Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Harmonic Vibration Analysis with Coherent-Nodal-Cluster Two-stage Model Reduction

Petr Krysl*[1] | Raghavendra Sivapuram[2]

[1]Structural Engineering Department, University of California, San Diego, California, USA

[2]Dassault Systèmes Simulia Corp, Johnston, Rhode Island, USA

Correspondence

*Petr Krysl. Email: pkrysl@ucsd.edu

Summary

Linear finite element harmonic vibration analysis is an important tool when investigating structural performance. In order to construct the requested frequency response functions by sweeping through the frequencies, the full finite element model is typically reduced by a Galerkin projection, using either the solution of a free vibration problem, or with load-dependent Ritz vectors. Here we construct an approximation of the model reduction matrix using an intermediate reduced-model stage based on the Coherent Nodal Clusters technique. Thus we arrive at a two-stage model-reduction procedure. There is an additional approximation, but the reduced models are obtained at a much lower cost than the one-stage models. In order to ameliorate the loss of accuracy involved in the additional approximation, we introduce enhancement of the reduced bases by inexpensive residual-minimizing iterations.

KEYWORDS:

model reduction, reduced basis, coherent node cluster, free vibration, load-dependent Ritz basis, steady-state dynamics

INTRODUCTION

Linear finite element harmonic vibration analysis (also known as steady-state dynamics, SSD) is a standard tool in engineering analysis of structures. It serves to construct frequency response functions (FRF): various quantities such as displacement, velocity, acceleration magnitudes, stress measures, and so on, presented as functions of frequency.

The matrix equations of motion are easily constructed, but since the dynamic stiffness is complex and depends on the frequency, the cost of computing the value of the frequency response function for a large number of frequencies is prohibitively expensive when performed for the original finite element model. We shall call the solution of the full finite element equations the direct approach.

Hence, a commonly accepted device is to trade accuracy for efficiency by performing a transformation of the original degrees of freedom of the finite element model into a new set of significantly fewer generalized degrees of freedom. The transformed system of equations may or may not have a diagonal matrix, but being of considerably smaller dimension than the original finite element model, the solution of the balance equations for each frequency is significantly reduced.

Thus the following has been accomplished: the sweep through the frequencies is now much faster than for the direct approach. On the other hand, the basis for the reduced space needs to be constructed. This typically becomes the determining cost of the procedure. So whereas for the direct approach the sweep dominates the computing cost, with the reduction of the model to the generalized degrees of freedom the cost has now shifted to the construction of the reduced model.

One of the first model-order reduction methods were proposed by Guyans[1] and Irons[2] in which the structural degrees of freedom are divided into master (independent) and slave (dependent) degrees of freedom. The reduced model involves only the master degrees of freedom and the transformation matrix for the reduction is computed using only the stiffness of the structure. The Guyan condensation is thus only accurate near zero-frequency and the approximation grows worse as the frequency of interest increases. Kidder[3] and O'Callahan[4] included the inertia terms alongside the stiffness terms in the computation of transformation matrices at a chosen frequency. Just like the Guyan condensation being accurate only around zero-frequency, these dynamic condensation methods are decently accurate only around the frequency chosen for the computation of the transformation matrix. The dynamic condensation methods also involve the expensive inversion of the dynamic stiffness matrix to compute the terms of a Neumann series expansion, and the approximation is accurate only within a small range of frequencies around the frequency used for this expansion[5]). The dynamic condensations can be iteratively improved[6], however, the transformation matrix involves the inversion of large and potentially singular/indefinite matrices. The System Equivalent Expansion Reduction Process (SEREP) uses eigenmodes of the full model and a pseudo inverse for model-reduction[7], however, the accuracy depends greatly on the choice of the selected eigenmodes[8].

Substructuring is a model-order reduction method where the reduced basis is computed by dividing the structure into many substructures (or superelements). One of the most popular substructuring methods, the Craig-Bampton method[9,10] uses two sets of modes in the reduced basis: the constraint modes which represent the static deformation of all the individual substructures, and fixed interface modes representing the normal modes of each substructure when the retained (master) degrees of freedom of the substructure are restrained. Many variants of this substructuring technique exist in literature[11,12]. The number of degrees of freedom in the reduced model is given by the number of retained degrees of freedom, and the number of normal modes included in the reduced basis. The accuracy of the substructures depends on the choice of substructures, the selected retained degrees of freedom, and the added normal modes. The number of constraint modes in the reduced model can be reduced by solving extra eigenproblems to obtain the characteristic deformations of the interfaces (Castanier et al[13]). Since these techniques depend on the components of the structure, they are also called Component Mode Synthesis (CMS) techniques. Bennighof and Lehoucq[14] developed an Automated Multi-Level Substructuring method (AMLS) where the structure is divided into a hierarchy of substructures and the substructure modes are used to construct the reduced bases. A detailed review of many other substructuring methods is given in de Klerk et al[15].

Krylov subspace reductions involve bases that are computed using the Arnoldi or Lanczos processes. Although they yield accurate approximations, the bottleneck of these processes involve solving many large-scale linear systems, and LU decompositions are usually required for every shift frequency involved in the reductions[16]. The balanced truncation model order reduction methods aim at eliminating the states which are difficult to reach and difficult to observe. Such states correspond to the smallest Hankel singular values (obtained using Singular Value Decomposition) of the controllability and observability Gramian matrices[17]. These matrices are computed by solving Lyapunov equations and solving them exactly is expensive. The Lyapunov equations can be approximately solved using low-rank Cholesky factor based methods[18]. The computational complexity of this process is that it involves solving large-scale linear systems. The solving of Lyapunov equations is done at multiple shift parameters. The Krylov subspace reduction and balanced truncation reduction methods are computationally expensive because they involve the solving of multiple large linear systems.

Some data-driven reduction methods based on Proper Orthogonal Decomposition (POD) have been developed for parametric dynamical systems. They offer great advantage in parametrized and multi-query applications. During the offline phase, the solutions of the large-scale system are computed at various parameter values and are called snapshots. The transformation matrix is created using the left singular vectors of the snapshot matrix corresponding to high-magnitude singular values[19,20]. The system matrices are parametrized so that the computation of the matrices in the online phase only involves a linear combination of some offline computed matrices. Rational interpolation methods involving local transformation matrices are also available[21,22]. Evidently, the efficiency of the procedure is determined by the cost of the offline phase.

The most commonly adopted set of generalized coordinates corresponding to the reduced basis modes are the mode shapes of the free vibration of the structure. These have the advantage of potentially producing diagonalized reduced system matrices for each frequency. Note that we shall not take advantage of such diagonalized systems here, as we will assume a general damping matrix, which precludes complete diagonalization of the reduced matrices for the sweep.

In this work we propose a two-stage model reduction technique. In the first stage, we reduce the full finite element model using the Coherent Nodal Clustering (CoNC) reduction[23], wherein nodes are partitioned into coherent clusters, and Legendre polynomial-based bases vectors are used in each of the nodal clusters of the finite element mesh for reduction. The number of nodal clusters and the polynomial order for the same are heuristically estimated as proposed in Krysl et al[23]. The second stage

involves either the free-vibration analysis, or the computation of load-dependent Ritz vectors of the CoNC reduced model to yield a much smaller second-stage reduced model. Enhancement of the basis can be computed based on residual minimizing algorithms to improve the accuracy of the second-stage model.

We introduce the basic ideas of the SSD computation in Section 1. Section 2 reviews the existing approaches to the construction of the transformation matrix between the full finite element models and the reduced versions. In Section 3 were describe the two-stage model reduction in detail, including a brief description of the Coherent Nodal Cluster model reduction. In Section 3 we describe the two-stage model reduction in detail, including a brief description of the Coherent Nodal Cluster model reduction. Section 4 illustrates the performance of the two-stage model reduction, both in terms of accuracy and in terms of the computing cost. Finally, in Section 5 we introduce iterative enhancement of the two-stage reduced basis to lower the approximation error relative to the one-stage reduced models.

1 | PROJECTION MODEL REDUCTION

The discrete model of a solid or structure vibrating with frequency ω can be expressed with this balance equation

$$\left(-\omega^2 \boldsymbol{M} + \mathrm{i}\omega \boldsymbol{C} + \mathrm{i}\boldsymbol{D} + \boldsymbol{K}\right) \boldsymbol{U} = \boldsymbol{F} \ , \tag{1}$$

where $\boldsymbol{M}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{K}$ are the mass, viscous damping, structural damping, and stiffness matrices, and $\boldsymbol{U}$ and $\boldsymbol{F}$ are the time-independent complex vectors of displacements and forces, and $\mathrm{i} = \sqrt{-1}$. In this work we consider the damping matrices in a completely general form, in particular we do not assume they are diagonalizable complex vectors of displacements and forces, and $\mathrm{i} = \sqrt{-1}$. Structural damping occurs because of a type of sliding friction between the molecular layers of the structure, and is out-of-phase with stiffness (and hence, the imaginary coefficient)[24]. In the time-domain, structural damping may be added as an equivalent viscous damping which induces a similar energy loss in the structure[25]. In this work we consider the damping matrix in a completely general form, in particular we do not assume it is diagonalizable by the free vibration modes of the undamped system.

The response of the system is computed for some frequency values in a given range, $\omega \in \langle \omega_\ell, \omega_u \rangle$. Thus equation (1) must be solved repeatedly, once for each new frequency value, of which there are typically hundreds or thousands. Solving the complex linear algebraic equations (1) for the displacement vector $\boldsymbol{U}$ is the so-called "direct approach".

In this work we focus on continuum finite element models of vibrating solids, such as three-dimensional bodies discretized with hexahedral or tetrahedral finite elements. The dimensions of the matrices in (1) are often quite substantial, and even though these matrices are symmetric and sparse, the direct approach tends to be prohibitively expensive.

The established solution is to reduce the dimension of the model, and a particularly popular route is to use a Galerkin projection: the original linear algebraic equation is converted to an analogous equation of appreciably smaller dimensions compared to the original finite element system (1)

$$\left(-\omega^2 \boldsymbol{M}_r + \mathrm{i}\omega \boldsymbol{C}_r + \mathrm{i}\boldsymbol{D}_r + \boldsymbol{K}_r\right) \boldsymbol{U}_r = \boldsymbol{F}_r \ , \tag{2}$$

where

$$\boldsymbol{M}_r = \boldsymbol{\Phi}^T \boldsymbol{M} \boldsymbol{\Phi} \ , \quad \boldsymbol{C}_r = \boldsymbol{\Phi}^T \boldsymbol{C} \boldsymbol{\Phi} \ , \quad \boldsymbol{D}_r = \boldsymbol{\Phi}^T \boldsymbol{D} \boldsymbol{\Phi} \ , \quad \boldsymbol{K}_r = \boldsymbol{\Phi}^T \boldsymbol{K} \boldsymbol{\Phi} \ , \quad \boldsymbol{F}_r = \boldsymbol{\Phi}^T \boldsymbol{F} \ . \tag{3}$$

and $\dim \boldsymbol{U}_r \ll \dim \boldsymbol{U}$. With the solution $\boldsymbol{U}_r$ in hand we can reconstruct an approximation of the solution of the full system as

$$\boldsymbol{U} \approx \widetilde{\boldsymbol{U}} = \boldsymbol{\Phi} \boldsymbol{U}_r \ . \tag{4}$$

The rectangular transformation matrix $\boldsymbol{\Phi}$ (tall and narrow) typically consists of globally supported vectors. The number of columns of the transformation matrix $\boldsymbol{\Phi}$ is the dimension of the reduced matrices, i.e. the number of the generalized degrees of freedom. When the original finite element model degrees of freedom number in the millions, the total of the globally supported vectors is still typically many orders of magnitude smaller (tens, perhaps hundreds, in exceptional cases thousands). The construction of this matrix is the focal point in this paper, and details will be provided later.

