4 Modal methods for reduced order modeling

Abstract: In this chapter, we present an overview of the so-called *modal methods* for reduced order modeling. The naming is loosely referring to techniques that aim at constructing the reduced order basis for reduction without resorting to data, typically obtained by full order simulations. We focus primarily on linear and nonlinear mechanical systems stemming from a finite element discretization of the underlying strong form equations. The nonlinearity is of a geometric nature, i. e. due to redirection of internal stresses due to large displacements. Intrusive vs non-intrusive techniques (i. e. requiring or not access to the finite element formulation to construct the reduced order model) are discussed, and an overview of the most popular methods is presented.

Keywords: Galerkin projection, geometric nonlinearities, modal derivatives, non-intrusive reduction, nonlinear manifolds

MSC 2010: 74H15, 74H45

4.1 Introduction

Structural dynamics often relies on Finite Element (FE) models leading to large system of equations, which need to be integrated in time to predict time histories of displacements, strain and stresses. Even in the case of linear models, this task is often prohibitive in a design context, as the large size of the systems and the large number of simulations required result in excessive time and storage requirements. This situation is exacerbated by the presence of nonlinearities, which essentially implies the evaluation and factorization of configuration-dependent residuals and Jacobians, and an iterative process for convergence at each time step. Clearly, model order reduction is a must in these cases.

Historically, linear structural dynamics have been resorting to modal decomposition since its dawn. Essentially, the displacement field is approximated as a superposition of few eigenmodes of the system to achieve reduction.

When nonlinearities are present, a plethora of phenomena might occur, namely a richer harmonic spectrum in the response than that of the applied forcing, internal resonances, bifurcation, etc. The challenge for a nonlinear reduced order model (NL-ROM) is to efficiently capture all these phenomena. As the type of nonlinearity strongly affects the response, it would make sense to use all the knowledge of the FE model,

Paolo Tiso, Morteza Karamooz Mahdiabadi, ETH Zürich, Institute for Mechanical Systems, Zurich, Switzerland

Jacopo Marconi, Department of Mechanical Engineering, Politecnico Di Milano, Milan, Italy

Open Access. © 2021 Paolo Tiso et al., published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

usually referred to as the High Fidelity Model (HFM) to devise methods to construct the NLROM. This strategy, which we name here *model-driven* or *system-driven*, is in some sense opposed to a *data-driven* approach, which constructs the NLROM essentially from data coming from simulations of the HFM. This chapter attempts to give an overview to methods belonging to the former category, and focuses on structural systems featuring geometric nonlinearities.

Essentially, a NLROM technique tackling such systems faces two challenges. First, the derivation of a compact reduction basis on which to project the HFM. Second, an efficient calculation of the projected nonlinear terms. As we will discuss, the nonlinear terms of a FE based HFM are never available in an explicit form at system level, and therefore their reduction and projection face some computational bottlenecks that need to be addressed.

An often employed categorization of NLROMs distinguishes between *intrusive* and *non-intrusive* methods. Respectively, the construction of the NLROM requires, or not, the access to element level formulation, and FE assembly. Clearly, a non-intrusive technique is preferable when one has only the availability of a commercial FE program, which typically only release forces and, in some cases, Jacobians, at the assembled level. On the other hand, intrusive methods are typically more systematic and theoretically sound, at the price of the mentioned need to access codes at the element level.

In this contribution, we discuss the main methods and point out recent developments and trends. We first recap the main concepts applicable to linear systems, and then tackle the relevant cases of geometric nonlinearities.

4.2 Linear systems

Let us first consider the linear system resulting from a FE discretization of second-order ordinary differential equations describing an elastic continuum. This can be written

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{p}(t), \quad (4.1)$$

where $\mathbf{u} \in \mathbb{R}^n$ is the vector of nodal generalized displacements, $\mathbf{M} \in \mathbb{R}^{n \times n}$ is the mass matrix, $\mathbf{C} \in \mathbb{R}^{n \times n}$ is the damping matrix, $\mathbf{K}$ is the stiffness matrix and $\mathbf{p}(t) \in \mathbb{R}^n$ is a forcing vector. The system (4.1) implies a linearization about an equilibrium point. For many structural dynamics applications, n is typically very large due to a required fine discretization.

A well established reduction technique in linear structural dynamics is based on the so-called *vibration modes*, i. e. eigenvectors associated to the free undamped vibration problem, obtained by setting $\mathbf{C} = \mathbf{0}$ and $\mathbf{p}(t) = \mathbf{0}$ in (4.1),

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{0}. \quad (4.2)$$

The solution of (4.2) reads

$$\mathbf{u}(t) = \boldsymbol{\phi} e^{i\omega t}. \quad (4.3)$$

When (4.3) is inserted in (4.2), we obtain the eigenvalue problem

$$(\mathbf{K} - \omega^2 \mathbf{M})\boldsymbol{\phi} = \mathbf{0}, \quad (4.4)$$

which admits eigenpairs ω_j , $\boldsymbol{\phi}_j$, $j = 1, \dots, n$ as solutions. Physically, each mode $\boldsymbol{\phi}_j$ represents the shape at which the system freely vibrates at a frequency ω_j . It can be easily shown that the vibration modes are orthogonal with respect to the mass and the stiffness matrices. Compactly, we can write

$$\boldsymbol{\phi}_j^T \mathbf{M} \boldsymbol{\phi}_k = \delta_{jk}, \quad \boldsymbol{\phi}_j^T \mathbf{K} \boldsymbol{\phi}_k = \omega_j^2 \delta_{jk}, \quad (4.5)$$

where $\delta_{jk} = 1$ for $j = k$, and zero otherwise. In the above, we also use a so called *mass normalization* for the vibration mode, i.e. a scaling $\boldsymbol{\phi}_j$ that guarantees that $\boldsymbol{\phi}_j^T \mathbf{M} \boldsymbol{\phi}_j = 1$, $j = 1, \dots, n$. Since the eigenmodes of the system constitute an orthonormal basis spanning $\mathbb{R}^n$, we can operate the following coordinate transformation, known as *Mode Superposition*:

$$\mathbf{u}(t) = \sum_{k=1}^n \boldsymbol{\phi}_k q_k(t) = \boldsymbol{\Phi} \mathbf{q}(t), \quad (4.6)$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_n]$ is a matrix containing all the eigenmodes and $\mathbf{q}(t) = [q_1(t), \dots, q_n(t)]^T$ is the vector of modal coordinates. By substituting (4.6) in (4.1) and projecting onto $\boldsymbol{\Phi}$, we obtain

$$\boldsymbol{\Phi}^T \mathbf{M} \boldsymbol{\Phi} \ddot{\mathbf{q}} + \boldsymbol{\Phi}^T \mathbf{C} \boldsymbol{\Phi} \dot{\mathbf{q}} + \boldsymbol{\Phi}^T \mathbf{K} \boldsymbol{\Phi} \mathbf{q} = \boldsymbol{\Phi}^T \mathbf{p}. \quad (4.7)$$

It is often the case, in structural dynamics applications, that the damping matrix $\mathbf{C}$ is given as linear combination of $\mathbf{K}$ and $\mathbf{M}$,

$$\mathbf{C} = \alpha \mathbf{K} + \beta \mathbf{M}, \quad (4.8)$$

where α and β are user defined coefficients. This is done for two reasons. First, the actual damping is due to several mechanisms (joint friction, internal heating, interaction with surrounding fluid, etc.) which cannot be physically described by a simple viscous term as in (4.1), but rather by much more complex, nonlinear laws. The term $\mathbf{C}\dot{\mathbf{u}}(t)$ is then designed to provide a velocity dependent linear term in (4.1) by globally representing the dissipation occurring during motion. Second, the form (4.8), also known as *Rayleigh Damping*, allows one to make use of the orthogonality property (4.5), and yields a decoupled set of equations. By virtue of mode orthogonality and mass normalization of modes, (4.7) becomes

$$\ddot{\mathbf{q}} + \mathbf{D} \dot{\mathbf{q}} + \boldsymbol{\Omega}^2 \mathbf{q} = \boldsymbol{\Theta}, \quad (4.9)$$

where $\Theta = [\phi_1^T \mathbf{p}, \dots, \phi_n^T \mathbf{p}]^T \in \mathbb{R}^n$ is the *load participation vector*,

$$\Omega^2 = \begin{bmatrix} \omega_1^2 & & \\ & \ddots & \\ & & \omega_n^2 \end{bmatrix} \quad (4.10)$$

is a diagonal matrix containing the eigenvalues ω_j^2 , $j = 1, \dots, n$ along its diagonal, and

$$\mathbf{D} = \Phi^T \mathbf{C} \Phi = \Phi^T (\alpha \mathbf{K} + \beta \mathbf{M}) \Phi = \begin{bmatrix} 2\xi_1 \omega_1 & & \\ & \ddots & \\ & & 2\xi_n \omega_n \end{bmatrix}, \quad (4.11)$$

with the damping ratios ξ_j found as

$$\xi_j = \frac{1}{2} \left(\alpha \omega_j + \frac{\beta}{\omega_j} \right). \quad (4.12)$$

Equation (4.9) indicates that the response of a damped linear system to a given excitation can be computed from a set of n uncoupled single DOF equations. Each of these uncoupled equations can be separately integrated using for instance the Laplace transform and convolution products (see [18]). However, the computation of all eigenmodes for a large dynamical system with several DOFs is computationally expensive in practice. It is anyway possible to approximate the response of the system using only few modes in (4.6), as explained in the next section.

Modal displacement method

To reduce the computational costs for obtaining the response of a system using mode superposition, one could only include the modes in (4.6) which effectively participate in the response. This technique is known as *Modal Displacement Method* (MDM). If this strategy holds, not only just a few eigenmodes of the system have to be calculated, but also the number of uncoupled modal equations that need to be solved is significantly reduced with respect to the size n of the original problem (4.1).

Let us first decompose the applied external load to a spatial distribution vector and a time-varying function as

$$\mathbf{p} = \mathbf{P}_0 \mathbf{g}(t), \quad (4.13)$$

where $\mathbf{P}_0 \in \mathbb{R}^{n \times p}$ is a matrix containing p spatial distributions of the applied load and $\mathbf{g}(t) \in \mathbb{R}^p$ represents their time-varying functions. According to the MDM, the motion of the system (4.1) can be approximated as a superposition of truncated number of modes by

$$\mathbf{u}(t) \approx \sum_{k=1}^m \phi_k \tilde{q}_k(t) = \tilde{\Phi} \tilde{\mathbf{q}}(t), \quad (4.14)$$

where $\tilde{\Phi} \in \mathbb{R}^{n \times m}$ is a matrix containing the truncated set of $m \ll n$ kept eigenmodes of the system, and $\tilde{\mathbf{q}}(t)$ are the corresponding modal amplitudes. By substituting (4.14) in (4.1) and projecting onto $\tilde{\Phi}$, the reduced set of modal equations reads

$$\ddot{\tilde{\mathbf{q}}} + \tilde{\mathbf{D}}\dot{\tilde{\mathbf{q}}} + \tilde{\mathbf{\Omega}}^2\tilde{\mathbf{q}} = \tilde{\Phi}^T \mathbf{P}_0 \mathbf{g}(t) = \tilde{\mathbf{P}}_0 \mathbf{g}(t), \quad (4.15)$$

where $\tilde{\mathbf{D}} = \tilde{\Phi}^T \mathbf{D} \tilde{\Phi}$ and $\tilde{\mathbf{\Omega}}^2 = \tilde{\Phi}^T \mathbf{K} \tilde{\Phi}$. Once the reduced system in (4.15) is integrated, the approximate displacement $\tilde{\mathbf{u}}$ is recovered as

$$\tilde{\mathbf{u}}(t) = \sum_{k=1}^m \boldsymbol{\phi}_k \tilde{q}_k(t). \quad (4.16)$$

The MDM can significantly reduce the computational cost of time integration. Typically, the eigenmodes that are used for the reduction are those whose frequencies spans the frequency content of $\mathbf{g}(t)$. While representing the spectral content of $\mathbf{g}(t)$, one should also strive for an accurate projection of the spatial distribution of the load $\tilde{\mathbf{P}}_0$. If this is not guaranteed, a significant error can arise. To quantify this error, let us define the residual spatial distribution vector, $\mathbf{P}_r$, as the portion of the load distribution vector $\mathbf{P}_0$, which is neglected after projecting it on the truncation modes basis. This residual can be obtained:

$$\mathbf{P}_r = \mathbf{P}_0 - \mathbf{P}_t, \quad (4.17)$$

where

$$\mathbf{P}_t = \mathbf{M} \tilde{\Phi} \tilde{\Phi}^T \mathbf{P}_0; \quad (4.18)$$

see [14] for details of the derivation. Note that, as m increases, $\tilde{\Phi} \tilde{\Phi}^T \rightarrow \Phi \Phi^T = \mathbf{M}^{-1}$ and therefore $\mathbf{P}_r \rightarrow \mathbf{0}$, but the required computational effort increases as well. In order to keep m low and compensate for the effect of modal truncation onto the load, two main methods are outlined here.

Mode acceleration correction

The MDM is based on the assumption that the reduction basis spans the whole frequency content of the applied load. Therefore, the modes that are left out would contribute to the solution in a quasi-static manner, i. e. without changing the velocity and the acceleration of the full response appreciably. The Mode Acceleration Correction (MAC) method computes the quasi-static effect of truncated modes and augments it to the displacement response obtained from reduced system using the Modal Displacement method. This is done as follows. From (4.1) we can write

$$\mathbf{K}\mathbf{u} = \mathbf{p} - \mathbf{M}\ddot{\mathbf{u}} - \mathbf{C}\dot{\mathbf{u}}. \quad (4.19)$$

Since the acceleration of the system is assumed not to be affected by modal reduction, we substitute (4.16) into (4.19) and solve it for the displacement, giving

$$\mathbf{u} \approx \bar{\mathbf{u}} = \mathbf{K}^{-1} \mathbf{p} - \sum_{k=1}^m \frac{\boldsymbol{\phi}_k \ddot{q}_k(t)}{\omega_k^2} - \sum_{k=1}^m \frac{2\xi_k \boldsymbol{\phi}_k \dot{q}_k(t)}{\omega_k}. \quad (4.20)$$

The acceleration of the k th generalized coordinate can be calculated from (4.9) as

$$\ddot{q}_k = \boldsymbol{\phi}_k^T \mathbf{p} - \omega_k^2 q_k - 2\xi_k \omega_k \dot{q}_k. \quad (4.21)$$

Equation (4.21) is then introduced to (4.20), and rearranged to introduce the final form of the displacement:

$$\bar{\mathbf{u}} = \sum_{k=1}^m \boldsymbol{\phi}_k q_k + \left(\mathbf{K}^{-1} - \sum_{k=1}^m \frac{\boldsymbol{\phi}_k \boldsymbol{\phi}_k^T}{\omega_k^2} \right) \mathbf{p}. \quad (4.22)$$

The last term in (4.22) represents a quasi-static correction of the displacement using MAC. In other words, MAC represents a statically complete alternative for the MDM by augmenting the static response of the deleted modes to the system. Note that the missing modes in the truncation are never computed; simply, their static contribution is added a posteriori.

