\mathbf{M} + s\,\mathbf{V} + \mathbf{K}\right)^{-1}\mathbf{g} \tag{80}$$

where $\mathbf{h}$ is a row vector with all zero entries except the last that is set to 1. The corresponding frequency response is shown in Fig. 3, where a 50-DOF finite element model has been employed as the beam full model.

3.2 Comparison of Some Reduced FE Models

Employing the following coordinate transformation

$$\mathbf{z}(t) = \mathbf{T}\,\mathbf{q}(t) \tag{81}$$

where $\mathbf{T}$ is a $50 \times m$ rectangular matrix (with $m < 50$), the reduced model transfer function $W_R(s) \in \mathbb{C}$ is given by

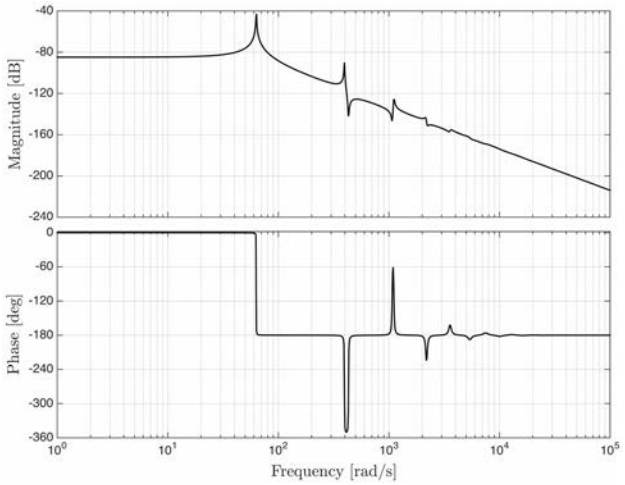

Fig. 3 Bode diagram of the beam transfer function $W(s)$ between the applied torque $\tau(t)$ and the beam tip displacement $y(L, t)$ from the undeformed configuration

Table 2 Cases for selection of mode shapes in modal truncation reduction

Case	Retained modes
1	1
2	2
3	3
4	1 and 2
5	1 and 3
6	2 and 3
7	1, 2 and 3
8	2, 3 and 4

$$W_R(s) = \mathbf{h}\,\mathbf{T}\left(s^2\,\mathbf{M}_R + s\,\mathbf{V}_R + \mathbf{K}_R\right)^{-1}\mathbf{T}^T\mathbf{g}. \tag{82}$$

In the following sections the magnitude frequency responses of the TFs of the beam reduced models obtained by means of the aforementioned techniques is selected to check their accuracy.

3.2.1 Modal Truncation

Here modal truncation reduction is applied to the beam FE model in Eq. (78). Different cases of selection of mode shapes retained in the reduced model are considered, as listed in Table 2. In Figs. 4, 5, 6, 7, 8 and 9 the corresponding frequency responses are shown. It is observed that:

- The system static gain is generally not preserved;
- The reduced model is near to the full model at the frequencies corresponding to the retained modes;

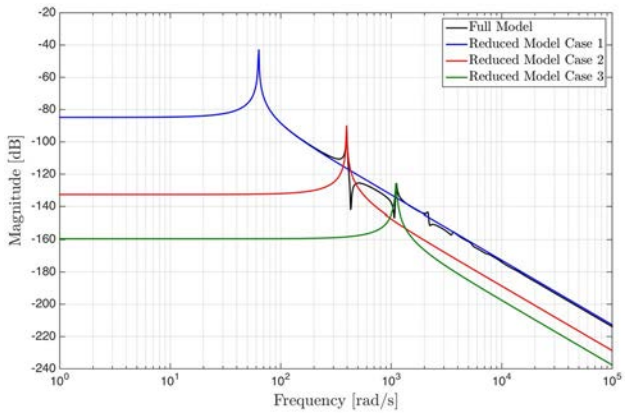

Fig. 4 Magnitude Bode diagrams of the full model TF and three TFs of reduced models defined in Cases 1 through 3 in Table 2