The square reduced matrices may or may not be diagonal, that depends on the particular form of the dynamic matrices of the system and the transformation matrix $\boldsymbol{\Phi}$. Even when the reduced matrices are not diagonal, the solution of the system (2) takes much less time compared to the original system (1), as the number of the generalized degrees of freedom is typically reduced by orders of magnitude relative to the dimension of the original system (i.e. the number of the original degrees of freedom).

2 | TRANSFORMATION MATRIX: REVIEW OF EXISTING APPROACHES

In this section we shall discuss some techniques that can be used to construct the transformation matrix (referred to in this paper as $\boldsymbol{\Phi}$). In particular, we shall focus on two approaches, the method of normal modes (vibration eigenvector expansion) and the method of the Krylov basis (Ritz vectors).

2.1 | Normal modes

An often-adopted approach relies on the normal modes. For instance, in conjunction with the need to control dynamical systems, where the complexity and the costly design of a control system is often a limiting factor, a normal-mode basis can be very effective[26]. A true modal expansion can be used for diagonalization of the system mass and stiffness matrices. More importantly, the normal modes express some important characteristics of the model well, such as the vibration shapes near certain frequencies.

The normal modes are usually computed for an undamped system. Thus the columns of the matrix $\boldsymbol{\Phi}$ must be obtained by solving the generalized eigenvalue problem

$$\left(-\omega_j^2 \boldsymbol{M} + \boldsymbol{K}\right) c_j(\boldsymbol{\Phi}) = \boldsymbol{0} \,, \tag{5}$$

where $c_j(\boldsymbol{\Phi})$ means the $j-$th column of the matrix $\boldsymbol{\Phi}$ (i.e. the $j-$th eigenvector). The solution of (5) represents a significant cost because (a) the dimension of the matrix pencil is large, and (b) the number of required eigenvectors is typically also large because the number of retained eigenvalues must be proportional to the largest frequency of interest ω, and the largest computed eigenvalue must exceed ω by a significant margin (often taken as 50%).

The normal modes would, by definition, diagonalize the mass and the stiffness matrix. However, unless the damping is rather special, the so-called Rayleigh damping, or proportional damping as it is sometimes referred to, the damping matrix would not be diagonalized. In this work we indeed assume that the damping model is in no way limited, and hence no form of the normal modes can be found the diagonalize all the matrices.

2.2 | Load-dependent Ritz vectors

A broad class of methods uses ideas related to the notion of Krylov subspaces. Wilson et al. proposed an algorithm for generating mass-orthogonal Krylov vectors by starting from the static displacement[27]. The computation of the mass-orthogonal Krylov vectors is in this work carried out as shown in Algorithm 1, which is rephrased from Reference 27. In particular, the pseudocode corresponds line-by-line to the actual implementation in the Julia programming language[28,29].

Note that the step of solving the projected eigenvalue problem is omitted. The orthogonality of the vectors resulting from that step is not needed in our approach. The solution of the eigenvalue problem was also subsequently omitted in some publications by Wilson and collaborators[30]. We shall refer to the basis vectors generated by Algorithm 1 as the WYD Ritz basis (in recognition of the authors of Reference 27).

Algorithm 1 Algorithm WYD Ritz

Require: $\boldsymbol{K}, \boldsymbol{M}, \boldsymbol{F}$: the stiffness and mass matrix, the load vector; n_m: number of modes

$\boldsymbol{p} = \boldsymbol{K}^{-1}\boldsymbol{F}$ ▷ Use factorization instead of an inverse

$c_1(\boldsymbol{\Phi}) = \boldsymbol{p}/\|\boldsymbol{p}\|_M$

for $i = 2, \ldots, n_m$ **do**

  $\boldsymbol{p} = \boldsymbol{K}^{-1}\boldsymbol{M}c_{i-1}(\boldsymbol{\Phi})$

  $\boldsymbol{q} = \boldsymbol{M}\boldsymbol{p}$

  for $j = 1, \ldots, i-1$ **do**

    $\boldsymbol{p} = \boldsymbol{p} - \left(\boldsymbol{q} \cdot c_j(\boldsymbol{\Phi})\right) c_j(\boldsymbol{\Phi})$

  end for

  $c_i(\boldsymbol{\Phi}) = \boldsymbol{p}/\|\boldsymbol{p}\|_M$

end for

The notation $\|\boldsymbol{p}\|_M$ denotes the $\boldsymbol{M}$-induced norm of the vector $\boldsymbol{p}$

$$\|\boldsymbol{p}\|_M^2 = \boldsymbol{p} \cdot \boldsymbol{M}\boldsymbol{p} \, . \tag{6}$$

As an alternative, one could also generate a Krylov basis using the Lanczos procedure[31]. While a two-term recurrence should theoretically suffice to produce an orthogonal basis, in practice either full or periodical (occasional) re-orthogonalization of all vectors generated so far is necessary[32]. As a result, in our experiments we did not find substantial difference either in accuracy or computational cost between the Lanczos procedure or the WYD Ritz procedure. Hence we will focus in this paper on the more robust WYD Ritz procedure.

The advantage of the Krylov basis/Ritz vectors is that by definition the basis includes response to the static loading. The disadvantages include poor ability to resolve high-frequency modes. The Ritz vectors include the loading in the computation of basis, unlike the loading-independent normal modes.

The vectors of the Krylov basis are often referred to as Ritz vectors. (This use of the term should not be confused with the so-called Ritz vectors as approximations of vibration mode shapes computed during Rayleigh-Ritz iteration.)

The success of the Ritz-vector methods depends on the quality of the generated vectors. Kline explored truncation errors for linear systems approximated with reduced bases consisting of exact vibration modes and load-dependent Ritz vectors[33], and found that the residual error was composed of two parts: the first due to the inability of the truncated basis to reproduce the loading, and the second stemming from the failure of the reduced basis to reproduce the exact vibration response of the full model. The frequency truncation criteria in the process that generated the Ritz vectors for the Krylov space was presented by Ibrahimbegovic and Wilson[34].

3 | TWO-STAGE MODEL REDUCTION

As regards speeding up the computation of the FRF curves with reduced models, the key is clearly the construction of the transformation matrix $\boldsymbol{\Phi}$. Here we present a two-stage model reduction procedure. In the first stage, the original finite element model is reduced using the concept of coherent nodal clusters. In the second stage, either the free vibration problem or the WYD Ritz approach is applied to the first stage reduced system. The result of the second stage is an approximation of the matrix $\boldsymbol{\Phi}$.

3.1 | Coherent Nodal Cluster model reduction

This idea expands upon the basic tenet of the finite element method: The role of a finite element is to streamline the solution to the problem of finding suitable functions to describe the field of interest (such as displacement). The shape of a finite element is very simple, and, due to its size being typically much smaller than the dimensions of the entire domain, the variation of the field of interest within the finite element is much simpler than that within the whole domain. Hence, it becomes possible to write down a basis in which the field of interest can be expressed over the extent of a single finite element.

If one considers the goal of approximating the nodal degrees of freedom within a *cluster of nodes* that are in *close proximity* to each other, simple basis functions may be easily discovered. The variation of the quantity of interest is then patched together from the individual representations over each node cluster. The crucial observation is that the degrees of freedom at nodes that are in close proximity are *coherent*. Hence, the technique underlying the present approach may be referred to as the ***Coherent Node Clusters*** (CoNC). For the benefit of the reader we shall briefly introduce the basic notions of the coherent nodal cluster basis. More details may be found in Reference 23.

A vector $\boldsymbol{f} \in \mathrm{R}^N$ can be approximated with a reduced model as

$$\boldsymbol{f} \approx \sum_{i=1}^{m} \widetilde{d}_i \boldsymbol{b}_i \, , \tag{7}$$

where $\boldsymbol{b}_i$ are suitably selected basis vectors, $\widetilde{d}_i$ are generalized degrees of freedom, and the number of such vectors m is considerably smaller than the number of the nodes N. The question then is how to choose the basis vectors. In the CoNC technique we will choose as the basis vectors the values of the cluster basis functions b_i (these are *not* the *finite element* basis functions) at the locations of the nodes

$$\boldsymbol{b}_i = \left[\, b_i(x_1), \; b_i(x_2), \; \ldots, \; b_i(x_N) \, \right]^T \, . \tag{8}$$

For instance, for a one-dimensional mesh, one can work with the Legendre polynomial basis functions[35]. For convenience, normalized coordinates $-1 \leq \xi \leq +1$ would be preferred instead of the physical coordinates x

$$\xi = \frac{2x - x_s - x_e}{x_e - x_s} \,, \tag{9}$$

where $x_s \leq x \leq x_e$, and x_s, x_e are the minimum and maximum coordinate.

The crucial step is to divide the mesh into clusters of adjacent nodes and to construct the basis functions for each cluster separately. Again in one dimension, this simply means dividing the interval $\langle x_s, x_e \rangle$ into a number of sub-intervals, and the clusters can then be taken as the nodes located in each of the sub-intervals. An example is provided in Figure 1, where the 19 nodes of a mesh of 18 two-node elements are divided into three clusters. The basis functions in each cluster are taken to be the constant, linear, quadratic, and cubic Legendre polynomial. Hence, there will be $3 \times 4 = 12$ basis vectors. Collecting those vectors as columns of a matrix $\boldsymbol{\Psi}$ results in the reduction

$$\boldsymbol{f} \approx \boldsymbol{\Psi} \widetilde{\boldsymbol{f}} \,. \tag{10}$$

Obviously, this example is purely an illustration: reduction from 19 degrees of freedom to 12 degrees of freedom is hardly worthwhile.

Concerning the properties of the transformation matrix $\boldsymbol{\Psi}$: Note that the resulting matrix $\boldsymbol{\Psi}$ is blocked and sparse (this is an indication that the basis is local). By way of contrast, in typical applications of the Rayleigh-Ritz method this matrix would be dense, which would correspond to the basis vectors being global. Furthermore, if the basis vectors are linearly independent within each cluster, the columns of the transformation matrix $\boldsymbol{\Psi}$ will also be linearly independent.

3.1.1 | One-dimensional basis functions

The technique of the CoNC is in this work demonstrated on three-dimensional solid finite element meshes. Consequently the transformation matrices need to be constructed so that they reflect the 3 degrees of freedom per node and the three spatial coordinates. The former is incorporated by the correct assembly of the columns of the transformation matrix based on the numbering of the degrees of freedom, for the displacements u, v, w separately, while the latter requires the basis functions to be expressed on a three-dimensional domain.

By using the scaling of the coordinates for each cluster analogous to (9), the basis functions can be expressed on the bi-unit cube, $-1 \leq \xi \leq +1$, $-1 \leq \eta \leq +1$, and $-1 \leq \zeta \leq +1$. Therefore it is natural to think of the decomposition of the three-dimensional basis functions $B(\xi, \eta, \zeta)$ as products $b_i(\xi) b_j(\eta) b_k(\zeta)$ of one-dimensional basis functions on the bi-unit interval. A set of prototype basis functions $b_q(s)$ is formulated, where $1 \leq q \leq n_1$, and in the examples below these are the Legendre polynomials, sometimes called Legendre functions of the first kind, $b_i(\xi) = P_{i-1}(\xi)$. It is assumed that n_1 is a suitably chosen number of basis functions to incorporate in the construction of the transformation matrix along each spatial direction. How to choose n_1 will be addressed below.