Modal truncation augmentation

The inaccuracies of the applied load's spatial distribution, caused by mode displacement projection, can be also improved by adding few correction vectors to the reduction basis. This technique goes under the name of *Modal Truncation Augmentation* (MTA) [15]. Let us first compute the static response of the system due to the residual spatial forces, which is simply

$$\mathbf{K}\mathbf{X} = \mathbf{P}_r, \quad (4.23)$$

where $\mathbf{X}$ is a matrix collecting Ritz vectors columnwise. This matrix is then employed to reduce the mass and stiffness matrices of the system as

$$\hat{\mathbf{M}} = \mathbf{X}^T \mathbf{M} \mathbf{X}, \quad \hat{\mathbf{K}} = \mathbf{X}^T \mathbf{K} \mathbf{X}. \quad (4.24)$$

Here, $\hat{\mathbf{M}}$ and $\hat{\mathbf{K}}$ are the reduced mass and stiffness matrices projected onto the space spanned by the columns of $\mathbf{X}$. In the next step, an eigenvalue problem of the form

$$(\hat{\mathbf{K}} - \omega_k^2 \hat{\mathbf{M}}) \mathbf{w}_k = \mathbf{0} \quad (4.25)$$

is solved to obtain the eigenvectors $\mathbf{w}_k$. Finally, the modal truncation augmentation vectors are found by projecting the Ritz vectors onto the space spanned by the eigenvectors calculated from (4.25). This is simply

$$\boldsymbol{\Phi}_{mt} = \mathbf{X} \mathbf{W}, \quad (4.26)$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_p]$ is a matrix containing the eigenvectors obtained from (4.25) in its columns. The final step is to append the modal truncation augmentation vectors to the reduction basis of truncated eigenmodes as

$$\mathbf{\Psi} = [\tilde{\mathbf{\Phi}} \quad \mathbf{\Phi}_{mt}], \quad (4.27)$$

where $\mathbf{\Psi}$ is the enrichment to the basis of mode displacement method. From here, the MTA procedure follows the MDM, namely, the matrix $\mathbf{\Psi}$ can be used instead of $\tilde{\mathbf{\Phi}}$ in (4.14) to reduce the size of the system to also take into account the quasi-static contribution of the truncated mode in the system.

Once the response of the generalized coordinates $\mathbf{q}$ is obtained the physical displacement can be retrieved as

$$\hat{\mathbf{u}} \approx \mathbf{\Psi} \hat{\mathbf{q}}. \quad (4.28)$$

Likewise, the physical acceleration is

$$\ddot{\mathbf{u}} \approx \mathbf{\Psi} \ddot{\mathbf{q}}. \quad (4.29)$$

As can be seen from (4.29), the effect of truncated modes is also reflected on the acceleration of the system through the basis $\mathbf{\Psi}$. This is different from the MAC method, which attempts only to correct for the displacement of the mode displacement reduced system in a strictly static manner.

4.3 Substructuring and component mode synthesis

Before discussing nonlinear problems, it is worth to briefly mention here another practical strategy for model reduction, namely *substructuring*. A thorough discussion of this topic is beyond the scope of this chapter, and the interested reader should refer to [1] for a comprehensive overview. The reduction methods discussed in this contribution tackle the whole high fidelity model monolithically, meaning that only one reduction basis is constructed to reduce the full solution. It is possible, however, to decompose the system into *substructures*, which communicate with each other through common interfaces. Then each substructure could be reduced independently from the others, while preserving the interface DOFs to allow for assembly. This approach was originated by the seminal work of Hurty in [25] and then later developed by Craig and Bampton in [12].

The division of the system into substructures is often dictated by the practical reason of dealing with the detailed design and the analysis of single components independently. The assembled model will then feature common interface DOFs and few, reduced generalized coordinates relative to each substructure, thus achieving model order reduction. All techniques aiming at reducing the substructures are based on

what discussed in the previous section. However, significant differences will arise depending on how the interface compatibility across substructures is enforced, namely whether a *primal* or *dual* assembly is adopted. In the former, common interface degrees of freedom are eliminated, and therefore the interface forces are not present in the assembled system. In the latter case, interface forces appear as Lagrange's multipliers, and thus compatibility is not a priori satisfied. For primal assembly, the Hurty–Graig–Bampton method [12, 25] achieves reduction for each substructure with a basis formed by static modes obtained by applying unit displacements at the interface, and vibration modes relative to the substructure clamped at the interface. In the case of dual assembly, it is natural to resort to the method of McNeal [42] and Rubin [56], which construct a reduction basis using vibration modes of the *free* substructure complemented with so-called attachment modes obtained by applying unit forces at the interface DOFs.

In case of several substructures featuring interfaces with many DOFs, the resulting assembled reduced order model could still be of large size. Then a second reduction could be performed on the interface DOFs, which could be applied before assembly to each component independently, or on the global interface after assembly. A recent review of the available methods can be found in [39].

4.4 Geometrically nonlinear structural dynamics

When the displacements about the equilibrium point cannot be considered as small, nonlinear geometrical effects might arise and (4.1) loses its validity. In particular, internal stresses due to deformations are rotated with respect to the reference, undeformed configuration. Typical engineering applications prone to geometric nonlinearities are thin-walled structures, where low-frequency and bending-dominated modes cause in-plane stretching when the displacements are in the order of the thickness of the structure. Then the system (4.1) is modified as

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{f}(\mathbf{u}(t)) = \mathbf{p}(t), \quad (4.30)$$

where $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the function of internal nodal forces. The nonlinearity here included can take various forms, depending on the kinematic and material model adopted. For instance, for linear elastic constitutive law and Green–Lagrange strain tensor, the resulting internal forces are linear, quadratic and cubic in $\mathbf{u}$ [43]. The same holds for the approximate kinematic model due to von Karman, which captures the behavior of beams and plates undergoing out-of-plane deflections in the order of one thickness of the structure [54]. Other kinematic descriptions, i. e. obtained by means of the co-rotational formulation, lead to more complex forms of $\mathbf{f}(\mathbf{u})$ [10, 13].

4.4.1 Galerkin projection

In order to achieve a low order version of (4.30), the generalized displacement vector $\mathbf{u}(t)$ can be approximated by a linear combination of vectors, contained in the columns of the matrix $\mathbf{V}$, as

$$\mathbf{u}(t) \approx \mathbf{V}\mathbf{q}(t) \quad (4.31)$$

where $\mathbf{q}(t)$ are the generalized modal amplitudes, unknowns of the Nonlinear Reduced Order Model (NLROM). By substituting (4.31) in equation (4.30), we get

$$\mathbf{M}\mathbf{V}\ddot{\mathbf{q}}(t) + \mathbf{C}\mathbf{V}\dot{\mathbf{q}}(t) + \mathbf{f}(\mathbf{V}\mathbf{q}(t)) - \mathbf{p}(t) = \mathbf{r}(t), \quad (4.32)$$

where $\mathbf{r}(t)$ is the residual resulting from the approximation. We can enforce this residual to be orthogonal to the reduction subspace, by requiring that

$$\mathbf{V}^T \mathbf{r} = \mathbf{0}. \quad (4.33)$$

This implies

$$\tilde{\mathbf{M}}\ddot{\mathbf{q}}(t) + \tilde{\mathbf{C}}\dot{\mathbf{q}}(t) + \tilde{\mathbf{f}}(\mathbf{V}\mathbf{q}(t)) = \tilde{\mathbf{p}}(t), \quad (4.34)$$

where

$$\tilde{\mathbf{M}} = \mathbf{V}^T \mathbf{M} \mathbf{V}, \quad \tilde{\mathbf{C}} = \mathbf{V}^T \mathbf{C} \mathbf{V}, \quad \tilde{\mathbf{p}}(t) = \mathbf{V}^T \mathbf{p}(t) \quad (4.35)$$

and

$$\tilde{\mathbf{f}}(t) = \mathbf{V}^T \mathbf{f}(\mathbf{V}\mathbf{q}(t)). \quad (4.36)$$

The above technique is known as Galerkin projection, according to which the subspaces on which the solution is sought and on which the residual is projected coincide. This preserves the symmetry—and thus stability—properties of the structural dynamics equations. In other cases, when symmetry is not featured in the full model, the right and left subspaces do not have to be equal. This latter procedure goes under the name of Petrov–Galerkin projection. Note that the reduced operators $\tilde{\mathbf{M}}$, $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{p}}(t)$ can be pre-computed *offline* before time-integrating the ROM. The nonlinear reduced force $\tilde{\mathbf{f}}$, however, still requires evaluation over the whole mesh, and hinders any significant speed-up that may be achieved from the ROM. In short, an accurate and efficient reduction relies on

1. a good choice of $\mathbf{V}$, and
2. an efficient computation of $\tilde{\mathbf{f}}$.

There is a variety of methods that aim at constructing a proper basis $\mathbf{V}$. In general, $\mathbf{V}$ should fulfill the following, in some sense related, features:

1. *Span the HFM solution subspace.* The choice of the reduction basis vectors is usually based on the expected spectrum of the response with respect to the spectrum of the excitation. While in the linear case, because of the superposition principle, this is an easily fulfilled requirement, nonlinear systems pose several challenges related to the raise of sub and super-harmonics in the response.
2. *Can be cheaply computed, in terms of computational time.*
3. *Can handle different cases.* It is for instance desirable to construct a NLRM able to provide accurate responses for different forcing amplitudes and frequencies, or different material and geometric parameters.

4.4.1.1 Evaluation of nonlinear terms

For FE applications, the nonlinear forces are evaluated as follows:

$$\tilde{\mathbf{f}}(\mathbf{q}(t)) = \sum_{e=1}^{n_e} \mathbf{V}_e^T \mathbf{f}_e(\mathbf{V}_e \mathbf{q}(t)) \quad (4.37)$$

where $\mathbf{f}_e(\mathbf{u}_e) \in \mathbb{R}^{N_e}$ is the contribution of the element e for the vector $\mathbf{f}(\mathbf{u})$ (N_e being the number of DOFs for the element e), $\mathbf{V}_e$ is the restriction of $\mathbf{V}$ to the rows indexed by the DOFs corresponding to element e , and n_e is the total number of elements of the mesh. Since the reduced nonlinear term $\tilde{\mathbf{f}}(\mathbf{q}(t))$ is evaluated in the space of full variables, the computational cost associated to its evaluation does not scale with m alone. Indeed, (4.37) shows that this cost scales linearly with the number of elements in the structure, and can hence be high for large systems. Thus, despite the reduction in dimensionality achieved in (4.34), the evaluation of $\tilde{\mathbf{f}}(\mathbf{q}(t))$ hinders any fast prediction of system response using the NLRM. Different strategies are available to overcome this problem.

4.4.1.2 Exact reduced forces

In the case of polynomial nonlinearities, as the one arising from von Karman kinematic model of linear elastic bodies, or solid continuum FE described in a total Lagrangian framework, the i th component of the HFM nonlinear force $\mathbf{f}(\mathbf{u})$ can be written as

$$f_i = K_{ij}u_j + B_{ijk}u_ju_k + C_{ijkl}u_ju_ku_l, \quad i, j, k, l = 1, \dots, n, \quad (4.38)$$

where K_{ij} , B_{ijk} and C_{ijkl} are constant tensors of second, third and fourth order, respectively. Note that the form (4.38) is never available as written, i. e. the tensors B_{ijk} and

C_{ijkl} are typically never assembled within a FE model, while K_{ij} is the linear stiffness matrix. When a reduction $\mathbf{u} \approx \mathbf{V}\mathbf{q}$ is inserted in (4.38) and a projection on $\mathbf{V}$ is performed, then $\tilde{f}_i$ can be expressed *exactly* by

$$\tilde{f}_i = \tilde{K}_{ij}u_j + \tilde{B}_{ijk}q_jq_k + \tilde{C}_{ijkl}q_jq_kq_l, \quad i, j, k, l = 1, \dots, m. \quad (4.39)$$

Once (4.39) is available, no connection to the HFM is needed any longer. For details of the computation of the reduced tensors $\tilde{K}_{ij}$, $\tilde{B}_{ijk}$ and $\tilde{C}_{ijkl}$ see for instance [66].

4.4.1.3 Approximate reduced forces

When the nonlinearity is not polynomial, for instance because of a nonlinear material law, or geometric nonlinearities described by a co-rotational formulation [13], one is facing essentially two possibilities:

1. assuming a certain form of $\tilde{\mathbf{f}}(\mathbf{q})$ and identifying its coefficients, or
 2. trying to compute $\tilde{\mathbf{f}}(\mathbf{q})$ affordably, but still querying the underlying HFM mesh.
- This strategy is addressed by the so-called *hyper-reduction* methods, which are discussed, at least to some extent, in Section 4.4.1.4.

For option 1., in the case of NLROMs addressing geometrically nonlinear problems, a typical choice is

$$\tilde{f}_i = \tilde{K}_{ij}^{(1)}q_j + \tilde{K}_{ijk}^{(2)}q_jq_k + \tilde{K}_{ijkl}^{(3)}q_jq_kq_l, \quad i, j, k, l = 1, 2, \dots, m, \quad (4.40)$$

where $\tilde{K}_{ij}^{(1)}$, $\tilde{K}_{ijk}^{(2)}$ and $\tilde{K}_{ijkl}^{(3)}$ are two, three and four dimensional matrices, respectively. Note the formal equivalence between (4.40) and (4.39). However, (4.40) is simply a pre-defined, convenient model of something which is essentially unknown. We will discuss in the following that this form is typically used in non-intrusive methods, in which the access to element information (which would enable (4.39)) is not possible. The problem then shifts to designing an efficient identification technique for the coefficients $\tilde{K}_{ijk}^{(2)}$ and $\tilde{K}_{ijkl}^{(3)}$, as the reduced stiffness matrix $\tilde{K}_{ij}^{(1)}$ is typically available.

4.4.1.4 Hyper-reduction

In the cases described before, the nonlinear term is written directly with respect to the modal reduced coordinates, in order to avoid any element level function evaluations during the numerical time integration of the NLROM. Another option is possible, which aims at scaling the computation of the nonlinear reduced terms with the size of the reduced coordinate vector m , rather than with n (size of the HFM). This is achievable only if the nonlinear terms are evaluated in a sparse manner, and if the missing

contributions are in some sense compensated for. Methods pursuing this strategy go under the name of *hyper-reduction*, meaning with this that they introduce a further reduction upon the one arising from projection. Among the various proposals, two methods (and numerous variants) are gaining popularity, namely the *Discrete Empirical Interpolation Method* (DEIM) [11] and the *Energy Conserving Sampling and Weighting* (ECSW) [16]. These two methods are briefly outlined now.