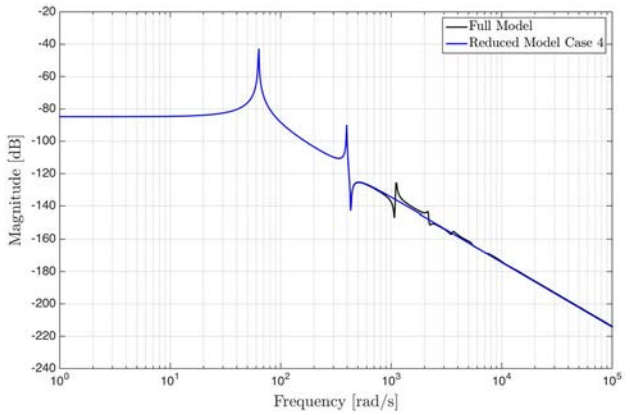

Fig. 5 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 4 in Table 2

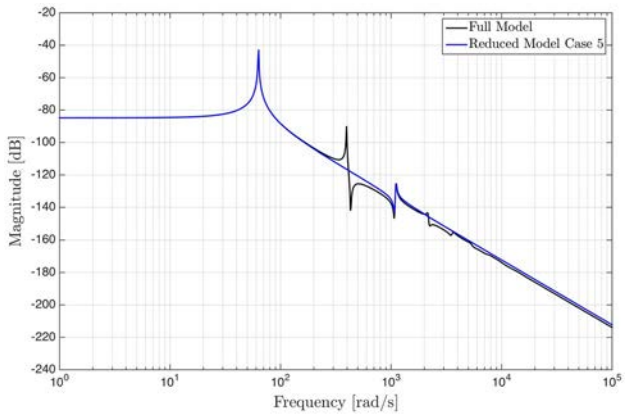

Fig. 6 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 5 in Table 2

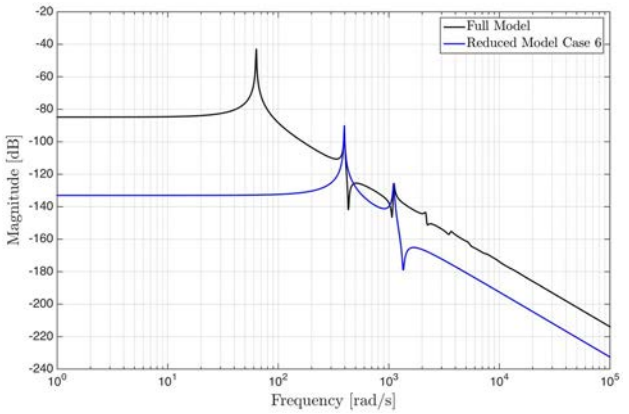

Fig. 7 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 6 in Table 2

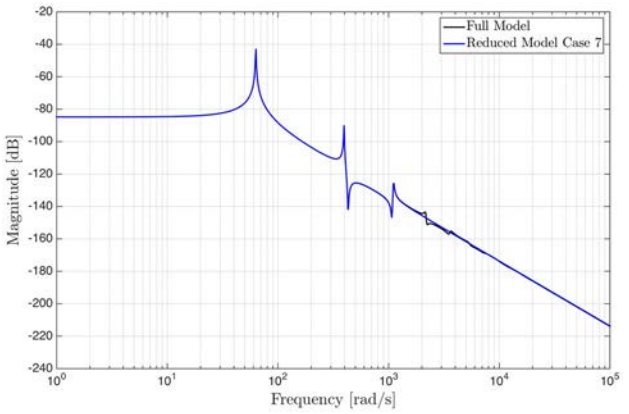

Fig. 8 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 7 in Table 2

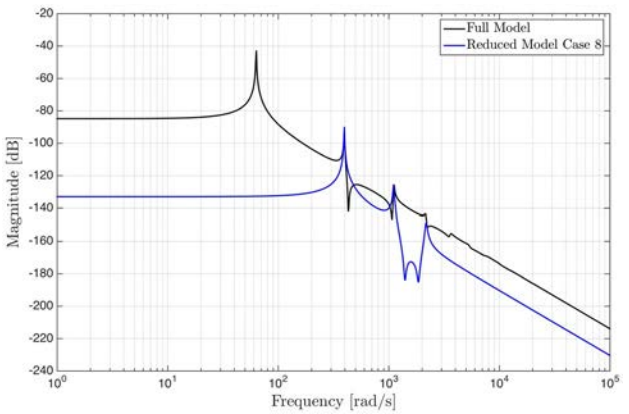

Fig. 9 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 8 in Table 2

Table 3 Cases for selection of centering frequencies and the corresponding number of Ritz vectors in the recurrence group

Case	Centering frequency [rad/s]–N. of Ritz vectors
1	63.11–1
2	395.52–1
3	1107.50–1
4	63.11–1 and 395.52–1
5	63.11–1 and 1107.50–1
6	395.52–1 and 1107.50–1
7	63.11–1, 395.52–1 and 1107.50–1
8	395.52–1, 1107.50–1 and 2170.20–2

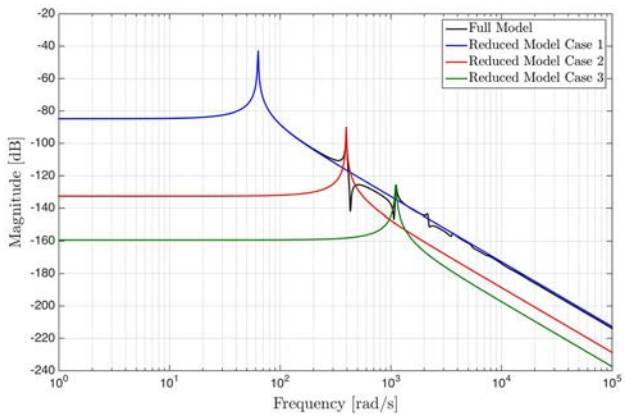

Fig. 10 Magnitude Bode diagrams of the full model TF and three TFs of reduced models defined in Cases 1 through 3 in Table 3

- Between two frequencies corresponding to two retained modes, a good accuracy of the reduced model is generally not guaranteed;
- Retaining just the first three modes (Case 7) in the reduced model, its response has high accuracy in a wide range of frequencies.

In Table 1 the reduction procedure has been considered of medium level of automation because, although the reduction methods from structural dynamics are heuristic, in modal truncation reduction the computational procedure is fully automatic as soon as a choice is specified for the frequency range of interest. Moreover, the computational burden has been considered of medium level due to the eigenproblem solution involved in the reduction procedure, which is computationally expensive when compared to the dynamic stiffness matrix inversion involved in Guyan condensation and Ritz vector methods.

