Chapter 2
Learning Projection-Based Reduced-Order Models

2.1 Motivation and Basic Assumptions

In this chapter, we introduce the solution space for high-fidelity models based on partial differential equations and the finite element model. The manifold learning approach to model order reduction requires simulated data. Hence, learning projection-based reduced order models (ROM) has two steps: (i) an offline step for the computation of simulated data and for consecutive machine learning tasks, (ii) an online step where the reduced order model is used as a surrogate for the high fidelity model. The offline step generates a train set and a validation set of simulated data. The accuracy and the generalisation of the reduced order model is evaluated in the online step by using a test set of data forecast by the high-fidelity model. The test set aims also to check the computational speedups of the reduced-order model compare to the high-fidelity model.

Learning projection-based reduced order model makes sense only if there is a significant computational speedup at the price of an acceptable loss of accuracy in predictions. The longer the computational time of the high-fidelity model, the smaller the acceptable speed up, if we save hours or days of numerical simulations. Regarding the acceptable accuracy of reduced predictions, engineers working in industry and scientists working in laboratories do not have the same expectations. In our own experience, learning projection-based reduced order model has the capability to adapt to engineering tasks, high-fidelity models elaborated in laboratories, in terms of accuracy and computational time. This approach contributes to data continuity between physics-based knowledge developed in laboratories and practical applications for engineering tasks.

From the mathematical point of view, Céa's lemma gives an overview of manifold learning for model order reduction applied to elliptic equations. Let $\mathcal{V}$ be a real Hilbert space, such that the weak form of the elliptic equation reads: find $\widetilde{u} \in \mathcal{V}$, such that:

$$a(\widetilde{u}, \widetilde{v}) = L(\widetilde{v}), \quad \forall \, \widetilde{v} \in \mathcal{V}, \tag{2.1}$$

where $a(\cdot, \cdot)$ is a bilinear form, with coercivity constant $\beta > 0$ and continuity continuity constant $C^a > 0$. $L(\cdot)$ is a linear form. Section 2.2 gives more details on weak forms. Let us $\mathcal{V}_h$ a finite dimensional subspace of $\mathcal{V}$. Here, $\mathcal{V}_h$ is an approximate solution space for the elliptic equation. The approximate solution of the elliptic equation is denoted by $u \in \mathcal{V}_h \subset \mathcal{V}$. The Galerkin projection of the elliptic equation onto the solution space reads: find $u \in \mathcal{V}_h \subset \mathcal{V}$ such that:

$$a(u, v) = L(v), \quad \forall v \in \mathcal{V}_h, \tag{2.2}$$

where $\mathcal{V}_h$ has been substituted for $\mathcal{V}$. Céa's lemma states that:

$$\|\widetilde{u} - u\| \leq \frac{C^a}{\beta} \|\widetilde{u} - v\|, \quad \forall v \in \mathcal{V}_h, \tag{2.3}$$

where $\| \cdot \|$ is a norm in $\mathcal{V}$. The related scalar product is denoted by $\langle \cdot \rangle$. In Finite Element models, $\mathcal{V}_h$ is the span of finite element shape functions. But Céa's lemma holds for all finite-dimensional subspace of $\mathcal{V}$. The closer $\widetilde{u}$ to the solution space $\mathcal{V}_h$, the smaller the right-hand term of Céa's lemma (2.3), and therefore one can expect a better prediction u in Eq. (2.2), although we do not know $\widetilde{u}$.

In few words, a reduced-order model is obtained by introducing a smaller solution space $\mathcal{V}_n \subset \mathcal{V}_h$ of smaller dimension $n < N$. A projection-based reduced order model can be achieved by using the Galerkin projection (2.2), where $\mathcal{V}_n$ is substituted for $\mathcal{V}_h$. The conclusion of Céa's lemma holds again. The closer $\widetilde{u}$ to the solution space $\mathcal{V}_n$, the better the prediction $\widehat{u} \in \mathcal{V}_n$. Manifold learning comes into play when we are given a set of predictions $(u^{(i)})_{i=1,\ldots,m}$ in a common ambient space $\mathcal{V}_h$, related to a given finite element mesh. The basic assumptions in manifold learning for model order reduction are:

- a latent space of reduced dimension is hidden in the data $(u^{(i)})_{i=1,\ldots,m}$, its dimension is denoted by n,
- a machine learning algorithm is available to learn this latent space by using a train set of simulated data extracted from $(u^{(i)})_{i=1,\ldots,m}$,
- the distance between $\widetilde{u}$ and this latent space is small enough, although we do not know $\widetilde{u}$,
- a numerical scheme enables the projection of the elliptic equation onto the latent space, in order to set up the reduced order model,
- the computational complexity of the solution of the reduced order model is smaller that the computational complexity of the finite element prediction,
- the computational complexity of the reduced order model is an increasing function of n.

As explained above, when the latent space is a vector subspace $\mathcal{V}_n$, both Galerkin projection and Céa's lemma hold, but simulation speedup may not be achieved. The study of more complex situations is the purpose of this Chapter. An estimation of computational complexity of projection-based reduced order model is proposed in Sect. 2.3.6.

Remarks:

- In the Rayleigh-Ritz method a small set of trial functions that satisfy the boundary conditions for $\widetilde{u}$ is introduce to span a solution space. This solution space is not related to any finite element model. The inclusion of the latent space in the ambient space $\mathcal{V}_h$ is essential for model order reduction.
- The finite element ambient space $\mathcal{V}_h$ incorporates homogeneous Dirichlet boundary conditions ($\widetilde{u} = 0$) on a boundary of the domain where the partial differential equations are set up. When such boundary conditions may change in the instances $(\widetilde{u}_N^{(i)})_{i=1,\ldots,m}$ these conditions must be taken into account as a linear constraint that supplement the partial differential equation. Such an issue appears when considering contact problems [38] for instance.
- Important limitations of projection-based model reduction methods includes situations where the geometry has to be handled in the exploitation phase of the reduced-order models, for instance when the problem features contact boundary conditions, crack propagation or when the geometry is a variability of the problem to learn. Geometrical variabilities are handled in the authors' works [1, 2, 22, 60, 61, 92].

2.2 High-Fidelity Model (HFM)

Consider an abstract partial differential equation in a domain Ω, with a μ-variability:

$$\mathcal{D}(\widetilde{u}; \boldsymbol{\xi}, \boldsymbol{\mu}) = 0, \quad \boldsymbol{\xi} \in \Omega, \ \widetilde{u} \in \mathcal{V}. \tag{2.4}$$

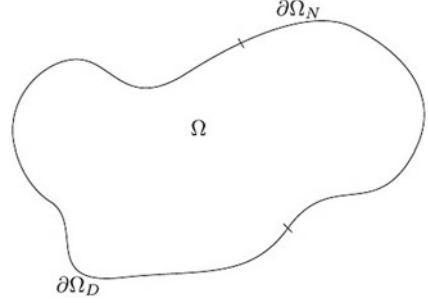The weak form of this partial differential equation reads: find $\widetilde{u} \in \mathcal{V}$ such that

$$\int_{\Omega} \widetilde{v}\, \mathcal{D}(\widetilde{u}; \boldsymbol{\xi}, \boldsymbol{\mu})\, d\boldsymbol{\xi} = 0, \quad \forall\, \widetilde{v} \in \mathcal{V}. \tag{2.5}$$

As an illustration, the concepts of this chapter are illustrated on a nonlinear structural mechanics problem, for which details on the high-fidelity model are provided in this Section. For an example with another physics, we refer to the authors' work [21], where a nonlinear transient thermal problem is considered.

The mechanical structure occupies the domain Ω, whose boundary $\partial\Omega$ is partitioned as $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ such that $\partial\Omega_D^{\mathrm{o}} \cap \partial\Omega_N^{\mathrm{o}} = \emptyset$, see Fig. 2.1.

The structure is subjected to a quasi-static time-dependent loading, composed of an homogeneous Dirichlet boundary conditions on $\partial\Omega_D$ and Neumann boundary conditions on $\partial\Omega_N$ in the form of a prescribed traction T_N, as well as a volumic force f. The setting depends on some variability μ, which can be a parameter vector, or represent some nonparametrized variability. The evolution of the displacement $\widetilde{u}^{\boldsymbol{\mu}}(\boldsymbol{\xi}, t)$ over $(\boldsymbol{\xi}, t) \in \Omega \times [0, T]$ is the solution of the following equations:

Fig. 2.1 Schematic representation of the structure of interest [18]

$$\boldsymbol{\epsilon}(\widetilde{u}^{\mu}) = \frac{1}{2}\left(\nabla\widetilde{u}^{\mu} + (\nabla\widetilde{u}^{\mu})^{T}\right), \qquad \text{in } \Omega \times [0, T], \quad \text{(compatibility equation)}$$

$$\text{div}\left(\sigma_{c}^{\mu}\right) + f^{\mu} = 0, \qquad \text{in } \Omega \times [0, T], \quad \text{(equilibrium equation)}$$

$$\sigma_{c}^{\mu} = \sigma_{c}(\boldsymbol{\epsilon}(\widetilde{u}^{\mu}), y^{\mu}), \qquad \text{in } \Omega \times [0, T], \quad \text{(constitutive law)}$$

$$\widetilde{u}^{\mu} = 0, \qquad \text{in } \partial\Omega_{D} \times [0, T], \quad \text{(prescribed zero displacement)}$$

$$\sigma_{c}^{\mu} \cdot n_{\partial\Omega} = T_{N}^{\mu}, \qquad \text{in } \partial\Omega_{N} \times [0, T], \quad \text{(prescribed traction)}$$

$$\widetilde{u}^{\mu} = 0, y^{\mu} = 0, \qquad \text{in } \Omega \text{ at } t = 0, \quad \text{(initial condition)}$$

$$(2.6)$$

where $\boldsymbol{\epsilon}$ is the linear strain tensor, σ_{c}^{μ} is the Cauchy stress tensor, y^{μ} denotes the internal variables of the constitutive law and $n_{\partial\Omega}$ is the outward normal vector on $\partial\Omega$. We precise that the evolution of the internal variables y^{μ} is updated when the constitutive law is solved.

Define $H_{0}^{1}(\Omega) = \{v \in L^{2}(\Omega)|\ \frac{\partial v}{\partial \xi_{i}} \in L^{2}(\Omega),\ 1 \leq i \leq 3 \text{ and } v|_{\partial\Omega_{D}} = 0\}$. Denote $\{\varphi_{i}\}_{1 \leq i \leq N} \in \mathbb{R}^{N \times N}$, a finite element basis whose span, denoted $\mathcal{V}_{h}$, constitutes an approximation of $H_{0}^{1}(\Omega)^{3}$; N is the number of finite element basis functions, hence the number of degrees of freedom of the discretized prediction. A discretized weak formulation reads: find $u^{\mu} \in \mathcal{V}_{h}$ such that for all $v \in \mathcal{V}_{h}$,

$$\int_{\Omega} \sigma_{c}(\boldsymbol{\epsilon}(u^{\mu}), y) : \boldsymbol{\epsilon}(v) = \int_{\Omega} f^{\mu} \cdot v + \int_{\partial\Omega_{N}} T_{N}^{\mu} \cdot v, \qquad (2.7)$$

that we denote for concision $\mathbf{F}^{\mu}(\mathbf{u}^{\mu}) = 0$, where $\mathbf{u}^{\mu}$ is the vector of N coordinates for $u^{\mu} \in \mathcal{V}_{h}$. A Newton algorithm can be used to solve this nonlinear global equilibrium problem at each time step:

$$\frac{D\mathbf{F}^{\mu}}{Du}\left(\mathbf{u}^{\mu,k}\right)\left(\mathbf{u}^{\mu,k+1} - \mathbf{u}^{\mu,k}\right) = -\mathbf{F}^{\mu}\left(\mathbf{u}^{\mu,k}\right), \qquad (2.8)$$

