

Chapter 3

Error Estimation



3.1 Confidence and Trust in Model-Based Engineering Assisted by AI

Consider first data-based machine learning techniques. They rely on large sets of examples provided during the training stage and do not learn with equations. Dealing with a situation that do not belong to the training set variability, namely an out-of-distribution sample, can be very challenging for these techniques. Trusting them could imply being able to guarantee that the training set covers the operational domain of the system to be trained. Besides, data-based AI can lack in robustness: examples have been given of adversarial attacks in which a classifier was tricked to infer a wrong class only by changing a very small percentage of the pixels of the input image. These models often also lack explainability: it is hard to understand what is exactly learned, what phenomenon occurs through the layers of a neural network. In some cases, information on the background of a picture is used by the network in the prediction of the class of an object, or bias present in the training data will be learned by the AI model, like gender bias in recruitment processes. Addressing these limitations is the subject of the Program Con fiance.ai,¹ regrouping French academic as well as industrial partners from defense, transports, manufacturing and energy sectors.

Model-based engineering enjoys better explainability—since the reference-model equations are known and used, and robustness—when the reference-model is well-posed. Concerning trust, in our projection-based reduced-order modeling context, the prediction is in general deterministic, and strict error bounds can be derived in particular cases. This is a main difference with AI-based models, which use notions like confidence intervals and predictive uncertainties, expressed as probability results. In

¹ <https://www.confiance.ai/>.

the remainder of this chapter, we present error bounds and indicators in projection-based reduced-order modeling, depending on the complexity of the underlying equations.

3.2 In Linear Elasticity and for Linear Problems

Parts of this section has been inspired from the authors previous works [7] and [9].

We suppose that the problem of interest has the following discrete variational form, depending on a parameter μ in a parameter set $\mathcal{P}$: for a finite-dimensional space $\mathcal{V}$ of dimension N (with $N \gg 1$ resulting, e.g., from discretization), find $u_\mu \in \mathcal{V}$ such that

$$E_\mu : a_\mu(u_\mu, v) = b(v), \quad \forall v \in \mathcal{V}, \quad (3.1)$$

where a_μ is an inf-sup stable bounded sesquilinear form on $\mathcal{V} \times \mathcal{V}$ and b is a continuous linear form on $\mathcal{V}$. We define the Riesz isomorphism J from $\mathcal{V}$ to $\mathcal{V}$ such that for all $l \in \mathcal{V}'$ and all $u \in \mathcal{V}$, $(Jl, u)_{\mathcal{V}} = l(u)$, where $(\cdot, \cdot)_{\mathcal{V}}$ denotes the inner product of $\mathcal{V}$ with associated norm $\|\cdot\|_{\mathcal{V}}$ and $\mathcal{V}'$ the dual of $\mathcal{V}$. We denote $\beta_\mu := \inf_{u \in \mathcal{V}} \sup_{v \in \mathcal{V}} \frac{|a_\mu(u, v)|}{\|u\|_{\mathcal{V}} \|v\|_{\mathcal{V}}} > 0$ the inf-sup constant of a_μ and $\tilde{\beta}_\mu$ a computable positive lower bound of β_μ . For simplicity, we consider that the linear form b is independent of the parameter μ . The extension to μ -dependent b is straightforward.

Applying the Galerkin method on the linear problem (3.1), using a ROB $(\psi_i)_{1 \leq i \leq n} \in \mathbb{R}^{n \times N}$ as done in Sect. 2.3.2 leads to finding $\hat{u}_\mu \in \mathcal{V}_n$ such that

$$\hat{E}_\mu : a_\mu(\hat{u}_\mu, u_j) = b(u_j), \quad \forall j \in \{1, \dots, n\}. \quad (3.2)$$

The approximate solution on the ROB is written as (2.16):

$$\hat{u}_\mu = \sum_{i=1}^n \gamma_i^\mu \psi_i. \quad (3.3)$$

We assume that the sesquilinear form a_μ depends on μ in an affine way, namely there exist d functions $\alpha_k(\mu) : \mathcal{P} \rightarrow \mathbb{C}$ and d μ -independent sesquilinear forms a_k bounded on $\mathcal{V} \times \mathcal{V}$ such that

$$a_\mu(u, v) = \sum_{k=1}^d \alpha_k^\mu a_k(u, v), \quad \forall u, v \in \mathcal{V}, \quad (3.4)$$

which enables the ROM to be online-efficient.