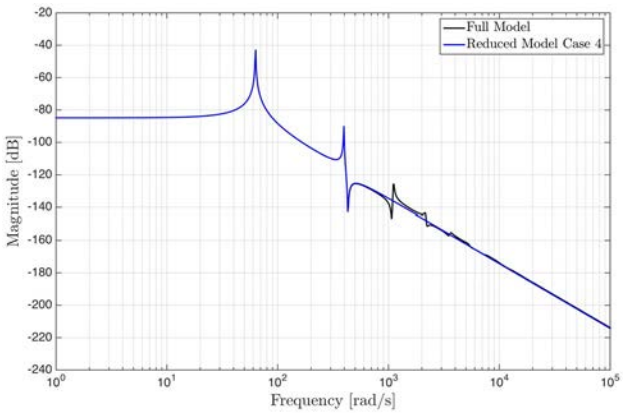

Fig. 11 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 4 in Table 3

3.2.2 Quasistatic Ritz Vector Method

Here the QSRV algorithm is applied to the beam FE model in Eq. (78). Different cases of selection of centering frequencies and the corresponding number of Ritz vectors in the recurrence group are considered, as listed in Table 3. In Figs. 10, 11, 12, 13, 14 and 15 the corresponding frequency responses are shown.

It is observed that, setting the centering frequencies to the system natural frequencies and employing one Ritz vector in each corresponding recurrence group, the reduced model obtained has the same accuracy as the reduced model obtained using modal truncation reduction. Nevertheless, the generation of Ritz vectors is less computationally expensive than the computation of eigenvectors. Moreover, increasing the number of Ritz vectors in a recurrence group, the valid frequency range of the reduced model close to the corresponding centering frequency becomes wider. This is shown in Fig. 15 where two Ritz vectors have been employed in the recurrence group corresponding to the fourth natural frequency. It is observed that with just one vector more, the accuracy of the reduced model is much higher than that of the reduced model obtained in Case 8 in Table 2 (whose frequency response is shown in Fig. 9).

In Table 1 the reduction procedure has been considered of high level of automation since, as for modal truncation reduction, the computational procedure is fully automatic as soon as a choice is specified for the frequency range of interest. Moreover, the definition of a participation factor p_i for each Ritz vector allows to automatically terminate their generation procedure as soon as the value of p_i drops below a given threshold. In conclusion, it is noted that the static response can always be preserved just adding one static Ritz vector (that is a QSRV with a centering frequency set to zero) for each loading pattern $\mathbf{g}_i$.

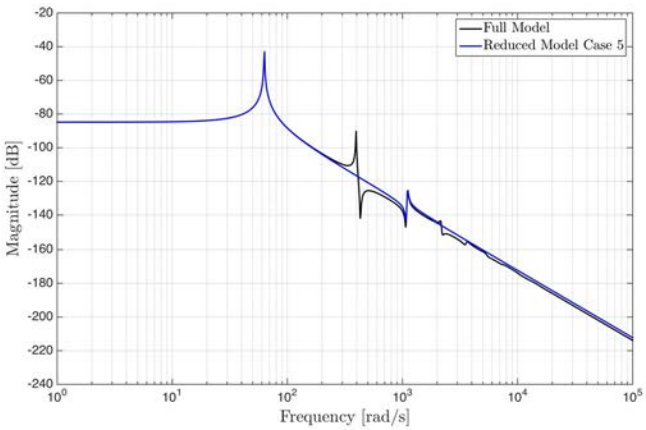

Fig. 12 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 5 in Table 3

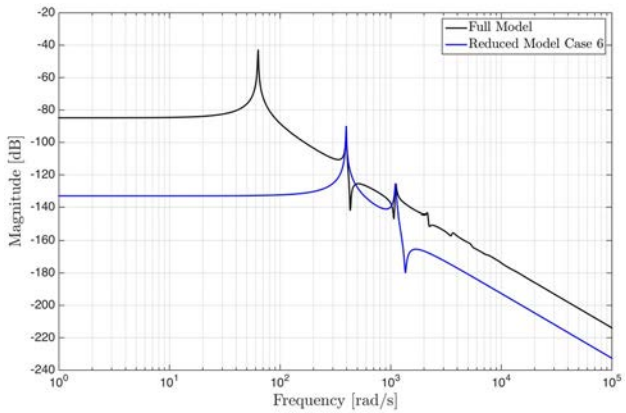

Fig. 13 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 6 in Table 3

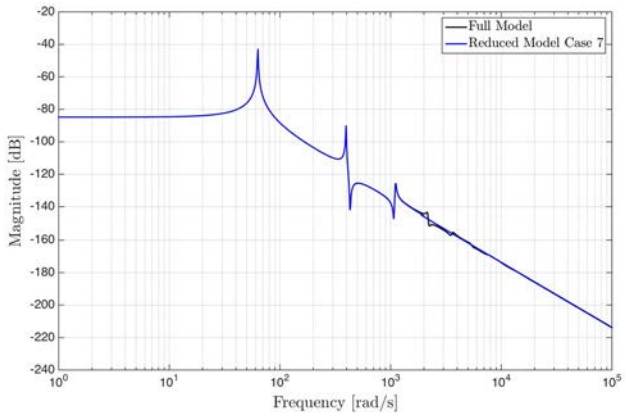

Fig. 14 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 7 in Table 3

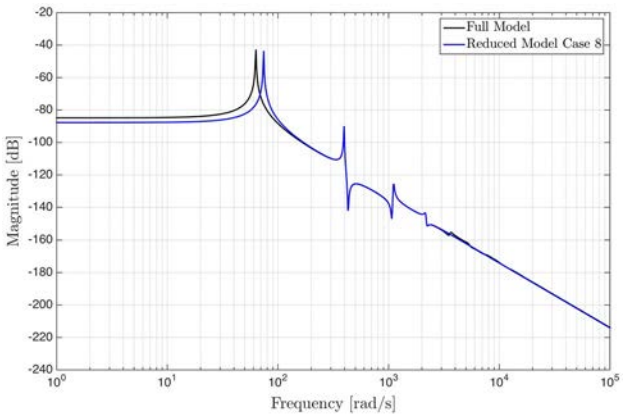

Fig. 15 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 8 in Table 3

Table 4 Cases for selection of centering frequencies and masters in Guyan condensation