where

$$\frac{D\mathbf{F}^{\mu}}{Du}\left(\mathbf{u}^{k}\right)_{ij} = \int_{\Omega} \boldsymbol{\epsilon}\left(\varphi_{j}\right) : \mathcal{K}\left(\boldsymbol{\epsilon}(u^{\mu,k}), y^{\mu}\right) : \boldsymbol{\epsilon}\left(\varphi_{i}\right), \qquad (2.9)$$

and

Linear data compression
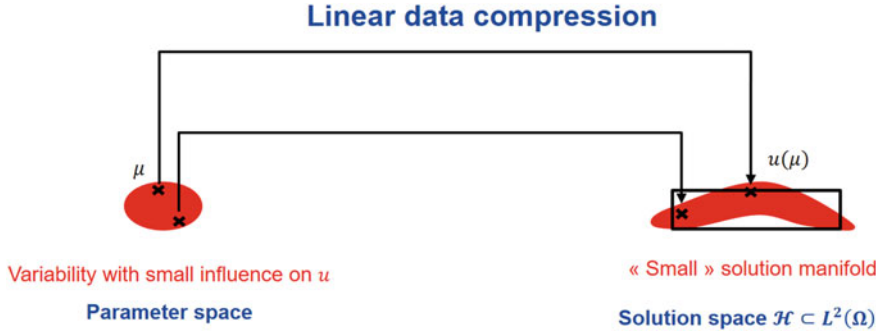
Fig. 2.2 Linear manifold learning

$$\mathbf{F}^{\mu}\left(\mathbf{u}^{\mu,k}\right)_{i} = \int_{\Omega} \boldsymbol{\sigma}_{c}\left(\boldsymbol{\epsilon}(u^{\mu,k}), y^{\mu}\right) : \boldsymbol{\epsilon}\left(\varphi_{i}\right) - \int_{\Omega} f^{\mu} \cdot \varphi_{i} - \int_{\partial\Omega_{N}} T_{N}^{\mu} \cdot \varphi_{i}. \quad (2.10)$$

In the two relations above, $\mathcal{K}\left(\boldsymbol{\epsilon}(u^{\mu,k}), y^{\mu}\right)$ is the local tangent operator, $u^{\mu,k} \in \mathcal{V}$ is the k-th iteration of the discretized displacement field at the current time-step, and $\mathbf{u}^{\mu,k} = \left(u_{i}^{\mu,k}\right)_{1\leq i\leq N} \in \mathbb{R}^{N}$ is such that $u^{\mu,k} = \sum_{i=1}^{N} u_{i}^{\mu,k} \varphi_{i}$. In particular, $f^{\mu}, T_{N}^{\mu}, u^{\mu,k}$ and y^{μ} are known and enforce the time-dependence of the solution. Depending on the constitutive law, the computation of the functions $\left(u^{\mu,k}, y\right) \mapsto \boldsymbol{\sigma}_{c}\left(\boldsymbol{\epsilon}(u^{\mu,k}), y^{\mu}\right)$ and $\left(u^{\mu,k}, y^{\mu}\right) \mapsto \mathcal{K}\left(\boldsymbol{\epsilon}(u^{\mu,k}), y^{\mu}\right)$ can require the resolution of a complex differential-algebraic system of equations.

2.3 Linear Manifold Learning for Projection-Based Reduced-Order Modeling

We start by explaining the online phase. Since we want to construct and solve the reduced-order model (ROM) in the most efficient way, the offline phase is dedicated to precompute as many steps as possible, under the considered variability.

Linear manifold learning means that the solution manifold is approximated by a vector subspace of the ambient solution space, as illustrated in Fig. 2.2.

2.3.1 Approaches Preceding the Use of Machine Learning

In structural mechanics, normal modes have been introduced for the analysis of vibrations in structures. When considering free vibrations, without external force, the solution of linear hyperbolic equation is sought by using the separation of space and

time variables: $\mathbf{u}_N(\mathbf{x}, t) = \boldsymbol{\psi}(\mathbf{x})\,\widehat{u}(t)$, where $\mathbf{x} \in \Omega$ is the space variable and $t \in \mathbb{R}$ the time variable. The hyperbolic equation of free vibration reads: find $\boldsymbol{\psi}(\mathbf{x}) \in \mathcal{V}_N$ and $\widehat{u}(t) \in \mathbb{R}$ such that $\mathbf{u}_N(\mathbf{x}, t) = \boldsymbol{\psi}(\mathbf{x})\,\widehat{u}(t)$ and

$$\langle \rho\,\ddot{\mathbf{u}}_N, \mathbf{v}_N \rangle + a(\mathbf{u}_N, \mathbf{v}_N) = 0, \quad \forall\,\mathbf{v}_N \in \mathcal{V}_N. \tag{2.11}$$

It follows that $\widehat{u}(t)$ is an harmonic function of frequency denoted by f and $\boldsymbol{\psi}$ is the eigenvector related to the eigenvalue $\lambda = (2\,\pi\,f)^2$ of the following generalized eigenproblem: find $\boldsymbol{\psi} \in \mathcal{V}_N$ and λ such that

$$a(\boldsymbol{\psi}, \mathbf{v}_N) - \lambda\,\langle \rho\,\boldsymbol{\psi}, \mathbf{v}_N \rangle = 0, \quad \forall\,\mathbf{v}_N \in \mathcal{V}_N, \tag{2.12}$$

where the rank of this system of equation is supposed to be $N - 1$ in order to find non zero eigenmodes. This eigenproblem admits N orthogonal normal modes that span $\mathcal{V}_N$. Therefore a selection of n normal modes span a reduced subspace of dimension n. In the beginning of the 21st century, model reduction using variable separation scheme in partial differential equations has been extended to an arbitrary number of variable by using low-rank approximations such as the Proper Generalized Decomposition [5]. Eigenmodes are global functions in contrast to finite element shape functions that have a local support. For dynamical problems involving nonlinear contributions to the PDE, adaptive computations of reduced subspaces have been proposed by Almroth et al. [4] and Noor et al. [79], by using Rayleigh-Ritz global functions. The set of these global functions is a reduced basis of the finite-element approximation-space.

The idea of using statistics to generate a solution space for differential equations has been proposed in the seminal work of Lorenz in [69] (in page 31), by using empirical orthogonal functions. The Galerkin projection of PDEs on empirical modes have been first developed in [13], where a reduced basis is computed via the proper orthogonal decomposition [70] of observational data. This was the first step towards manifold learning for projection-based model order reduction, which we now present. For other presentations of this technologies, the reader can refer to [81, 87].

2.3.2 Online Phase: Galerkin Projection

The reduced-order model is constructed in the form of a Galerkin method written on a Reduced-Order Basis (ROB). In the present case, it consists in assembling the physical problem in the same fashion as the HFM in Sect. 2.2, with the difference that the finite element basis $(\varphi_i)_{1 \leq i \leq N} \in \mathbb{R}^{N \times N}$, is replaced by a ROB $(\psi_i)_{1 \leq i \leq n} \in \mathbb{R}^{n \times N}$, with $n \ll N$. Hence, the reduced Newton algorithm is constructed as

$$\frac{D\mathcal{F}_\mu}{D\widehat{u}}\left(\widehat{u}_\mu^k\right)\left(\boldsymbol{\gamma}^{\mu, k+1} - \boldsymbol{\gamma}^{\mu, k}\right) = -\mathcal{F}_\mu\left(\widehat{u}_\mu^k\right), \tag{2.13}$$

where

$$\frac{D\mathcal{F}_\mu}{D\hat{u}}\left(\hat{u}_\mu^k\right)_{ij} = \int_\Omega \epsilon\left(\psi_j\right) : \mathcal{K}\left(\epsilon(\hat{u}_\mu^k), y_\mu\right) : \epsilon\left(\psi_i\right) \tag{2.14}$$

and

$$\mathcal{F}_\mu\left(\hat{u}_\mu^k\right)_i = \int_\Omega \sigma\left(\epsilon(\hat{u}_\mu^k), y\right) : \epsilon\left(\psi_i\right) - \int_\Omega f_\mu \cdot \psi_i - \int_{\partial\Omega_N} T_{\mu,N} \cdot \psi_i. \tag{2.15}$$

In the two relations above, $\hat{u}_\mu^k \in \hat{\mathcal{V}} := \text{Span}\,(\psi_i)_{1 \le i \le n}$ is the k-th iteration of the reduced displacement field for the current time-step and $\boldsymbol{\gamma}^{\boldsymbol{\mu},k} = \left(\gamma_i^{\boldsymbol{\mu},k}\right)_{1 \le i \le n} \in \mathbb{R}^n$ is such that

$$\hat{u}_\mu^k = \sum_{i=1}^n \gamma_i^{\boldsymbol{\mu},k} \psi_i. \tag{2.16}$$

Notice that the use of the Galerkin method is made possible by the linearity of the tangent problem (2.13) and the choice of a linear dimensionality reduction technique in (2.16).

The online stage is called *efficient* if the reduced problem can be constructed and solved in computational complexity independent of N. When the variability μ is parametrized, efficiency is possible by precomputing various terms. With non-parametrized variability, depending on its nature, some assembling task with a linear complexity in N may be required at the beginning of the online stage (for instance for a boundary condition). All these scenarios are handled by genericROM, the ROM library developed at Safran and presented in Sect. 4.2.

The offline phase contains three steps, for which we present below the method-ological choices made in genericROM.

2.3.3   Offline Phase

2.3.3.1   Data Generation

This step corresponds to the generation of the snapshots by solving the high-fidelity model. In parametric contexts, the simplest workflows consist in choosing param-eter values a priori, following Design of Experiments (DoE, see [50] for a recent technique), and computing the corresponding snapshots by solving the high-fidelity model.

2.3.3.2   Data Compression: Dimensionality Reduction

This step corresponds to the generation of the ROB $(\psi_i)_{1 \le i \le n}$. One of the most classical and simple method is the snapshot Proper Orthogonal Decomposition (POD) [25, 86], detailed below:

1. Choose a tolerance ϵ_{POD}.
2. Compute the correlation matrix $C_{i,j} = \int_\Omega u_i \cdot u_j$, $1 \le i, j \le N_s$, where N_s is the total number of HFM snapshots.
3. Compute the ϵ_{POD}-truncated eigendecomposition of C: $\xi_i \in \mathbb{R}^{N_c}$ and $\lambda_i > 0$, where $1 \le i \le n$, are the n first eigenvector and eigenvalues.
4. Compute the reduced order basis $\psi_i(x) = \dfrac{1}{\sqrt{\lambda_i}} \sum_{j=1}^{N_s} u_j(x) \xi_{ij}$, $1 \le i \le n$.

The advantages of the snapshot-POD are a reasonable computational complexity when the number of degree of freedom of the high-fidelity model are much larger than the number of snapshots, and the fact that this algorithm can be easily parallelized.

Variants can be used, for instance in the Reduced Basis [84] and the POD-greedy [41] methods, with respectively an orthonormalization of the computed high-fidelity snapshots and the incremental Singular Value Decomposition (SVD) [15].

2.3.3.3   Operator Compression

A ROM is called online-efficient if in the online stage, the reduced problems can be constructed and solved in computational complexity independent of N. The operator compression step consists in additional treatments required for the efficiency of the online stage, by pre-processing some computationally demanding integration tasks over the high-fidelity domain Ω and $\partial\Omega_N$. We notice that without any additional treatment, the numerical integration involved in the assembling of Eq. (2.13) strongly limits in practice the efficiency of the ROM: no speedup with respect to the high-fidelity model can be obtained in practice. The complexity of such an additional treatment depends on the type of parameter-dependence of the problem. This step is actually needed for all classes of problems reduced by projection-based methods.