ECSW

In ECSW, the nonlinear projected terms are approximated as

$$\tilde{\mathbf{f}}(t) = \sum_{e=1}^{n_e} \mathbf{V}_e^T \mathbf{f}_e(\mathbf{V}_e \mathbf{q}(t)) \approx \sum_{e \in E} \xi_e \mathbf{V}_e^T \mathbf{f}_e(\mathbf{V}_e \mathbf{q}(t)), \quad (4.41)$$

where $\xi_e \in \mathbb{R}^+$ are positive weights, and $|E| < n_e$. In other words, (4.41) weights the element forces of the element set E in order to match the work done by the internal forces onto the displacements induced by the reduction basis $\mathbf{V}$. This procedure shows striking similarities with Gauss quadrature, where a defined integral is approximated by the evaluation of the integrand function at specific points, and weighted with pre-defined coefficients. In ECSW, the weights are determined by matching the work done by the nonlinear forces onto the reduction basis for a set of n_t sampled training forces. This translates into the minimization problem

$$\boldsymbol{\xi} : \arg \min_{\boldsymbol{\xi} \in \mathbb{R}^{n_e}, \boldsymbol{\xi} \geq 0} \|\mathbf{G}\tilde{\boldsymbol{\xi}} - \mathbf{b}\|_2, \quad (4.42)$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_{11} & \cdots & \mathbf{g}_{1n_e} \\ \vdots & \ddots & \vdots \\ \mathbf{g}_{n_t 1} & \cdots & \mathbf{g}_{n_t n_e} \end{bmatrix} \quad (4.43)$$

and

$$\mathbf{g}_{ie} = \mathbf{V}_e^T \mathbf{f}_e(\mathbf{V}_e \mathbf{q}^{(i)}), \quad \mathbf{b}_i = \tilde{\mathbf{f}}_i(\mathbf{q}^{(i)}) = \sum_{e=1}^{n_e} \mathbf{g}_{ie}, \quad (4.44)$$

where $i = 1, \dots, n_t$, $e = 1, \dots, n_e$. A sparse solution to (4.42) returns a sparse set of elements $E : \{e : \xi_e > 0\}$. An optimally sparse solution to (4.42) can be obtained by using a greedy-approach-based algorithm [50]. For the sake of compactness, this algorithm is not reported here. As an illustrative example, Figure 4.1 shows the ECSW hyper-reduction of the dynamic response of a slightly curved panel, subjected to a bi-harmonic loading (for details, refer to [29]). The ECSW selects only 19 elements out of a mesh of 400, thus decreasing the evaluation cost of the nonlinear terms significantly. The resulting speed-up (ratio between the computing time of the full and reduced solution) is about 50 in this example.

Figure 4.1: (Hyper-)reduction on a simply supported slightly curved panel subject to a time-varying load; see [29] for the details. On the left, the full response (thick black line) is compared to two Proper Orthogonal Decomposition (POD) reductions with 5 and 4 modes in the reduction basis, respectively for POD-1 and POD-2, and to the ECSW response. On the right, corresponding elements and their weights selected by the ECSW. In this case, 19 elements (out of 400) are picked. For a brief discussion of the POD method, see Section 4.5.4.

DEIM

The idea underlying the DEIM is somewhat different. Here, the nonlinear force $\mathbf{f}$ is approximated as a superposition of few force modes collected into the matrix $\mathbf{U} \in \mathbb{R}^{n \times m}$ as

$$\mathbf{f} \approx \mathbf{U}\mathbf{c}, \quad (4.45)$$

where $\mathbf{c} \in \mathbb{R}^m$ are unknown factors. Then a Boolean matrix $\mathbf{P}$ selects m rows of (4.45), so that the factors $\mathbf{c}$ can be found as

$$\mathbf{P}^T \mathbf{f} = \mathbf{P}^T \mathbf{U}\mathbf{c} \Rightarrow \mathbf{c} = (\mathbf{P}^T \mathbf{U})^{-1} \mathbf{P}^T \mathbf{f}. \quad (4.46)$$

The force vector $\mathbf{f}$ can be then approximated as

$$\mathbf{f} \approx \mathbf{U}(\mathbf{P}^T \mathbf{U})^{-1} \mathbf{P}^T \mathbf{f}, \quad (4.47)$$

and therefore the reduced forces are

$$\tilde{\mathbf{f}} = \mathbf{V}^T \mathbf{U}(\mathbf{P}^T \mathbf{U})^{-1} (\mathbf{P}^T \mathbf{f}). \quad (4.48)$$

Note that, in the above, the term $\mathbf{P}^T \mathbf{f}$ is grouped within brackets, to highlight that only the components of $\mathbf{f}$ picked by $\mathbf{P}$ have to be computed. Moreover, the matrix $\mathbf{V}^T \mathbf{U}(\mathbf{P}^T \mathbf{U})^{-1}$ can be pre-computed offline once for all. The issue then shifts to finding $\mathbf{U}$ and $\mathbf{P}$. Typically, $\mathbf{U}$ is formed by the left vectors of a Single Value Decomposition (SVD) of sampled training forces, i. e.

$$\mathbf{U}\mathbf{\Sigma}\mathbf{W}^T = \mathbf{F}, \quad (4.49)$$

where $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_{n_t}]$ contains columnwise all the sampled n_t forces, while $\mathbf{P}$ is selected via a greedy procedure; see [11] for details.

While ECSW and DEIM aim both at reducing the evaluation cost of the nonlinear terms, they differ significantly. ECSW operates directly on elements, while DEIM picks DOFs. This feature could lead to a decrease of efficiency for DEIM, as the computation of a specific component of $\mathbf{f}$ requires the evaluation of all the elements connected to such a DOF. A workaround has been proposed in [65], where DEIM is applied on the unassembled mesh. In this manner, each DOF picked by the algorithm maps to one element only. To counteract the increased cost of the SVD of training forces (now unassembled), it is shown that a surrogate quantity per element is sufficient to reproduce the nonlinear contribution. This variant goes under the name of *Surrogate Unassembled* DEIM (SU-DEIM). Note that both methods rely on training data from the HFM. As such, the hyper-reduced NLROM is optimal with respect to the training set used for its generation. At the same time, there is no guaranty that the same NLROM provides an accurate response for a set of parameter values (i. e. load, material properties, etc.) far from those related to the training set. For this reason, hyper-reduction is typically employed to generate NLROM which are local with respect to parameter values. Then there exist techniques that interpolate such NLROM over the domain of the parameters. This topic is the subject of other contributions in this book. In any case, we outline the most common strategies in Section 4.6.

4.5 Reduction methods for geometrically nonlinear systems

As briefly discussed above, geometric nonlinearities induce stretching for bending and twisting deformations. Think of the beam, pinned at both ends (i. e. the extremities of the beam cannot move, both axially and transverse) depicted in Figure 4.2. If it is bent by a pressure load, the bending will cause axial stretching. This effect is not present when the displacements are assumed infinitesimal. As such, a good reduction basis should also include vectors able to produce such effects in the reduced solution. This fact is exemplified when considering the discretized equation of a flat structure, made of linear elastic and isotropic material, as discussed in the next section.

4.5.1 Static condensation

For thin-walled structures excited in the nonlinear range, one can argue that the dominant behavior is still represented by the low-frequency dynamics spanned by bending/twisting vibration modes. Also, the eigenfrequency associated to axially dominated modes is typically of much higher frequency, as the axial to bending stiffness

Figure 4.2: Illustration of the bending–stretching coupling due to geometric nonlinearity for a planar beam. An out-of-plane displacement $w(x, t)$ causes the beam to stretch, and induce an in-plane displacement field $v(x, t)$.

ratio scales with h^{-2} , h being the thickness of the component. As such, axial modes are likely not to be excited dynamically, but rather have a nonlinear contribution to the overall displacement. In many cases indeed, it is possible to identify *slow* and *fast* DOFs and have an effective reduction of the problem size. As an example, consider here the discretized FE equations for an Euler–Bernoulli beam modeled with von Karman kinematics. The equations of motion read

$$\begin{bmatrix} \mathbf{M}_{ww} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{vv} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{w}} \\ \ddot{\mathbf{v}} \end{bmatrix} + \begin{bmatrix} \mathbf{K}_{ww} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{vv} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix} + \begin{bmatrix} {}_2\mathbf{f}_w(\mathbf{w}, \mathbf{v}) \\ {}_2\mathbf{f}_v(\mathbf{w}, \mathbf{w}) \end{bmatrix} + \begin{bmatrix} {}_3\mathbf{f}_w(\mathbf{w}, \mathbf{w}, \mathbf{w}) \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{p}_w \\ \mathbf{0} \end{bmatrix}, \quad (4.50)$$

where $\mathbf{v}$ and $\mathbf{w}$ denote the in-planes (axial) and the out-of-plane (bending) displacements. The order of the nonlinear forces ${}_2\mathbf{f}_w$, ${}_2\mathbf{f}_v$ and ${}_3\mathbf{f}_w$ is denoted by the left subscript. As is well known, the in-plane and out-of-plane dynamics is not coupled through the linear operators ($\mathbf{M}_{uw} = \mathbf{M}_{wu}^T = \mathbf{0}$, $\mathbf{K}_{vw} = \mathbf{K}_{wv}^T = \mathbf{0}$). Moreover, for many structural applications, the load is applied only in the transverse direction, i. e. $\mathbf{p}_v = \mathbf{0}$. For slender structures like the beam under consideration, the in-plane dynamics is characterized by much higher frequencies than the bending dynamics. This allows one to neglect the inertial term of the in-plane block, and define $\mathbf{v}$ as a function of $\mathbf{w}$ as

$$\ddot{\mathbf{v}} \approx \mathbf{0} \Rightarrow \mathbf{v} = -\mathbf{K}_{vv}^{-1} {}_2\mathbf{f}_v(\mathbf{w}, \mathbf{w}). \quad (4.51)$$

In other words, the in-plane DOFs $\mathbf{v}$ are a quadratic function of the out-of-plane generalized DOFs $\mathbf{w}$. (4.51) is usually referred to as *static condensation*. When (4.51) is inserted into (4.50) one obtains

$$\mathbf{M}_{ww} \ddot{\mathbf{w}} + \mathbf{K}_{ww} \mathbf{w} + {}_2\mathbf{f}_w(\mathbf{w}, \mathbf{K}_{vv}^{-1} {}_2\mathbf{f}_v(\mathbf{w}, \mathbf{w})) + {}_3\mathbf{f}_w(\mathbf{w}, \mathbf{w}, \mathbf{w}) = \mathbf{p}_w. \quad (4.52)$$

Note that the static condensation of $\mathbf{v}$ goes into forming a third-order term which “corrects” the cubic stiffness term ${}_3\mathbf{f}_w$. This physically corresponds to allow the beam

to properly compensate for the correct in-plane displacements arising from finite displacements. Clearly, this simple procedure is only effective when one could identify the slow and fast DOFs from the full order model—something possible only in the case of simple geometries [30]. For more complex thin-walled structural components, as curved stiffened panels, the dynamics is still characterized by some low-frequency modes, which statically trigger geometrically nonlinear coupling effects. In these cases, however, it is in general not possible to detect slow and fast DOFs directly from the governing discretized equation. More general methods to tackle these cases are discussed next.

4.5.2 Modal derivatives

Note that equation (4.51) implies a nonlinear (in this case quadratic) constraint between master (i. e. kept) coordinates $\mathbf{w}$ and slave (i. e. condensed) coordinates $\mathbf{v}$. On the contrary, what proposed in Section 4.4.1 relies on a linear combination of modes over the entire set of DOFs, without partitioning them into slow and fast, which can be difficult for complex geometries. As already discussed, the reduction basis $\mathbf{V}$ should span the solution accurately. For the cases we are tackling, this would mean to properly represent the stretching induced by the out-of-plane dominated modes. One could then think of introducing in $\mathbf{V}$ high-frequency axial modes to enrich the subspace. While straightforward, this strategy presents two issues, namely: (i) ideally, all modes of the system need to be computed, this is computationally very intensive for large systems, (ii) there is no criterion to select the proper vibration modes to reproduce the nonlinear coupling effectively. Typically, a reduction basis of large size results from this approach; see for instance [19].

One alternative way to achieve reduction and properly account for nonlinear stretching effects is through the use of *Modal Derivatives* (MDs) in the reduction basis.

MDs were originally proposed in [26] and [27], and more recently in [67]. Essentially, the MDs are modes shapes stemming from a pre-selected basis of vibration modes, assumed to accurately represent the dynamics of the underlying linearized system. The MDs reproduce the most important deformation shapes resulting from finite deflections in the direction of the dominant vibration modes. As such, they complement the set of vibration modes initially selected in the MDM. At first, MDs are computed by differentiating the eigenvalue problem (4.4) with respect to modal amplitudes, as

$$(\mathbf{K} - \omega^2 \mathbf{M}) \frac{\partial \boldsymbol{\phi}_i}{\partial q_j} + \left(\frac{\partial \mathbf{K}}{\partial q_j} - \frac{\partial \omega^2}{\partial q_j} \mathbf{M} - \frac{\partial \mathbf{M}}{\partial q_j} \omega^2 \right) \boldsymbol{\phi}_i = \mathbf{0}, \quad (4.53)$$

where $\boldsymbol{\theta}_{ij} = \frac{\partial \boldsymbol{\phi}_i}{\partial q_j}$ is the “MD of vibration mode $\boldsymbol{\phi}_i$ with respect to mode $\boldsymbol{\phi}_j$ ”. The above definition requires the computation of the derivative of the eigenfrequency. Moreover,

the system (4.53) is singular and requires special treatment to be solved; see [31] for a detailed discussion. Regardless of the method adopted to solve (4.53), a high dimensional matrix needs to be factorized for each MD, and this leads to high computational cost. A computationally cheaper and more theoretically sound definition of MDs resort to a static version of (4.53), where all the mass related contributions are ignored; see [67] and [69] for details. When the inertial terms are ignored, (4.53) becomes

$$\mathbf{K} \frac{\partial \boldsymbol{\phi}_i}{\partial q_j} = - \frac{\partial \mathbf{K}}{\partial q_j} \boldsymbol{\phi}_i. \quad (4.54)$$

Note that the coefficient matrix $\mathbf{K}$ is already available and therefore each MDs requires only a new right-hand side. MDs computed according to (4.54) are usually referred to as *static* MDs (SMDs). It can be proven that SMDs are symmetric, i. e.

$$\frac{\partial \boldsymbol{\phi}_i}{\partial q_j} = \frac{\partial \boldsymbol{\phi}_j}{\partial q_i}. \quad (4.55)$$

The first five vibration modes and the corresponding MDs for a flat rectangular plate are shown in Figures 4.3 and 4.4, taken from [31]. Note that, while the vibration modes (associated with the lowest eigenfrequencies) feature out-of-plane displacements only, the MDs are in-plane, and represent the displacement field that one has to add to a given vibration mode to account for the geometric nonlinearity. This separation is of course due to the flatness of the system. For more complex systems for which it is not possible to distinguish between in-plane and out-of-plane DOFs, however, the interpretation of MDs still holds. For instance, one could look at [62] for an application of MDs to the case of a shallow arch structure.

Figure 4.3: First three vibration modes of a rectangular plate, simply supported at its short edges [31]. Note that all modes feature out-of-plane displacement components only.