Case	Centering frequency [rad/s]	Master DOF
1	0	5
2	0	15
3	0	50
4	0	25 50
5	0	1 through 20
6	0	15, 35 and 50
7	200	50
8	400, 600, 1100 and 2100	50

3.2.3 Guyan Condensation Methods

Here the quasistatic Guyan condensation and generalized quasistatic Guyan condensation are applied to the beam FE model in Eq. (78). Different cases of selection of centering frequencies and masters are considered for quasistatic Guyan condensation and generalized quasistatic Guyan condensation, as listed in Tables 4 and 5, respectively. In Figs. 16, 17, 18, 19, 20, 21, 22, 23 and 24 the corresponding frequency responses are shown.

Table 5 Cases for selection of centering frequencies and masters in generalized Guyan condensation

Case	Centering frequency (rad/s)	Master DOF
9	0	5
10	0	1 and 2
11	200	50
12	400, 800 and 1400	50

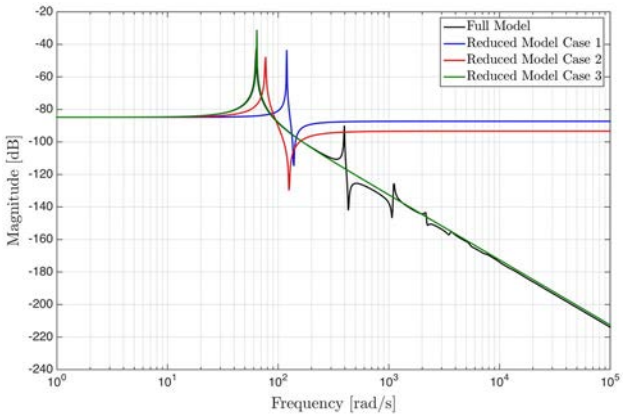

Fig. 16 Magnitude Bode diagrams of the full model TF and three TFs of reduced models defined in Cases 1 through 3 in Table 4

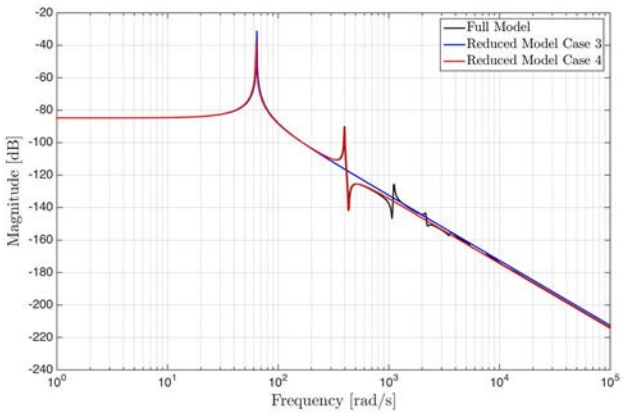

Fig. 17 Magnitude Bode diagrams of the full model TF and two TFs of reduced models defined in Cases 3 and 4 in Table 4

From Fig. 16 it is clear that the accuracy of the reduced model highly depends on the analyst's choice of masters. In particular, it is evident that when the 50th node is chosen as master the slave model is stiffer than when the 5th or 15th is chosen. This means that the frequency range where the reduced model is near to the full model is wider in Case 3 than in Cases 1 and 2. Figure 18 shows that a 3-DOF model can be much more accurate than a 20-DOF model if the

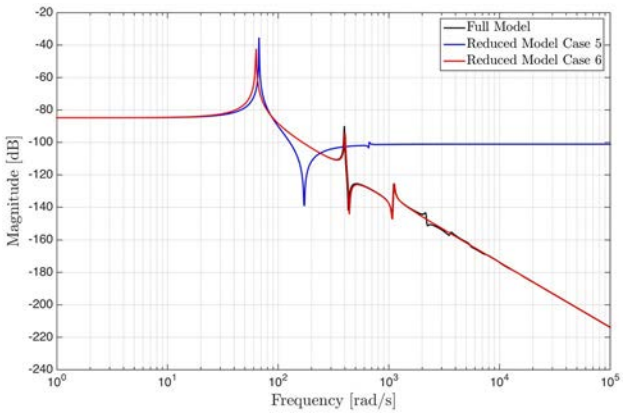

Fig. 18 Magnitude Bode diagrams of the full model TF and two TFs of reduced models defined in Cases 5 and 6 in Table 4