Consider the simplest case: a linear problem with an affine dependence in the parameter μ, for instance $A_\mu \mathbf{u} = \mathbf{c}$, where $A_\mu = A_0 + \mu A_1$. Denote $\mathbf{V}$, the matrix whose columns are the vectors of the ROB evaluated at the high-fidelity degrees of freedom. The obtained ROM writes $\mathbf{V}^T A_\mu \mathbf{V} \hat{\mathbf{u}} = \mathbf{V}^T \mathbf{c}$: it is not assembled in the online phase, but rather the matrices $\mathbf{V}^T A_0 \mathbf{V}$ and $\mathbf{V}^T A_1 \mathbf{V}$ and the vector $\mathbf{V}^T \mathbf{c}$ are precomputed in the offline stage so that the reduced problem is constructed without approximation and efficiently by summing two small matrices. The operator compression step consists in this case in the construction of $\mathbf{V}^T A_0 \mathbf{V}$ and $\mathbf{V}^T A_1 \mathbf{V}$ in the offline stage.

Actually, there exist linear problems for which the operation compression step require an additional approximation, and nonlinear problems that can be carried-

out exactly. In the first case, consider $A_\mu \mathbf{u} = \mathbf{c}$ with $A_{ij} = \int_\Omega \nabla \left(g(x, \mu)\varphi_j(x) \right) \cdot \nabla\varphi_i(x)$ and $c_i = \int_\Omega f(x)\varphi_i(x)$, where u is the unknown, f a known loading and $g(x, \mu)$ a known function with variables x and μ that cannot be separated: the previous precomputation of reduced matrices cannot be applied, and a treatment is required to, for example, approximately separate the dependencies in x and μ of g as $g(x, \mu) \approx \sum_{k=1}^d g_k^a(x) g_k^b(\mu)$. Then, $\mathbf{V}^T A_\mu \mathbf{V} \approx \sum_{k=1}^d g_k^b(\mu) A_k$ where $(A_k)_{ij} = \int_\Omega \nabla \left(g_k^a(x)\varphi_j(x) \right) \cdot \nabla\varphi_i(x)$, so that the efficiency of the online stage is recovered; the Empirical Interpolation Method has been proposed in [14, 71] for this purpose. A case of linear problem in harmonic aeroacoustics with nonaffine dependence with respect to the frequency is available in [20]. Conversely, nonlinearities can be handled without approximation in some cases, for instance, the advection term in fluid dynamics can be precomputed in the form of an order-3 tensor: $\int_\Omega \psi_i \cdot \left(\psi_j \cdot \nabla \right) \psi_k, 1 \le i, j, k \le n$; see [3] for the reduction of the nonlinear Navier-Stokes equations with an exact operator compression step. Other examples are found in structural dynamics with geometric nonlinearities, where order-2 and -3 tensors can also be precomputed, see [58, Sect. 3.2] and [73].

When additional approximations are required, the methods proposed for the operator compression step are call "hyper-reduction" in the literature. This term was coined by the seminal method proposed in [88] in 2005, but has been extended to refer to all the methods proposing a such second reduction stage. Hyper-reduction methods include the Empirical Interpolation Method (EIM, [14]), the Missing Point Estimation (MPE, [12]), the Best Point Interpolation Method (BPIM, [78]), the Discrete Empirical Interpolation Method (DEIM, [23]), the Gauss-Newton with Approximated Tensors (GNAT, [17]), the Energy-Conserving Sampling and Weighting (ECSW, [36]), the Empirical Cubature Method (ECM, [45]), and the Linear Program Empirical Quadrature Procedure (LPEQP, [100]). The reader can find an algorithmic comparison of the Hyper-Reduction and the Discrete Empirical Interpolation Method for a nonlinear thermal problem in [39]. A particular focus is given on hyper-reduction techniques via oblique projection and empirical cubature in the following sections.

2.3.4 Hyper-Reduction via a Reduced Integration Domain

Hyper-reduction via a reduced integration domain has been proposed in [88]. It requires a train set of displacement predictions, so that a reduced approximation vector space can be trained. The finite elements simulations that generate the train set of displacement fields are also predicting stresses fields $\boldsymbol{\sigma}$ and internal variables $\mathbf{y}$. Hence, additional reduced bases can be trained for these variables, by using these simulation results [89]. Heuristically, we found it more accurate to include a reduced basis for the stresses $\boldsymbol{\sigma}$, as an additional reduced basis for this hyper-reduction

scheme. Such a reduced basis is also very convenient for error estimation [91]. The finite element shape functions for displacement fields are denoted by $(\boldsymbol{\varphi}_i)_{i=1,...,N}$. For stress fields, we also need to introduce a related finite element representation. We denote by $(\boldsymbol{\varphi}_i^\sigma)_{i=1,...,N^\sigma}$ the dedicated shape functions. In the linear framework of manifold learning, we assume that the same finite element mesh is used for the target simulation, in the online step or for the test set of data, and all simulations used to generate the train set of data.

In practice, the implementation of the hyper-reduction follows the manifold learning step that trains reduced bases for displacements and stresses. We recall that they are respectively denoted by $\mathbf{V} \in \mathbb{R}^{N \times n}$ and $\mathbf{V}^\sigma \in \mathbb{R}^{N^\sigma \times n^\sigma}$ in their matrix form, and $(\boldsymbol{\psi}_k)_{k=1,...,n}$ $(\boldsymbol{\psi}_k^\sigma)_{k=1,...,n^\sigma}$ in their continuous form:

$$\boldsymbol{\psi}_k(\mathbf{x}) = \sum_{i=1}^{N} \boldsymbol{\varphi}_i(\mathbf{x}) V_{ik}, \quad \forall \, \mathbf{x} \in \Omega, \ k = 1, \dots, n, \tag{2.17}$$

$$\boldsymbol{\psi}_k^\sigma(\mathbf{x}) = \sum_{i=1}^{N^\sigma} \boldsymbol{\varphi}_i^\sigma(\mathbf{x}) V_{ik}^\sigma, \quad \forall \, \mathbf{x} \in \Omega, \ k = 1, \dots, n^\sigma. \tag{2.18}$$

The reduced displacement reads:

$$\widehat{\mathbf{u}}(\mathbf{x}) = \sum_{k=1}^{n} \boldsymbol{\psi}_k(\mathbf{x}) \, \gamma_k, \ \forall \mathbf{x} \in \Omega. \tag{2.19}$$

The hyper-reduction method proposed in [88] aims at computing reduced coordinates $(\gamma_k)_{k=1,...,n}$ introduced in Eq. (2.19), by projecting the equilibrium equation on $\mathbf{V}$, via a restriction of the domain Ω to a Reduced Integration Domain (RID) denoted by Ω_R. By following the empirical interpolation method [14], interpolation points are computed for column vectors in $\mathbf{V}$ and $\mathbf{V}^\sigma$ separately [48]. The set of respective interpolation points are denoted by $\mathcal{P}^u$ and $\mathcal{P}^\sigma$. We follow Algorithm 2.1, proposed for the Discrete Empirical Interpolation Method [23].

The RID Ω_R is such that it contains the interpolation points related to $\mathbf{V}$ and $\mathbf{V}^\sigma$. For engineering applications, the RID can also include a zone of interest in Ω that is user-defined, by using a subset of finite elements. This zone of interest is denoted by $\Omega_{ZI} \subset \Omega$. By construction, for contact-free problems, the RID is the following:

$$\Omega_R = \Omega_{ZI} \cup_{i \in \mathcal{P}^u} \text{supp}(\boldsymbol{\varphi}_i) \cup_{i \in \mathcal{P}^\sigma} \text{supp}(\boldsymbol{\varphi}_i^\sigma), \tag{2.20}$$

where $\text{supp}(f) \in \Omega$ is the support of function f. In practice, Ω_R has its own finite-element mesh. It is a reduced mesh involving much less elements than the original finite element mesh of Ω. It can be enlarge by adding a layer of connected element in the original mesh. Integration of constitutive equations are performed on this reduced mesh, without any intrusive operation on the original finite element solver. A similar hyper-reduction scheme has been developed in [38] for contact problems.

Algorithm 2.1: Interpolation points of the Discrete Empirical Interpolation Method (DEIM) [23]

Input : reduced basis vectors $\mathbf{V}[:, k] \in \mathbb{R}^N, k = 1, \ldots M$
Output: interpolation point index set $I := \{i_1, \ldots, i_M\}$

1   set $I_0 := \emptyset$;   `// initialization`
2   **for** $l = 1 \ldots, M$ **do**
3     $\mathbf{U}_{l-1} \leftarrow \mathbf{V}[:, 1 : (l-1)]$;   `// truncated basis`
4     $\mathbf{A} \leftarrow (\mathbf{U}_{l-1}[I_{l-1}, :]^T \mathbf{U}_{l-1}[I_{l-1}, :])^{-1} \mathbf{U}_{l-1}[I_{l-1}, :]^T$;   `// projector`
5     $\mathbf{q}_l \leftarrow \mathbf{V}[:, l] - \mathbf{U}_{l-1} \mathbf{A} \mathbf{V}[I_{l-1}, l]$;   `// interpolation residual`
6     $i_l \leftarrow \arg \max_{i \in \{1, \ldots, N\}} |(\mathbf{q}_l)_i|$;   `// maximum of residual`
7     $I_l := I_{l-1} \cup \{i_l\}$;   `// extend interpolation points`
8   **end**
9   set $I := I_M$.

Once the RID is obtained, a set of test functions is set up in order to restrain the balance equations to Ω_R. They are denoted by $\boldsymbol{\psi}_{Rj}$:

$$\mathcal{P} = \left\{ i \in \{1, \ldots, N\}, \int_{\Omega \setminus \Omega_R} (\boldsymbol{\varphi}_i)^2 \, d\Omega = 0 \right\}, \tag{2.21}$$

$$\boldsymbol{\psi}_{Rj}(\mathbf{x}) = \sum_{i \in \mathcal{P}} \boldsymbol{\varphi}_i(\mathbf{x}) \, V_{ij}, \quad \forall \mathbf{x} \in \Omega, \; j = 1, \ldots, n, \tag{2.22}$$

where $\mathcal{P}$ is the set of all degrees of freedom in Ω_R excepted those belonging to the interface between Ω_R and its counterpart. This interface is denoted by $\mathcal{I}_R$. As explained in [94], the test functions are null on the interface $\mathcal{I}_R$, as if Dirichlet boundary conditions were imposed to the RID. On this interface, displacements follow the shape of the modes $\boldsymbol{\psi}_k$ according to Eq. (2.19). The hyper-reduction method gives access to reduced coordinates $(\gamma_k)_{k=1,\ldots,n}$ that fulfill the following balance equations, for contactless problems:

$$\widehat{\mathbf{u}}(\mathbf{x}) = \sum_{k=1}^{n} \boldsymbol{\psi}_k(\mathbf{x}) \, \gamma_k, \; \forall \mathbf{x} \in \Omega_R \tag{2.23}$$

$$\int_{\Omega_R} \boldsymbol{\varepsilon}(\boldsymbol{\psi}_{Rj}) \, : \, \boldsymbol{\sigma}(\boldsymbol{\varepsilon}(\widehat{\mathbf{u}})) \, d\Omega \tag{2.24}$$

$$- \int_{\Omega_R} \boldsymbol{\psi}_{Rj} \, f_\mu \, d\Omega - \int_{\partial \Omega_R \cap \partial \Omega_N} \boldsymbol{\psi}_{Rj} \, T_\mu, N \, dS = 0. \tag{2.25}$$

$$\forall j = 1, \ldots, n \tag{2.26}$$

The matrix form of the hyper-reduced balance equations reads: find $\boldsymbol{\gamma} \in \mathbb{R}^n$ such that

$$\widehat{\mathbf{u}}(\mathbf{x}) = \sum_{i=1}^{N} \boldsymbol{\varphi}_i(\mathbf{x})\, \widehat{q}_i, \ \forall \mathbf{x} \in \Omega_R, \tag{2.27}$$

$$\widehat{\mathbf{q}} = \mathbf{V}\,\boldsymbol{\gamma}, \tag{2.28}$$

$$\mathcal{F}^{HR}(\boldsymbol{\gamma}) := \mathbf{V}[\mathcal{P}, :]^T \, \mathcal{F}(\mathbf{V}\,\boldsymbol{\gamma})[\mathcal{P}], \tag{2.29}$$

$$\mathcal{F}^{HR}(\boldsymbol{\gamma}) = 0, \tag{2.30}$$

where $\mathbf{V}[\mathcal{P}, :]$ denotes a row restriction of matrix $\mathbf{V}$ to indices in $\mathcal{P}$. The reduced Newton-Raphson step reads:

$$\widehat{\mathbf{u}}^k(\mathbf{x}) = \sum_{i=1}^{N} \boldsymbol{\varphi}_i(\mathbf{x})\, \widehat{q}_i^k, \ \forall \mathbf{x} \in \Omega_R, \tag{2.31}$$

$$\widehat{\mathbf{q}}^k = \mathbf{V}\,\boldsymbol{\gamma}^k, \tag{2.32}$$

$$\mathbf{K}^{HR} := \mathbf{V}[\mathcal{P}, :]^T \, \frac{D\mathcal{F}_\mu}{D\widehat{\mathbf{u}}}(\widehat{\mathbf{u}}^{k-1})[\mathcal{P}, :]\, \mathbf{V}, \tag{2.33}$$

$$\mathbf{K}^{HR}\,(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^{k-1}) = -\mathbf{V}[\mathcal{P}, :]^T \, \mathcal{F}(\widehat{\mathbf{u}}^{k-1})[\mathcal{P}], \tag{2.34}$$

where the reduced stiffness matrix $\mathbf{K}^{HR}$ is computed by using solely the elements of the RID Ω_R. We assume that the matrix $\mathbf{K}^{HR}$ is full rank. This assumption is always checked in numerical solutions of hyper-reduced equations. Rank deficiency may appear when the RID construction do not account for the contribution of a reduced basis dedicated to stresses.