Under the current assumptions, the following error bound holds (see [16, Sect.4.3.2]): for all $\mu \in \mathcal{P}$,

$$\begin{aligned} \|u_\mu - \hat{u}_\mu\|_{\mathcal{V}} &\leq \mathcal{E}_1(\mu) := \tilde{\beta}_\mu^{-1} \|G_\mu \hat{u}_\mu\|_{\mathcal{V}}, \\ &= \tilde{\beta}_\mu^{-1} \left\| G_{00} + \sum_{i=1}^{\hat{N}} \sum_{k=1}^d \alpha_k^\mu \gamma_i^\mu G_k u_i \right\|_{\mathcal{V}}, \end{aligned} \quad (3.5)$$

with G_μ the linear map from $\mathcal{V}$ to $\mathcal{V}$ such that $\mathcal{V} \ni u \mapsto G_\mu u := J(a_\mu(u, \cdot) - b) \in \mathcal{V}$; G_μ inheriting the affine dependence of a_μ on μ since, for all $u \in \mathcal{V}$,

$$G_\mu u = -Jb + \sum_{k=1}^d \alpha_k^\mu J a_k(u, \cdot) = G_{00} + \sum_{k=1}^d \alpha_k^\mu G_k u, \quad (3.6)$$

where $G_{00} := -Jb \in \mathcal{V}$ and $G_k u := J a_k(u, \cdot) \in \mathcal{V}$ for all $k \in \{1, \dots, d\}$.

The error bound (3.5) can be rewritten in an equivalent way as

$$\begin{aligned} \mathcal{E}_2(\mu) &= \tilde{\beta}_\mu^{-1} \left((G_{00}, G_{00})_{\mathcal{V}} + 2\text{Re} \sum_{i=1}^{\hat{N}} \sum_{k=1}^d \gamma_i^\mu \alpha_k^\mu (G_k u_i, G_{00})_{\mathcal{V}} \right. \\ &\quad \left. + \sum_{i,j=1}^{\hat{N}} \sum_{k,l=1}^d \gamma_i^\mu \alpha_k^\mu \gamma_j^{*\mu} \alpha_l^{*\mu} (G_k u_i, G_l u_j)_{\mathcal{V}} \right)^{\frac{1}{2}}, \\ &= \tilde{\beta}_\mu^{-1} \left(\delta^2 + 2\text{Re}(s^t \hat{x}_\mu) + \hat{x}_\mu^{*t} S \hat{x}_\mu \right)^{\frac{1}{2}}, \end{aligned} \quad (3.7)$$

where $\delta := \|G_{00}\|_{\mathcal{V}}$, s and $\hat{x}_\mu$ are vectors in $\mathbb{C}^{dn}$ with components $s_I := (G_k u_i, G_{00})_{\mathcal{V}}$ and $(\hat{x}_\mu)_I := \alpha_k^\mu \gamma_i^\mu$, and S is a matrix in $\mathbb{C}^{dn, dn}$ with coefficients $S_{I,J} := (G_k u_i, G_l u_j)_{\mathcal{V}}$ (with I and J re-indexing respectively (k, i) and (l, j) , for all $1 \leq k, l \leq d$ and all $1 \leq i, j \leq n$). The t superscript denotes the transposition. The vector s and the matrix S depend on the ROB $(\psi_i)_{1 \leq i \leq n}$ but are independent of μ , hence can be precomputed; while the vector $\hat{x}_\mu$ depends on the reduced basis approximation $\hat{u}_\mu$ via the coefficients γ_i^μ . A lower bound $\tilde{\beta}_\mu$ of the stability constant of a_μ is also computed in complexity independent of N (which is possible, for example, by the Successive Constraint Method, see [10, 13]).

We would like to stress that $\mathcal{E}_1(\mu) = \mathcal{E}_2(\mu)$ (in infinite precision arithmetic): the indices 1 and 2 are used to denote two different ways to compute the same quantity. In particular, $\mathcal{E}_1(\mu)$ is not online efficient, while $\mathcal{E}_2(\mu)$ is.

3.3 In Nonlinear Mechanics of Materials

The remaining of this section is inspired from the authors work [8].

We look here for an efficient error indicator in the context of general nonlinearities and nonparametrized variabilities in nonlinear structural mechanics. The generality

of the assumptions and the complexity of the model lead us to search for quantities correlated to the error made by the ROM with respect to the HFM, instead of rigorous error bounds as considered in the previous section.

The problem of interest is the same as described in Sect. 2.2, and we suppose that we have constructed a ROM following the methods described in Sects. 2.3 and 2.3.5, namely used POD for data compression and ECM for operator compression.

The quantity of interest is the accumulated plastic strain over the complete structure. Since this is a dual quantity, the ROM do not provide directly a prediction over the structure, but only at the reduced quadrature points selected by ECM, see Sect. 2.3.5. The Gappy-POD can be used for to recover the dual quantity of interest over the rest of the structure, see [8, Algorithms 3 and 4] for a presentation on the present context, and [11] for in seminal paper on Gappy-POD.