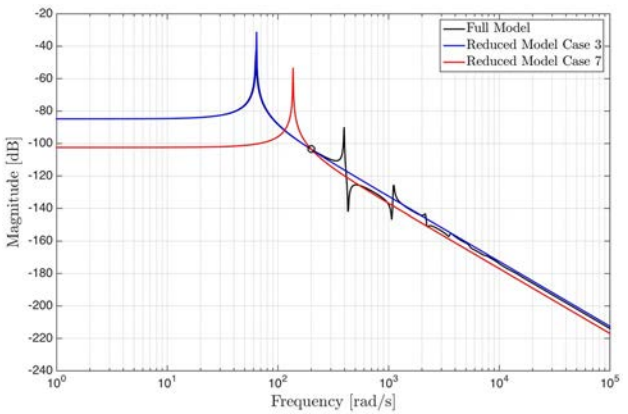

Fig. 19 Magnitude Bode diagrams of the full model TF and two TFs of reduced models defined in Cases 3 and 7 in Table 4

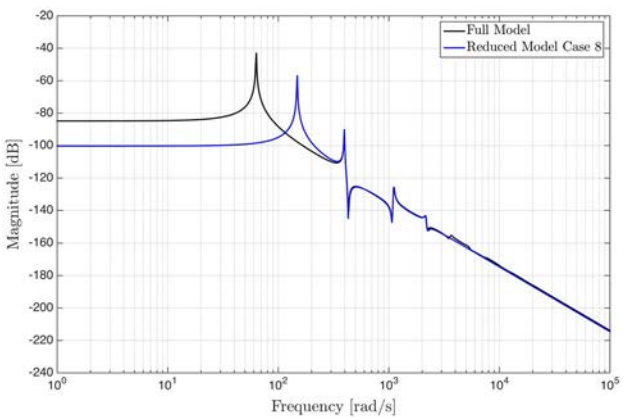

Fig. 20 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 8 in Table 4

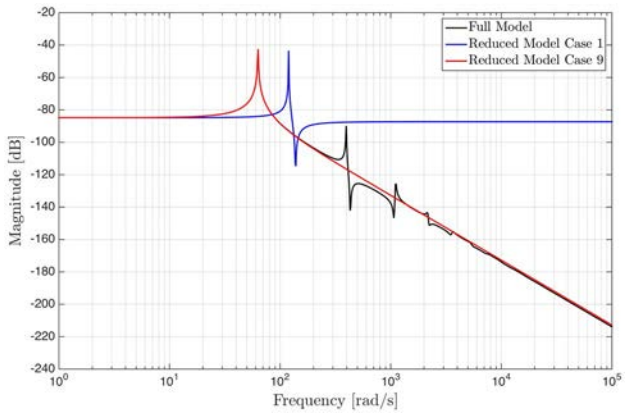

Fig. 21 Magnitude Bode diagrams of the full model TF and two TFs of reduced models defined in Cases 1 and 9 in Tables 4 and 5, respectively

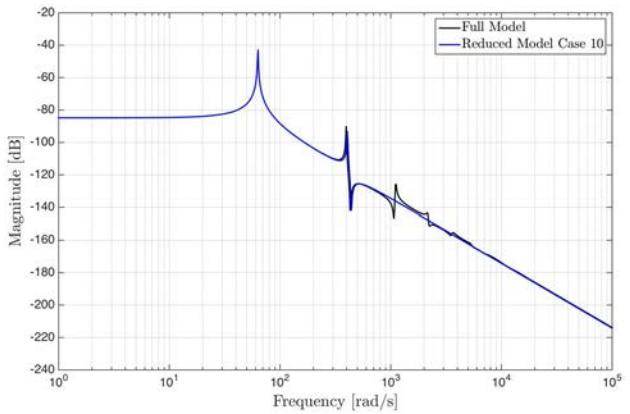

Fig. 22 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 10 in Table 5

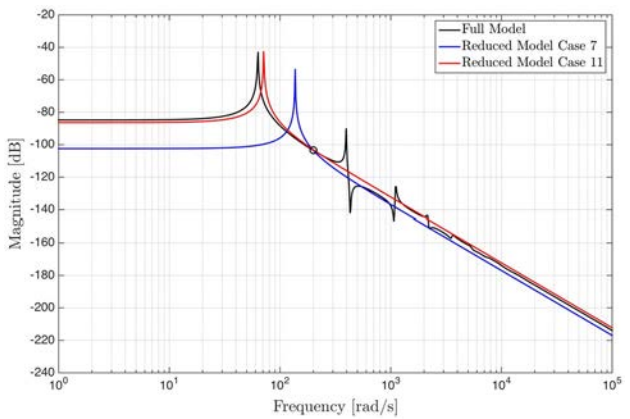

Fig. 23 Magnitude Bode diagrams of the full model TF and two TFs of reduced models defined in Cases 7 and 11 in Tables 4 and 5, respectively

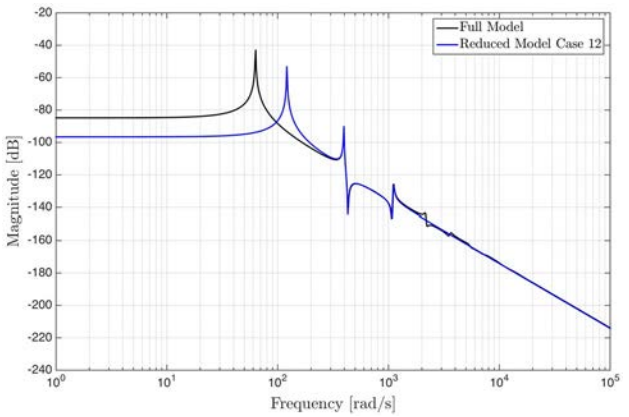

Fig. 24 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 12 in Table 5

masters are correctly chosen. For this reason, as reported in Table 1, Guyan condensation methods have a very low level of automation. Figure 17 shows that the valid frequency range of the reduced model grows when the number of masters is increased. Figure 19 shows that quasistatic Guyan condensation may leads to a reduced model with a very low accuracy. In particular the reduced model has the same behaviour of the full model solely at the centering frequency (200 Hz in Case 7).