Once the RID is represented as a finite element mesh, this hyper-reduction scheme is intrusive solely for the linear solver involved in the Newton-Raphson step and its related convergence criterion. Nevertheless, the mesh of the RID has to include labels for the set $\mathcal{P}$ or its counterpart $\mathcal{I}_R$. This counterpart is the set of degrees of freedom connected to elements of the original mesh that are not in the reduced mesh.

Remarks:

- Here the most complex operations are indeed the computation of $\mathbf{K}^{HR}$ and the solution of the reduced linear system of equations. They respectively scale linearly with card($\mathcal{P}$) n^2 and n^3. Hence n^3 has to be small enough compared to N if we consider the computational complexity for the solution of sparse linear systems in the finite element method.
- Because of the spreading nature of interpolation points, most of the time, the RID is not a compact subdomain.
- The hyper-reduced order model is a kind of submodel where the displacements at the interface $\mathcal{I}_R$ follow the shape of the modes $\boldsymbol{\psi}_k$ according to Eq. (2.19).
- Finite element corrections for displacements and stresses can be easily computed over the RID once the reduced prediction have been achieved. This scheme is termed Hybrid Hyper-Reduction in [46].
- A parallel programming of the hyper-reduction method has been proposed in [95].

- Reduced order models not only save computational time, they save computational resources including energy consumption savings as explained in [90] and memory footprint [46].

Property 1: In linear elasticity, if $\mathbf{K}^{HR}$ is full rank, the hyper-reduced balance equations are equivalent to an oblique projection of the finite element prediction $\mathbf{q} \in \mathbb{R}^N$:

$$\mathbf{\Pi}^T := \mathbf{V}[\mathcal{P}, :]^T \, \mathbf{K}[\mathcal{P}, :], \tag{2.35}$$

$$\widehat{\mathbf{q}} = \mathbf{V} \, (\mathbf{\Pi}^T \, \mathbf{V})^{-1} \mathbf{\Pi}^T \, \mathbf{q}, \tag{2.36}$$

$$\text{and} \quad \mathbf{\Pi}^T \, \widehat{\mathbf{q}} = \mathbf{\Pi}^T \, \mathbf{q}, \tag{2.37}$$

with $\mathbf{K} \, \mathbf{q} = \mathbf{F}$. Hence the hyper-reduced prediction of the reduced vector $\boldsymbol{\gamma}$ is a minimizer for $f(\boldsymbol{\beta})$:

$$\boldsymbol{\gamma}^\star \in \mathbb{R}^n, \ f(\boldsymbol{\gamma}^\star) = \|\mathbf{\Pi}^T \, \left(\mathbf{V} \, \boldsymbol{\gamma}^\star - \mathbf{q}\right)\|_2^2. \tag{2.38}$$

Here $\mathbf{\Pi}$ is a projector for elastic stresses in Ω_R according to the reduced test functions:

$$\sum_{i=1}^{N} \Pi_{ik} \, (\mathbf{V} \, \boldsymbol{\gamma} - \mathbf{q})_i = \int_{\Omega_R} \boldsymbol{\varepsilon}(\boldsymbol{\psi}_{Rk}) \, : \, (\boldsymbol{\sigma}(\widehat{\mathbf{q}}) - \boldsymbol{\sigma}(\mathbf{q})) \, d\Omega. \tag{2.39}$$

The proof is straightforward. Here, $\mathbf{K}^{HR} = \mathbf{\Pi}^T \, \mathbf{V}$. The Jacobian matrix for f reads $\mathbf{J} = \mathbf{V}^T \, \mathbf{\Pi} \, \mathbf{\Pi}^T \, \mathbf{V} = (\mathbf{K}^{HR})^T \, \mathbf{K}^{HR}$. If $\mathbf{K}^{HR}$ is full rank, then $\mathbf{J}$ is symmetric definite positive and $\mathbf{J}^{-1} = (\mathbf{K}^{HR})^{-1} \, (\mathbf{K}^{HR})^{-T}$. Then, both the minimization problem and the hyper-reduced equation have a unique solution. The solution of the minimization problem is:

$$\mathbf{q}^f = \mathbf{V} \, \mathbf{J}^{-1} \mathbf{V}^T \, \mathbf{\Pi} \, \mathbf{\Pi}^T \, \mathbf{q}, \tag{2.40}$$

$$= \mathbf{V} \, (\mathbf{K}^{HR})^{-1} \, \mathbf{\Pi}^T \, \mathbf{q}, \tag{2.41}$$

$$= \mathbf{V} \, (\mathbf{K}^{HR})^{-1} \, \mathbf{V}[\mathcal{P}, :]^T \, \mathbf{K}[\mathcal{P}, :] \, \mathbf{q}, \tag{2.42}$$

$$= \mathbf{V} \, (\mathbf{K}^{HR})^{-1} \, \mathbf{V}[\mathcal{P}, :]^T \, \mathbf{F}[\mathcal{P}], \tag{2.43}$$

$$= \widehat{\mathbf{q}}. \tag{2.44}$$

As an intermediate result, Eq. (2.42) is the oblique projection.

In linear elasticity, the Céa's lemma holds. Let us denote $\mathbf{q}^\circ \in \mathbb{R}^N$ the minimizer of the upper bound in Eq. (2.3) related to this lemma:

$$\mathbf{q}^\circ = \arg \min_{\mathbf{q}^\star \in \mathbb{R}^N} \|\widetilde{u} - \sum_{i=1}^N \boldsymbol{\varphi}_i q_i^\star\|, \tag{2.45}$$

$$\widetilde{v}^\circ = \sum_{i=1}^N \boldsymbol{\varphi}_i q_i^\circ. \tag{2.46}$$

The best projection of the minimizer $\mathbf{q}^\circ$ in the approximation space is denoted by $\boldsymbol{\gamma}_P$:

$$\boldsymbol{\gamma}^P = \mathrm{argmin}_{\mathbf{g} \in \mathbb{R}^n} (\mathbf{q}^\circ - \mathbf{V}\,\mathbf{g})^T \mathbf{M}(\mathbf{q}^\circ - \mathbf{V}\,\mathbf{g}), \tag{2.47}$$

$$\widetilde{v}^P = \sum_{i=1}^N \boldsymbol{\varphi}_i (\mathbf{V}\,\boldsymbol{\gamma}^P)_i. \tag{2.48}$$

Let us introduce an ideal reduced basis $\mathbf{V}^\circ \in \mathbb{R}^{N \times n}$ (It assumes that n is an ideal reduced dimension) such that: $\mathbf{q}^\circ = \mathbf{V}^\circ\,\boldsymbol{\gamma}^\circ$, and $\mathbf{V}^{\circ T}\mathbf{M}\mathbf{V}^\circ = \mathbf{I}$, where $M_{ij} = \langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \rangle$. Hence $\boldsymbol{\gamma}^P = \mathbf{V}^T \mathbf{M} \mathbf{V}^\circ\,\boldsymbol{\gamma}^\circ$.

Property 2: In linear elasticity, the upper bound of approximation error is increased by a Chordal distance [101] between $\mathbf{V}$ and the ideal reduced basis $\mathbf{V}^\circ$:

$$\|\widetilde{u} - \widetilde{v}^\circ\| \le \|\widetilde{u} - \widetilde{v}^P\| + \|\boldsymbol{\gamma}^\circ\|_2\, d^{Ch}(\mathbf{V}^\circ, \mathbf{V}), \tag{2.49}$$

where $d^{Ch}(\mathbf{V}^\circ, \mathbf{V})$ is the Chordal distance between $\mathbf{V}^\circ$ and $\mathbf{V}$.

Hence, the smaller the Chordal distance between the sub-space spanned by $\mathbf{V}$ and $\mathbf{V}^\circ$, the better the reduced prediction by using a Galerkin projection (When the RID covers the full domain). A certification of the reduced projection can be achieved, when all errors admit an upper bound, by following the constitutive relation error proposed in [47, 55].

The Chordal distance uses the principal angles $\boldsymbol{\theta} \in \mathbb{R}^n$, $\theta_k \in [0, \pi/2[$ for $k = 1, \ldots, n$, computed via a full singular value decomposition:

$$\mathbf{V}^T\,\mathbf{M}\,\mathbf{V}^\circ = \mathbf{U}\,\cos(\boldsymbol{\theta})\mathbf{U}^{\circ T}, \quad \mathbf{U}^T\mathbf{U} = \mathbf{U}^{\circ T}\mathbf{U}^\circ = \mathbf{I}, \tag{2.50}$$

$$d^{Ch}(\mathbf{V}^\circ, \mathbf{V}) = \|\sin(\boldsymbol{\theta})\|_F, \tag{2.51}$$

$$\|\mathbf{U}^\circ\|_F^2 = n, \tag{2.52}$$

where $\|\cdot\|_F$ is the Frobenius norm. Here, $\cos(\boldsymbol{\theta})$ and $\sin(\boldsymbol{\theta})$ are cosine and sine diagonal matrices. In addition the following property holds when a full SVD is computed:

$$\mathbf{U}^\circ\,\mathbf{U}^{\circ T} = \mathbf{I}, \quad \mathbf{U}\,\mathbf{U}^T = \mathbf{I}. \tag{2.53}$$

The proof of the previous property is straightforward by using the triangular inequality. We just need to prove that:

$$\|\widetilde{v}^\circ - \widetilde{v}^P\| \le \|\boldsymbol{\gamma}^\circ\|_2\, d^{Ch}(\mathbf{V}^\circ, \mathbf{V}).$$

Hence, the proof is the following:

$$\|\widetilde{v}^\circ - \widetilde{v}^P\|^2 = \boldsymbol{\gamma}^{\circ T} \, (\mathbf{V}^\circ - \mathbf{V}\,\mathbf{V}^T\mathbf{M}\mathbf{V}^\circ)^T \, \mathbf{M} \, (\mathbf{V}^\circ - \mathbf{V}\,\mathbf{V}^T\mathbf{M}\mathbf{V}^\circ) \, \boldsymbol{\gamma}^\circ$$
$$= \boldsymbol{\gamma}^{\circ T} \, (\mathbf{I} - \mathbf{V}^{\circ T}\mathbf{M}\mathbf{V}\,\mathbf{V}^T\mathbf{M}\mathbf{V}^\circ) \, \boldsymbol{\gamma}^\circ. \tag{2.54}$$