A quantification for the prediction relative error of the accumulated plastic strain is defined as

$$E_\mu^p := \begin{cases} \frac{\|p_\mu - \tilde{p}_\mu\|_{L^2(\Omega)}}{\|p_\mu\|_{L^2(\Omega)}} & \text{if } \|p_\mu\|_{L^2(\Omega)} \neq 0, \\ \frac{\|p_\mu - \tilde{p}_\mu\|_{L^2(\Omega)}}{\max_{\mu \in \mathcal{P}_{\text{off}}} \|p_\mu\|_{L^2(\Omega)}} & \text{otherwise,} \end{cases} \quad (3.8)$$

where p_μ and $\tilde{p}_\mu$ are respectively the high-fidelity and reduced predictions for the accumulated plasticity field at the variability μ , and $\mathcal{P}_{\text{off}}$ is the set of variabilities encountered during the offline stage. We underline the fact that $\tilde{p}_\mu$ is the reduced prediction over the complete structure, hence after applying the Gappy-POD reconstruction.

Define the ROM-Gappy-POD residual as

$$\mathcal{E}_\mu^p := \begin{cases} \frac{\|\tilde{\mathbf{p}}_\mu - \hat{\mathbf{p}}_\mu\|_2}{\|\hat{\mathbf{p}}_\mu\|_2} & \text{if } \|\hat{\mathbf{p}}_\mu\|_2 \neq 0, \\ \frac{\|\tilde{\mathbf{p}}_\mu - \hat{\mathbf{p}}_\mu\|_2}{\max_{\mu \in \mathcal{P}_{\text{off}}} \|\hat{\mathbf{p}}_\mu\|_2} & \text{otherwise,} \end{cases} \quad (3.9)$$

where $\tilde{\mathbf{p}}_\mu$ is the reduced prediction (after applying the Gappy-POD) taken at the reduced quadrature points ($\tilde{\mathbf{p}}_{\mu,k} = \tilde{p}_\mu(\hat{x}_k)$, $1 \leq k \leq m^p$), $\hat{\mathbf{p}}_\mu$ is the vector of the accumulated plastic strain as computed by the constitutive law solver at the reduced quadrature points during the online stage, and $\|\cdot\|_2$ denotes the Euclidean norm. Notice that in the general case, $\tilde{\mathbf{p}}_\mu \neq \hat{\mathbf{p}}_\mu$: this discrepancy is at the base of our proposed error indicator.

Notice that the relative error E_μ^p involves fields and L^2 -norms whereas the ROM-Gappy-POD residual $\mathcal{E}_\mu^p$ involves vectors of dual quantities in the set of reduced integration points and Euclidean norms. In (3.9), $\|\tilde{\mathbf{p}}_\mu - \hat{\mathbf{p}}_\mu\|_2$ is the error between the online evaluation of the accumulated plastic strain by the behavior law solver: $\hat{\mathbf{p}}_\mu$, and the reconstructed prediction at the reduced integration points $\hat{x}_k$: $\tilde{\mathbf{p}}_\mu$, $1 \leq k \leq m^p$. It is explained in [8, Sect. 4.1] that $\|\tilde{\mathbf{p}}_\mu - \hat{\mathbf{p}}_\mu\|_2$ is also the residual of the least-square optimization involved in the online stage of the Gappy-POD.

Let $B \in \mathbb{R}^{m^p \times n^p}$ such that $B_{k,i} = \psi_i^p(\hat{x}_k)$, $1 \leq k \leq m^p$, $1 \leq i \leq n^p$, $K := \{p_\mu$, for all possible variabilities $\mu\}$ and $d(K, W)_{L^2(\Omega)} := \sup_{v \in K} \inf_{w \in W} \|v - w\|_{L^2(\Omega)}$, with

W a finite-dimensional subspace of $L^2(\Omega)$. The following propositions and corollary are proven in [8, Sect. 4.1].

Proposition 3.1 *There exist two positive constants C_1 and C_2 independent of μ (but dependent on n^p) such that*

$$\|p_\mu - \tilde{p}_\mu\|_{L^2(\Omega)}^2 \leq C_1 \|Bz_\mu - \hat{p}_\mu\|_2^2 + C_1 \|p_\mu - \hat{p}_\mu\|_2^2 + C_2 d(K, \text{Span}\{\psi_i^p\}_{1 \leq i \leq n^p})_{L^2(\Omega)}^2. \quad (3.10)$$

Proposition 3.2 *There exist two positive constants K_1 and K_2 independent of μ such that*

$$\|\tilde{p}_\mu - \hat{p}_\mu\|_2^2 \leq K_1 \|p_\mu - \tilde{p}_\mu\|_{L^2(\Omega)}^2 + K_2 \|p_\mu - \hat{p}_\mu\|_2^2. \quad (3.11)$$