An improvement to quasistatic Guyan condensation, which allows to shift the valid frequency range of the reduced model, is proposed in this work. It consists in considering multiple centering frequencies in the reduction procedure. This leads to the following transformation matrix

$$\overline{\mathbf{T}}_G = \begin{bmatrix} \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} \\ \overline{\mathbf{R}}_G(\omega_1) & \overline{\mathbf{R}}_G(\omega_2) & \cdots & \overline{\mathbf{R}}_G(\omega_r) \end{bmatrix} \tag{83}$$

where $\overline{\mathbf{R}}_G(\omega_i)$ is the quasistatic Guyan condensation matrix at the shifting frequency ω_i, and r is the number of shifting frequencies. The main drawback of this reduction procedure is that the physical meaning of the reduced coordinates is lost. Stated differently, the reduced coordinates are no longer a subset of the full coordinates. An example of this reduction procedure is shown in Figure 20, where four centering frequencies have been employed.

Figure 21 shows that, choosing the same master DOF, generalized Guyan condensation leads to a reduced model which has higher accuracy than the reduced model obtained by Guyan condensation. As shown in Fig. 22, even when the first two nodes are chosen as masters (which is the worst choice) the accuracy of the reduced model is high. The same improvement as in Eq. (83), which allows to shift the valid frequency range of the reduced model, is proposed for generalized quasistatic Guyan condensation. This leads to the following transformation matrix

Table 6 Cases for selection of masters and number of eigenvectors of the slave model in fixed interface method

Case	Master DOF	Number of retained modes
1	50	1
2	50	2

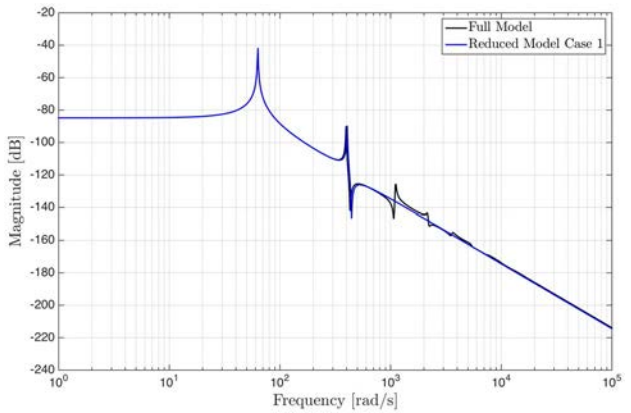

Fig. 25 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 1 in Table 6

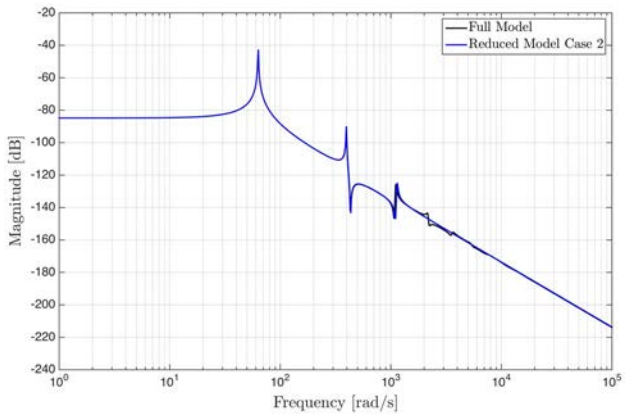

Fig. 26 Magnitude Bode diagrams of the full model TF and reduced model TF defined in Case 2 in Table 6

$$\overline{\mathbf{T}}_{Gg} = \begin{bmatrix} \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} \\ \overline{\mathbf{R}}_{Gg}(\omega_1) & \overline{\mathbf{R}}_{Gg}(\omega_2) & \cdots & \overline{\mathbf{R}}_{Gg}(\omega_r) \end{bmatrix} \tag{84}$$

where $\overline{\mathbf{R}}_{Gg}(\omega_i)$ is the generalized quasistatic Guyan condensation matrix at the shifting frequency ω_i, and r is the number of shifting frequencies. An example of this reduction procedure is shown in Fig. 24 where three centering frequencies have been employed. Furthermore, it is also observed that all Guyan condensation procedures may preserve the full model static gain.