Therefore

$$\|\widetilde{v}^\circ - \widetilde{v}^P\|^2 = \boldsymbol{\gamma}^{\circ T} \, (\mathbf{I} - \mathbf{U}^\circ \cos(\boldsymbol{\theta})^2 \, \mathbf{U}^{\circ T}) \, \boldsymbol{\gamma}^\circ \tag{2.55}$$
$$= \boldsymbol{\gamma}^{\circ T} \, \mathbf{U}^\circ (\mathbf{I} - \cos(\boldsymbol{\theta})^2) \, \mathbf{U}^{\circ T} \, \boldsymbol{\gamma}^\circ \tag{2.56}$$
$$= \boldsymbol{\gamma}^{\circ T} \, \mathbf{U}^\circ \sin(\boldsymbol{\theta})^2 \, \mathbf{U}^{\circ T} \, \boldsymbol{\gamma}^\circ \tag{2.57}$$
$$= \|\sin(\boldsymbol{\theta}) \, \mathbf{U}^{\circ T} \, \boldsymbol{\gamma}^\circ\|_2^2. \tag{2.58}$$

For all matrices $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{B} \in \mathbb{R}^{m \times n}$ the following property holds:

$$\|\mathbf{A}\mathbf{B}\|_F \le \|\mathbf{A}\|_F \, \|\mathbf{B}\|_F,$$

and for $\mathbf{a} \in \mathbb{R}^n$: $\|\mathbf{a}\|_F = \|\mathbf{a}\|_2$.
Thus:

$$\|\widetilde{v}^\circ - \widetilde{v}^P\|^2 \le \|\sin(\boldsymbol{\theta})\|_F^2 \, \|\mathbf{U}^{\circ T} \, \boldsymbol{\gamma}^\circ\|_2^2 \le \|\sin(\boldsymbol{\theta})\|_F^2 \, \|\boldsymbol{\gamma}^\circ\|_2^2. \tag{2.59}$$

Property 3: When the identity matrix is substituted for $\mathbf{K}$ in Eqs. (2.35) and (2.36) is known as the Gappy POD reconstruction [35] of truncated variables $\mathbf{q}[\mathcal{P}]$. The reconstructed vector in $\mathbb{R}^N$ is:

$$\widetilde{\mathbf{q}} = \mathbf{V} \, (\mathbf{V}[\mathcal{P}, :]^T \, \mathbf{V}[\mathcal{P}, :])^{-1}\mathbf{V}[\mathcal{P}, :]^T \, \mathbf{q}[\mathcal{P}]. \tag{2.60}$$

This Gappy POD reconstruction is useless for displacement variables because the oblique projection in Eq. (2.36) is a direct outcome of the hyper-reduced prediction. But such a reconstruction is very convenient for stress variables that the hyper-reduced scheme forecasts only on Ω_R. The reconstructed stress variables reads:

$$\widetilde{\mathbf{q}}^\sigma = \mathbf{V}^\sigma \, (\mathbf{V}^\sigma[\overline{\mathcal{P}}^\sigma, :]^T \, \mathbf{V}^\sigma[\overline{\mathcal{P}}^\sigma, :])^{-1}\mathbf{V}^\sigma[\overline{\mathcal{P}}^\sigma, :]^T \, \mathbf{q}^\sigma[\overline{\mathcal{P}}^\sigma], \tag{2.61}$$

where $\overline{\mathcal{P}}^\sigma$ is the set of all stress indices available in Ω_R. Since the RID contains interpolation points for $\mathbf{V}^\sigma$, these points are included in $\overline{\mathcal{P}}^\sigma$, therefore the truncated matrix $\mathbf{V}^\sigma[\overline{\mathcal{P}}^\sigma, :]$ is full column rank and the reconstruction is a well posed problem.

Remark about the RID construction and the DEIM: If the RID contains solely the elements connected to interpolation points related to the reduced basis $\mathbf{V}$, such that $\mathcal{P} = \mathcal{P}^u$, then the Gappy POD gives the interpolation scheme of the DEIM:

$$\widetilde{\mathbf{q}}^{DEIM} = \mathbf{V} \, (\mathbf{V}[\mathcal{P}^u, :])^{-1} \, \mathbf{q}[\mathcal{P}^u]. \tag{2.62}$$

But, when considering the hyper-reduction scheme, one can observe overfitting in the sense that the train set of displacement is very well approximated by the DEIM

reconstruction, but the hyper-reduced predictions are not accurate. For this reason, we recommend the use of the additional reduced basis $\mathbf{V}^\sigma$ and the related interpolation points.

Various applications of the hyper-reduction method using a RID have been developed for:

- thermal problems in structures or solids, in [88],
- boundary element models [93],
- reduced simulations of sintering processes [99],
- ductile damage predictions, including unstable localisation of strains [97],
- reduction of multidimensional domains, when space variables are an Euclidean space of arbitrary dimension $D > 3$ [98],
- simulation of viscoelastic-viscoplastic composites materials [74],
- model calibration in plasticity of materials [37, 46, 96],
- contact problems using Lagrange multipliers [38, 62],
- arc length algorithm for buckling problems or strain localisation [59],
- micromorphic continua including higher order stress fields [48].

2.3.5 Hyper-Reduction via Empirical Cubature

To assemble the linearized equations of the reduced Newton algorithm (2.13) when using the ROM in the online phase, hyper-reduction techniques via empirical cubature aim to compute the costly integrals over the high-fidelity domain by replacing the high-dimensional quadrature formula by a low-dimensional reduced quadrature with positive weights. The ECSW [36], ECM [45] and LPEQP [100] are methods implementing such reduced quadratures. In this section, we present the ECM, more details are available in [19].

We consider the high-fidelity model described in Sect. 2.2. The integrals involved in the assembling of the linearized Eq. (2.7) make use of high-fidelity quadrature formulas. Apply such quadrature to the reduced internal forces vector:

$$
\begin{aligned}
\hat{F}_i^{\text{int}}(t) &:= \int_\Omega \sigma \left(\epsilon(\hat{u}), y \right)(x, t) : \epsilon(\psi_i)(x) \\
&= \sum_{e \in E} \sum_{k=1}^{n_e} \omega_k \sigma \left(\epsilon(\hat{u}), y \right)(x_k, t) : \epsilon(\psi_i)(x_k), \, 1 \le i \le n,
\end{aligned}
\tag{2.63}
$$

where E denotes the set of elements of the mesh, n_e the number of quadrature points for the element e, ω_k and x_k are the quadrature weights and points associated to e. The total number of quadrature points is denoted N_G.

The ECM aims to approximate the high-fidelity quadrature by a reduced quadrature with positive weights, which, when applied to the reduced internal forces vector, writes

$$\hat{F}_i^{\text{int}}(t) \approx \sum_{k'=1}^{n_g} \hat{\omega}_{k'} \sigma\left(\epsilon(\hat{u}), y\right)(\hat{x}_{k'}, t) : \epsilon\left(\psi_i\right)(\hat{x}_{k'}), 1 \leq i \leq n, \qquad (2.64)$$

where $\hat{\omega}_{k'} > 0$ and $\hat{x}_{k'}$ are respectively the reduced quadrature weights and points, and $n_g \ll N_G$ is the length of the reduced quadrature.

Denote $f_q := \sigma\left(\epsilon(u_{(q//n)+1}), y\right) : \epsilon\left(\psi_{(q\%n)+1}\right)$, $1 \leq q \leq nN_c$. where $//$ and $\%$ are the quotient and the remainder of the Euclidean division. Denote as well $\mathcal{Z}^{n_G}$ a subset of $[1; N_G]$ of size n_G and $J_{\mathcal{Z}^{n_G}} \in \mathbb{R}^{nN_c \times n_G}$ and $\boldsymbol{b} \in \mathbb{N}^{nN_c}$ such that for all $1 \leq q \leq nN_c$ and all $1 \leq q' \leq n_G$,

$$J_{\mathcal{Z}^{n_G}} = \left(f_q(x_{\mathcal{Z}_{q'}^{n_G}})\right)_{1 \leq q \leq nN_c,\ q' \in \mathcal{Z}^{n_G}}, \qquad \boldsymbol{b} = \left(\int_\Omega f_q\right)_{1 \leq q \leq nN_c}, \qquad (2.65)$$

where $\mathcal{Z}_{q'}^{n_G}$ denotes the q'-th element of $\mathcal{Z}^{n_G}$. We remind that n is the number of POD modes, see Sect. 2.3.3.2. Let $\hat{\boldsymbol{\omega}} \in \mathbb{R}^{+n}_G$, $\left(J_{\mathcal{Z}^{n_G}} \hat{\boldsymbol{\omega}}\right)_q = \sum_{q'=1}^{n_G} \hat{\omega}_{q'} \sigma\left(\epsilon(u_{(q//n)+1}), y\right)$ $(x_{\mathcal{Z}_{q'}^{n_G}}) : \epsilon\left(\psi_{(q\%n)+1}\right)(x_{\mathcal{Z}_{q'}})$, $1 \leq q \leq nN_c$, is a candidate approximation for $\int_\Omega \sigma\left(\epsilon(u_{(q//n)+1}), y\right) : \epsilon\left(\psi_{(q\%n)+1}\right) = b_q$, $1 \leq q \leq nN_c$. The problem of finding the more accurate reduced quadrature formula of length n_G for the reduced internal forces vector is:

$$\left(\hat{\omega}, \mathcal{Z}^{n_G}\right) = \arg \min_{\hat{\omega}' \in \mathbb{R}^{+n_G}, \mathcal{Z}^{n_G} \subset [1; N_G]} \left\| J_{\mathcal{Z}^{n_G}} \hat{\omega}' - b \right\|, \qquad (2.66)$$

where $\|\cdot\|$ denotes the Euclidean norm. Minimizing the length of the reduced quadrature formula as well leads to a NP-hard problem, which solution can be approximated using a Nonnegative Orthogonal Matching Pursuit algorithm, see Algorithm 2.2.

Algorithm 2.2: Nonnegative Orthogonal Matching Pursuit.

Input : J, b, tolerance $\epsilon_{\text{Op.comp.}}$, x_k, $1 \leq k \leq N_G$
Output: $\hat{\omega}_k$, $\hat{x}_k$, $1 \leq k \leq d$.
1 Set $\mathcal{Z} = \emptyset$, $k' = 0$, $\hat{\omega} = 0$ and $r_0 = b$;   // initialization
2 **while** $\|r_{k'}\|_2 > \epsilon_{\text{Op.comp.}} \|b\|_2$ **do**
3 $\quad$ $\mathcal{Z} \leftarrow \mathcal{Z} \cup$ max index $\left(\left(J_{[1;N_G]}\right)^T r_{k'}\right)$
4 $\quad$ $\hat{\omega} \leftarrow \arg \min_{\hat{\omega}'>0} \|b - J_{\mathcal{Z}}\hat{\omega}'\|_2$
5 $\quad$ $r_{k'+1} \leftarrow b - J_{\mathcal{Z}}\hat{\omega}$
6 $\quad$ $k' \leftarrow k' + 1$
7 **end**
8 $d \leftarrow k'$
9 $\hat{x}_k := x_{\mathcal{Z}_k}$, $1 \leq k \leq d$

In Algorithm 2.2, $J_{[1;N_G]}$ satisfies the definition (2.65) with $\mathcal{Z}^{n_G} = [1; N_G]$. The positivity of the weights of the reduced quadrature preserves the spectral properties of the operator associated with the high-fidelity problem, see [19, Remark 1].

2.3.6 Computational Complexity

In this section we restrict our attention to elliptic problems or to linearized problems. The bilinear part of the weak form for finite-dimensional solution spaces is a matrix. When using Finite Element solution space, this matrix is sparse. But when using a reduced solution space, this matrix is usually a full matrix. Therefore, computational complexity of the finite element prediction is the complexity of the solution of sparse linear system. It scales linearly with $N \omega^2$, where ω is the band width of the sparse matrix. For the reduced prediction, the solution of a full linear system scales linearly with n^3. We recommend to restrict linear model reduction schemes, with or without hyper-reduction, to reduced dimension n lower than $N^{1/3}$, otherwise the solution of reduced equation will have a computational complexity similar to the complexity of the finite element model. This recommendation does not concern explicit solvers.