Once a set of m dominant vibration modes is selected, and the corresponding MDs are computed, the reduction basis $\mathbf{V}$ can be formed as

$$\mathbf{V} = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_i, \dots, \boldsymbol{\phi}_m, \boldsymbol{\theta}_{11}, \dots, \boldsymbol{\theta}_{ij}, \dots, \boldsymbol{\theta}_{mm}], \quad (4.56)$$

with $i = 1, \dots, m, j = i, \dots, m$. Due to the symmetry property (4.55), the number of MDs that can be generated is $m(m + 1)/2$. Therefore, the size of the reduction basis grows

Figure 4.4: Modal derivatives corresponding to the first five vibration modes depicted in Figure 4.3 [31]. Due to the flatness of the structure, the MDs feature in-plane displacements only, and bring the displacement field necessary to account for the geometric nonlinearity.

with m^2 . However, there are heuristic criteria to select the most relevant MDs and reduce the associated NLROM size significantly; see [64] and [31]. Note also that the computation of the MDs requires the Hessian of the stiffness matrix in the direction of the vibration modes. In case the element formulation is accessible, this can be computed at element level analytically and then assembled. This would then classify the MDs method as *intrusive*. However, it is also possible to compute MD through a finite difference procedure (and therefore *non-intrusively*) only by using FE assembly level information typically made available in commercial codes, as outlined in [61]. As a note, MDs have been successfully used for flexible multibody applications, where nonlinear elastic deflections are superimposed to large rigid body motions [68, 69].

4.5.3 Dual modes

Another way to account for the transverse-membrane coupling is by appending the so called *Dual Modes* (DMs) to the dominant linear vibration modes in the reduction basis [38]. The main idea of DMs is based on applying a series of nonlinear static forces to the HFM to obtain related nonlinear displacements. These displacements are then orthogonalized with respect to the linear vibration modes, to yield shapes that represent the nonlinear contribution. The most common procedure for computing the DMs is described in [47, 52] and it is here briefly outlined.

We start from a set of mass normalized vibration modes $\mathbf{V}_{\text{VM}} = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_m]$, $m \ll n$, typically selected by the MDM (see Section 4.2). The procedure then selects a *dominant* vibration mode $\boldsymbol{\phi}_d$, where $d \in \mathcal{M}$ is the index relative to the mode with the maximum load participation factor, i. e.

$$d = \operatorname{argmax}_{i \in \mathcal{M}} (|\mathbf{p}^T \boldsymbol{\phi}_i|), \quad (4.57)$$

where $\mathbf{p}$ is the external force in (4.30). Then we form a set of scaling factors $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_r]$, $\boldsymbol{\alpha} \in \mathbb{R}^r$ such that the static solutions $\mathbf{x}_i$ of

$$\mathbf{f}(\mathbf{x}_i) - \alpha_i \mathbf{K} \boldsymbol{\phi}_d = \mathbf{0} \quad (4.58)$$

trigger displacements ranging from linear to strongly nonlinear. That is to say, $\mathcal{O}(\|\mathbf{f}(\mathbf{x}_i) - \mathbf{K} \mathbf{x}_i\|) \ll \mathcal{O}(\|\mathbf{K} \mathbf{x}_i\|)$ for small $|\alpha_i|$ and $\mathcal{O}(\|\mathbf{f}(\mathbf{x}_i) - \mathbf{K} \mathbf{x}_i\|) \approx \mathcal{O}(\|\mathbf{K} \mathbf{x}_i\|)$ for large $|\alpha_i|$, respectively. Let $\mathcal{R} = [1, \dots, r]$ be the set of indices associated to $\boldsymbol{\alpha}$. It is now possible to create load sets,

$$\mathbf{G}^{(j)} = \operatorname{diag}(\boldsymbol{\alpha}) \mathbf{K} \left(\frac{\boldsymbol{\phi}_d + \boldsymbol{\phi}_j}{2} \right), \quad j \in \mathcal{M}, \quad (4.59)$$

and let $\mathbf{g}_i^{(j)}$ be the i th column of $\mathbf{G}^{(j)}$. Now, for each $j \in \mathcal{M}$:

1. Solve the static problems

$$\mathbf{f}(\mathbf{x}_i^{(j)}) - \mathbf{g}_i^{(j)} = \mathbf{0}, \quad i \in \mathcal{R}, \quad (4.60)$$

and collect the obtained solutions in the matrix $\mathbf{X}^{(j)}$ as

$$\mathbf{X}^{(j)} = [\mathbf{x}_1^{(j)}, \dots, \mathbf{x}_r^{(j)}]. \quad (4.61)$$

2. Mass-orthogonalize each column of $\mathbf{X}^{(j)}$ to $\mathbf{V}_{\text{VM}}$, i. e.

$$\mathbf{x}_{i_\perp}^{(j)} = \mathbf{x}_i^{(j)} - \sum_{k=1}^m (\boldsymbol{\phi}_k^T \mathbf{M} \mathbf{x}_i^{(j)}) \boldsymbol{\phi}_k, \quad i \in \mathcal{R}, \quad (4.62)$$

and collect them into $\mathbf{X}_\perp^{(j)}$ as

$$\mathbf{X}_\perp^{(j)} = [\mathbf{x}_{1_\perp}^{(j)}, \dots, \mathbf{x}_{r_\perp}^{(j)}]. \quad (4.63)$$

This step is performed to exclude the linear contribution from the static solutions $\mathbf{x}_i^{(j)}$.

3. Perform an SVD of $\mathbf{X}_\perp^{(j)}$, and retain the singular vectors associated to the lowest $k < r$ singular values, as

$$\bar{\boldsymbol{\Psi}}^{(j)} \mathbf{S}^{(j)} \mathbf{R}^{(j)} = \operatorname{SVD}(\mathbf{X}_\perp^{(j)}), \quad (4.64)$$

$$\bar{\boldsymbol{\Psi}}^{(j)} = [\hat{\boldsymbol{\psi}}_1^{(j)}, \dots, \hat{\boldsymbol{\psi}}_k^{(j)}] \mid \sigma_1 > \sigma_2 > \dots > \sigma_k > \sigma_{k+1} > \dots > \sigma_n, \quad (4.65)$$

where σ_i is the i th diagonal component of $\mathbf{S}^{(j)}$.

4. Compute the linearized strain energy measures $E_i^{(j)}$ of $\mathbf{x}_i^{(j)}$ associated to each column of $\bar{\Psi}^{(j)}$ (i. e. to each dual mode candidate), as

$$E_i^{(j)} = \bar{\psi}_i^{(j)T} \mathbf{K} \bar{\psi}_i^{(j)} \sum_{s=1}^r \beta_s^{(j)2}, \quad \beta_s^{(j)} = \frac{\bar{\psi}_i^{(j)T} \mathbf{x}_s^{(j)}}{\bar{\psi}_i^{(j)T} \bar{\psi}_i^{(j)}}, \quad (4.66)$$

which are collected in a vector $\mathbf{E}^{(j)} = [E_1^{(j)}, \dots, E_r^{(j)}]$.

5. Select the most relevant $s < k$ DMs $\psi_{p_l}^{(j)}$, $l = 1, \dots, s$ as the columns of $\bar{\Psi}^{(j)}$ associated to the largest entries of $\mathbf{E}^{(j)}$, i. e.

$$E_{p_1}^{(j)} > E_{p_2}^{(j)} > \dots > E_{p_s}^{(j)}, \quad (4.67)$$

and let $\mathbf{V}_{\text{DM}}^{(j)} = [\psi_{p_1}^{(j)}, \dots, \psi_{p_s}^{(j)}]$.

6. Move to the next load set: $j \leftarrow j + 1$.

The total collection of selected DMs is then $\mathbf{V}_{\text{DM}} = [\mathbf{V}_{\text{DM}}^{(1)} \dots \mathbf{V}_{\text{DM}}^{(j)} \dots \mathbf{V}_{\text{DM}}^{(m)}]$. The reduction basis $\mathbf{V}$ is formed as

$$\mathbf{V} = [\mathbf{V}_{\text{VM}} \quad \mathbf{V}_{\text{DM}}]. \quad (4.68)$$

Note that the computation of the DMs is intrinsically *non-intrusive*, as only nonlinear static solutions of the HFM are required. Also, the procedure just described is *load-dependent*, meaning that the information of the shape of the external load exerted is needed. Typically, the computation of DMs requires a careful selection of the scaling factors α to trigger the nonlinearity at the desired value.

4.5.4 Proper orthogonal decomposition

As alternative to model-based techniques, a reduction basis for Galerkin projection can be from data obtained by sampling HFM solutions. The most popular of such method is the *Proper Orthogonal Decomposition* (POD) [2, 35, 36, 41], also known as Karhunen–Loeve decomposition or principal component analysis (PCA). This topic is extensively treated in [8, Chapter 2]. Here we just discuss the main idea.

Let us assume to have collected n_t time snapshots of a HFM solution in a matrix $\mathbf{U} \in \mathbb{R}^{n \times n_t}$, $\mathbf{U} := [\mathbf{u}_1, \dots, \mathbf{u}_{n_t}]$. The POD seeks for a lower dimensional representation of $\mathbf{U}$ by a basis $\mathbf{V} \in \mathbb{R}^{n \times m}$, $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_m]$, $m \ll n_t$ by solving the following least squares problem:

$$\min_{\mathbf{v}_i \in \mathbb{R}^n} \sum_{j=1}^{n_t} \left\| \mathbf{u}_j - \sum_{i=1}^m (\mathbf{u}_j^T \mathbf{v}_i) \mathbf{v}_i \right\|_2^2. \quad (4.69)$$

It can be shown that the vectors $\mathbf{v}_i$ are the left singular vectors of $\mathbf{U}$, obtained by the Singular Value Decomposition (SVD) of $\mathbf{U}$ as

$$\mathbf{U} = \mathbf{L}\mathbf{S}\mathbf{R}^T, \quad (4.70)$$

where $\mathbf{L} = [\mathbf{l}_1, \dots, \mathbf{l}_{n_t}]$, $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_n]$, $\mathbf{L}\mathbf{L}^T = \mathbf{I} \in \mathbb{R}^{n_t \times n_t}$, $\mathbf{R}\mathbf{R}^T = \mathbf{I} \in \mathbb{R}^{n \times n}$ and $\mathbf{S}$ is a rectangular matrix containing the singular values on the diagonal, and zero otherwise. For details on algorithms to compute SVD, refer for instance to [23]. These singular values represent the relative importance of corresponding eigenvectors of $\mathbf{L}$ in forming the basis $\mathbf{V}$. If the singular values S_{ii} , and the corresponding left and right vectors $\mathbf{l}_i$ and $\mathbf{r}_i$ are sorted in a decreasing order, i. e. $S_{11} > S_{22} > \dots, S_{n_t}$, it can be shown that the error norm is bounded, as

$$\sum_{j=1}^{n_t} \left\| \mathbf{u}_j - \sum_{i=1}^m (\mathbf{u}_j^T \mathbf{l}_i)^T \mathbf{l}_i \right\|_2^2 = \sum_{i=m+1}^{n_t} S_{ii}^2. \quad (4.71)$$

In other words, the left singular vectors $\mathbf{l}_1, \dots, \mathbf{l}_m$ corresponding to the highest singular values are the most significant for constructing a reduction basis. In Figure 4.5, we show an example of a POD analysis for the nonlinear dynamic response of a plate. Note that, in this context, the left vectors $\mathbf{r}_i$ represent the (normalized) time evolution of the corresponding left singular vector. Since an SVD can be performed for any set of snapshots of any general nonlinear problem, POD is widespread as a versatile method to form a reduction basis. It should be noted, however, that the POD basis is optimal only for the set of snapshots which were used to generate it, while there is no guarantee on the accuracy of the NLROM outside the training data. For this reason, an NLROM based on POD are used in parametric contexts, i. e. when the HFM is parametrized with respect to material, geometric and load properties. In these cases, POD bases are obtained for different instances of the parameter set. Then parametric NLROMs are formed by either building a larger POD basis that spans the solution within the parameters range of interest [5], or by interpolating local NLROMs online for the desired value of the parameters which were not previously sampled [3]. In this framework, the high computational cost associated to the HFM solution required by the POD are amortized by the large number of queries to the NROM. For further discussion of this matter, we refer to Section 4.6.

4.5.5 Non-intrusive methods

As already mentioned, non-intrusive model order reduction approaches are beneficial when a HFM model is developed by means of a FE model with no access to element forces and Jacobians required for the analytical computations of reduced nonlinear internal forces. There are various strategies to non-intrusively develop a nonlinear reduced order model. We outline here the most common methods.

Figure 4.5: Example of a POD applied to a nonlinear structural dynamics case. A flat plate is simply supported on three sides, and loaded with step pressure, as shown. The HFM linear and nonlinear response the mid node of the free side of the plate are shown. The marked difference of the two responses highlight the significance of the nonlinear term. The first three left and right singular vectors, and the corresponding singular values are shown on the bottom of the figure. Note the resemblance of the dominant first left vector l_1 and its corresponding time evolution r_1 , with a random snapshot of the HFM (top left corner).

4.5.5.1 Enforced displacement

One way to identify the nonlinear stiffness coefficients of a reduced order model is to prescribe a set of nonlinear static displacements to the HFM and solve for the corresponding reaction forces that generate them. This method, which is known as *Enforced Displacement* (ED) or *Stiffness Evaluation Procedure* (STEP), was first developed by Muravyov et al. [48] and later extended by Kim et al. [38] for the case of unknown linear stiffness components. Essentially, ED first collects the reaction forces due to statically enforced displacements, projects them onto the modal space, and set linear equations to obtain the nonlinear stiffness coefficients $K_{ijl}^{(2)}$ and $K_{ijlp}^{(3)}$, $i, j, l, p = 1, \dots, m$.

Let us consider the reduced nonlinear restoring forces equation (4.40) already introduced in Section 4.4.1.3, which reads

$$\tilde{f}_i = \tilde{K}_{ij}^{(1)} q_j + \tilde{K}_{ijl}^{(2)} q_j q_l + \tilde{K}_{ijlp}^{(3)} q_j q_l q_p, \quad (4.72)$$

with $i = 1, 2, \dots, m$, where summation over repeated indices is implied. The second-order tensor $\tilde{K}_{ij}^{(1)}$ can be immediately found to be

$$\tilde{K}_{ij}^{(1)} = \mathbf{V}_i^T \mathbf{K} \mathbf{V}_j, \quad (4.73)$$

where $\mathbf{K}$ is the stiffness matrix of the HFM, usually obtainable from any FE program. Clearly, $\tilde{K}_{ij}^{(1)} = \tilde{K}_{ji}^{(1)}$.

The ED method sequentially set linear systems of equation whose unknowns are $K_{ijl}^{(2)}$ and $K_{ijlp}^{(3)}$. The first step consists of choosing a modal amplitude $\pm q_r$, $r = 1, \dots, m$ and assign it to the reduced static problem as

$$\tilde{K}_{ir}^{(1)} q_r + \tilde{K}_{irr}^{(2)} q_r^2 + \tilde{K}_{irr}^{(3)} q_r^3 = \tilde{p}_i^{(r_+)}, \quad (4.74a)$$

$$-\tilde{K}_{ir}^{(1)} q_r + \tilde{K}_{irr}^{(2)} q_r^2 - \tilde{K}_{irr}^{(3)} q_r^3 = \tilde{p}_i^{(r_-)}, \quad (4.74b)$$

with $i = 1, \dots, m$. The right-hand sides $\tilde{p}_i^{(r_+)}$ and $\tilde{p}_i^{(r_-)}$ are given by

$$\tilde{p}_i^{(r_+)} = \mathbf{V}_i^T \mathbf{f}(\mathbf{u}^{(r_+)}), \quad (4.75)$$

$$\tilde{p}_i^{(r_-)} = \mathbf{V}_i^T \mathbf{f}(\mathbf{u}^{(r_-)}), \quad (4.76)$$

where $\mathbf{u}^{(r_+)} = \mathbf{V}_r q_r$ and $\mathbf{u}^{(r_-)} = -\mathbf{V}_r q_r$. In words, $\mathbf{u}^{(r_+)}$ and $\mathbf{u}^{(r_-)}$ are imposed statically to the HFM, and the resulting reaction forces are projected onto $\mathbf{V}_i$.