Table 7 Cases for selection of the reduced model order in asymptotic waveform evaluation

Case	Reduced model order
1	1
2	2
3	3
4	4

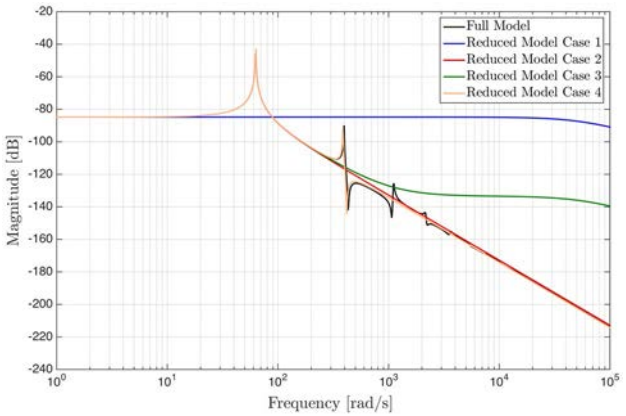

Fig. 27 Magnitude Bode diagrams of the full model TF and four TFs of reduced models defined in Cases 1 through 4 in Table 7

3.2.4 Craig–Bampton

Two cases of application of fixed interface reduction to the beam FE model in Eq. (78) are considered, in which the 50th node is set as master. The number of retained modes of the slave model are one and two in the first and second case, respectively (as listed in Table 6). In Figs. 25 and 26 the corresponding frequency responses are shown.

The level of automation of this reduction procedure is low since fixed interface method is based on Guyan condensation. Nevertheless, increasing the number of the eigenvectors employed, the valid frequency range of the reduced model is extended. This makes the level of automation of fixed interface method higher than that of Guyan condensation. Figures 25 and 26 show that the valid frequency range of the reduced model becomes wider when the number of eigenvectors employed is increased. It is also observed that the static response of the full model is always preserved. Furthermore, the computational burden has been considered of medium level due to the eigenproblem solution of the slave model involved in this reduction procedure.[5]

[5] Generally, the slave model has almost the same dimension as the full model.

3.2.5 Asymptotic Waveform Evaluation

Using asymptotic waveform evaluation reduction, four reduced beam models are obtained, with the expansion point is set to zero. They are listed in Table 7. In Fig. 27 are shown the corresponding frequency responses.

From this figure we observe that the accuracy of the reduced model in Case 1 is very low. More generally, considering the same order of the reduced model, asymptotic waveform evaluation leads to a reduced model with lower accuracy than that of the reduced models obtained by means of the above reduction procedures. In Table 1 asymptotic waveform evaluation has been considered of medium level of automation since, although this reduction method is heuristic, the computational procedure is fully automatic as soon as a choice is specified for the expansion point. It is also observed that the static gain of the full model is always preserved.

4 Conclusion

In this work general model order reduction techniques from the fields of structural dynamics, numerical mathematics and systems and control have been first briefly reviewed and then compared. Such comparison may facilitate the choice of a particular reduction technique, on the basis of the desired properties retained in the reduced model. Particular attention has been paid on reduction technique from structural dynamics.

A qualitative comparison among the aforementioned reduction methods has been presented. It has been based on the following aspects: first-order form versus second-order form; input–output behaviour versus global behaviour; automatic versus user-dependent model reduction, as well as the computational burden and the accuracy of the reduced model. Here, an important difference is due to the fact that the global dynamics is taken into account in the reduction techniques from structural dynamics, whereas reduction techniques from numerical mathematics and systems and control aim at the approximation of the input–output behaviour. This difference makes all reduction methods developed in the area of structural dynamics generally not suitable to be applied in other fields, whereas 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.

In Table 1 all the reduction techniques discussed in this work have been summarized. In this table, several features, which may be more or less present in each reduction technique and which have been considered to be the most important in the area of structural dynamics, are listed along its columns.

Finally, the differences among some of the hereby presented reduction techniques have been illustrated on a quantitative level by means of their application to the case of a slewing flexible beam. A consistent-mass finite element model, with only translational degrees of freedom, has been employed (as full model) in the application of the different reduction techniques. The frequency response has been employed to measure the accuracy of different reduced models. In conclusion, by analyzing the frequency responses of different reduced models, it has been verified that the results obtained are in good agreement with the information provided in the summary table.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Neggers J, Allix O, Hild F, Roux S (2018) Big data in experimental mechanics and model order reduction: todays challenges and tomorrows opportunities. Arch Comput Methods Eng 25(1):143–164
2. Salah Mohamed K (2018) Machine learning for model order reduction. Springer, Berlin. ISBN 978-3-319-75714-8
3. Qu Z-Q (2004) Model order reduction technique with applications in finite element analysis. Springer, Berlin
4. Rubin S (1975) Improved component-mode representation for structural dynamic analysis. AIAA J 3(8):995–1006
5. Papadrakakis M (1993) Solving large-scale problems in mechanics. Wiley, New York
6. Besselink B et al (2013) A comparison of model reduction techniques from structural dynamics, numerical mathematics and systems and control. J Sound Vib 332(19):4403–4422
7. Akgun MA (1993) A new family of mode superposition methods for response calculations. J Sound Vib 167(2):289–302
8. Gu J (2000) Efficient model reduction methods for structural dynamics analyses. Ph.D. thesis, University of Michigan
9. Qu Z-Q (2000) Hybrid expansion method for frequency responses and their sensitivities, part I: undamped systems. J Sound Vib 231(1):175–193
10. Rayleigh JWS, Lindsay RB (1945) The theory of sound, 2nd edn. Dover Publications, New York
11. Wilson EL, Yuan M-W, Dickens JM (1982) Dynamic analysis by direct superposition of Ritz vectors. Earthq Eng Struct Dyn 10(6):813–821
12. Qu Z-Q (2001) Accurate methods for frequency responses and their sensitivities of proportionally damped systems. Comput Struct 79(1):87–96
13. Gu J, Ma Z-D, Hulbert GM (2000) A new load-dependent Ritz vector method for structural dynamics analyses: quasi-static Ritz vectors. Finite Elem Anal Des 36(3–4):261–278
14. Nour-Omid B, Clough RW (1984) Dynamic analysis of structures using Lanczos co-ordinates. Earthq Eng Struct Dyn 12(4):565–577
15. Nour-Omid B, Clough RW (1985) Short communication block Lanczos method for dynamic analysis of structures. Earthq Eng Struct Dyn 13(2):271–275