2.4 Nonlinear Manifold Learning for Projection-Based Reduced-Order Modeling

Consider a parametrized variability, and a set of snapshots generated using the high-fidelity model over a sampling of the parameter domain. The parametrized problem is said nonreducible when applying a linear data compression over this set of snapshots leads to a ROB containing too many vectors for the online problem to feature an interesting speedup. Formally, this happens when the Kolmogorov n-width $d_n(\mathcal{M})$ decreases too slowly with respect to n, where we recall that n is the cardinality of the ROB,

$$d_n(\mathcal{M}) := \inf_{\mathcal{H}_n \in \mathrm{Gr}(n,\mathcal{H})} \sup_{u \in \mathcal{M}} \inf_{v \in \mathcal{H}_n} ||u - v||_{\mathcal{H}}, \tag{2.67}$$

with the Grassmannian $\mathrm{Gr}(n, \mathcal{H})$ being the set of all n-dimensional subspaces of $\mathcal{H}$ and $\mathcal{H}_n \in \mathrm{Gr}(n, \mathcal{H})$ the subspace spanned by the considered ROB. Qualitatively, the solution manifold $\mathcal{M}$ covers too many independent directions to be embedded in a low-dimensional subspace. To address this issue, several techniques have been developed:

- Problem-specific methods tackle the difficulties of some specific physics problems that are known to be nonreducible, such as advection-dominated problems which have been largely investigated, for instance in [16, 49, 85].

- Online-adaptive model reduction methods update the ROM in the exploitation phase by collecting new information online as explained in [102], in order to limit extrapolation errors when solving the parametrized governing equations in a region of the parameter space that was not explored in the training phase. The ROM can be updated for example by querying the high-fidelity model when necessary for basis enrichment [18, 44, 56, 80, 88].
- ROM interpolation methods [6–9, 24, 64–68, 75, 76] use interpolation techniques on Grassmann manifolds or matrix manifolds to adapt the ROM to the parameters considered in the exploitation phase by interpolating between two precomputed ROMs.
- Dictionaries of basis vector candidates enable building a parameter-adapted ROM in the exploitation phase by selecting a few basis vectors. This technique is presented in [54, 72] for the Reduced Basis method.
- Nonlinear manifold ROM methods [57, 63] learn a nonlinear embedding and project the governing equations onto the corresponding approximation manifold, by means of a nonlinear function mapping a low-dimensional latent space to the solution space.
- Dictionaries of ROMs rely on the construction of several local ROMs adapted to different regions of the solution manifold. These local ROMs can be obtained by partitioning the time interval [32, 33], the parameter space [33, 34, 42, 44, 51, 52, 82], or the solution space [10, 11, 27, 40, 77, 82, 92].

In the following Sects. 2.4.1 and 2.4.2, we provide more details on the last two entries of the previous list.

2.4.1 Nonlinear Dimensionality Reduction via Auto-Encoder

Nonlinear manifold learning means that the solution manifold is approximated by a domain in the ambient solution space that is not included in a low-dimensional vector subspace, as illustrated in Fig. 2.3.

Let us consider a formal representation of parabolic and nonlinear Partial Differential Equations (PDEs) that are parameterized with respect to some physical parameters of interest. We mean by physical parameters the parameters that appear directly within the equations such as the boundary conditions, the viscosity for fluid mechanics (henceforth the Reynolds number), the time step for dynamical systems of fluid flows or infectious diseases, etc. These parameters are denoted μ without any loss of generality as introduced in the preceding section. The formal representation of the equations is given as follows:

$$\frac{\partial \widetilde{u}}{\partial t} = f(\widetilde{u}, \mu). \tag{2.68}$$
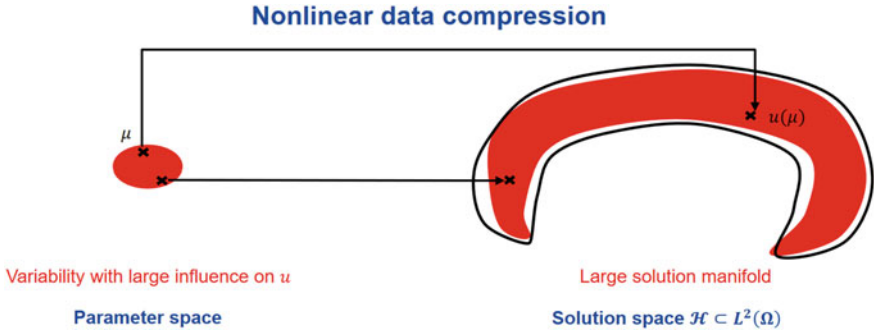
Fig. 2.3 Nonlinear manifold learning

Reduced order modeling based on nonlinear data compression techniques might be a solution for example when the described physical fields by the model equations require a large number of vectors in the ROM as specified above. Nevertheless there are cases where even if the physical solution fields are completely reducible, the Galerkin projection may not be appropriate for the model equations describing this physics.

Convection-diffusion PDEs have this stability issue even more generally when considering Galerkin projection using finite element basis functions. In the literature, it is proved that the coherent structures of a turbulent, unsteady and in-compressible fluid flow are reproducible by a small number of POD basis functions. However, if these functions are used to solve the Newton-Raphson problem of the associated reduced order Galerkin dynamical system, then an instability appears as a function of the time. In the literature, many solutions are proposed to tackle this difficulty while keeping the reduced order approximation in a linear space spanned by the POD basis functions. We can refer to the Petrov-Galerkin technique, the least square minimization of the equations residual, the variational finite element method, etc. Recently, nonlinear approximations of the solution fields in a manifold of reduced dimension start to gain importance in the literature. In this case, the reduced order model is said to be a nonlinear projection based reduced model. Some authors introduced nonlinear approximations using Deep Learning approaches for projection based reduced models. We find in the literature more classical nonlinear approximations based on the Kernel POD technique.

In what follows, we make a focus from the literature on Deep Learning projection based reduced models and their different possible formulations. More precisely, we are talking about Deep AutoEncoders (DAEs) from the domain of Deep Learning. DAEs are artificial neural networks formed of layers of spatial convolutions, nonlinear activation functions and linear systems called fully connected functions. These architectures are used to perform nonlinear dimensional reduction, following unsupervised data compression. Henceforth, the DAE allows to determine latent features within a set of given inputs.

We denote by h and g respectively the encoder mapping and the decoder mapping of a DAE. In general g is the transpose mapping of h. We denote by $\widehat{\alpha}$ the reduced latent features inferred by h. The dimension of the latent features is equal to the intrinsic dimensionality of the manifold as stated in Remark 2.1 in [63]. This intrinsic dimensionality is in the current case the dimension N_μ of the vector of parameters μ, which may include the time variable also.

We note that the reduced latent features of a DAE are not parameterized variables in general. In other words they can be seen as non-parametric features associated with a given set of inputs. This formulation is interesting in the framework of projection-based model reduction, where the associated parameters are given straightforwardly by the physical equations. Hence, knowing the variable parameters within the inputs data helps only with the determination of the intrinsic dimension of the manifold.

In the literature, we find two different formulations for DAEs projection based reduced models.

The first formulation was proposed by Kashima [53] and Hartman et al. [43]. It is very analogical to the Galerkin formulation of projection based reduced models: given h and g such that $g \circ h : \widetilde{u} \to \widehat{u}$ and $\widehat{u} \approx \widetilde{u}$, then the reduced model is formulated as follows:

$$\frac{\partial}{\partial t}\widehat{\alpha}(\mu) = h \circ f \circ g(\widehat{\alpha}(\mu)) \quad , \quad \widehat{\alpha}(t=0) = h(\widetilde{u}(t=0)) \qquad (2.69)$$

$$= h\left(f(g(\widehat{\alpha}(\mu)), \mu)\right). \qquad (2.70)$$

In the above formulation, the authors relied on the following three points in order to set the time derivative of the latent features equal to the right hand side of Eq. (2.69).

- $\dfrac{\partial}{\partial t}g(\widehat{\alpha}(\mu))$ belongs to the manifold described by $\mu \to \widetilde{u}(\mu)$,
- $h\left(\dfrac{\partial}{\partial t}g(\widehat{\alpha}(\mu))\right) = \dfrac{\partial}{\partial t}h\left(g(\widehat{\alpha}(\mu))\right)$,
- $h \circ g = I_{N_\mu}$.

Remark 2.1 The first two above items are hypothesis that are fulfilled in the case where h and g are linear or affine functions. The last item is fulfilled theoretically by the inputs data compression using parameters optimisation of the DAE architecture.

The second formulation of DAEs projection based reduced models is proposed in [63], where a least square minimisation of the residual of parabolic PDEs because of the decoder approximation is performed. Then, the reduced model is formulated as follows in order to determine the reduced latent features:

$$\frac{\partial}{\partial t}\widehat{\alpha}(\mu) = \mathrm{argmin}_{\widehat{v}(\mu)\in\mathbb{R}^{N_\mu}} \|J(\widehat{a}(\mu))\widehat{v}(\mu) - f(g(\widehat{a}(\mu)), \mu)\|_2^2, \qquad (2.71)$$

where $\widehat{v}(\mu)$ is the time derivative of $\widehat{a}(\mu)$, $\|.\|_2$ denotes the mean square norm or the euclidean norm and, J is the Jacobian matrix of the decoder mapping which belong

element-wise to the tangent space to the solutions manifold at a given point. J is expressed as follows:

$$J : \widehat{a} \rightarrow \nabla g(\widehat{a}).$$

In this second formulation, the authors do not suppose that the velocity of the decoder approximation is in the manifold of the solutions because mathematically it belongs to the tangent space to the manifold at a given point. Hence, they claim that encoding the decoder approximation will produce a poor approximation by the reduced model.

2.4.2 Piecewise Linear Dimensionality Reduction via Dictionary-Based ROM-Nets

Parts of this section has been inspired from the authors previous work [30].

Piecewise linear manifold learning means that the solution manifold is approximated by a dictionary of local linear subspaces, as illustrated in Fig. 2.4, where we denote $\mathcal{M}$ the solution manifold.

The solution manifold is partitioned to get a collection of subsets $\mathcal{M}_k \subset \mathcal{M}$ that can be covered by a dictionary of low-dimensional subspaces, enabling the use of local linear ROMs. If $\{\mathcal{M}_k\}_{k \in [\![1; K]\!]}$ is a partition of $\mathcal{M}$, then:

$$\forall k \in [\![1; K]\!], \ \forall N \in \mathbb{N}^*, \quad d_N(\mathcal{M}_k) \leq d_N(\mathcal{M}). \tag{2.72}$$

The concepts of ROM-net and dictionary-based ROM-net are introduced in [27], which we present in this section. Suppose we dispose of an already computed dictionary of ROMs for the parametrized problem (2.4), where each element of the dictionary is a ROM that can approximate the problem on a subset of the solution manifold $\mathcal{M}$. A dictionary-based ROM-net is a machine learning algorithm trained to assign the parameter $\mu \in \mathcal{X}$ to the ROM of the dictionary leading to the most accu-
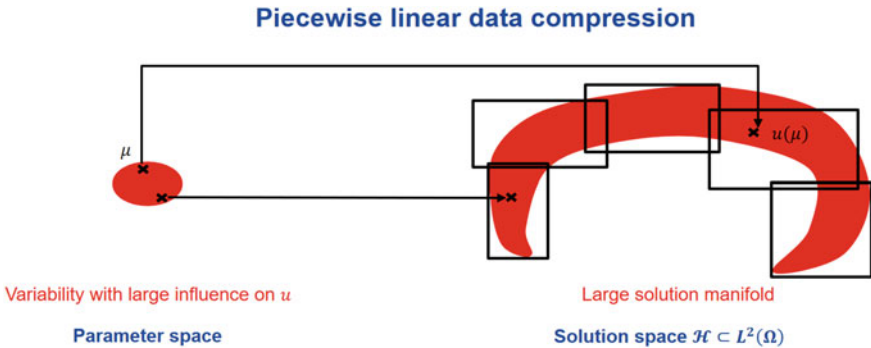

Fig. 2.4 Piecewise linear manifold learning

Fig. 2.5 Exploitation phase of a dictionary-based ROM-net. K local ROMs, combined with a classifier C_K for automatic ROM $C_K(\mu)$ recommendation, are used to predict the quantity of interest $Z(\mu)$

rate reduced prediction. This assignment, called model recommendation in [77], is a classification task, see Fig. 2.5.