3.1.2 | Multi-dimensional basis functions

The one-dimensional prototype basis functions can be used to construct basis functions for multiple space dimensions using the concept of tensor product functions. For simplicity only isotropic sets of basis functions are considered: for each space dimension the same number of the same kind of basis functions is used.

How to construct the basis functions in two space dimensions is illustrated in Table 1. In two spatial dimensions the two-dimensional basis functions are constructed as tensor products $b_i(\eta) b_j(\xi)$ of the one-dimensional basis functions $b_i(\eta)$ and $b_j(\xi)$. Only such i and j are admitted that satisfy $2 \leq i + j \leq n_1 + 1$.

The structure of the basis functions in three dimensions is analogous to that shown in Table 1, except that one needs a three-dimensional array to express the entries of the table. In three dimensions, the entries of the array are $B_{ijk}(\xi, \eta, \zeta) = b_k(\zeta) b_i(\eta) b_j(\xi)$, where $3 \leq i + j + k \leq n_1 + 2$. The total number of the multi-dimensional basis functions for a cluster, m, is determined by the number of the prototype one-dimensional functions, n_1. Take S to be the number of spatial dimensions, then

$$m = n_1 \text{ for } S = 1, \quad m = \frac{n_1(n_1+1)}{2} \text{ for } S = 2, \quad \text{and } m = \frac{n_1(n_1+1)(n_1+2)}{6} \text{ for } S = 3. \tag{11}$$

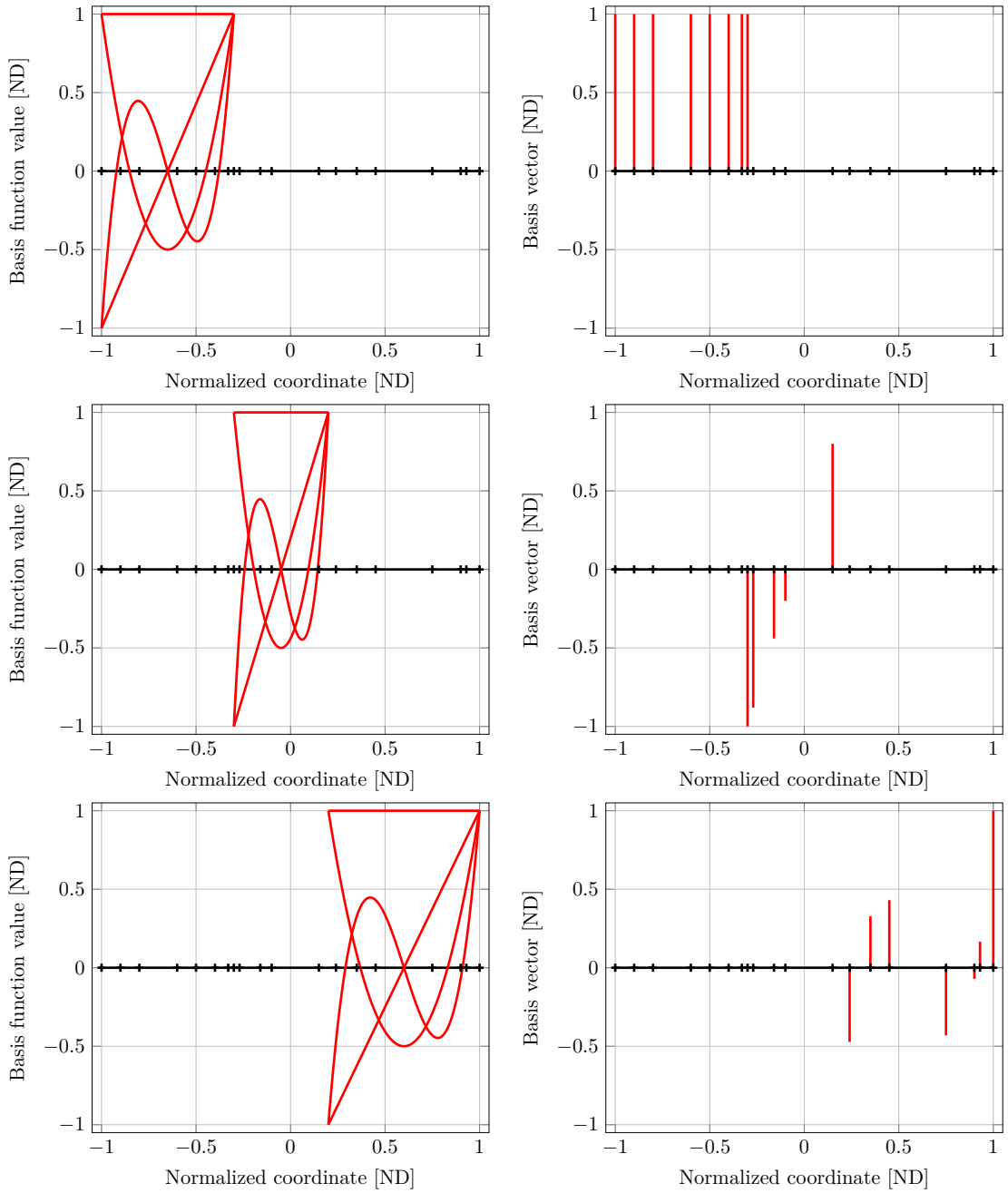

FIGURE 1 Basis functions and three selected basis vectors out of the total of 12 vectors for a one-dimensional mesh with 19 nodes. Left column: basis functions $b_j(\xi)$, $j = 1, 2, 3, 4$ in the form of Legendre polynomials, rows top to bottom: cluster 1, cluster 2, cluster 3. Right column: basis vector corresponding to the constant function within cluster 1 in row 1; basis vector corresponding to the linear function within cluster 2 in row 2; and basis vector corresponding to the cubic function within cluster 3 in row 3.

3.1.3 | Remarks

The proposed transformation is an algebraic operation that reduces the physical degrees of freedom to generalized degrees of freedom. As such, this algebraic operation does not affect in any way the variational structure of the finite element model.

TABLE 1 Construction of two-dimensional basis functions, $B_{ij}(\xi,\eta) = b_i(\eta)b_j(\xi)$, from the prototype 1-D basis functions $b_i(\eta), b_j(\xi)$. In this particular case $n_1 = 4$ is taken, and therefore there are $m = n_1(n_1+1)/2 = 10$ basis functions B_{ij}.

	$b_1(\xi)$	$b_2(\xi)$	$b_3(\xi)$	$b_4(\xi)$
$b_1(\eta)$	$b_1(\eta)b_1(\xi)$	$b_1(\eta)b_2(\xi)$	$b_1(\eta)b_3(\xi)$	$b_1(\eta)b_4(\xi)$
$b_2(\eta)$	$b_2(\eta)b_1(\xi)$	$b_2(\eta)b_2(\xi)$	$b_2(\eta)b_3(\xi)$	
$b_3(\eta)$	$b_3(\eta)b_1(\xi)$	$b_3(\eta)b_2(\xi)$		
$b_4(\eta)$	$b_4(\eta)b_1(\xi)$			

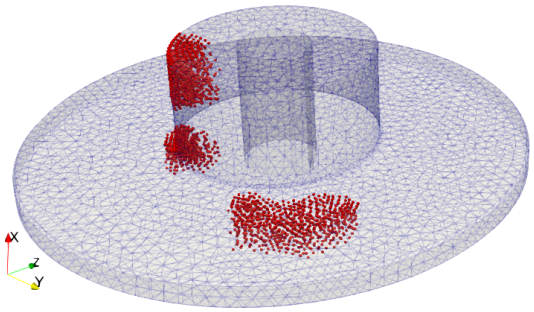

FIGURE 2 Tetrahedral mesh of a brake disc with a few randomly selected node clusters visualized with spherical markers. There are 64 clusters in the mesh in total.

Again due to the algebraic character of the transformation of the degrees of freedom, the continuity of the underlying finite element formulation is preserved. The method is applied below to C^0 finite elements (linear and quadratic). This smoothness is preserved by the model reduction procedure.

3.1.4 | Choosing the number of the clusters and the number of the basis functions

How many clusters and how many basis functions per cluster should be used? This question has been studied in Reference 23, and we shall not repeat the details here. Rather, refer to Algorithm 2 from Reference 23, which provides an estimate from the basic properties of the vibrating solid and the frequency range of interest. Here we follow precisely this algorithm except for the clamping of the number of basis functions n_1. The number of basis functions n_1 is clamped from both below and from above to ensure both a sufficient number of such functions to generate a reasonably rich approximation, and to prevent an unnecessarily expensive computation with too many basis functions. Here we use $3 \leq n_1 \leq 6$.

3.2 | Generation of node clusters

Previously the simple Recursive Inertial Bisection (RIB) was adopted for the partitioning of the nodes into spatially coherent clusters. Some researchers trace this partitioning technique to the work of Williams [36]. The technique is very simple and relatively quick (for instance for a mesh with 100,000 nodes the partitioning can be executed within a small fraction of a second). It produces relatively compact node clusters. The disadvantage of RIB is that the connectivity of the finite element mesh is not taken into account. This means that nodes may be members of the same cluster even when separated by the outer boundary, e.g., the faces of a sharp crack. For our purpose, nodes on the opposite sides of the crack should not be part of the same cluster.

A more general technique is available in the unstructured graph partitioning library Metis [37,38]. In this work we partition the connectivity matrix using the k-way partitioning algorithm from Metis [37]. Figure 2 shows some nodal clusters obtained via k-way partitioning in a brake disc meshed with quadratic tetrahedra.

3.3 | Stage 1 of model reduction

We shall use the coherent-nodal-cluster transformation matrix to perform a Galerkin projection to complete the first stage. The original linear algebraic equation is reduced to considerably smaller dimensions using the CoNC transformation matrix $\boldsymbol{\Psi}$

$$\left(-\omega^2\overline{\boldsymbol{M}}_r + \mathrm{i}\omega\overline{\boldsymbol{C}}_r + \mathrm{i}\overline{\boldsymbol{D}}_r + \overline{\boldsymbol{K}}_r\right)\overline{\boldsymbol{U}}_r = \overline{\boldsymbol{F}}_r \ , \tag{12}$$

where

$$\overline{\boldsymbol{M}}_r = \boldsymbol{\Psi}^T\boldsymbol{M}\boldsymbol{\Psi} \ , \quad \overline{\boldsymbol{C}}_r = \boldsymbol{\Psi}^T\boldsymbol{C}\boldsymbol{\Psi} \ , \quad \overline{\boldsymbol{D}}_r = \boldsymbol{\Psi}^T\boldsymbol{D}\boldsymbol{\Psi} \ , \quad \overline{\boldsymbol{K}}_r = \boldsymbol{\Psi}^T\boldsymbol{K}\boldsymbol{\Psi} \ , \quad \overline{\boldsymbol{F}}_r = \boldsymbol{\Psi}^T\boldsymbol{F} \ . \tag{13}$$

The reduced system has coefficient matrices which are sparse and symmetric. Even though the number of modes (columns of $\boldsymbol{\Psi}$) may be in the thousands, it is still considerably smaller than the dimension of the original finite element model.

3.4 | Stage 2 of model reduction

Next the reduced model (13) is piped into the second stage, where the normal modes (Section 2.1) or the WYD Ritz modes (Section 2.2) are computed from the matrices $\overline{\boldsymbol{M}}_r$, $\overline{\boldsymbol{K}}_r$. We shall thereby construct a rectangular matrix $\overline{\boldsymbol{\Phi}}_r$, with the modes of the reduced system (13) as columns.