Likewise, we can select pairs of modal amplitudes (q_r, q_s) , $(-q_r, -q_s)$ and $(q_r, -q_s)$ and impose it to the reduced static problem as

$$\begin{aligned} \tilde{K}_{ir}^{(1)} q_r + \tilde{K}_{is}^{(1)} q_s + \tilde{K}_{irr}^{(2)} q_r^2 + \tilde{K}_{irs}^{(2)} q_r q_s + \tilde{K}_{iss}^{(2)} q_s^2 \\ + \tilde{K}_{irr}^{(3)} q_r^3 + q_r^2 q_s + 3\tilde{K}_{irss}^{(3)} q_r q_s^2 + \tilde{K}_{iss}^{(3)} q_s^3 = \tilde{p}_i^{(rs_{++})} \end{aligned} \quad (4.77a)$$

$$\begin{aligned} -\tilde{K}_{ir}^{(1)} q_r - \tilde{K}_{is}^{(1)} q_s + \tilde{K}_{irr}^{(2)} q_r^2 + \tilde{K}_{irs}^{(2)} q_r q_s + \tilde{K}_{iss}^{(2)} q_s^2 \\ - \tilde{K}_{irr}^{(3)} q_r^3 - \tilde{K}_{irs}^{(3)} q_r^2 q_s - \tilde{K}_{irss}^{(3)} q_r q_s^2 - \tilde{K}_{iss}^{(3)} q_s^3 = \tilde{p}_i^{(rs_{--})} \end{aligned} \quad (4.77b)$$

$$\begin{aligned} \tilde{K}_{ir}^{(1)} q_r - \tilde{K}_{is}^{(1)} q_s + \tilde{K}_{irr}^{(2)} q_r^2 - \tilde{K}_{irs}^{(2)} q_r q_s + \tilde{K}_{iss}^{(2)} q_s^2 \\ + \tilde{K}_{irr}^{(3)} q_r^3 - \tilde{K}_{irs}^{(3)} q_r^2 q_s + \tilde{K}_{irss}^{(3)} q_r q_s^2 - \tilde{K}_{iss}^{(3)} q_s^3 = \tilde{p}_i^{(rs_{+-})}, \end{aligned} \quad (4.77c)$$

where we restrict the case to $s \geq r$. The right-hand sides of (4.77a)–(4.77c) are given by

$$\tilde{p}_i^{(rs_{++})} = \mathbf{V}_i^T \mathbf{f}(\mathbf{u}^{(rs_{++})}), \quad (4.78a)$$

$$\tilde{p}_i^{(rs_{--})} = \mathbf{V}_i^T \mathbf{f}(\mathbf{u}^{(rs_{--})}), \quad (4.78b)$$

$$\tilde{p}_i^{(rs_{+-})} = \mathbf{V}_i^T \mathbf{f}(\mathbf{u}^{(rs_{+-})}), \quad (4.78c)$$

where, analogously to (4.75), we statically impose to the HFM $\mathbf{u}^{(rs_{++})} = \mathbf{V}_r q_r + \mathbf{V}_s q_s$, $\mathbf{u}^{(rs_{--})} = -\mathbf{V}_r q_r - \mathbf{V}_s q_s$ and $\mathbf{u}^{(rs_{+-})} = \mathbf{V}_r q_r - \mathbf{V}_s q_s$. Lastly, we need to compute $\tilde{K}_{irsp}^{(3)}$,

$r \neq s \neq p$. Similarly to the previous steps, we determine a triple (q_r, q_s, q_p) and insert it to the reduced static problem. This results in a single equation for $\tilde{K}_{irsp}^{(3)}$, as

$$\begin{aligned} & \tilde{K}_{ir}^{(1)} q_r + \tilde{K}_{is}^{(1)} q_s + \tilde{K}_{ip}^{(1)} q_p + \tilde{K}_{irr}^{(2)} q_r^2 + \tilde{K}_{iss}^{(2)} q_s^2 + \tilde{K}_{ipp}^{(2)} q_p^2 \\ & + \tilde{K}_{irs}^{(2)} q_r q_s + \tilde{K}_{irp}^{(2)} q_r q_p + \tilde{K}_{isp}^{(2)} q_s q_p + \tilde{K}_{irr}^{(3)} q_r^3 + \tilde{K}_{iss}^{(3)} q_s^3 + \tilde{K}_{ipp}^{(3)} q_p^3 \\ & + \tilde{K}_{irrs}^{(3)} q_r^2 q_s + \tilde{K}_{irrp}^{(3)} q_r^2 q_p + \tilde{K}_{irss}^{(3)} q_r q_s^2 + \tilde{K}_{irpp}^{(3)} q_r q_p^2 + \tilde{K}_{ipss}^{(3)} q_p q_s^2 + \tilde{K}_{ispp}^{(3)} q_s q_p^2 \\ & + \tilde{K}_{irsp}^{(3)} q_r q_s q_p = \tilde{p}_i^{(rsp)}, \end{aligned} \quad (4.79)$$

with $p \geq s \geq r$, and

$$\tilde{p}_i^{(rsp)} = \mathbf{V}_i^T \mathbf{f}(\mathbf{u}^{(rsp)}), \quad (4.80)$$

where $\mathbf{u}^{(rsp)} = \mathbf{V}_r q_r + \mathbf{V}_s q_s + \mathbf{V}_p q_p$. Note that the restriction $s > r$ in (4.77a)–(4.77c) and $p > s > r$ in (4.79) is in fact setting $\tilde{K}_{irs}^{(2)} = 0 \forall s < r$ and $\tilde{K}_{irsp}^{(2)} = 0 \forall p < s < r$. This is done to minimize the number of coefficients multiplying the same combinations of q_r, q_s and $q_r q_s q_p, \forall r, s, p = 1, \dots, m$. Alternatively, one could set symmetry properties as

$$K_{ijl}^{(2)} = K_{ijl}^{(2)} \quad (4.81)$$

and

$$K_{ijlp}^{(3)} = K_{ijpl}^{(3)} = K_{iljp}^{(3)} = K_{ilpj}^{(3)} = K_{ipjl}^{(3)} = K_{iplj}^{(3)}, \quad (4.82)$$

and modify (4.77a)–(4.77c) and (4.79) accordingly. The total number N_{ED} of the required nonlinear static evaluations to fully construct the nonlinear reduced restoring forces is

$$N_{ED} = 2m + 3 {}_m C_2 + {}_m C_3, \quad (4.83)$$

where

$${}_m C_r = \frac{m!}{(m-r)! r!}. \quad (4.84)$$

Since this number is $\mathcal{O}(m^3)$, the computational cost associated to ED grows quickly with the number of retained modes and severely hinders the applicability of the method.

4.5.5.2 Enhanced enforced displacement

To reduce the computational costs of the ED method, Perez et al. [52] proposed a way to identify the NLROM coefficients in the case that the tangent stiffness matrix is released

by the FE program at hand. We refer to this method as *Enhanced Enforced Displacements* (EED). The tangent stiffness matrix $\mathbf{K}^t(\mathbf{u})$ is the Jacobian of the nonlinear forces $\mathbf{f}$ with respect to the nodal displacements, as

$$\mathbf{K}^t(\mathbf{u}) = \frac{\partial \mathbf{f}}{\partial \mathbf{u}}. \quad (4.85)$$

Let us assume that $\mathbf{K}^t(\mathbf{u})$ is available. Then the reduced tangent stiffness matrix $\tilde{\mathbf{K}}^t$ can be obtained from $\mathbf{K}^t$ as

$$\tilde{\mathbf{K}}^t = \mathbf{V}^T \mathbf{K}^t(\mathbf{V}\mathbf{q})\mathbf{V}, \quad (4.86)$$

whose ir component is given by

$$\tilde{K}_{ir}^t = \mathbf{V}_i^T \mathbf{K}^t(\mathbf{V}\mathbf{q})\mathbf{V}_r, \quad (4.87)$$

and we pose $\mathbf{u} = \mathbf{V}\mathbf{q}$. The reduced tangent stiffness matrix $\tilde{\mathbf{K}}^t$ is the Jacobian of the reduced nonlinear internal force vector, $\tilde{\mathbf{f}}$, namely

$$\tilde{\mathbf{K}}^t(\mathbf{q}) = \begin{bmatrix} \frac{\partial \tilde{f}_1}{\partial q_1} & \dots & \frac{\partial \tilde{f}_1}{\partial q_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{f}_m}{\partial q_1} & \dots & \frac{\partial \tilde{f}_m}{\partial q_m} \end{bmatrix}. \quad (4.88)$$

Then, following from (4.40), $\tilde{K}_{ir}$ reads

$$\begin{aligned} \tilde{K}_{ir}^t &= \frac{\partial \tilde{f}_i}{\partial q_r} = \frac{\partial}{\partial q_r} [\tilde{K}_{ij}^{(1)} q_j + \tilde{K}_{ijl}^{(2)} q_j q_l + \tilde{K}_{ijlp}^{(3)} q_j q_l q_p] \\ &= \tilde{K}_{ir}^{(1)} + [\tilde{K}_{ijr}^{(2)} + \tilde{K}_{irj}^{(2)}] q_j + [\tilde{K}_{ijlr}^{(3)} + \tilde{K}_{ijrl}^{(3)} + \tilde{K}_{irjl}^{(3)}] q_j q_l. \end{aligned} \quad (4.89)$$

The EED method equates (4.89) and (4.87) for as many combinations of modal amplitudes as those required to solve for the reduced stiffness coefficients $\tilde{K}_{ijr}^{(2)}$ and $\tilde{K}_{ijrl}^{(3)}$, $i, j, r, l = 1, \dots, m$. Similarly to the ED method, first two static displacement vectors for each column of the reduction basis are constructed and inserted in (4.86). We can then set two equations as

$$\tilde{K}_{ir}^{t(a)} = \tilde{K}_{ir}^{(1)} + [\tilde{K}_{ijr}^{(2)} + \tilde{K}_{irj}^{(2)}] q_j^{(a)} + [\tilde{K}_{ijjr}^{(3)} + \tilde{K}_{ijrj}^{(3)} + \tilde{K}_{irjj}^{(3)}] (q_j^{(a)})^2, \quad a = 1, 2. \quad (4.90)$$

Since the elements $\tilde{K}_{ijl}^{(2)}$ and $\tilde{K}_{ijlp}^{(3)}$ were assumed to be zero unless $p \geq l \geq j$, (4.90) splits into three different cases, namely

$$\tilde{K}_{ir}^{t(a)} = \tilde{K}_{ir}^{(1)} + \tilde{K}_{ijr}^{(2)} q_j^{(a)} + \tilde{K}_{ijjr}^{(3)} (q_j^{(a)})^2, \quad \text{if } j < r, \quad (4.91a)$$

$$\tilde{K}_{ir}^{t(a)} = \tilde{K}_{ir}^{(1)} + 2\tilde{K}_{irr}^{(2)} q_j^{(a)} + 3\tilde{K}_{irrr}^{(3)} (q_j^{(a)})^2, \quad \text{if } j = r, \quad (4.91b)$$

$$\tilde{K}_{ir}^{(a)} = \tilde{K}_{ir}^{(1)} + \tilde{K}_{irj}^{(2)} q_j^{(a)} + \tilde{K}_{irjj}^{(3)} (q_j^{(a)})^2, \quad \text{if } j > r. \quad (4.91c)$$

The solutions of (4.90) and (4.91a)–(4.91c) yield the coefficients $\tilde{K}_{ijr}^{(2)}$, $\tilde{K}_{ijjr}^{(3)}$, $\tilde{K}_{ijrr}^{(3)}$, and $\tilde{K}_{ijjj}^{(3)}$, $j, r = 1, \dots, m$. At this point, the only set of coefficients to be identified is $\tilde{K}_{ijlr}^{(3)}$, $j \neq l \neq r$, $j, l, r = 1, \dots, m$. To calculate these coefficients one can use displacement vectors which are formed from combinations of two different columns of $\mathbf{V}$ by

$$\mathbf{u} = \mathbf{V}_j q_j + \mathbf{V}_l q_l, \quad j, l = 1, \dots, m, \quad (4.92)$$

where q_j and q_l are prescribed. These displacements are again inserted into (4.86) to obtain their corresponding tangent stiffness matrices. The ir th component of the associated reduced tangent stiffness matrix is then given by

$$\tilde{K}_{ir}^t = \tilde{K}_{ir}^{(1)} + [\tilde{K}_{ijr}^{(2)} q_j + \tilde{K}_{ilr}^{(2)} q_l] + [\tilde{K}_{ijlr}^{(3)} q_j q_l + \tilde{K}_{ijjr}^{(3)} q_j^2 + \tilde{K}_{illr}^{(3)} q_l^2]. \quad (4.93)$$

The only unknown in (4.93) is $\tilde{K}_{ijlr}^{(3)}$, which can be found without needing combinations of three columns of $\mathbf{V}$ as for the case of the ED method. As a result, EED is computationally more efficient than ED. In fact, the total number of required static solutions N_{EED} is then

$$N_{\text{EED}} = 2m + {}_m C_2, \quad (4.94)$$

which is $\mathcal{O}(m^2)$, as opposed to $\mathcal{O}(m^3)$ for the ED method. Note that the amplitudes q_i , $i = 1, \dots, m$ are user defined. Typically, they are selected such that the resulting displacements generate nonlinear forces of the desired magnitude. For shell-like structures, it is usually suggested to select q_i , $i = 1, \dots, m$ in the order of one thickness of the structure for transverse-dominated modes and 0.1 to 0.01 of that for membrane-dominated ones. A flowchart of the ED and EED method is shown in Figure 4.6.

4.5.5.3 Implicit condensation

In order to identify the coefficients of (4.40) via either the ED or the EED method, the reduction basis $\mathbf{V}$ must contain vibration modes and corresponding membrane-dominated vectors (e. g. MDs, DMs or manually selected high-frequency VMs). This increases the size of the basis significantly and impacts the efficiency of the method. However, it is possible to avoid the use of membrane-dominated modes by using the method of *Implicit Condensation* (IC) or *Applied Force* method, developed by McEwan et al. [45, 46]. Let $\mathbf{p}^{(s)}$ be a generic load obtained by combining one, two and three columns of $\mathbf{V}$ with chosen scaling factors $\bar{q}_i$, $\bar{q}_j$, $\bar{q}_k$ as

$$\mathbf{p}^{(s)} = \mathbf{K}(\mathbf{V}_i(\pm\bar{q}_i) + \mathbf{V}_j(\pm\bar{q}_j) + \mathbf{V}_k(\pm\bar{q}_k)), \quad i, j, k = 0, 1, 2, \dots, m, \quad i \neq j \neq k, \quad (4.95)$$

Figure 4.6: Flowchart of comparison between ED and EED steps to identify nonlinear stiffness coefficients. The EED method manages to identify all the nonlinear stiffness coefficients in two steps only, because of the availability of the full order tangent stiffness matrix [34].

where the index 0 implies that the corresponding term is ignored, i. e.

$$\mathbf{K}(\mathbf{V}_0 \bar{q}_0 + \mathbf{V}_1 \bar{q}_1 + \mathbf{V}_2 \bar{q}_2) = \mathbf{K}(\mathbf{V}_1 \bar{q}_1 + \mathbf{V}_2 \bar{q}_2).$$

One can show that the total number of cases N_{IC} that can be generated from (4.95) is

$$N_{IC} = 2m + 4_m C_2 + 8_m C_3, \quad (4.96)$$

and therefore $s = 1, \dots, N_{IC}$. The scaling factors $\bar{q}_i$, $i = 1, \dots, m$ should be chosen such that the resulting displacements are in the nonlinear regime. Gordon and Hollkamp [20] propose that a scaling factor $\bar{q}_i$ for (4.95) should generate a force that induces a desired maximum displacement $W_{i_{\max}}$ as

$$\bar{q}_i = \frac{W_{i_{\max}}}{V_{i_{\max}}} \omega_i^2, \quad (4.97)$$

where $V_{i_{\max}}$ denotes the component of $\mathbf{V}_i$ mode which has the maximum out-of-plane displacement, and ω_i is the eigenfrequency of mode $\mathbf{V}_i$. The maximum desired displacement $W_{i_{\max}}$, is usually chosen in the order of the thickness of the structure to properly exercise the nonlinearity.