The dictionary of ROMs is constructed in a clustering stage, during which snapshots are regrouped depending on their respective proximity on $\mathcal{M}$, in the sense of a particular dissimilarity measure we introduced in [29] and [26]. The dissimilarity between two parameter values $\mu, \mu' \in \mathcal{X}$, denoted by $\delta(\mu, \mu')$, involves the sine of the principal angles between subspaces associated to the solutions of the HFM $u(\mu), u(\mu') \in \mathcal{M}$, see [26, Definition 4.10]. Applying a k-medoids clustering algorithm on the solution manifold $\mathcal{M}$ using the dissimilarity δ leads to an optimal partitioning for a dictionary of local ROMs, in a sense introduced in [26, Property 4.13]. We refer to the remaining of [26] for the description of a practical efficiency criterion of the dictionary-based ROM-net, which enables to decide, before the computationally costly steps of the workflow, if a dictionary of ROMs is preferable to one global ROM, and how to calibrate the various hyperparameters of the ROM-net.

Remark 2.2 Importance of the classification. One could argue that the classification step can be replaced by choosing the cluster k for which the dissimilarity measure $\delta(\mu, \tilde{\mu}_k)$ between the parameter μ and the cluster medoid $\tilde{\mu}_k$ is the smallest. However, we recall that the computation of the dissimilarity measure requires solving the HFM at the parameter value μ, which would render the complete model reduction framework useless. Hence, the classification step enables to bypass this HFM solve and directly recommend the appropriate local ROM.

As briefly mentioned in the introduction of Sect. 2.4, local ROMs can be constructed by partitioning the parameter space [33, 34], in which case the classification step is not required: the cluster affectation is made by computing distances directly in the parameter space. In other cases, partitioning in the solution space can be considered without requiring a classification step [10]. Consider a time-dependent problems where the initial condition is not a parameter of the problem, and suppose an efficient computation of the clustering distance in the solution space based on the reduced solution at the previous time-step. Then, local basis affectation and switching is possible without requiring classification.

The training of the classifier can be difficult when working with physical fields: simulations are costly, data are in high dimension and classical data augmentation

techniques for images cannot be applied. Hence, we can consider replacing the HFM by an intermediate-fidelity solver for generating the data needed for the training of the classifier, by considering coarser meshes and fewer time steps. We point out that the HFM should be used at the end for generating the data required in the training of the local ROMs. We propose in [28] improvements for the training of the classifier in our context by developing a fast variant of the mRMR [83] feature selection algorithm, and new class-conserving transformations of our data, acting like a data augmentation procedure.

2.5 Iterative and Greedy Strategies

For the sake of the presentation, we have separated the offline and online phases, where the Reduced-Order Model is learnt, then exploited. Actually, more involved strategies exist, where the ROM is constructed in a iterative fashion. The Reduced Basis Method [84] is a greedy method, where the ROB is constructed by a single snapshot, corresponding to a randomly chosen parameter value, and the ROB is enriched by the parameter value that maximizes the error made by the current ROM. In complex parameter dependencies, the hyper-reduction scheme can be simulateously constructed as the ROB grows, see [31] for such a scheme, with the EIM as hyper-reduction. This greedy construction as been extended to time-dependant problems in the POD-greedy method [41], and simultaneous hyper-reduction construction have also been proposed [88].

Such iterative strategies rely on a efficient computation of the error made by the ROM. Error estimation is investigated in the next chapter.

References

1. N. Akkari, F. Casenave, E. Hachem, D. Ryckelynck, A bayesian nonlinear reduced order modeling using variational autoencoders. Fluids **7**(10), (2022)
2. N. Akkari, F. Casenave, D. Ryckelynck, C. Rey, An updated Gappy-POD to capture non-parameterized geometrical variation in fluid dynamics problems. Adv. Model. Simul. Eng. Sci. **9**(1), 1–34 (2022)
3. N. Akkari, R. Mercier, G. Lartigue, V Moureau, Stable pod-galerkin reduced order models for unsteady turbulent incompressible flows, in *55th AIAA Aerospace Sciences Meeting*, Grapevine, Texas (2017)
4. B.O. Almroth, P. Stern, F.A. Brogan, Automatic choice of global shape functions in structural analysis. AIAA J. **16**(5), 525–528 (1978)
5. A. Ammar, B. Mokdad, F. Chinesta, R. Keunings, A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. J. Non Newton. Fluid Mech. **139**(3), 153–176 (2006)
6. D. Amsallem, J. Cortial, C. Farhat, Towards real-time computational-fluid-dynamics-based aeroelastic computations using a database of reduced-order information. AIAA J. **48**(9), 2029–2037 (2010)

7. D. Amsallem, C. Farhat, Interpolation method for adapting reduced-order models and application to aeroelasticity. AIAA J. **46**(7), 1803–1813 (2008)
8. D. Amsallem, C. Farhat, An online method for interpolating linear parametric reduced-order models. SIAM J. Sci. Comput. **33**(5), 2169–2198 (2011)
9. D. Amsallem, R. Tezaur, C. Farhat, Real-time solution of linear computational problems using databases of parametric reduced-order models with arbitrary underlying meshes. J. Comput. Phys. **326**, 373–397 (2016)
10. D. Amsallem, M. Zahr, C. Farhat, Nonlinear model order reduction based on local reduced-order bases. Int. J. Numer. Methods Eng. 1–31 (2012)
11. D. Amsallem, M. Zahr, K. Washabaugh, Fast local reduced basis updates for the efficient reduction of nonlinear systems with hyper-reduction. Adv. Comput. Math. **41**, 02 (2015)
12. P. Astrid, S. Weiland, K. Willcox, T. Backx, Missing point estimation in models described by proper orthogonal decomposition. Proc. IEEE Conf. Decis. Control. **53**(10), 1767–1772 (2005)
13. N. Aubry, P. Holmes, J.L. Lumley, E. Stone, The dynamics of coherent structures in the wall region of a turbulent boundary-layer. J. Fluid Mech. **192**, 115–173 (1988)
14. M. Barrault, Y. Maday, N.C. Nguyen, A.T. Patera, An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. Comptes Rendus Mathematiques **339**(9), 666–672 (2004)
15. M. Brand, Incremental singular value decomposition of uncertain data with missing values, in *Computer Vision — ECCV 2002*, ed. by A. Heyden, G. Sparr, M. Nielsen, P. Johansen (Springer, Berlin, Heidelberg, 2002), pp. 707–720
16. N. Cagniart, Y. Maday, B. Stamm, Model order reduction for problems with large convection effects, in *Contributions to Partial Differential Equations and Applications*, ed. by B. Chetverushkin, W. Fitzgibbon, Y. Kuznetsov, P. Neittaanmäki, J. Periaux, O. Pironneau. Computational Methods in Applied Sciences, vol. 47 (2019)
17. K. Carlberg, F. Charbel, J. Cortial, D. Amsallem, The gnat method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows. J. Comput. Phys. **242**, 623–647 (2013)
18. F. Casenave, N. Akkari, An error indicator-based adaptive reduced order model for nonlinear structural mechanics—application to high-pressure turbine blades. Math. Comput. Appl. **24**(2), (2019)
19. F. Casenave, N. Akkari, F. Bordeu, C. Rey, D. Ryckelynck, A nonintrusive distributed reduced-order modeling framework for nonlinear structural mechanics-application to elastoviscoplastic computations. Int. J. Numer. Methods Eng. **121**(1), 32–53 (2020)
20. F. Casenave, A. Ern, T. Lelièvre, A nonintrusive reduced basis method applied to aeroacoustic simulations. Adv. Comput. Math. **41**(5), 961–986 (2015)
21. F. Casenave, A. Gariah, C. Rey, F. Feyel, A nonintrusive reduced order model for nonlinear transient thermal problems with nonparametrized variability. Adv. Model. Simul. Eng. Sci. **7**(1), 1–19 (2020)
22. F. Casenave, B. Staber, X. Roynard, MMGP: a Mesh Morphing Gaussian Process-based machine learning method for regression of physical problems under non-parameterized geometrical variability (2023)
23. S. Chaturantabut, D. Sorensen, Discrete empirical interpolation for nonlinear model reduction, in *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference, CDC/CCC 2009*, pp. 4316–4321 (2010)
24. Y. Choi, G. Boncoraglio, S. Anderson, D. Amsallem, C. Farhat, Gradient-based constrained optimization using a database of linear reduced-order models. J. Comput. Phys. **423**, 109787 (2020)
25. L. Cordier, M. Bergmann, Proper Orthogonal Decomposition: an overview, in *Lecture series 2002-04, 2003-03 and 2008-01 on post-processing of experimental and numerical data*, Von Karman Institute for Fluid Dynamics (2008)
26. T. Daniel, F. Casenave, N. Akkari, A. Ketata, D. Ryckelynck, Physics-informed cluster analysis and a priori efficiency criterion for the construction of local reduced-order bases. J. Comput. Phys. **458**, 111120 (2022)

27. T. Daniel, F. Casenave, N. Akkari, D. Ryckelynck, Model order reduction assisted by deep neural networks (ROM-net). Adv. Model. Simul. Eng. Sci. **7**(16), (2020)
28. T. Daniel, F. Casenave, N. Akkari, D. Ryckelynck, Data augmentation and feature selection for automatic model recommendation in computational physics. Math. Comput. Appl. **26**(1), (2021)
29. T. Daniel, F. Casenave, N. Akkari, D. Ryckelynck, Optimal piecewise linear data compression for solutions of parametrized partial differential equations (2021)
30. T. Daniel, F. Casenave, N. Akkari, D. Ryckelynck, C. Rey, Uncertainty quantification for industrial numerical simulation using dictionaries of reduced order models. Mech. & Ind. **23**, 3 (2022)
31. C. Daversin, C. Prud'Homme, Simultaneous empirical interpolation and reduced basis method for non-linear problems. Comptes Rendus Mathematique **353**(12), 1105–1109 (2015)
32. M. Dihlmann, M. Drohmann, B. Haasdonk, Model reduction of parametrized evolution problems using the reduced basis method with adaptive time partitioning, vol. 01 (2011)
33. M. Drohmann, B. Haasdonk, M. Ohlberger, Adaptive reduced basis methods for nonlinear convection-diffusion equations, vol. 4, pp. 369–377, 12 (2010)
34. J. Eftang, A. Patera, E. Ronquist, An "hp" certified reduced basis method for parametrized elliptic partial differential equations. SIAM J. Sci. Comput. **32**, 3170–3200, 09 (2010)
35. R. Everson, L. Sirovich, Karhunen-Loève procedure for gappy data. J. Opt. Soc. Am. A **12**(8), 1657–1664 (1995)
36. C. Farhat, P. Avery, T. Chapman, 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)
37. H. Farooq, D. Ryckelynck, S. Forest et al., A pruning algorithm preserving modeling capabilities for polycrystalline data. Comput. Mech **68**, 1407–1419 (2021)
38. J. Fauque, I. Ramiere, D. Ryckelynck, Hybrid hyper-reduced modeling for contact mechanics problems. Int. J. Numer. Methods Eng. **115**(1), 117–139 (2018)
39. F. Fritzen, B. Haasdonk, D. Ryckelynck, S. Schöps, An algorithmic comparison of the hyper-reduction and the discrete empirical interpolation method for a nonlinear thermal problem. Math. Comput. Appl. **23**(1), (2018)
40. S. Grimberg, C. Farhat, R. Tezaur, C. Bou-Mosleh, Mesh sampling and weighting for the hyperreduction of nonlinear Petrov-Galerkin reduced-order models with local reduced-order bases, vol. 08 (2020)
41. B. Haasdonk, Convergence rates of the pod-greedy method. ESAIM Math. Model. Numer. Anal. **47**(3), 859–873 (2013)
42. B. Haasdonk, M. Dihlmann, M. Ohlberger, A training set and multiple bases generation approach for parametrized model reduction based on adaptive grids in parameter space. Math. Comput. Model. Dyn. Syst. **17**, 423–442, 08 (2011)
43. D. Hartman, L.K. Mestha, A deep learning framework for model reduction of dynamical systems, in *2017 IEEE Conference on Control Technology and Applications (CCTA)* (IEEE, 2017), pp. 1917–1922
44. W. He, P. Avery, C. Farhat, In-situ adaptive reduction of nonlinear multiscale structural dynamics models (2020)
45. J.A. Hernandez, M.A. Caicedo, A. Ferrer, Dimensional hyper-reduction of nonlinear finite element models via empirical cubature. Comput. Methods Appl. Mech. Eng. **313**, 687–722 (2017)
46. W. Hilth, D. Ryckelynck, C. Menet, Data pruning of tomographic data for the calibration of strain localization models. Math. Comput. Appl. **24**(1), (2019)
47. K.C. Hoang, P. Kerfriden, S.P.A. Bordas, A fast, certified and tuning free two-field reduced basis method for the metamodelling of affinely-parametrised elasticity problems. Comput. Methods Appl. Mech. Eng. **298**, 121–158 (2016)
48. M. Horák, D. Ryckelynck, S. Forest, Hyper-reduction of generalized continua. Comput. Mech **59**, 753–778 (2017)