Therefore, the transformation matrix $\boldsymbol{\Phi}$ produced in the second stage can be detailed as

$$\boldsymbol{\Phi} = \boldsymbol{\Psi}\overline{\boldsymbol{\Phi}}_r \ . \tag{14}$$

The transformation matrix $\boldsymbol{\Phi}$ is an *approximation* of the modes that could be constructed by applying the normal-mode or the WYD Ritz procedures directly to the original finite element system. It also consists of globally-supported vectors, typically numbering tens or hundreds, in some applications possibly more (we presented a use case where the required number of modes exceeded 1000 in Reference 23).

When the transformation matrix $\boldsymbol{\Phi}$ is applied to (2), the resulting system with dense complex matrices can be solved relatively quickly, as demonstrated in our examples below. The main contribution of the present paper is the reduction of the work that goes into constructing the transformation matrix $\boldsymbol{\Phi}$. By eliminating the major cost of this computation, the entire calculation of the frequency response functions can be significantly accelerated.

4 | ILLUSTRATION OF THE TWO-STAGE MODEL REDUCTION PERFORMANCE

The examples used in this work have real symmetric system matrices, and so the WYD basis and normal vectors are real-valued. The example considered in this section is a cantilevered twisted bar of aluminum (refer to Figure 3). The cantilever length is 250 mm and cross-section dimensions are 15 × 25 mm. The Young's modulus, the Poisson ratio, and the mass density are taken as 70000 MPa, 0.33, 2700 $\mathrm{kg/m^3}$. Figure 3 shows the coarsest mesh considered. The finite element model is based on the nodally-integrated four-node tetrahedra with stabilization as detailed in Reference 39. The meshes used in this example are given in Table 2. The FRF was computed at 200 frequencies between 100 and 20,000 Hz. The discrete finite element model has 18 natural frequencies within the frequency range.

The damping model adopted in this example was the proportional Rayleigh viscous damping, but based upon the one-point tetrahedral rule executed over the elements, not the nodal integration rule. Hence the reduced matrices were not diagonal under any circumstances.

The clusters in this example were computed using the RIB partitioning of Section 3.2. The parameters of the reduced models are shown in Table 3. In particular, the number of clusters was fixed, and the number of one-dimensional functions was progressively increased as the number of nodes per cluster increased (up to a fixed maximum of 6).

4.1 | Direct versus reduced approach

First, let us look at the run times of the direct method compared with the usual one-stage reduced models (free vibration and WYD Ritz): The full finite element model for mesh 8 with ≈140,000 degrees of freedom was considered. The direct approach

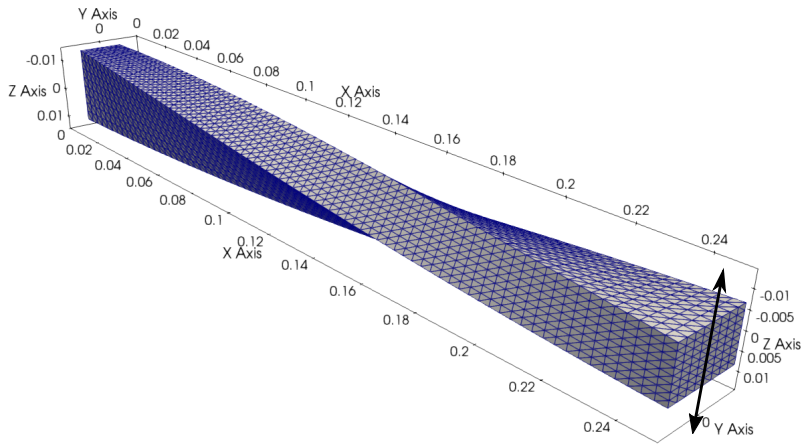

FIGURE 3 Coarsest mesh of the twisted bar with 6561 nodes (undeformed). The arrow indicates the applied loading uniformly distributed in the cross-section. The opposite face is clamped. The dimensions shown are in meters.

TABLE 2 Twisted bar. Meshes for the nodally-integrated four-node tetrahedra with stabilization[39].

Mesh number [ND]	Number of nodes [ND]
4	6561
6	20449
8	46529
10	88641
12	150625
14	236321
16	349569

TABLE 3 Twisted bar. Parameters of the stage 1 reduced model.

Mesh number [ND]	Number of 1D basis functions n_1 [ND]	Number of clusters [ND]
4	3	32
6	5	32
8	6	32
10	6	32
12	6	32
14	6	32
16	6	32

ran for more than 10 hours (36240 seconds), where essentially all of this time was spent in the sweep. Clearly here we could save significantly with the mode-based transformation: the eigenvalue problem seeking 200 modes for the full model could be solved in 518 seconds, so almost 70 times faster than the sweep for the direct approach. The sweep with the reduced model based on 200 modes then took negligible 13 seconds. The WYD Ritz Reduced model for the same number of modes constructed the transformation matrix in 192 seconds, so almost three times faster than the free-vibration normal-mode basis construction. And the sweep was again executed in 13 seconds, because the dimension of the reduced model was the same. Hence we can conclude that direct approach took roughly two orders of magnitude more time then the reduced model to compute the FRF curve, where in the cost of the reduced model we include the construction of the transformation matrix from the full to the reduced model and the cost of the sweep itself.

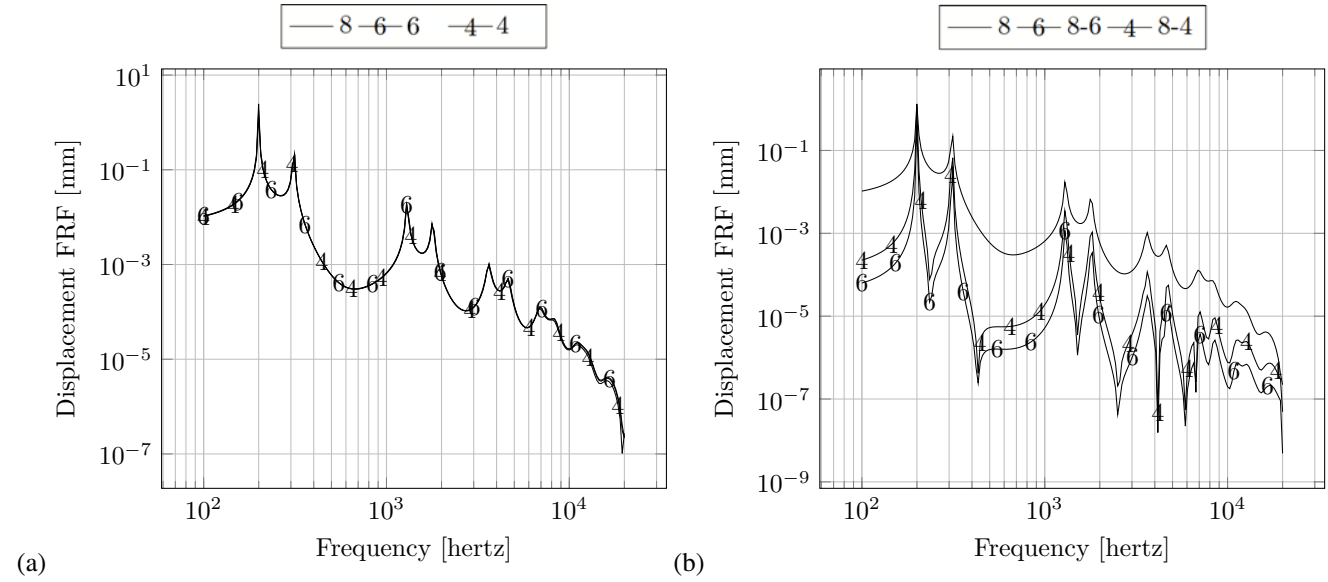

FIGURE 4 Twisted bar. Direct solutions with meshes 4, 6, and 8. FRF of the y - displacement of the center of the loaded face. (a) Amplitudes, and (b) differences between the solutions for meshes 8 and 6 and 8 and 4. The FRF curve for mesh 8 is shown without markers.

4.2 | First look at accuracy

It is necessary at this point to take to account the difficulties associated with comparing FRF curves. We may consider the sources of error to be of two kinds: errors due to the mismatch of the amplitudes of the peaks in the response, and errors due to the shifts of those peaks in frequency. Both of these sources of errors can contribute to the perceived error in amplitude: Consider a curve with a single sharp peak and a version of this curve slightly shifted in frequency. There will be a mismatch between amplitudes at any given frequency, even though the curves may be visually virtually identical.

Figure 4 shows the FRF curves for three progressively more refined meshes of the twisted bar. All have been obtained with the direct approach (i.e. no model reduction was involved). First, in the subfigure ad (a) we look at the three curves superimposed in a single chart: While some differences may be discovered between the three curves for higher frequencies, for lower frequencies the curves are nearly visually indistinguishable.

The subfigure ad (b) presents the *differences* between the FRF curves computed with different meshes. First we show the curve obtained with the mesh 8, and then the differences of the results obtained with meshes 6 and 4 and with the mesh 8. The difference curves display non-negligible amplitudes, especially at the first two peaks. Since the amplitudes in the subfigure (a) can be seen to be nearly identical, most of the differences between the models are due to the peaks being (slightly) shifted in frequency.

Next we look at the FRF curves obtained with the reduced models. The differences now will be taken between the direct solution with mesh 8 and the reduced models for the same mesh. The reduced models employed in Stage 1 are described by the parameters of Table 3. The number of clusters is the same for all the meshes. The number of the one-dimensional basis functions is fixed from mesh 8 onwards.

The following key is used in the figures to reference a particular reduced model:

FV The reduction matrix is computed by solving the free-vibration eigenvalue problem for the original finite element model.

W The WYD Ritz basis for the transformation matrix is computed for the original finite element model.

2SV The two-stage free vibration model reduction: The reduction matrix is computed by solving the free-vibration eigenvalue problem of the first-stage reduced model.

2SW The two-stage WYD Ritz construction of the reduced model: The reduction matrix is computed by applying the WYD Ritz algorithm to the first-stage reduced model.

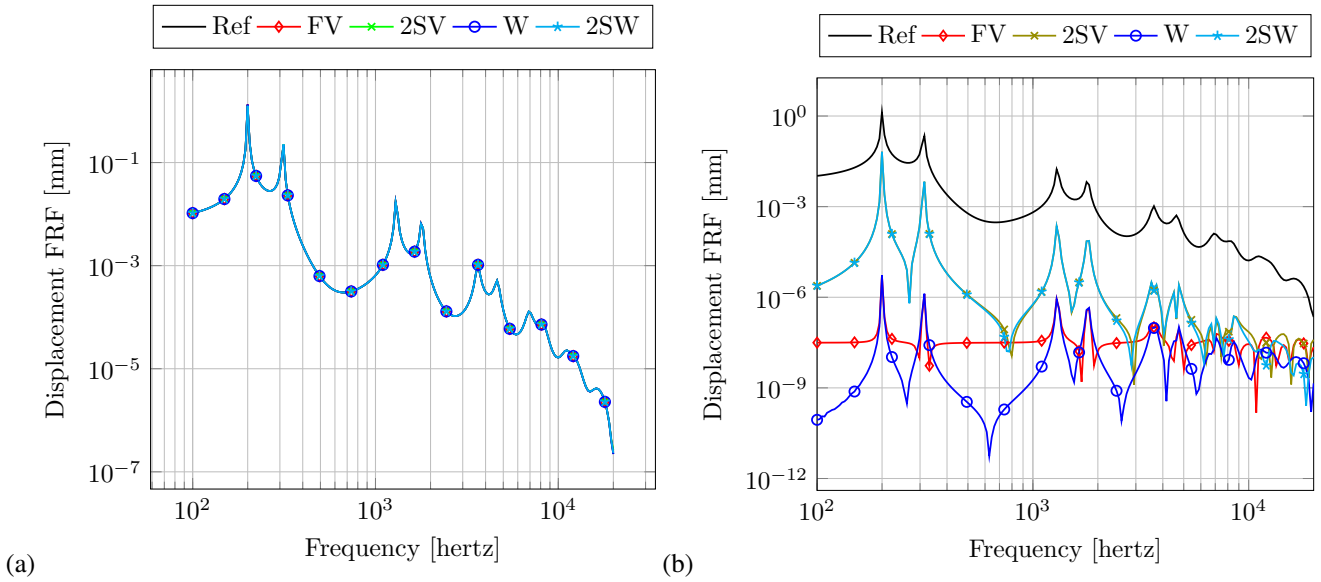

FIGURE 5 Twisted bar. Finite element model based on mesh 8. Reduced models with 200 modes. FRF of the y - displacement of the center of the loaded face. (a) All solutions, direct and reduced, overlaid. (b) Solid line without markers: direct method. The curves with markers for the *reduced* models are the difference between the direct solution and the mode-based reduced model.