Each load $\mathbf{p}^{(s)}$ is applied to the static HFM as

$$\mathbf{f}(\mathbf{u}^{(s)}) = \mathbf{p}^{(s)}. \quad (4.98)$$

The resulting displacement $\mathbf{u}^{(s)}$ is then expressed in the subspace spanned by $\mathbf{V}$ through the modal amplitudes $\mathbf{q}^{(s)}$ as

$$\mathbf{u}^{(s)} = \mathbf{V}\mathbf{q}^{(s)} + \hat{\mathbf{u}}^{(s)} \Rightarrow \mathbf{q}^{(s)} = \mathbf{V}^T \mathbf{M}\mathbf{u}^{(s)}, \quad (4.99)$$

where $\hat{\mathbf{u}}^{(s)}$ is the remainder and it is assumed that $\mathbf{V}^T \mathbf{M}\hat{\mathbf{u}}^{(s)} = \mathbf{0}$. Likewise, $\mathbf{p}^{(s)}$ is projected on $\mathbf{V}$ to give the modal loads $\tilde{\mathbf{p}}^{(s)}$ as

$$\tilde{\mathbf{p}}^{(s)} = \mathbf{V}^T \mathbf{p}^{(s)}. \quad (4.100)$$

Then each pair $(\tilde{\mathbf{p}}^{(s)}, \mathbf{q}^{(s)})$ is inserted in the static reduced order problem to obtain

$$\tilde{K}_{ijl}^{(2)} q_j^{(s)} q_l^{(s)} + \tilde{K}_{ijlp}^{(3)} q_j^{(s)} q_l^{(s)} q_p^{(s)} = \tilde{p}_i^{(s)} - \omega_i^2 q_i^{(s)}, \quad i = 1, 2, \dots, m, \quad (4.101)$$

where it is further assumed that the symmetry properties (4.81) and (4.82) hold. All equations resulting from (4.101) can be written in matrix form as

$$\mathbf{G}_{nl} \mathbf{K}_{nl} = \mathbf{P}_{nl}, \quad (4.102)$$

where

$$\mathbf{G}_{nl}^T = \begin{bmatrix} (q_1^{(1)})^2 & \dots & (q_1^{(N_{lc})})^2 \\ \vdots & \ddots & \vdots \\ q_i^{(1)} q_j^{(1)} & \ddots & q_i^{(N_{lc})} q_j^{(N_{lc})} \\ \vdots & \ddots & \vdots \\ (q_m^{(1)})^2 & \dots & (q_m^{(N_{lc})})^2 \\ (q_1^{(1)})^3 & \dots & (q_1^{(N_{lc})})^3 \\ \vdots & \ddots & \vdots \\ q_i^{(1)} q_j^{(1)} q_k^{(1)} & \ddots & q_i^{(N_{lc})} q_j^{(N_{lc})} q_k^{(N_{lc})} \\ \vdots & \ddots & \vdots \\ (q_m^{(1)})^3 & \dots & (q_m^{(N_{lc})})^3 \end{bmatrix}, \quad (4.103)$$

$$\mathbf{K}_{nl}^T = \begin{bmatrix} \tilde{K}_{111}^{(2)} & \dots & \tilde{K}_{1ij}^{(2)} & \dots & \tilde{K}_{1mm}^{(2)} & \tilde{K}_{1111}^{(3)} & \dots & \tilde{K}_{1ijk}^{(3)} & \dots & \tilde{K}_{1mmmm}^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{K}_{m11}^{(2)} & \dots & \tilde{K}_{mij}^{(2)} & \dots & \tilde{K}_{mmm}^{(2)} & \tilde{K}_{m111}^{(3)} & \dots & \tilde{K}_{mijk}^{(3)} & \dots & \tilde{K}_{mmmm}^{(3)} \end{bmatrix}, \quad (4.104)$$

and

$$\mathbf{P}_{nl} = \begin{bmatrix} \tilde{p}_1^{(1)} - \omega_1^2 q_1^{(1)} & \dots & \tilde{p}_1^{(N_{IC})} - \omega_1^2 q_1^{(N_{IC})} \\ \vdots & \ddots & \vdots \\ \tilde{p}_m^{(1)} - \omega_m^2 q_1^{(1)} & \dots & \tilde{p}_m^{(N_{IC})} - \omega_m^2 q_1^{(N_{IC})} \end{bmatrix}. \quad (4.105)$$

The system of equations (4.102) can be solved by means of any regression technique, for instance the least square method.

The main advantage of the IC method is the fact that the resulting NROM is based on vibration modes only, and thus is very compact as compared to the ED and EED method, which require also modes (e. g. DMs or MDs,) that represent the membranal behavior due to the large displacements. In the case of ED and EED is then possible to directly retrieve a full displacement field, and from it compute strain and stresses. This feature is crucial when the final aim of the analysis is to perform stress/strain investigations (for instance for fatigue life estimation). On the contrary, the IC method does not offer this possibility, as membranal behavior is not present in the basis.

To overcome this issue, Holkkamp and Gordon [24] developed a post-processing expansion procedure, which approximate the response of these DOFs without increasing the number of required static solutions in IC. In this method, the total displacement vector $\mathbf{u}^{(s)}$ relative to load case s can be written as

$$\mathbf{u}^{(s)} = \mathbf{V}\mathbf{q}^{(s)} + \hat{\mathbf{u}}^{(s)}, \quad (4.106)$$

where $\mathbf{u}^{(s)}$ is the solution of (4.98), and $\hat{\mathbf{u}}^{(s)}$ is the correction vector for in-plane DOFs, to be found by the expansion procedure. We can express $\hat{\mathbf{u}}^{(s)}$ in a modal fashion as

$$\hat{\mathbf{u}}^{(s)} = \hat{\mathbf{V}}\hat{\mathbf{q}}^{(s)} \quad (4.107)$$

where $\hat{\mathbf{V}}$ is the transformation matrix and $\hat{\mathbf{q}}^{(s)}$ is the vector of modal membrane coordinates. The whole displacement set for $s = 1, \dots, N_{IC}$ can be compactly written as

$$\mathbf{U} = \mathbf{V}\mathbf{Q} + \hat{\mathbf{V}}\hat{\mathbf{Q}}, \quad (4.108)$$

where $\mathbf{U} = [\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(N_{IC})}]$, $\mathbf{Q} = [\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(N_{IC})}]$ while $\hat{\mathbf{V}}$ and $\hat{\mathbf{Q}}$ are yet to be determined. It is assumed that $\hat{\mathbf{V}}^T \mathbf{M} \mathbf{V} = \mathbf{0}$, that is to say, the in-plane displacement to be retrieved lies in a subspace which is $\mathbf{M}$ -orthogonal to $\mathbf{V}$. Therefore,

$$\mathbf{Q} = \mathbf{V}^T \mathbf{M} \mathbf{U}. \quad (4.109)$$

It is then assumed that each column $\hat{\mathbf{q}}^{(s)}$ of $\hat{\mathbf{Q}}$ is given by

$$\hat{\mathbf{q}}^{(s)} = [(q_1^{(s)})^2, q_1^{(s)} q_2^{(s)}, \dots, q_i^{(s)} q_j^{(s)}, \dots, (q_m^{(s)})^2]^T; \quad (4.110)$$

see [49] for more details. Then, by virtue of (4.110) and (4.109), we finally obtain

$$\hat{\mathbf{V}} = (\mathbf{I} - \mathbf{V}\mathbf{V}^T \mathbf{M}) \mathbf{U} \hat{\mathbf{Q}}^+, \quad (4.111)$$

where $\hat{\mathbf{Q}}^+$ is the pseudo-inverse of $\hat{\mathbf{Q}}$ and $\hat{\mathbf{V}} \in \mathbb{R}^{n \times \frac{m(m+1)}{2}}$.

In practice, the NLROM obtained by IC is integrated in time and the vectors of modal coordinates $\mathbf{q}$ at time samples $t_1, t_2, \dots, t_{n_T}$ are collected in $\mathbf{Q}_T$ as

$$\mathbf{Q}_T = [\mathbf{q}(t_1), \dots, \mathbf{q}(t_{n_T})]. \quad (4.112)$$

The total displacement $\mathbf{U}_T = [\mathbf{u}(t_1), \dots, \mathbf{u}(t_{n_T})]$ is retrieved by

$$\mathbf{U}_T = \mathbf{V}\mathbf{Q}_T + \hat{\mathbf{V}}\hat{\mathbf{Q}}_T, \quad (4.113)$$

where the column $\hat{\mathbf{q}}(t_i)$ of $\hat{\mathbf{Q}}_T$ relative to time step t_i is given by

$$\hat{\mathbf{q}}(t_i) = [(q_1(t_i))^2, q_1(t_i)q_2(t_i), \dots, q_i(t_i)q_j(t_i), \dots, (q_m(t_i))^2]. \quad (4.114)$$

A comparison of the number of evaluations required by ED, EED and IC is shown in Figure 4.7.

Figure 4.7: Number of required nonlinear static solution for ED, EED and IC versus the number of modes in the reduction basis [34].

4.5.6 Nonlinear normal modes

As briefly discussed, the presence of nonlinearity may alter the stiffness of the system (for geometric nonlinearities), the damping forces (i. e. with friction), or both. In fact, it was shown that linear reduction techniques based on vibration modes only are not sufficient to capture the nonlinear behavior, and therefore other vectors (e. g. MDs or DMs) have to be added to the reduction basis. One could also wonder how a certain eigensolution (i. e. a vibration mode oscillating at its own eigenfrequency) changes

when the displacements trigger the nonlinearity. For conservative autonomous systems (i. e. systems with no damping and no external excitation), there still exist periodic solutions, usually called *Nonlinear Normal Modes* (NNMs), which can be thought of as the nonlinear equivalent of a vibration mode for a linear system. This concept was first defined by Rosenberg [55] as “any vibration-in-unison of a conservative nonlinear system, i. e. where the coordinates of the system pass through the equilibrium and reach their extrema simultaneously”, and later generalized to include any non-necessarily synchronous periodic motion [37]. The latter definition allows the system to exhibit internal resonance, i. e. the coexistence of two or more modes which feature periodic motions at periods at commensurate ratios. Pioneering work on NNMs was done by Shaw and Pierre [59, 60].

NNMs are instrumental in assessing the impact of the nonlinearity on the dynamic response. In [40], NNMs are proposed as tool to assess the quality of a NLROM. In this work, it is in fact suggested to directly compare NNMs computed with NLROMs, as the NNMs of the HFM are too costly to be computed for large systems. Usually, the effect of a NNM is summarized by means of a *Frequency–Energy Plot* (FEP), where the frequency of the periodic motion is plotted against the energy associated to the motion itself, for instance resulting from initial conditions. Note that, in such representation, the actual shape of the motion throughout the period is hidden. In Figure 4.8 (taken from [62]), the NNM associated to the first vibration mode of a shallow arch pinned at both ends is shown. The response stays linear (i. e. of constant frequency) for a wide range of energy. Then the frequency decreases due to a softening¹ behavior, followed by an internal resonance tongue due to the interaction with the NNM associated to the third vibration mode. At this point, in fact, the two NNMs evolve with a period ratio of 4:1, enabling the existence of a periodic solution at which both contributions can coexist. The motion triggers then a hardening behavior at higher energy. A FEP as the one shown in Figure 4.8 is computationally expensive to construct. In fact, every point of the FEP requires the knowledge of a periodic motion for a given energy level. The frequency and the initial conditions of such motion are not known a priori. Numerical techniques for solving this problem are discussed in detail in [51] and [28]. Here, is it worth to highlight the benefits of resorting to NLROMs when computing NNMs. In [62], for instance, MDs-based ROMs were assessed for the computation of NNMs of FE-discretized planar beam structures. In this work, the shooting method was used to obtain periodic solutions of the autonomous conservative systems of interest. This method essentially integrates the system in time for trial initial conditions (i. e. initial deformed shape imposed to the structure) and a trial period, and then correct them via

¹ The term “softening” (“hardening”) usually refers to a decrease (increase) of the slope of the restoring nonlinear forces. For instance, a flat beam, cantilevered at both ends and loaded by a transverse force at its mid-span features a hardening stiffness: the bending associated to finite displacements induces stretching in the cross-section, which in turn generates an additional restoring bending moment to that due to linear effects.

Figure 4.8: First NNM of a shallow arch. Note that the NNM first exhibits softening (i. e. decrease in frequency due to a decrease in stiffness), followed by an internal resonance (4:1 interaction with the third NNM), and then hardening. The FEP computed for the full model (blue solid line), a ROM containing $\mathbf{V} = [\boldsymbol{\phi}_1]$, a ROM containing $\mathbf{V} = [\boldsymbol{\phi}_1 \boldsymbol{\phi}_3]$, and a ROM including $\mathbf{V} = [\boldsymbol{\phi}_1 \boldsymbol{\phi}_3 \boldsymbol{\theta}_{11} \boldsymbol{\theta}_{13} \boldsymbol{\theta}_{33}]$. Plots (a) to (d) show the spatial shape of the response (solid line) for the corresponding points on the FEP. The undeformed configuration is shown with dashed line [62].

Newton–Raphson iterations until a periodicity condition is met with pre-defined accuracy. Once a solution is obtained (i. e. a point of the FEP), continuation is employed to estimate the new guess at higher energy level. It becomes clear then that each point of an FEP requires multiple time integration of the system, and therefore, the tracing of a FEP for each mode of interest can be a daunting task. In [62], it was demonstrated that MD-based NLROMs capture well the nonlinear behavior and internal resonances by only including the vibration modes—and corresponding MDs, which are expected to interact. This is achieved at a smaller computational cost as the one required for the HFM. Moreover, a NLROM so constructed filters a physically meaningless contribution of high-frequency dynamics, which is only due to the resolution of the spatial discretization. In doing so, the convergence of numerical implicit time integration is significantly improved.