16. Leger P, Wilson EL, Clough RW (1986) The use of load dependent vectors for dynamic and earthquake analyses. Technical report, Earthquake Engineering Research Center, University of California
17. Balms Etienne (1996) Parametric families of reduced finite element models. Theory and applications. Mech Syst Signal Process 10(4):381–394. https://doi.org/10.1006/mssp.1996.0027. ISSN 0888-3270
18. Ma Z-D, Hagiwara I (1991) Improved mode-superposition technique for modal frequency response analysis of coupled acoustic-structural systems. AIAA J 29(10):1720–1726
19. Guyan RJ (1965) Reduction of stiffness and mass matrices. AIAA J 3(2):380
20. Irons BM (1965) Structural eigenvalue problems—elimination of unwanted variables. AIAA J 3(6):961–962
21. Weaver W, Johnston PR (1987) Structural dynamics by finite elements. Prentice Hall, Englewood Cliffs
22. Argyris JH, Mlejnek H-P (1991) Dynamics of structures. North-Holland, Amsterdam
23. Petyt M (1990) Introduction to finite element vibration analysis. Cambridge University Press, Cambridge
24. Bathe KJ (1996) Finite element procedures. Prentice Hall, Englewood Cliffs
25. Johnson CP (1979) Computational aspects of a quadratic eigenproblem. In: Conference on electronic computation. Washington University
26. Bampton MCC, Craig Jr RR (1968) Coupling of substructures for dynamic analyses. In: AIAA
27. Bouhaddi N, Fillod R (1996) Model reduction by a simplified variant of dynamic condensation. J Sound Vib 191(2):233–250
28. Moore BC (1981) Principal component analysis in linear systems: controllability, observability, and model reduction. IEEE Autom Control 26(1):17–32
29. Pernebo L, Silverman LM (1982) Model reduction via balanced state space representations. IEEE Autom Control 27(2):382–387
30. Enns DF (1984) Model reduction with balanced realizations: an error bound and a frequency weighted generalization. In: The 23rd IEEE conference on decision and control, Las Vegas, Nevada, USA, pp 127–132
31. Glover K (1984) All optimal Hankel-norm approximations of linear multivariable systems and their l^{∞}. Int J Control 39(6):1115–1193
32. Friswell MI, Penny JET, Garvey SD (1996) The application of IRS and balanced realization methods to obtain reduced model of structures with local non-linearities. J Sound Vib 196(4):453–468
33. Baur U, Benner P, Feng L (2014) Model order reduction for linear and nonlinear systems: a system-theoretic perspective. Arch Comput Methods Eng 21(4):331–358
34. Laub AJ (1980) Computation of "balancing" transformation. In: American control conference. IEEE
35. Gu ZQ, Ma KG, Chen WD (1997) Active control of vibration. China National Defense Industry Press, Beijing
36. Freund RW (2004) SPRIM: structure-preserving reduced-order interconnect macromodeling. In: Proceedings of the IEEE/ACM international conference on computer aided design, pp 80–87
37. Odabasioglu A, Celik M, Pileggi LT (1998) PRIMA: passive reduced-order interconnect macromodeling algorithm. IEEE Trans Comput-Aid Des Integr Circuits Syst 17(8):58–65
38. Liew H-L, Pinsky PM (2010) Matrix-padé via Lanczos solutions for vibrations of fluid–structure interaction. Int J Numer Methods Eng 84(10):1183–1204
39. Chiprout E, Nakhla MS (1994) Asymptotic waveform evaluation and moment matching for interconnect analysis. Kluwer, Dordrecht
40. Pillage LT, Rohrer RA (1990) Asymptotic waveform evaluation for timing analysis. IEEE Comput-Aid Des Integr Circuits Syst 9(4):352–366
41. Feldmann P, Freund RW (1995) Efficient linear circuit analysis by Padé approximation via the Lanczos process. IEEE Comput-Aided Des Integr Circuits Syst 14(5):639–649
42. Saad Y (1996) Iterative methods for sparse linear systems. Society for Industrial and Applied Mathematics (SIAM), vol 82. 2nd edn
43. Villemagne CD, Skelton RE (1987) Model reductions using a projection formulation. Int J Control 46(6):2141–2169
44. Kim HM, Craig RR (1988) Structural dynamics analysis using an unsymmetric block Lanczos algorithm. Int J Numer Methods Eng 26(10):2305–2318
45. Grimme E (1997) Krylov projection methods for model reduction. Ph.D. thesis, University of Illinois at Urbana-Champaign
46. Schilders WHA, van der Vorst HA, Rommes J (2008) Model order reduction: theory, research aspects and applications. Springer, Berlin
47. Lohmann B, Salimbahrami B (2003) Introduction to Krylov subspace methods in model order reduction. In: Methods and applications in automation, pp 1–13
48. Heres PJ (2005) Robust and efficient Krylov subspace methods for model order reduction. Ph.D. thesis, Eindhoven University of Technology
49. Giovagnoni M (1993) Linear decoupled models for a slewing beam undergoing large rotations. J Sound Vib 164(3):485–501