Figure 5(a) shows the reduced-model FRF curves superimposed upon the direct solution curve (marked **Ref**). There are no major visual differences between the direct solution and the reduced models.

Figure 5(b) shows that the **W** one-stage model produces the smallest errors relative to the direct solution. The expansion into the normal modes of the full model (**FV**) gives errors relative to the direct solution that are comparable to the **W** reduced model at the peaks of the FRF curve; elsewhere the error is worse, but insignificant compared to the differences between the direct solutions with different meshes (Figure 4).

The two-stage reduced models **2SV** and **2SW** result in differences relative to the direct solution more than three orders of magnitude larger than the one-stage reduced models. This may be due to slight shifts in frequency as well as amplitude mismatch: The vibration spectrum of the reduced models is known to be different from the spectrum of the original finite element model. Clearly, the small changes in frequency of the resonance will produce noticeable differences in the FRF curves. Nonetheless, all of the computed results may be good enough for practical purposes (as evidenced in Figure 5(a)).

So, the results of the two-stage reduced models indicate that some accuracy is being lost. What is then gained? Figure 6 presents the time necessary to construct the transformation matrix between the reduced model for the sweep, either one-stage or two-stage, and the original full finite element model. Since the model reduction procedure depends on the number of modes (i.e. the dimension of the reduced model) we present the timing for 25 modes ad (a) and 400 modes ad (b). The computed points are labeled with the base-10 logarithms of the timing in seconds.

The **W** one-stage model is consistently faster than the **FV** free-vibration model (as reported already in the original publication[27]). For instance, for the full finite element model with more than one million degrees of freedom, when the reduced model has 400 modes, the **W** one-stage model is more than twice as fast ($10^{4.07} = 11,749$ seconds) as the **FV** free-vibration model ($10^{4.43} = 26,915$ seconds). Both two-stage models are more than an order of magnitude faster than the one-stage reduced finite element models: for instance **2SV** takes $10^{2.54} = 347$ seconds, i.e. it is more than thirty times faster than the **W** one-stage model.

5 | ADAPTIVE IMPROVEMENT OF THE TWO-STAGE MODEL ACCURACY

The loss of accuracy of the two-stage models relative to the one-stage models is usually most pronounced at resonances. A simple strategy can be used for the free vibration two-stage models to improve the fit near the resonant peaks of the FRF curve.

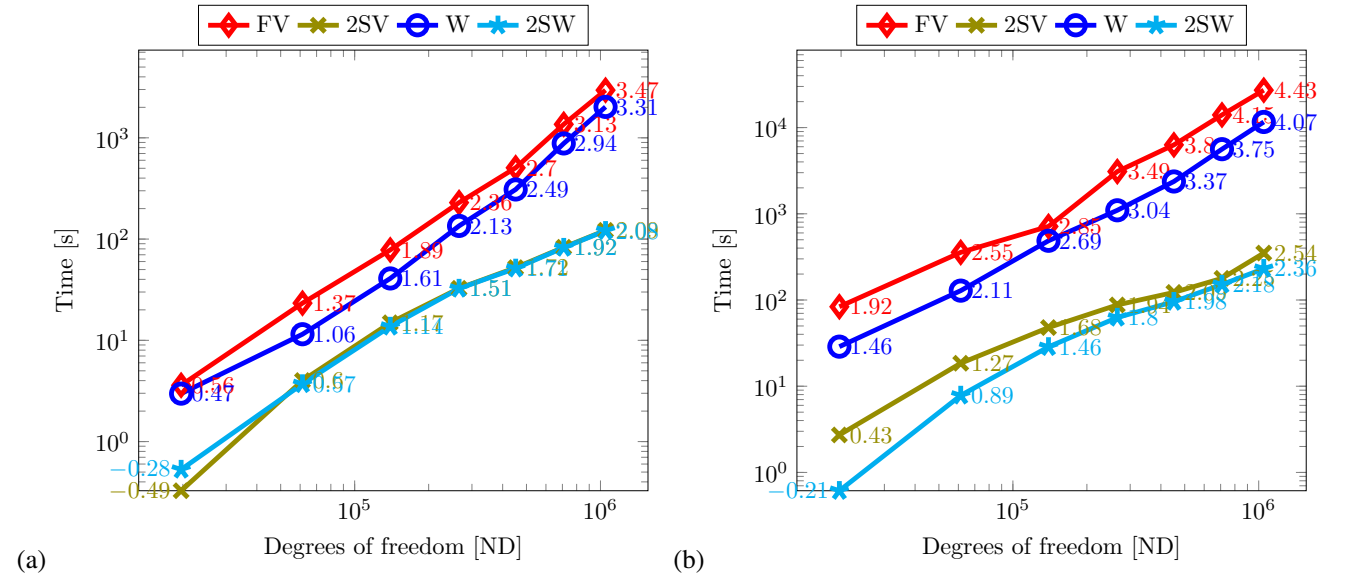

FIGURE 6 Twisted Bar. Comparison of time needed to construct the transformation matrix $\boldsymbol{\Phi}$ (i.e. the modes) for meshes 4, 6, 8, 10, 12, 14, 16. (a) Reduced models based on 25 modes. (b) Reduced models based on 400 modes.

5.1 | Two-stage free vibration procedure

Consider the equation of free resonant undamped motion of the full finite element model

$$\left(-\omega_j^2 \boldsymbol{M} + \boldsymbol{K}\right) \boldsymbol{\phi}_j = \boldsymbol{S}(\omega_j)\boldsymbol{\phi}_j = \boldsymbol{0}\,, \tag{15}$$

where ω_j and $\boldsymbol{\phi}_j$ are the natural angular frequency and the mode shape of the full system, and $\boldsymbol{S}(\omega_j)$ is the dynamic stiffness. We can think of the mode shape $\boldsymbol{\phi}_j$ as the eigenvector of the dynamic stiffness $\boldsymbol{S}(\omega_j)$ associated with a zero eigenvalue.

The two-stage reduced model provides us with an *approximation* of the two resonance quantities, $\widetilde{\omega}_j$ and $\widetilde{\boldsymbol{\phi}}_j$. Hence we get the residual

$$\boldsymbol{S}(\widetilde{\omega}_j)\widetilde{\boldsymbol{\phi}}_j = \boldsymbol{r}(\widetilde{\boldsymbol{\phi}}_j) \neq \boldsymbol{0}\,, \tag{16}$$

since $\boldsymbol{S}(\widetilde{\omega}_j)$ is not singular, and $\widetilde{\boldsymbol{\phi}}_j$ is not an eigenvector of $\boldsymbol{S}(\widetilde{\omega}_j)$, and of course it is not an eigenvector of $\boldsymbol{S}(\omega_j)$.

While both $\widetilde{\omega}_j$ and $\widetilde{\boldsymbol{\phi}}_j$ are inaccurate (approximate), our concern is here with the ability to represent response of the system for frequencies that are close to resonances, not just at the resonances themselves. Therefore it would be helpful to obtain a better approximation of $\boldsymbol{\phi}_j$. If we had the true ω_j, we could apply the algorithm of inverse power iteration as

$$\boldsymbol{S}(\omega_j)\widetilde{\boldsymbol{\phi}}_j^{(1)} = \widetilde{\boldsymbol{\phi}}_j^{(0)}\,, \tag{17}$$

to obtain $\widetilde{\boldsymbol{\phi}}_j^{(1)}$ as an improved approximation of the eigenvector, using $\widetilde{\boldsymbol{\phi}}_j^{(0)} = \widetilde{\boldsymbol{\phi}}_j$ as the initial guess.

The true angular frequency ω_j is not available, but we could approximate (17) as

$$\boldsymbol{S}(\widetilde{\omega}_j)\widetilde{\boldsymbol{\phi}}_j^{(1)} = \widetilde{\boldsymbol{\phi}}_j^{(0)}\,. \tag{18}$$

Convergence to the true eigenvector of $\boldsymbol{S}(\widetilde{\omega}_j)$ could be quick, because $\boldsymbol{S}(\widetilde{\omega}_j)$ is nearly singular, but we cannot afford more than one iteration of (18). In fact, we cannot afford an accurate solution of (18), since the matrix $\boldsymbol{S}(\widetilde{\omega}_j)$ has the dimensions of the original finite element model. Fortunately, we can obtain a decent approximation of $\widetilde{\boldsymbol{\phi}}_j^{(1)}$ from (18) using just a few iterations of a suitable algorithm. Given the properties of the coefficient matrix – it is not positive definite, although it is symmetric – the MINRES algorithm without preconditioning is suitable[40], and we can get a decent approximation of $\widetilde{\boldsymbol{\phi}}_j^{(1)}$ with just $O(10)$ iterations, which is quite inexpensive.

The approximation $\widetilde{\boldsymbol{\phi}}_j^{(1)}$ can be used as an additional basis vector (column) of the transformation matrix $\boldsymbol{\Phi}$. Alternatively, this basis vector can be swapped for a column of $\boldsymbol{\Phi}$ that corresponds to a high frequency (which usually has only a slight effect on the FRF curve at lower frequencies, and hence is "dispensable").

The third option is to replace the approximate mode shape $\widetilde{\boldsymbol{\phi}}_j$ with $\widetilde{\boldsymbol{\phi}}_j^{(1)}$. In other words, one column of the transformation matrix, the one with the mode shape j, is updated. That is the choice adopted here.

5.2 | Two-stage WYD vector procedure

The enhancement of the basis suggested above is well-suited for the **2SV** reduction because **2SV** estimates approximate natural frequencies during the computation of the basis. We use a different enhancement for the **2SW** reduction by using dynamic damped response vectors at some frequency values improved by iteration.

The 2SW reduced solution of the steady-state dynamics problem at some frequency of interest ω is computed from (2) as

$$\widetilde{\boldsymbol{U}}(\omega) \approx \boldsymbol{\Phi}\boldsymbol{U}_r(\omega)\,, \tag{19}$$

where $\widetilde{\boldsymbol{U}}(\omega)$ is the 2SW reduced approximation of the steady-state dynamic response of the full model at the frequency ω. Because of the approximate nature of $\widetilde{\boldsymbol{U}}(\omega)$, we get the residual from the balance equation of the full finite element model

$$\boldsymbol{F} - \left(\boldsymbol{K} - \omega^2\boldsymbol{M} + \mathrm{i}\omega\boldsymbol{C} + \mathrm{i}\boldsymbol{D}\right)\widetilde{\boldsymbol{U}}(\omega) = \boldsymbol{R}\left(\omega, \widetilde{\boldsymbol{U}}(\omega)\right) \neq \boldsymbol{0}\,. \tag{20}$$

By using the **2SW** reduced solution as the initial solution, the residual at the frequency of interest can be decreased by using a suitable iteration algorithm. Algorithms such as MINRES can be employed to compute the direct solution to the SSD problem starting from the 2SW reduced solution (19) as the initial guess. Evidently, iterating the MINRES algorithm until convergence (i.e. when the residual $\boldsymbol{R}\,(\omega, \boldsymbol{U})$ vanishes) is computationally expensive, but a few iterations of MINRES applied to (20) improve the SSD response vectors significantly, and the resulting response vectors can be used as enhancement vectors.