4.5.7 Nonlinear manifolds for reduction

Up to now, we discussed techniques that approximate the HFM solution as a linear combination of modes spanning a low-dimensional subspace. This is not the only possibility, as many dynamical systems evolve on *manifolds*, rather than subspaces [44]. This is typically the case for systems characterized by *slow* and *fast* dynamics. An example of this is static condensation (see Section 4.5.1), where the fast dynamics of the in-plane motion is statically enslaved to the slow dynamics of the bending deformations. The resulting mapping between fast and slow dynamics is in this case quadratic, see equation (4.51). Once a condensed system is obtained through slow–fast decom-

position, a further reduction is of course possible. One could apply the techniques discussed in Sections 4.5.2, 4.5.3 and 4.5.4 (i. e. based on linear subspaces for reduction) or seek for a manifold on which the solution is embedded. Such a manifold passes through the equilibrium point and is there tangent to the corresponding linear mode. Recently, the concept of *Spectral Submanifold* (SSM) was introduced by Haller and Ponsioen in [21]. The SSM was introduced as an invariant manifold asymptotic to a NNM, and it is the smoothest nonlinear continuation of a modal subspace emanating from the equilibrium. A discussion on the potential of SSM for model reduction can be found in [53], where SSMs are computed for a linear FE-discretized beam with a cubic nonlinear spring acting at the tip. An interesting application of slow–fast decomposition combined with SSM reduction is presented in [30] for the case of a geometrically nonlinear FE discretized beam.

An approximate method to construct a manifold for reduction, related to the MDs method, is presented in [31] and then generalized in [58]. There, it is postulated that the HFM displacements $\mathbf{u}(t)$ are mapped to the modal coordinates $\mathbf{q}(t)$, i. e. $\mathbf{q} \in \mathbb{R}^m \rightarrow \Gamma(\mathbf{q}) \in \mathbb{R}^n$ as

$$\mathbf{u}(t) \approx \Gamma(\mathbf{q}(t)) := \mathbf{\Phi} \cdot \mathbf{q}(t) + \frac{1}{2}(\mathbf{\Omega} \cdot \mathbf{q}(t) \cdot \mathbf{q}(t)), \quad (4.115)$$

where $\mathbf{\Phi} \in \mathbb{R}^{n \times m}$ contains selected vibration modes, and $\mathbf{\Omega} \in \mathbb{R}^{n \times m \times m}$ is a third-order tensor constructed with the MDs stemming from $\mathbf{\Phi}$. The resulting manifold is therefore quadratic in $\mathbf{q}$, hence the naming *Quadratic Manifold* (QM). The concept is illustrated in Figure 4.9, for the case of a cantilevered plate. Instead of constituting additional DOFs for the NLRM, the MDs provide the curvature of the manifold. The NLRM is then obtained by inserting (4.115) into (4.30) and projecting onto the tangent manifold $\mathbf{P}_\Gamma(\mathbf{q})$, obtained:

$$\mathbf{P}_\Gamma(\mathbf{q}) = \frac{\partial \Gamma(\mathbf{q})}{\partial \mathbf{q}}. \quad (4.116)$$

Again, the interested reader is referred to [31] for details. The method was tested on a fairly large FE model of a wing subject to representative gust load. This approach outperformed a POD reduction with the same number of modes (5 vibration modes vs 5 POD modes), in terms of accuracy. It is worth mentioning that the QM technique needs to be equipped with a compression technique of the resulting nonlinear reduced terms. An exact tensorial form as outlined in Section 4.4.1.2 is cumbersome, as the resulting nonlinear terms would contain tensors up to seventh order. Likely, the higher order terms are not relevant for accuracy, and yet require the most intensive computational and storage effort. A possible strategy could be therefore to neglect some of the highest terms. Another possibility is hyper-reduction. It was shown in [32] that ECSW can be extended to the case of nonlinear manifolds for reduction. In this work, the same example tackled in [31] is considered, and better speed-ups than traditional, linear basis POD reduction are obtained, as less elements are picked by the algorithm.

Figure 4.9: Illustrative picture of a quadratic manifold for reduction of a cantilevered plate. The manifold is a function of the amplitudes of the first two vibration modes. The three MDs provide the curvature of the manifold.

Arguably, this is due to the fact that the QM more compactly represents the physics of the problem than the POD basis.

Still, if hyper-reduction is the way to go for reduction with nonlinear manifolds, one would desire to form the training sets without recurring to expensive HFM simulations. Attempts in this sense are made by [57] and [29]. In the latter of these contributions, the QM is in fact used to “lift” inexpensive linear modal solutions obtained as described in Section 4.2. The obtained displacements are then used to generate nonlinear forces to train the ECSW selection. The method proved very effective for geometrically nonlinear problems, for which the QM provides a good description of the physical behavior. The NLROM, though, is still obtained by projection on linear subspaces containing vibration modes and modal derivatives. The extension of this technique to QM-based reduction is yet to be tried, but it is foreseen not to pose major hurdles.

4.6 Overview on parametric nonlinear ROMs

In the previous sections, projection-based NLROMs strategies to select a suitable basis and to evaluate the nonlinear system of equations have been discussed. As it has already been pointed out, most of the showcased methods do provide very high speed-ups in the online phase, but usually they require some kind of overhead costs involving offline training and/or pre-computations. This often heavily reduces the effective gain provided by the ROM. Indeed, calling for convenience t_{on} and t_{off} the online and offline times of a ROM and t_f the time of a full order analysis, the effective speed-up factor can be computed as $SP = (t_{\text{on}} + t_{\text{off}})/t_f$. It is clear that, if the same reduced model is used for multiple (say M) queries requiring different analyses, this factor increases

to $SP_M = (Mt_{\text{on}} + t_{\text{off}})/(Mt_f)$. In other words, the offline costs become less important as the number of evaluations increases. Parametric ROMs (PROMs) are then a natural extension of this concept, as the number M would include the sum of the number of queries for each combination of parameters, making the reduction even more profitable.

For linear systems parametric reduced order modeling (PMOR) techniques are well developed. Loosely speaking, the main challenges in PMOR reside in (i) the construction of the model itself, (ii) identification of a projection basis that can be valid over a *space of parameters*, and (iii) selection of an interpolation and sampling strategies. The model can be constructed following a *local* approach (e. g. building several ROMs for different parameters and interpolating the bases/ROM/solutions) or a *global* approach, where the full order system is parametrized and a single basis is selected. For the latter case, again many options are available. Moment-matching (discussed extensively in [7, Chapter 3]) is a popular strategy, where the basis is constructed requiring the n th derivatives (moments) of the reduced and full order systems with respect to the parameters to match [17]. Another approach is to construct the basis repeatedly applying POD to a set of simulations selected through some kind of parameter-space sampling procedures (e. g. Latin Hyper-cube, Smolyak sparse grid). For an extensive survey of existing methods the reader is referred to [6].

On the other hand, projection-based PMORs for nonlinear systems feature a much less mature theory, and constitute a broad and rapidly evolving area of research. Many of the methods available in the literature try to carry over concepts from linear analysis to the nonlinear case. In this sense, POD represents the most widely exploited method due to its versatility and properties, such as error bound and optimality of the reduction basis it provides. In [4] for instance, a PROM was developed for contact problems by performing a number of full order simulations sampling the parameter space, then taking the POD of each and clustering all the resulting bases. Using a computationally cheap error estimator, this process is repeated in order to refine the approximation. A lot of other strategies involving interpolation of such full order simulations and/or POD bases also exist. In [70] the data coming from the full order simulations is manipulated using a two-level radial basis function interpolation method to obtain hyper-surfaces in the parameters, while in [22] the interpolation step is carried out using neural networks to compute the interpolation coefficients. All the aforementioned strategies belong to the family of (non-intrusive) data-driven methods, and they all demand a huge upfront cost to be paid to develop the ROM itself. Nevertheless, these approaches might probably be the only viable ways to deal with complex problems such as the ones in fluid-mechanics, which often provides the benchmarks for these algorithms.

In the context of geometric nonlinearities in solid mechanics, however, we showed that under some circumstances the nonlinear internal forces take specific shapes (e. g. third-order polynomials), and thus some information from the model can be exploited for the construction of the reduction basis (e. g. MDs) without the

need of full simulations. In [43] for instance, such information is used to construct a PROM to describe *shape defects* of a structure in a parametric fashion and relying solely on the model of the system. Even if probably model-driven approaches seem to go against the trend of a consistent share of the scientific community that gradually, but steadily, moves towards the machine-learning area, we deem such strategies still worth of investigation not only for the possible advantages over other techniques, but also for their engineering value in terms of interpretability and ease of use.

The parametrization could also be intended as time-dependent, as done for instance in [9] and [63] for the reduction of models of meshing gears. Here, a reduction basis which is used to reproduce the contact behavior between gears teeth is made function of a time-dependent load position parameter, and appended to a constant basis that spans the global behavior of the system. The local, parametrized modes can be either residual attachment modes [63], or static solutions at specific configuration [9]. The time variation of the reduction basis is then generating additional, state dependent terms in the ROM, which are shown to be important for stress recovery. In all this contributions, the time variation of the parameter is known a priori. Likewise, a prescribed time-dependent law could parameterize the equilibrium of a given system. This is the case of thermo-mechanical systems, where the temperature field is typically determined by solving the heat equation, and then applied to the mechanical problem. Typically, there is a large characteristic time scale difference between the thermal and mechanical problem, being the latter characterized by much faster dynamics. In this case indeed, the slow variation of the applied thermal field justifies a *parametric equilibrium* and a *parametric reduction basis*, which can be conveniently interpolated online; see [33]. For the same reason, state dependent terms that are otherwise present (see again [9, 63]) can be neglected.

4.7 Conclusions

We briefly overviewed the most popular methods for model order reduction of linear and geometrical nonlinear mechanical systems. The approaches here considered are *system-driven*, meaning with this that the reduced order model is built from quantities that can be derived directly from the discretized equations governing the high fidelity model. This is in contrast with *data-driven* techniques, which in fact build the reduced order model from solutions of the high fidelity model. The main appeal of system-driven method is their independence from computationally expensive solutions and their exploitation of the system characteristics.

In structural dynamics applications, the *modal displacement method* has become a standard already for decades. The full order solution is approximated by a combination of a few carefully selected eigenmodes, which are also used to span a space

where the system is projected. The resulting reduced order model consists of an uncoupled system of modal equations which can be easily solved. Enhancements to the modal displacement method aim at retrieving the contribution of discarded vectors by quasi-static corrections.

The same strategy, namely modal expansion and projection, can be employed also for nonlinear systems. Here, we restricted to systems with geometric nonlinearities, which are typically modeled by smooth nonlinear functions of displacements. Systems affected by geometric nonlinearities usually feature a *slow* dominant behavior, which is complemented by a *fast* dynamics. The slow dynamics—typically featuring out-of-plane motion for the case of thin-walled structures—can still be developed by a set of dominant vibration modes selected by the modal displacement method. The fast dynamics is usually *enslaved* to the slow one, as it is given by the bending–stretching coupling due to the geometric nonlinearity. In fact, for flat systems where the in-plane and out-of-plane dynamics can be clearly separated, a nonlinear mapping between out-of-plane (slow) and in-plane (fast) dynamics—usually referred to as *static condensation*—can be established by neglecting the in-plane inertial terms.

For more geometrically complex systems, this separation of the degrees of freedom might be hard or impossible to perform. In this case, one can of course rely on expansion and projection onto a subspace that is able to reproduce the nonlinear coupling effects. Two methods to form such basis gained popularity over the past two decades, namely *modal derivatives* and *dual modes*. Both methods find the most relevant in-plane contribution to best span the nonlinear behavior. While the dual modes method is inherently *non-intrusive*, as it requires several nonlinear static solutions of the high fidelity model, the modal derivative strategy can be applied also *intrusively* if the directional derivative of the tangent stiffness matrix along the dominant linear modes could be calculated.

Indeed, the wording *non-intrusive* and *intrusive* refer to two distinct classes of reduction methods. The distinction lies on whether or not it is possible to access elemental functions to compute the projected reduced order terms. In non-intrusive methods, the structure of these terms is first postulated and then identified, typically by probing the system with sufficient nonlinear static cases, where either displacements or forces are imposed on the high fidelity model. We discussed the methods of *enforced displacements* and *enhanced enforced displacements*, which identify the coefficients of the projected nonlinear forces by statically applying displacements determined by combinations of the modes used for reduction. The number of such evaluations is $\mathcal{O}(m^3)$ and $\mathcal{O}(m^2)$ for the enforced displacement and the enhanced enforced displacement method, respectively. The latter gain is due to the assumed availability of the tangent stiffness matrix from the high fidelity model. The method of *implicit condensation and expansion* differs from the former two as it relies only on dominant vibration modes—and not on either dual modes or modal derivatives—to construct the reduced order model. A recovery of the full displacement field is obtained through an expansion step that assumes a quadratic dependency of the in-plane modes to the dominant

vibration modes. In intrusive methods, on the contrary, the reduced nonlinear forces are computed exactly by evaluating and projecting at element level. This is efficient if the resulting reduced terms are functions of modal coordinates only—i. e. no access to the mesh is required any longer, once the reduced order model is constructed. This is the case indeed of polynomial terms, resulting from lagrangian formulation and elastic material of continuum elements, or plates modeled with the von Karman kinematic assumptions. In more complex cases, the evaluation at element level is unavoidable, and the speed-up associated to a reduced order model drops significantly. In these cases, one can resort to *hyper-reduction*, which is extensively treated in [8, Chapter 5]. Essentially, hyper-reduction scales the computation of the reduced terms with the size of the reduction basis and not with the one of the high fidelity model, and thus delivers huge speed-ups. Here, we outlined the Discrete Empirical Interpolation Method (DEIM) and the Energy Conserving Sampling and Weighting (ECSW) method. Both methods rely on *training forces*, which typically come from the Proper Orthogonal Decomposition of high fidelity solutions. As such, they could be classified as data-driven methods and would therefore not belong to this chapter. However, recent trends target the construction of such training sets without recurring to expensive full solution. For instance, in the case of geometric nonlinearities, one can *lift* computationally linear modal solutions on a quadratic manifold centered around the equilibrium, and whose curvature is provided by modal derivatives. The resulting displacements generate forces which can be used for training of the hyper-reduction.

The slow–fast dynamics dichotomy of the behavior of nonlinear systems in fact favors reduction methods based on nonlinear manifolds, rather than projection on linear subspaces. In other words, the solution is assumed to evolve on curved, rather than flat, manifolds, which can be constructed via asymptotic expansions around the equilibrium point. Along this direction, the recently proposed *spectral submanifold* is defined as the smoothest nonlinear continuation of a modal subspace at the equilibrium point. Likewise, we briefly discussed the reduction on *quadratic manifolds* constructed with vibration modes and modal derivatives. As the manifold can be seen as a nonlinear constraint to the solution, a configuration-dependent mass matrix and convective generalized forces are generated by this reduction method. Reduction through nonlinear manifolds exacerbates the issue of efficient computation of reduced nonlinear terms. In the case of polynomial nonlinearities for the high fidelity model, the resulting reduced tensors are of higher order as the ones corresponding to linear projection. This could quickly exhaust memory resources and slow down the time integration of the reduced system significantly. Recent contributions are tackling this problem by applying for instance hyper-reduction.