49. A. Iollo, D. Lombardi, Advection modes by optimal mass transfer. Phys. Rev. E **89**, 022923 (2014)
50. V.R. Joseph, E. Gul, S. Ba, Maximum projection designs for computer experiments. Biometrika **102**(2), 3 (2015)
51. M.G. Kapteyn, D.J. Knezevic, K.E. Willcox, Toward predictive digital twins via component-based reduced-order models and interpretable machine learning (2020)
52. M.G. Kapteyn, K.E. Willcox, From physics-based models to predictive digital twins via interpretable machine learning (2020)
53. K. Kashima, Nonlinear model reduction by deep autoencoder of noise response data, in *2016 IEEE 55th Conference on Decision and Control (CDC)* (IEEE, 2016), pp. 5750–5755
54. S. Kaulmann, B. Haasdonk, Online greedy reduced basis construction using dictionaries (2012)
55. P. Kerfriden, J.J. Rodenas, S.P.-A. Bordas, Certification of projection-based reduced order modelling in computational homogenisation by the constitutive relation error. Int. J. Numer. Methods Eng. **97**(6), 395–422 (2014)
56. T. Kim, D.L. James, Skipping steps in deformable simulation with online model reduction. ACM Trans. Graph. **28**(5), 1–9 (2009)
57. Y. Kim, Y. Choi, D. Widemann, T. Zohdi, A fast and accurate physics-informed neural network reduced order model with shallow masked autoencoder (2020)
58. R.J. Kuether, Nonlinear modal substructuring of geometrically nonlinear finite element models. Ph.D. thesis, The University of Wisconsin-Madison, 2014
59. H. Launay, J. Besson, D. Ryckelynck, F. Willot, Hyper-reduced arc-length algorithm for stability analysis in elastoplasticity. Int. J. Solids Struct. **208–209**, 167–180 (2021)
60. H. Launay, D. Ryckelynck, L. Lacourt, J. Besson, A. Mondon, F. Willot, Deep multimodal autoencoder for crack criticality assessment. Int. J. Numer. Methods Eng. **123**(6), 1456–1480 (2022)
61. H. Launay, F. Willot, D. Ryckelynck, J. Besson, Mechanical assessment of defects in welded joints: morphological classification and data augmentation. J. Math. Ind. **11**(8), 18 (2021)
62. S. Le Berre, I. Ramière, J. Fauque, D. Ryckelynck, Condition number and clustering-based efficiency improvement of reduced-order solvers for contact problems using Lagrange multipliers. Mathematics **10**(9), (2022)
63. K. Lee, K.T. Carlberg, Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. J. Comput. Phys. **404**, 108973 (2020)
64. T. Lieu, C. Farhat, Adaptation of POD-based aeroelastic ROMs for varying Mach number and angle of attack: application to a complete F-16 configuration, in *AIAA Paper 2005-7666*, (2005)
65. T. Lieu, C. Farhat, Adaptation of aeroelastic reduced-order models and application to an F-16 configuration. AIAA J. **45**, 1244–1257 (2007)
66. T. Lieu, C. Farhat, M. Lesoinne, POD-based aeroelastic analysis of a complete F-16 configuration: ROM adaptation and demonstration, in *AIAA Paper 2005-2295* (2005)
67. T. Lieu, C. Farhat, M. Lesoinne, Reduced-order fluid/structure modeling of a complete aircraft configuration. Comput. Methods Appl. Mech. Eng. **195**, 5730–5742 (2006)
68. T. Lieu, M. Lesoinne, Parameter adaptation of reduced order models for three-dimensional flutter analysis, in *AIAA Paper 2004-0888* (2004)
69. E.N. Lorenz, Empirical orthogonal functions and statistical weather prediction. *MIT, Department of Meteorology, Scientific Report N1, Statistical Forecasting Project* (1956)
70. J.L. Lumley, The structure of inhomogeneous turbulence, in *Atmospheric Turbulence and Wave Propagation*, pp. 166–177 (1967)
71. Y. Maday, N.-C. Nguyen, A.T. Patera, S.H. Pau, A general multipurpose interpolation procedure: the magic points. Commun. Pure Appl. Anal. **8**(1), 383–404 (2009)
72. Y. Maday, B. Stamm, Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces. SIAM J. Sci. Comput. **35**(6), A2417–A2441 (2013)
73. M.P. Mignolet, A. Przekop, S.A. Rizzi, S.M. Spottswood, A review of indirect/non-intrusive reduced order modeling of nonlinear geometric structures. J. Sound Vib. **332**(10), 2437–2460 (2013)

74. B. Miled, D. Ryckelynck, S. Cantournet, A priori hyper-reduction method for coupled viscoelastic-viscoplastic composites. Comput. & Struct. **119**, 95–103 (2013)
75. R. Mosquera, A. El Hamidi, A. Hamdouni, A. Falaize, Generalization of the Neville-Aitken interpolation algorithm on Grassmann manifolds : applications to reduced order model (2019). arXiv:1907.02831
76. R. Mosquera, A. Hamdouni, A. El Hamidi, C. Allery, POD basis interpolation via inverse distance weighting on Grassmann manifolds. Discret. Contin. Dyn. Syst. S **12**, 1743–1759, 01 (2018)
77. F. Nguyen, S.M. Barhli, D.P. Muñoz, D. Ryckelynck, Computer vision with error estimation for reduced order modeling of macroscopic mechanical tests. Complexity (2018)
78. N.C. Nguyen, A.T. Patera, J. Peraire, A best points interpolation method for efficient approximation of parametrized functions. Int. J. Numer. Methods Eng. **73**, 521–543 (2008)
79. A.K. Noor, J.M. Peters, Reduced basis technique for nonlinear analysis of structures. AIAA J. **18**(4), 455–462 (1980)
80. M. Ohlberger, F. Schindler, Error control for the localized reduced basis multiscale method with adaptive on-line enrichment. SIAM J. Sci. Comput. **37**(6), A2865–A2895 (2015)
81. A.T. Patera, G. Rozza, Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations. MIT Pappalardo Graduate Monographs in Mechanical Engineering (2007)
82. B. Peherstorfer, D. Butnaru, K. Willcox, H.J. Bungartz, Localized discrete empirical interpolation method. SIAM J. Sci. Comput. **36**, 01 (2014)
83. H. Peng, F. Long, C. Ding, Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 1226–38, 09 (2005)
84. C. Prud'Homme, D.V. Rovas, K. Veroy, L. Machiels, Y. Maday, A.T. Patera, G. Turinici, Reliable real-time solution of parametrized partial differential equations: reduced-basis output bound methods. J. Fluids Eng. **124**(1), 70–80 (2001)
85. J. Reiss, P. Schulze, J. Sesterhenn, V. Mehrmann, The shifted proper orthogonal decomposition: a mode decomposition for multiple transport phenomena. SIAM J. Sci. Comput. **40**(3), A1322–A1344 (2018)
86. C. Rowley, T. Colonius, R. Murray, Model reduction for compressible flow using POD and Galerkin projection. Phys. D Nonlinear Phenom. **189**, 115–129, 01 (2003)
87. G. Rozza, D. Huynh, A. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. Arch. Comput. Methods Eng. **15**, 1–47, 09 (2007)
88. D. Ryckelynck, A priori hyperreduction method: an adaptive approach. J. Comput. Phys. **1**(202), 346–366 (2005)
89. D. Ryckelynck, Hyper reduction of mechanical models involving internal variables. Int. J. Numer. Methods Eng. (2009)
90. D. Ryckelynck, D.M. Benziane, A. Musienko, G. Cailletaud, Toward "green" mechanical simulations in materials science. Eur. J. Comput. Mech. **19**(4), 365–388 (2010)
91. D. Ryckelynck, L. Gallimard, S Jules, Estimation of the validity domain of hyper-reduction approximations in generalized standard elastoviscoplasticity. Adv. Model. Simul. Eng. Sci. (2015)
92. D. Ryckelynck, T. Goessel, F. Nguyen, Mechanical dissimilarity of defects in welded joints via Grassmann manifold and machine learning. Comptes Rendus. Mécanique **348**(10–11), 911–935 (2020)
93. D. Ryckelynck, L. Hermanns, F. Chinesta, E. Alarcón, An efficient 'a priori' model reduction for boundary element models. Eng. Anal. Bound. Elem. **29**(8), 796–801 (2005)
94. D. Ryckelynck, K. Lampoh, S. Quilicy, Hyper-reduced predictions for lifetime assessment of elasto-plastic structures. Meccanica **51**(2), 309–317 (2016)
95. D. Ryckelynck, D. Missoum Benziane, Multi-level a priori hyper-reduction of mechanical models involving internal variables. Comput. Methods Appl. Mech. Eng. **199**(17), 1134–1142 (2010)

96. D. Ryckelynck, D. Missoum Benziane, Hyper-reduction framework for model calibration in plasticity-induced fatigue. Adv. Model. Simul. Eng. Sci. **3**, (2016)
97. D. Ryckelynck, D. Missoum Benziane, S. Cartel, J. Besson, A robust adaptive model reduction method for damage simulations. Comput. Mater. Sci. **50**(5), 1597–1605 (2011)
98. D. Ryckelynck, F. Vincent, S. Cantournet, Multidimensional a priori hyper-reduction of mechanical models involving internal variables. Comput. Methods Appl. Mech. Eng. **225–228**, 28–43 (2012)
99. B. Sarbandi, S. Cartel, J. Besson et al., Truncated integration for simultaneous simulation of sintering using a separated representation. Arch. Comput. Methods Eng. **17**, 455–463
100. M. Yano, A.T. Patera, An LP empirical quadrature procedure for reduced basis treatment of parametrized nonlinear PDEs. Comput. Methods Appl. Mech. Eng. **344**, 1104–1123 (2019)
101. K. Ye, L.-H. Lim, Schubert varieties and distances between subspaces of different dimensions. SIAM J. Matrix Anal. Appl. **37**(3), 1176–1197 (2016)
102. R. Zimmermann, B. Peherstorfer, K. Willcox, Geometric subspace updates with applications to online adaptive nonlinear model reduction. SIAM J. Matrix Anal. Appl. **39**, 11 (2017)