The iteration results in $\widetilde{\boldsymbol{U}}(\omega) + \Delta\widetilde{\boldsymbol{U}}(\omega)$, where the incremental change $\Delta\widetilde{\boldsymbol{U}}(\omega)$ includes portion of the response that the **2SW** basis could not capture, and so the MINRES iterated response vectors are linearly independent from the **2SW** basis $\boldsymbol{\Phi}$. Both the real and imaginary parts of the increment of the improved response $\Delta\widetilde{\boldsymbol{U}}(\omega)$ can be considered for the enhancement vectors as columns of an add-on matrix $\boldsymbol{\Xi}$, whose column space is now outside $\texttt{range}(\boldsymbol{\Phi})$. These enhancement vectors are next orthogonalized by solving a small eigenproblem,

$$\left(\boldsymbol{\Xi}^T\boldsymbol{\Xi}\right)\boldsymbol{C} = \boldsymbol{C}\boldsymbol{\Lambda}\,, \tag{21}$$

and the vectors corresponding to near-to-zero eigenvalues (10^{-12} herein) in $\boldsymbol{\Lambda}$ are eliminated to obtain a linearly independent set of enhancement vectors in the matrix of enhancement vectors $\boldsymbol{\Xi}$. The transformation matrix corresponding to the **2SWE** reduction (E for enhanced) then results as

$$\widehat{\boldsymbol{\Phi}} = \left[\boldsymbol{\Phi}\,,\ \boldsymbol{\Xi}\right]\,, \tag{22}$$

where $\boldsymbol{\Phi}$ is the 2SW basis, and $\boldsymbol{\Xi}$ is the matrix of enhancement vectors.

Computing the additional enhancement vectors adds to the computational cost. However, some computational cost can be avoided by saving some projected matrices. A matrix $\boldsymbol{A}$ (representative of the global stiffness/mass/viscous damping/structural damping operators) is first projected during the computation of enhancement vectors as

$$\boldsymbol{A}_{r1} = \boldsymbol{\Phi}^T\boldsymbol{A}\boldsymbol{\Phi}\,. \tag{23}$$

Such reduced matrices are used to compute the **2SW** reduced responses at a handful of frequencies chosen for computing the enhancement vectors. The MINRES algorithm is applied to compute the incremental improvements of the response vectors at the selected frequencies, which are then orthogonalized as shown in (21) to form the matrix of enhancement vectors $\boldsymbol{\Xi}$. The projections made during the computation of enhancement vectors (23) are reused in the SSD using the **2SWE** basis. The projection of $\boldsymbol{A}$ using the **2SWE** basis is computed as

$$\boldsymbol{A}_r = \begin{bmatrix} \boldsymbol{A}_{r1} & \boldsymbol{\Phi}^T\boldsymbol{A}\boldsymbol{\Xi} \\ \texttt{SYMM} & \boldsymbol{\Xi}^T\boldsymbol{A}\boldsymbol{\Xi}\,. \end{bmatrix} \tag{24}$$

This implies that we only need to compute $\boldsymbol{\Phi}^T\boldsymbol{A}\boldsymbol{\Xi}$ and $\boldsymbol{\Xi}^T\boldsymbol{A}\boldsymbol{\Xi}$ projections to enhance the reduced basis for the harmonic vibration analysis of the **2SWE** reduced model, which is relatively fast when the number of enhancement vectors is much less in comparison with the number of WYD modes used for the **2SW** reduction. The **2SWE** SSD response is then computed as

$$\begin{aligned} \left(\boldsymbol{K}_r - \omega^2\boldsymbol{M}_r + \mathrm{i}\omega\boldsymbol{C}_r + \mathrm{i}\boldsymbol{D}_r\right)\boldsymbol{U}_r &= \boldsymbol{F}_r\,, \\ \boldsymbol{U}(\omega) &\approx \widehat{\boldsymbol{\Phi}}\boldsymbol{U}_r(\omega)\,. \end{aligned} \tag{25}$$

where the reduced quantities were obtained with $\widehat{\boldsymbol{\Phi}}$.

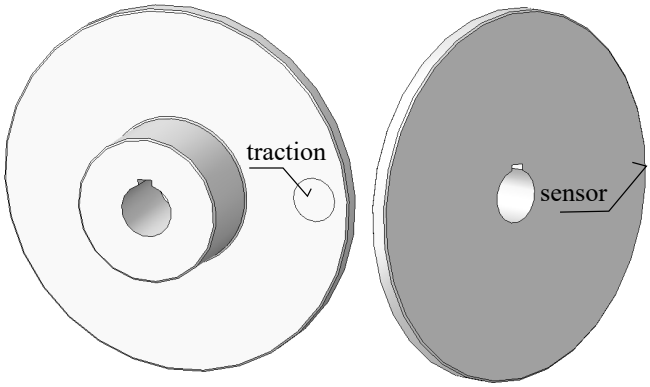

FIGURE 7 Brake disc. Visualization of the circular patch loaded by circumferential, in-plane traction, and the location of the sensor of the axial displacement on the right.

TABLE 4 Brake disc. Meshes of quadratic ten-node tetrahedra.

Mesh number [ND]	Number of nodes [ND]
3	52463
4	130051
5	246934

TABLE 5 Brake disc. Parameters of the stage 1 reduced model

Mesh number [ND]	Number of 1D basis functions n_1 [ND]	Number of clusters [ND]
3	4	278
4	5	278
5	6	278

5.3 | Examples of improvements of the accuracy of two-stage reductions

The mechanical part in this example is a brake disc B 250/12,5 bore ⌀ 36 (Figure 7). It was downloaded as a Solidworks CAD part[41], part number 10-08012001-119755. The disc was loaded on a circular patch on one of the faces, and the response of one of the points on the circumference of the disc out of the plane of the disc was monitored. We enhance the subspace spanned by the **2SV** and **2SW** methods to achieve improved results without much of an increase in the computational time. Figure 8 illustrates the FRF curve in the sweep range between 1000 and 20000 Hz, computed at 10^4 frequencies.

5.3.1 | Enhancement of the two-stage free vibration basis

The structural damping model was employed in this case, with a loss factor of 0.001. The finite element model used standard ten-node quadratic tetrahedra (Table 4). The frequency response curves were computed between 1000 Hz and 20,000 Hz at 10,000 frequency samples.

The clusters in this example were computed using the k-way graph partitioning using Metis of Section 3.2. The parameters of the reduced models are shown in Table 5. In particular, the number of clusters was fixed, and the number of one-dimensional functions was progressively increased as the number of nodes per cluster increased.

Figure 9(a) zooms in on the range 1200 Hz to 3200 Hz. The two-stage model captures the overall character of the curve quite well, but there is noticeable shift in frequency: Here the example considered is a brake disc. We enhance the subspace spanned by the **2SV** and **2SW** methods to achieve improved results without much increase in the computational time. Figure 8 illustrates the FRF curve in the sweep range between 1000 and 20000 Hz, computed at 10^4 frequencies.

5.3.2 | Enhancement of the two-stage free vibration basis

Figure 9(a) zooms in on the range 1200 Hz to 3200 Hz. The two-stage model captures the overall character of the curve quite well, but there is noticeable shift in frequency.

The full finite element model predicts the natural frequencies as [1443.2, 1443.8, 1478.3, 1685.2, 1685.4] Hz, whereas the one-stage reduced model introduces errors because it predicts the frequencies as [1482.6, 1486.0, 1518.9, 1721.8, 1723.8] Hz. In order to control the error around the 1480 Hz peak, we can use the above approach by applying the correction to the mode shapes at the frequencies 1-5 (FRF curve peaks 1, 2, and 3).

Interestingly, this can be done *after* the free-vibration two-stage model has computed the solution. For the selected modes we simply compute the improved modes (18) (and hence an improved transformation matrix), and we update either the entire FRF curve, or just a part of it as in Figure 9(b): The solution was recalculated with the improved transformation matrix $\mathbf{\Phi}$ for 800 frequencies between 1300 and 3100 Hz.

In this case the mode shape corrections were obtained by 20 iterations with the MINRES solver (without preconditioning)[42]. The shift in frequency at the first three peaks was successfully mitigated. Consequently, the magnitude of the residual in (16) decreased due to the improved transformation by more than an order of magnitude (approximately to 5% of the original residual norm) compared to the residual for the 2SV basis.

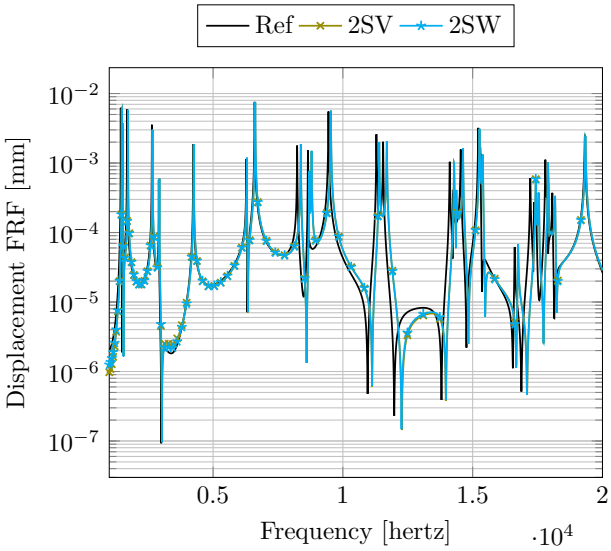

FIGURE 8 Brake disc. Mesh 3. FRF curve between 1000 and 20000 Hz, solution computed at 10^4 points, using 100 modes in the transformation matrix. Reference solution: one-stage normal-mode reduced model. Accelerated solution with the two-stage model based on normal modes shown in color with markers.