Lastly, we briefly discussed parametric model order reduction. In this context, the general trend is to rely on data-driven methods, usually based on POD, to construct the parametrized reduced order model. In general this could be done by either forming a large reduction basis that spans the parameter space of interest, or by constructing local reduced order models and interpolate across the parameter space. In either case,

a greedy procedure often drives the generation of the parametric reduced order model. However, for the case of geometric nonlinearity, it is possible to exploit the structure of the governing equations and devise a reduction basis containing *sensitivities* of modes with respect to parameters—for instance, in the case of shape imperfections. This procedure is completely system-driven and it is shown to span the parameter space in the neighborhood of the nominal parameter values.

Bibliography

- [1] M. S. Allen, D. Rixen, M. van der Seijs, P. Tiso, T. Abrahamsson, and R. L. Mayes. *Substructuring in Engineering Dynamics: Emerging Numerical and Experimental Techniques*. CISM International Centre for Mechanical Sciences (Courses and Lectures). Springer, Cham, 2020.
- [2] M. Amabili, A. Sarkar, and M. P. Paidoussis. Reduced-order models for nonlinear vibrations of cylindrical shells via the proper orthogonal decomposition method. *J. Fluids Struct.*, 18(2):227–250, 2003. Axial and Internal Flow Fluid-Structure Interactions.
- [3] D. Amsallem, J. Cortial, K. Carlberg, and C. Farhat. A method for interpolating on manifolds structural dynamics reduced-order models. *Int. J. Numer. Methods Eng.*, 80(9):1241–1258, 2009.
- [4] M. Balajewicz, D. Amsallem, and C. Farhat. Projection-based model reduction for contact problems. *Int. J. Numer. Methods Eng.*, 106:644–663, 2015.
- [5] M. Balajewicz, D. Amsallem, and C. Farhat. Projection-based model reduction for contact problems. *Int. J. Numer. Methods Eng.*, 106(8):644–663, 2016. nme.5135.
- [6] P. Benner, S. Gugercin, and K. Willcox. A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems. *SIAM Rev.*, 57(4):483–531, 2015.
- [7] P. Benner, S. Griwet-Talocia, A. Quarteroni, G. Rozza, W. H. A. Schilders and L. M. Silveira. *Model Order Reduction. Volume 1: System- and Data-Driven Methods and Algorithms*. De Gruyter, Berlin, 2020.
- [8] P. Benner, S. Griwet-Talocia, A. Quarteroni, G. Rozza, W. H. A. Schilders and L. M. Silveira. *Model Order Reduction. Volume 2: Snapshot-Based Methods and Algorithms*. De Gruyter, Berlin, 2020.
- [9] B. Blockmans, T. Tamarozzi, F. Naets, and W. Desmet. A nonlinear parametric model reduction method for efficient gear contact simulations. *Int. J. Numer. Methods Eng.*, 102(5):1162–1191, 2015.
- [10] J. Bonet, A. J. Gil, and R. D. Wood. *Nonlinear Solid Mechanics for Finite Element Analysis: Statics*. Cambridge University Press, 2016.
- [11] S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM J. Sci. Comput.*, 32(5):2737–2764, 2010.
- [12] R. R. Craig and M. C. C. Bamton. Coupling of substructures for dynamic analyses. *AIAA J.*, 6(7):1313–1319, 1968.
- [13] R. de Borst, M. A. Crisfield, J. J. C. Remmers, and C. V. Verhoosel. *Non-Linear Finite Element Analysis of Solids and Structures*, 2nd edition, 2012.
- [14] J. M. Dickens, J. M. Nakagawa, and M. J. Wittbrodt. A critique of mode acceleration and modal truncation augmentation methods for modal response analysis. *Comput. Struct.*, 62(6):985–998, 1997.
- [15] J. M. Dickens and K. V. Pool. Modal truncation vectors and periodic time domain analysis applied to a cyclic symmetry structure. *Comput. Struct.*, 45(4):685–696, 1992.

- [16] C. Farhat, P. Avery, T. Chapman, and J. Cortial. Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting for computational efficiency. *Int. J. Numer. Methods Eng.*, 98(9):625–662, 2014.
- [17] B. Fröhlich, J. Gade, F. Geiger, M. Bischoff, and P. Eberhard. Geometric element parameterization and parametric model order reduction in finite element based shape optimization. *Comput. Mech.*, 2018.
- [18] M. Geradin and D. J. Rixen. *Mechanical Vibrations: Theory and Application to Structural Dynamics*. Wiley, 2015.
- [19] A. Givois, A. Grolet, O. Thomas, and J.-F. Deü. On the frequency response computation of geometrically nonlinear flat structures using reduced-order finite element models. *Nonlinear Dyn.*, 97(2):1747–1781, 2019.
- [20] R. W. Gordon and J. J. Hollkamp. Reduced-Order Models for Acoustic Response Prediction. Technical Report July, AFRL-RB-WP-TR-2011-3040, 2011.
- [21] G. Haller and S. Ponsioen. Nonlinear normal modes and spectral submanifolds: existence, uniqueness and use in model reduction. *Nonlinear Dyn.*, 86(3):1493–1534, 2016.
- [22] J. S. Hesthaven and S. Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *J. Comput. Phys.*, 363:55–78, 2018.
- [23] L. Hogben. *Handbook of Linear Algebra*. Chapman & Hall/CRC, 2007.
- [24] J. J. Hollkamp and R. W. Gordon. Reduced-order models for nonlinear response prediction: Implicit condensation and expansion. *J. Sound Vib.*, 318(4–5):1139–1153, 2008.
- [25] W. C. Hurty. Dynamic analysis of structural systems using component modes. *AIAA J.*, 3(4):678–685, 1965.
- [26] S. R. Idelsohn and A. Cardona. A load-dependent basis for reduced nonlinear structural dynamics. In A. K. Noor and R. J. Hayduk, editors, *Advances and Trends in Structures and Dynamics*, pages 203–210. Pergamon, 1985.
- [27] S. R. Idelsohn and A. Cardona. A reduction method for nonlinear structural dynamic analysis. *Comput. Methods Appl. Mech. Eng.*, 49(3):253–279, 1985.
- [28] S. Jain, T. Breunung, and G. Haller. Fast computation of steady-state response for high-degree-of-freedom nonlinear systems. *Nonlinear Dyn.*, 97(1):313–341, 2019.
- [29] S. Jain and P. Tiso. Simulation-Free Hyper-Reduction for Geometrically Nonlinear Structural Dynamics: A Quadratic Manifold Lifting Approach. *J. Comput. Nonlinear Dyn.*, 13(7):071003, 2018.
- [30] S. Jain, P. Tiso, and G. Haller. Exact nonlinear model reduction for a von kármán beam: Slow-fast decomposition and spectral submanifolds. *J. Sound Vib.*, 423:195–211, 2018.
- [31] S. Jain, P. Tiso, J. B. Rutzmoser, and D. J. Rixen. A quadratic manifold for model order reduction of nonlinear structural dynamics. *Comput. Struct.*, 188:80–94, 2017.
- [32] S. Jain and P. Tiso. Hyper-Reduction Over Nonlinear Manifolds for Large Nonlinear Mechanical Systems. *J. Comput. Nonlinear Dyn.*, 14(8):081008, 2019.
- [33] S. Jain and P. Tiso. Model order reduction for temperature-dependent nonlinear mechanical systems: A multiple scales approach. *J. Sound Vib.*, 465:115022, 2020.
- [34] M. Karamooz Mahdiabadi. Nonlinear Model Order Reduction and Substructuring for Structural Dynamics Analysis: Non-intrusive Methods. Phd thesis, Technical University of Munich 2019.
- [35] P. Kerfriden, P. Gosselet, S. Adhikari, and S. P. A. Bordas. Bridging proper orthogonal decomposition methods and augmented newton–krylov algorithms: An adaptive model order reduction for highly nonlinear mechanical problems. *Comput. Methods Appl. Mech. Eng.*, 200(5):850–866, 2011.
- [36] G. Kerschen, J. Golinval, A. F. Vakakis, and L. A. Bergman. The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview. *Nonlinear Dyn.*, 41(1):147–169, 2005.

- [37] G. Kerschen, M. Peeters, J. C. Golinval, and A. F. Vakakis. Nonlinear normal modes, part I: A useful framework for the structural dynamicist. *Mech. Syst. Signal Process.*, 23(1):170–194, 2009. Special Issue: Non-linear Structural Dynamics.
- [38] K. Kim, A. G. Radu, X. Q. Wang, and M. P. Mignolet. Nonlinear reduced order modeling of isotropic and functionally graded plates. *Int. J. Non-Linear Mech.*, 49(2013):100–110, 2013.
- [39] D. Krattiger, L. Wu, M. Zacharczuk, M. Buck, R. J. Kuether, M. S. Allen, P. Tiso, and M. R. W. Brake. Interface reduction for hurty/craig-bampton substructured models: Review and improvements. *Mech. Syst. Signal Process.*, 114:579–603, 2019.
- [40] R. J. Kuether, B. J. Deaner, J. J. Hollkamp, and M. S. Allen. Evaluation of geometrically nonlinear reduced-order models with nonlinear normal modes. *AIAA J.*, 53(11):3273–3285, 2015.
- [41] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numer. Math.*, 90(1):117–148, 2001.
- [42] R. H. MacNeal. A hybrid method of component mode synthesis. *Comput. Struct.*, 1(4):581–601, 1971. Special Issue on Structural Dynamics.
- [43] J. Marconi, P. Tiso, and F. Braghin. A nonlinear reduced order model with parametrized shape defects. *Comput. Methods Appl. Mech. Eng.*, 360:112785, 2020.
- [44] H. G. Matthies and M. Meyer. Nonlinear galerkin methods for the model reduction of nonlinear dynamical systems. *Comput. Struct.*, 81(12):1277–1286, 2003. Advanced Computational Models and Techniques in Dynamics.
- [45] M. I. McEwan, J. R. Wright, J. E. Cooper, and A. Y. T. Leung. A combined modal/finite element analysis technique for the dynamic response of a non-linear beam to harmonic excitation. *J. Sound Vib.*, 243(4):601–624, 2001.
- [46] M. I. McEwan, J. R. Wright, J. E. Cooper, and A. Y. T. Leung. A finite element/modal technique for nonlinear plate and stiffened panel response prediction. In *Collection of Technical Papers – AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, volume 5, pages 3061–3070, 2001.
- [47] M. P. Mignolet, A. Przekop, S. A. Rizzi, and S. M. Spottswood. A review of indirect/non-intrusive reduced order modeling of nonlinear geometric structures. *J. Sound Vib.*, 332(10):2437–2460, 2013.
- [48] A. A. Muravyov and S. A. Rizzi. Determination of nonlinear stiffness with application to random vibration of geometrically nonlinear structures. *Comput. Struct.*, 81(15):1513–1523, 2003.
- [49] M. Nash. Nonlinear Structure Dynamics by Finite Element Modal Synthesis. PhD thesis, Imperial College, London 1977.
- [50] K. O’Hanlon and M. D. Plumbley. Learning overcomplete dictionaries with l0-sparse non-negative matrix factorisation. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 977–980, 2013.
- [51] M. Peeters, R. Vigné, G. Sérandour, G. Kerschen, and J.-C. Golinval. Nonlinear normal modes, part ii: Toward a practical computation using numerical continuation techniques. *Mech. Syst. Signal Process.*, 23(1):195–216, 2009. Special Issue: Non-linear Structural Dynamics.
- [52] R. Perez, X. Q. Wang, and M. P. Mignolet. Nonintrusive Structural Dynamic Reduced Order Modeling for Large Deformations: Enhancements for Complex Structures. *J. Comput. Nonlinear Dyn.*, 9:031008, 2014.
- [53] S. Ponsioen and G. Haller. Exact model reduction and fast forced response calculation in high-dimensional nonlinear mechanical systems, 2019. arXiv:1912.11399v1.
- [54] J. N. Reddy. *Introduction to the Finite Element Method*, 3rd edition. McGraw-Hill Education, New York, 2006.
- [55] R. M. Rosenberg. Normal Modes of Nonlinear Dual-Mode Systems. *J. Appl. Mech.*, 27(2):263–268, 1960.
- [56] S. Rubin. Improved component-mode representation for structural dynamic analysis. *AIAA J.*, 13(8):995–1006, 1975.

- [57] J. B. Rutzmoser and D. J. Rixen. A lean and efficient snapshot generation technique for the hyper-reduction of nonlinear structural dynamics. *Comput. Methods Appl. Mech. Eng.*, 325:330–349, 2017.
- [58] J. B. Rutzmoser, D. J. Rixen, P. Tiso, and S. Jain. Generalization of quadratic manifolds for reduced order modeling of nonlinear structural dynamics. *Comput. Struct.*, 192:196–209, 2017.
- [59] S. W. Shaw and C. Pierre. Non-linear normal modes and invariant manifolds. *J. Sound Vib.*, 150(1):170–173, 1991.
- [60] S. W. Shaw and C. Pierre. Normal modes of vibration for non-linear continuous systems. *J. Sound Vib.*, 169(3):319–347, 1994.
- [61] P. M. A. Slaats, J. de Jongh, and A. A. H. J. Sauren. Model reduction tools for nonlinear structural dynamics. *Comput. Struct.*, 54(6):1155–1171, 1995.
- [62] C. S. M. Sombroek, P. Tiso, L. Renson, and G. Kerschen. Numerical computation of nonlinear normal modes in a modal derivative subspace. *Comput. Struct.*, 195:34–46, 2018.
- [63] T. Tamarozzi, G. H. K. Heirman, and W. Desmet. An on-line time dependent parametric model order reduction scheme with focus on dynamic stress recovery. *Comput. Methods Appl. Mech. Eng.*, 268:336–358, 2014.
- [64] P. Tiso. Optimal second order reduction basis selection for nonlinear transient analysis. In T. Proulx, editor, *Modal Analysis Topics*, volume 3, pages 27–39. Springer New York, New York, NY, 2011.
- [65] P. Tiso, R. Dedden, and D. J. Rixen. A modified discrete empirical interpolation method for reducing non-linear structural finite element models. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Volume 7B: 9th International Conference on Multibody Systems, Nonlinear Dynamics, and Control*, 2013.
- [66] C. Touzé, M. Vidrascu, and D. Chapelle. Direct finite element computation of non-linear modal coupling coefficients for reduced-order shell models. *Comput. Mech.*, 54(2):567–580, 2014.
- [67] O. Weeger, U. Wever, and B. Simeon. On the use of modal derivatives for nonlinear model order reduction. *Int. J. Numer. Methods Eng.*, 108(13):1579–1602, 2016.
- [68] L. Wu and P. Tiso. Nonlinear model order reduction for flexible multibody dynamics: a modal derivatives approach. *Multibody Syst. Dyn.*, 36(4):405–425, 2016.
- [69] L. Wu, P. Tiso, K. Tatsis, E. Chatzi, and A. van Keulen. A modal derivatives enhanced rubin substructuring method for geometrically nonlinear multibody systems. *Multibody Syst. Dyn.*, 45(1):57–85, 2019.
- [70] D. Xiao, F. Fang, C. C. Pain, and I. M. Navon. A parameterized non-intrusive reduced order model and error analysis for general time-dependent nonlinear partial differential equations and its applications. *Comput. Methods Appl. Mech. Eng.*, 317:868–889, 2017.