5.3.3 | Enhancement of the two-stage WYD basis

Figure 10(a) zooms in on the range 1200 Hz to 3200 Hz. Just as in the case of **2SV**, the **2SW** reduced model predicts the overall FRF character well, albeit with a noticeable phase difference. Unlike the **2SV** reduction, **2SW** does not compute approximate natural frequencies of the model. Therefore, we enhance the **2SW** basis by adding additional vectors which are obtained from the iterates of the `MINRES` algorithm applied to (20). We evaluate such enhancement vectors at uniformly spaced frequencies [1000, 1400, 1800, 2200, 2600, 3000] Hz. The choice of the frequencies for enhancement is dependent purely on the

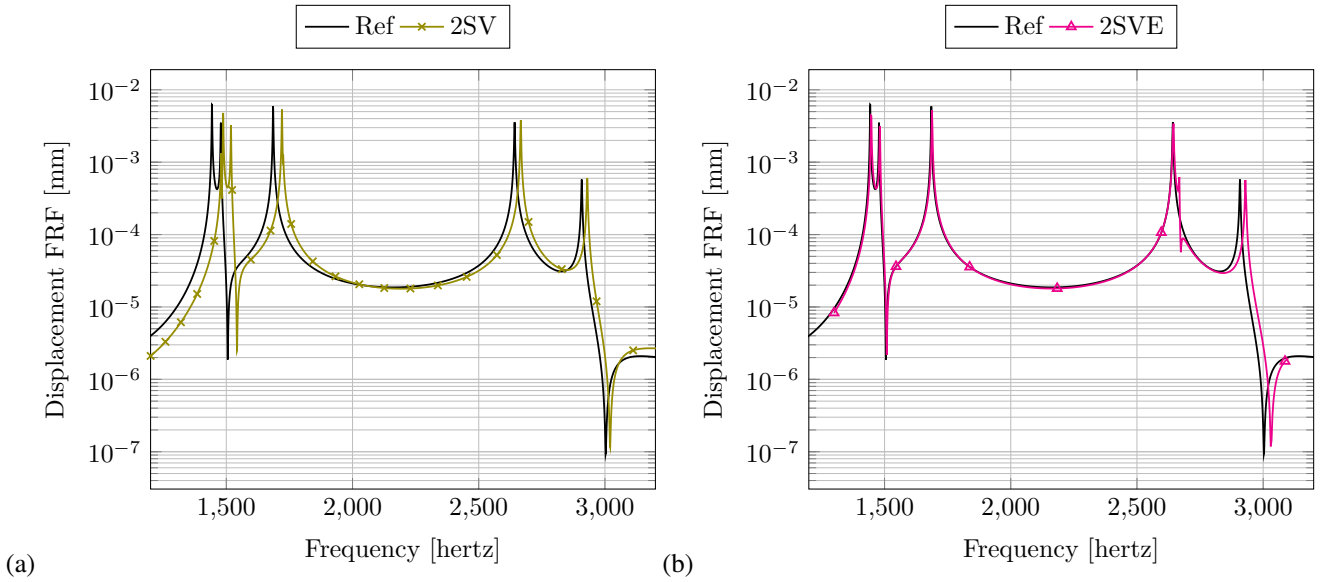

FIGURE 9 Brake disc. Detail of Figure 8: Zoom of the FRF curve between 1200 and 3200 Hz. (a) Comparison of the one-stage free-vibration model (**Ref**) with the free-vibration two-stage finite element model. (b) Comparison of the one-stage free-vibration model (**Ref**) with the free-vibration two-stage finite element model with enhancement of 5 modes corresponding to the three peaks around 1500 Hz (labeled 2SVE).

accuracy concerns of the user, and herein it is arbitrary. After orthogonalization (Equation 21) of the 12 vectors obtained from the MINRES-enhanced responses at these 6 frequencies, 7 eigenvectors corresponding to significant eigenvalues (i.e. non zero) were used for the enhancement matrix $\boldsymbol{\Xi}$.

Figure 10(b) shows the improvement in the FRF curve due to the addition of enhancement vectors to the **2SW** basis, and it can be observed that the phase difference is now negligible. Note that the improvement extends to 3000 Hz, because our choice of the frequencies at which to improve the basis includes that frequency.

5.3.4 | Performance of the enhanced two-stage reductions

Figure 11 presents the timing results when 400 modes are employed for the meshes 3, 4, and 5, comparing the one-stage free-vibration and WYD Ritz models, two-stage free-vibration and two-stage WYD Ritz, and the two-stage free-vibration with enhancement of the six modes corresponding to the first three peaks in the FRF curve, using 40 iterations of MINRES, labeled 2SVE, and two-stage WYD Ritz with enhancement vectors (real and imaginary parts of improved responses obtained using 40 MINRES iterations added from 6 chosen frequencies between 1000 Hz and 3000 Hz), labeled 2SWE.

The two-stage free-vibration required 912 seconds. The enhancement to construct **2SVE** cost further 134 seconds. Thus, the enhanced two-stage free-vibration (**2SVE**) required only half of the time of the one-stage WYD Ritz (**W**), and only $1/8$ of the time it took to execute the one-stage free-vibration procedure (**SV**). The **2SWE** procedure is even more cost effective. Note that the time reported includes the Galerkin projection, which is quite inexpensive because we can use the stored blocks of the projection employed before the enhancement.

It would be worth while to mention here that the enhancement of the individual mode shapes is embarrassingly parallel: all modes can be subjected to enhancement at the same time. This would further reduce the cost of an already relatively inexpensive procedure.

6 | CONCLUSIONS

We proposed a model-reduction method for harmonic vibration analysis using two-stage model-reduction procedures, where the first stage uses the Coherent Nodal Clustering (CoNC) reduction[23]. The proposed two-stage procedures are much cheaper than the conventional model-reduction procedures such as those which employ the solution of a free vibration problem, or which

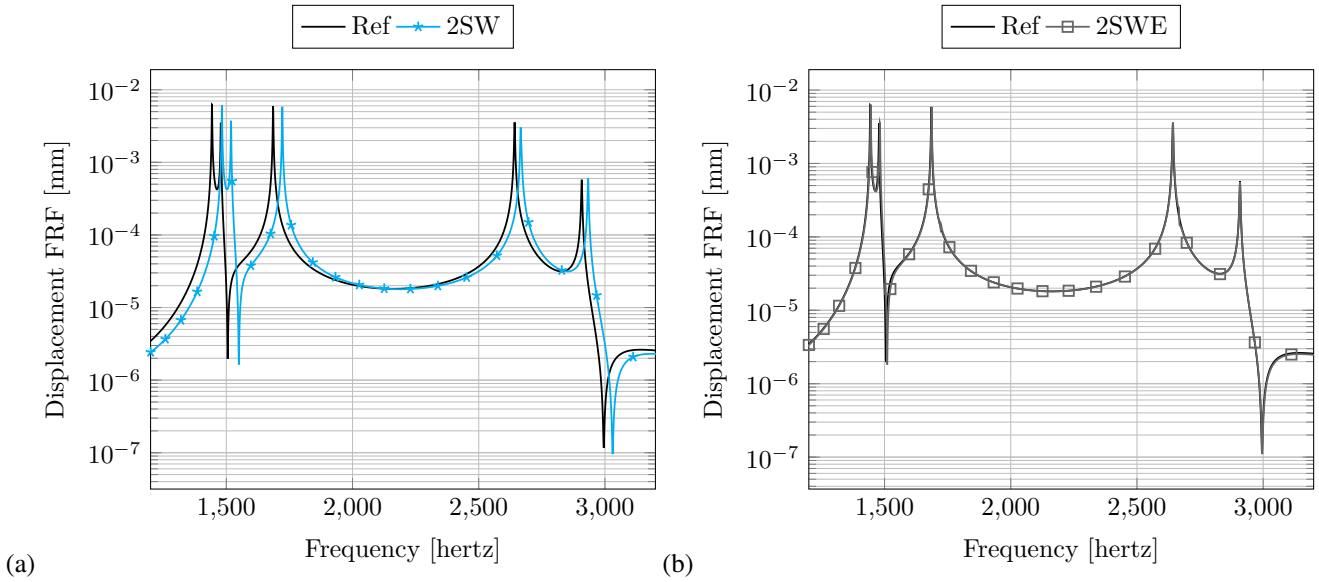

FIGURE 10 Brake disc. Detail of Figure 8: Zoom of the FRF curve between 1200 and 3200 Hz. (a) Comparison of the one-stage WYD finite element model (**Ref**) with the two-stage WYD finite element model. (b) Comparison of the one-stage WYD model (**Ref**) with the WYD two-stage model with enhancement at frequencies [1000, 1400, 1800, 2200, 2600, 3000] Hz (labeled 2SWE).

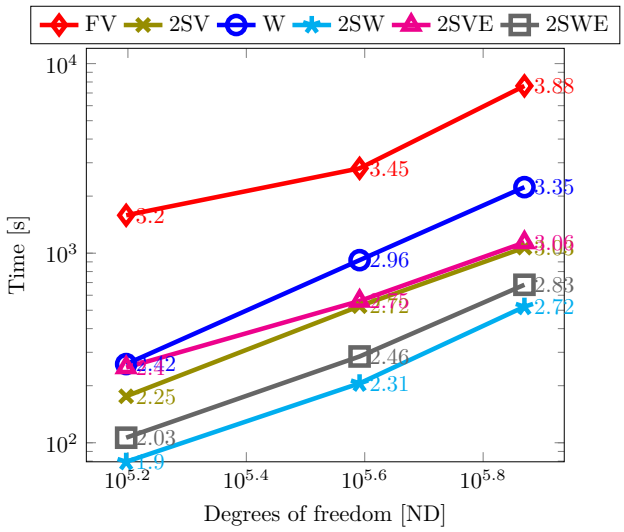

FIGURE 11 Brake disc. Model with 400 modes, meshes 3, 4, and 5. Comparison of time needed to construct the transformation matrix $\boldsymbol{\Phi}$ (2SVE), or the transformation matrix $\widehat{\boldsymbol{\Phi}}$ (2SWE).

construct load-dependent Ritz vectors. The proposed two-stage procedures first reduce the full finite element model using the CoNC reduction into a much smaller model with sparse reduced operators, and then perform either a free-vibration analysis (**2SV**) or compute load-dependent Ritz vectors (**2SW**) using the one-stage reduced model for further reduction in the second stage. It is shown through examples that the **2SV** and **2SW** bases can be computed very inexpensively as compared to the conventional (one-stage) model reductions for harmonic vibration analyses, whilst representing the characteristics of the FRF well.

To improve the accuracy of the FRF approximation relative to one-stage procedures, enhancements to the **2SV** and **2SW** reductions are proposed. The enhancement to the **2SV** reduction, i.e. **2SVE** reduction, involves improving the **2SV**-reduced

eigenmodes (columns of the **2SV** transformation matrix) by running the residual-minimizing MINRES algorithm for a few iterations of one step of inverse power method applied to the undamped free vibration of the full finite element model. The original eigenmodes are then replaced by the improved eigenmodes, and the results show a great improvement in FRF curve accuracy at very slight increase in cost. The user can pick the regions (in frequency domain) of enhancement based on their need for improved accuracy, and enhance the eigenmodes whose approximate eigenfrequencies (computed during the **2SV** reduction) lie in those regions of interest. The number of enhancements is typically much smaller than the number of **2SV** basis, thereby adding very little cost to the reduction.

The **2SW** reduction is improved by adding enhancement vectors to the **2SW** basis (and so is referred to as **2SWE**) such that these vectors do not belong to the column space of the **2SW** transformation matrix. The enhancement vectors are constructed by first picking the enhancement frequencies from the frequency domain in the regions where more accuracy is demanded by the user. A few iterations of the MINRES algorithm applied to the damped balance equations of the full model are then used to minimize the residual of the dynamic equilibrium of the full model at the enhancement frequencies. The **2SW** reduced approximations of the complex response act as the corresponding initial guesses. The improved complex responses do not belong to the column space of the **2SW** transformation matrix, and the real and imaginary parts of the improved responses are used as the enhancement vectors, after orthogonalization. The **2SW**-reduced operators computed during the computation of the improved complex responses are saved and reused during the projection of the global operators on the **2SWE** basis to achieve additional performance gains. The results demonstrate the effectiveness of the enhancement in terms of improved FRF prediction, whilst the performance is much better than for the conventional (one-stage) model reduction.

The **2SW** reduction is improved by adding enhancement vectors to the **2SW** basis (and so is referred to as **2SWE**) such that these vectors do not belong to the column space of the **2SW** transformation matrix. The enhancement vectors are constructed by first picking the enhancement frequencies from the frequency domain in regions where more accuracy is demanded by the user. Few iterations of the MINRES algorithm are then used to minimize the residual of the steady-state dynamic equilibrium of the full model at the enhancement frequencies using the **2SW**-reduced complex responses as the corresponding initial solutions. The improved complex responses do not belong to the column space of the **2SW** transformation matrix. The real and imaginary parts of the improved responses are used as the enhancement vectors, post-orthogonalization. The **2SW**-reduced operators computed during the computation of the improved complex responses are saved and reused during the projection of the global operators on the **2SWE** basis to achieve additional performance gains. The results demonstrate the effectiveness of the enhancement in terms of improved FRF prediction (almost no phase difference) whilst the performance is observed to be still much better than the conventional (one-stage) model-reductions.

The present algorithm addresses linear problems of the steady-state dynamics of models of continua. Since the coherent node cluster technique leads to an algebraic transformation applied to the degrees of freedom, there are no fundamental obstacles to the application of the technique in nonlinear models. Extensions of the proposed model reduction to shell structures and nonlinear analyses will be pursued in the near future.

ACKNOWLEDGMENTS

Continued support of the first author by the Office of Naval Research (program manager Michael J. Weise) is gratefully acknowledged.

Author contributions

PK and RS implemented the algorithms. All authors participated in the writing of the paper.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

APPENDIX

A HARDWARE AND SOFTWARE

Unless stated otherwise, the computations described below were performed on a Linux machine with 64 AMD Opteron(tm) Processor 6380 cores at 1.4 GHz, with 256 GB of DDR3 RAM, and L1 cache of 768KiB, L2 cache of 16MiB, and L3 cache of 12MiB.

The computations described below were implemented in the Julia programming language[28,29], in the framework of the `FinEtools.jl` Julia package[43]. The eigenvalue problem solutions for the full and reduced eigenvalue problems were obtained with Arpack[44].

The eigenvalue solver package Arpack[44] was accessed through the Julia package `Arpack.jl`[45]. In order to verify that Arpack was implemented correctly and efficiently in `Arpack.jl`, an experiment was performed to determine how the eigenvalue problem solution compared with an independent implementation of the Lanczos method[23].

The solutions for the free-vibration spectra of the full finite element model were obtained with the Lanczos algorithm[46]. By default Arpack terminates when the Ritz values are found with tolerance equal to the machine epsilon, $\mathrm{tol} = \mathrm{eps}(1.0)$ ($\sim 2.2e^{-16}$). A Ritz value θ is considered converged if the norm of the residual of the standard eigenvalue problem form

$$\boldsymbol{r} = (\boldsymbol{A} - \theta \mathbf{1})\, \boldsymbol{v} \tag{A1}$$

where θ and $\boldsymbol{v}$ are approximations of the eigenvalue and eigenvector, is

$$\|\boldsymbol{r}\| \leq \mathrm{tol} \times |\theta| \tag{A2}$$

The eigenvalue problem in the reduced model Stage 1 was also solved with the same eigenvalue solver package Arpack, also using the default error tolerance.

None of the algorithms exercised in this paper were explicitly formulated as parallel (even though there may be such opportunities). Parallel execution with shared-memory threading may nevertheless have been be applied within the linear-algebra libraries such as BLAS and LAPACK that underlie the various operations such as matrix multiplies, linear equation solves etc.

References

1. Guyan RJ. Reduction of stiffness and mass matrices. *AIAA Journal* 1965; 3(2): 380–380. doi: 10.2514/3.2874
2. Irons B. Structural eigenvalue problems - elimination of unwanted variables. *AIAA Journal* 1965; 3(5): 961–962. doi: 10.2514/3.3027
3. Kidder RL. Reduction of structural frequency equations. *AIAA Journal* 1973; 11(6): 892–892. doi: 10.2514/3.6852
4. O'Callahan JC. A procedure for an improved reduced system (IRS) model. In: Proc. 7. International Modal Analysis Conference. ; 1989.
5. Chen SH, Pan HH. Guyan Reduction. *Communications in Applied Numerical Methods* 1988; 4(4): 549–556. doi: 10.1002/cnm.1630040412
6. Friswell MI, Garvey SD, Penny JET. Model reduction using dynamic and iterated IRS techniques. *Journal of Sound and Vibration* 1995; 186(2): 311–323. doi: 10.1006/jsvi.1995.0451
7. O'Callahan JC, Avitabile P, Riemer R. System equivalent reduction expansion process (SEREP). In: Proc. 7. International Modal Analysis Conference. ; 1989.
8. Sastry C, Mahapatra DR, Gopalakrishnan S, Ramamurthy T. An iterative system equivalent reduction expansion process for extraction of high frequency response from reduced order finite element model. *Computer Methods in Applied Mechanics and Engineering* 2003; 192(15): 1821 - 1840. doi: https://doi.org/10.1016/S0045-7825(03)00204-4

9. Craig JRR, Bampton MCC. Coupling of substructures for dynamic analyses. *AIAA Journal* 1968; 6(7): 1313–1319. doi: 10.2514/3.4741

10. Craig JR. Coupling of substructures for dynamic analyses: an overview. In: 41st Structures, Structural Dynamics, and Materials Conference and Exhibit, Structures, Structural Dynamics, and Materials and Co-located Conferences. American Institute of Aeronautics and Astronautics; 2000

11. Gladwell GML. Branch mode analysis of vibrating systems. *Journal of Sound and Vibration* 1964; 1(1): 41 - 59. doi: https://doi.org/10.1016/0022-460X(64)90006-9

12. Hurty WC. Dynamic analysis of structural systems using component modes. *AIAA Journal* 1965; 3(4): 678-685. doi: 10.2514/3.2947

13. Castanier M, Tan YC, Pierre C. Characteristic Constraint Modes for Component Mode Synthesis. *AIAA Journal* 2001; 39: 1182 - 1187. doi: https://doi.org/10.1115/1.1338948

14. Bennighof J, Lehoucq R. An Automated Multilevel Substructuring Method for Eigenspace Computation in Linear Elastodynamics. *SIAM Journal on Scientific Computing* 2004; 25(6): 2084–2106. doi: 10.1137/S1064827502400650

15. Klerk DD, Rixen DJ, Voormeeren SN. General Framework for Dynamic Substructuring: History, Review and Classification of Techniques. *AIAA Journal* 2008; 46(5): 1169-1181. doi: 10.2514/1.33274

16. Salimbahrami B, Lohmann B. Order reduction of large scale second-order systems using Krylov subspace methods. *Linear Algebra and its Applications* 2006; 415(2): 385 - 405. Special Issue on Order Reduction of Large-Scale Systemsdoi: https://doi.org/10.1016/j.laa.2004.12.013

17. Antoulas A. *Approximation of Large-Scale Dynamical Systems*. Society for Industrial and Applied Mathematics . 2005

18. Nowakowski C, Kürschner P, Eberhard P, Benner P. Model reduction of an elastic crankshaft for elastic multibody simulations. *ZAMM Zeitschrift fur Angewandte Mathematik und Mechanik* 2013; 93(4): 198–216. doi: 10.1002/zamm.201200054

19. Bernard H, Mario O. Efficient reduced models and a posteriori error estimation for parametrized dynamical systems by offline/online decomposition. *Mathematical and Computer Modelling of Dynamical Systems* 2011; 17(2): 145-161. doi: 10.1080/13873954.2010.514703

20. Benner P, Gugercin S, Willcox K. A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems. *SIAM Review* 2015; 57(4): 483-531. doi: 10.1137/130932715

21. Baur U, Beattie C, Benner P, Gugercin S. Interpolatory Projection Methods for Parameterized Model Reduction. *SIAM Journal on Scientific Computing* 2011; 33(5): 2489-2518. doi: 10.1137/090776925

22. Nguyen NC, Patera AT, Peraire J. A 'best points' interpolation method for efficient approximation of parametrized functions. *International Journal for Numerical Methods in Engineering* 2008; 73(4): 521-543. doi: 10.1002/nme.2086

23. Krysl P, Sivapuram R, Abawi AT. Rapid free-vibration analysis with model reduction based on coherent nodal clusters. *International Journal for Numerical Methods in Engineering* 2020; 121(15): 3274-3299. doi: https://doi.org/10.1002/nme.6358

24. Chopra AK. *Dynamics of Structures*. Pearson . 2012.

25. Richardson M, Potter R. Viscous vs. Structural Damping in Modal Analysis. In: Proc. 46. Shock and Vibration Symposium. ; 1975.

26. Qu ZQ. *Model Order Reduction Techniques with Applications in Finite Element Analysis*. Springer . 2004.

27. Wilson E, Yuan M, Dickens J. Dynamic analysis by direct superposition of Ritz vectors. *Earthquake Engineering and Structural Dynamics* 1982; 10: 813-821.

28. The Julia Project . The Julia Programming Language. https://julialang.org/; Accessed 03/13/2021.

29. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing. *SIAM review* 2017; 59(1): 65–98.

30. Bayo EP, Wilson EL. Use of Ritz vectors in wave-propagation and foundation response. *Earthquake Engineering & Structural Dynamics* 1984; 12(4): 499-505. doi: 10.1002/eqe.4290120406

31. Nour-Omid B, Clough RW. Dynamic analysis of structures using Lanczos co-ordinates. *Earthquake Engineering & Structural Dynamics* 1984; 12(4): 565-577. doi: https://doi.org/10.1002/eqe.4290120410

32. Kim BW, Jung HJ, Kim WH, Lee IW. Dynamic analysis of structures using modified Lanczos co-ordinates. *Earthquake Engineering & Structural Dynamics* 2003; 32(9): 1469-1474. doi: https://doi.org/10.1002/eqe.284

33. Kline K. Dynamic analysis using a reduced basis of exact modes and Ritz vectors. *AIAA Journal* 1986; 24(12): 2022-2029.

34. Ibrahimbegovic A, Wilson EL. Automated truncation of Ritz vector basis in modal transformation. *Journal of Engineering Mechanics ASCE* 1990; 116(11): 2506-2520.

35. Abramowitz M, Stegun I. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Applied mathematics seriesDover Publications . 1965.

36. Williams RD. Performance of dynamic load balancing algorithms for unstructured mesh calculations. Technical Report C3P913, California Institute of Technology; Pasadena, California: 1990.

37. Karypis G, Kumar V. Metis. a software package for partitioning unstructured graphs, partitioning meshes, and computing fill-reducing orderings of sparse matrices. technical report version 4.0, University of Minnesota; Minneapolis, MN: 1998.

38. Metis.jl development team . Julia interface to Metis graph partitioning Metis.jl. https://github.com/JuliaSparse/Metis.jl; Accessed 04/11/2021.

39. Sivapuram R, Krysl P. On the energy-sampling stabilization of Nodally Integrated Continuum Elements for dynamic analyses. *Finite Elements in Analysis and Design* 2019; 167: 103322.

40. Orban D, Arioli M. *Iterative solution of symmetric quasi-definite linear systems*. SIAM . 2017

41. Siam Ringspann Brake disk B 250/12,5 bore diameter 36. https://www.traceparts.com/en/product/siam-ringspann-brake-disk-form-b?CatalogPath=TRACEPARTS%3ATP01003003003004002&Product=10-08012001-119755; 2020.

42. Montoison A, Orban D, contributors . Krylov.jl: A Julia Basket of Hand-Picked Krylov Methods. https://github.com/JuliaSmoothOptimizers/Krylov.jl; 2020

43. Petr Krysl . FinEtools: Finite Element tools in Julia. https://github.com/PetrKryslUCSD/FinEtools.jl; Accessed 06/11/2021.

44. ARPACK developers . Arpack official website. https://www.caam.rice.edu/software/ARPACK/; Accessed 04/11/2019.

45. Julia Linear algebra development team . The Arnoldi eigenvalue problem solvers Arpack.jl. https://github.com/JuliaLinearAlgebra/Arpack.jl; Accessed 04/11/2019.

46. Lehoucq RB, Sorensen DC, Yang C. *ARPACK Users' Guide: Solution of Large-scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. Software, Environments and Tools.Society for Industrial and Applied Mathematics . 1998.