Corollary 3.3.1 *Suppose that the reduced solution is exact up to the considered time step at the online variability μ : $p_\mu = \tilde{p}_\mu$ in $L^2(\Omega)$. In particular, the behavior law solver has been evaluated with the exact strain tensor and state variables at the integration points x_k , leading to $\hat{p}_\mu(\hat{x}_k) = p_\mu(\hat{x}_k)$, $1 \leq k \leq m^d$. From Proposition 3.2, $\|\tilde{p}_\mu - \hat{p}_\mu\|_2 = 0$, and $\mathcal{E}_\mu^p = 0$.*

We observe that in practice, the evaluations of the ROM-Gappy-POD residual $\mathcal{E}_\mu^p$ (3.9) and the error E_μ^p (3.8) are very correlated in our numerical simulations. The idea is to exploit this correlation by training a Gaussian process regressor for the function $\mathcal{E}_\mu^p \mapsto E_\mu^p$. At the end of the offline stage, we propose to compute reduced predictions at variability values $\{\mu_i\}_{1 \leq i \leq N_c}$ encountered during the data generation step, and the corresponding couples $(E_{\mu_i}^p, \mathcal{E}_{\mu_i}^p)$, $1 \leq i \leq N_c$. A Gaussian process regressor is trained on these values and we define an approximation function

$$\mathcal{E}_\mu^p \mapsto \text{Gpr}^p(\mathcal{E}_\mu^p), \quad (3.12)$$

for the error E_μ^p at variability μ as the mean plus 3 times the standard deviation of the predictive distribution at the query point $\mathcal{E}_\mu^p$: this is our proposed error indicator. If the dispersion around the learning data is small for certain values $\mathcal{E}_\mu^p$, then adding 3 times the standard deviation will not change very much the prediction, whereas for values with large dispersion of the learning data, this correction aims to provide an error indicator larger than the error. We use the GaussianProcessRegressor python class from scikit-learn [17]. Notice that although some operations in computational complexity dependent on N are carried-out, we are still in the offline stage, and they are much faster than the resolutions of the large size systems of nonlinear equations (2.8). If the offline stage is correctly carried-out and since $\mathcal{E}_\mu^p$ is highly correlated with the error, only small values for $\mathcal{E}_\mu^p$ are expected to be computed. Hence, in order to train the Gaussian process regressor correctly for larger values of the error, the reduced Newton algorithm (2.13) is solved with a large tolerance $\epsilon_{\text{Newton}}^{\text{ROM}} = 0.1$. We call these

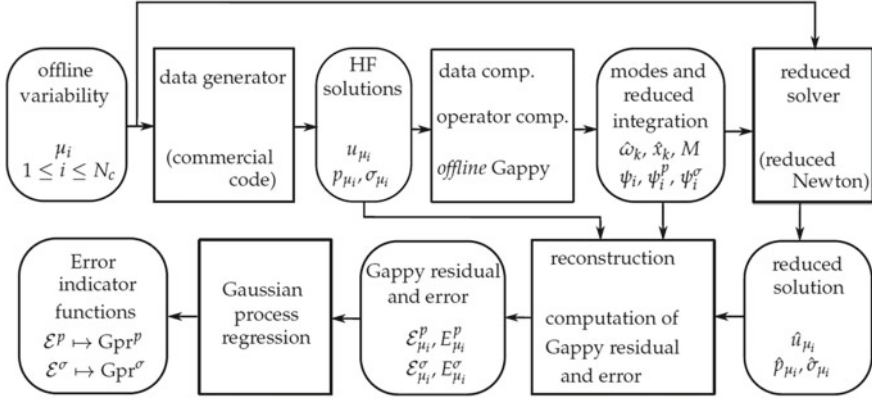


Fig. 3.1 Workflow for the offline stage with error indicator calibration [8]

operations “calibration of the error indication”, see Algorithm 3 for a description and Fig. 3.1 for a presentation of the workflow featuring this calibration step.

Algorithm 3: Calibration of the error indicator.

Input : Outputs of the data generation, data compression and operator compression steps of Section 2.3.3

Output: Approximation function $\mathcal{E}_\mu^p \mapsto \text{Gpr}^p(\mathcal{E}_\mu^p)$ of the error

- 1 Set $\mathcal{Z} = \emptyset$, $k' = 0$, $\hat{\omega} = 0$ and $r_0 = b$; // initialization
 - 2 **for** $i = 1 \dots, N_c$ **do**
 - 3 Construct and solve the reduced problem (2.13) with $\epsilon_{\text{Newton}}^{\text{ROM}} = 0.1$
 - 4 Compute the reconstructed plasticity $\tilde{p}_{\mu_i}$ using Gappy-POD and $\mathcal{E}_{\mu_i}^p$
 - 5 Compute the error $E_{\mu_i}^p$ using (3.8)
 - 6 $\mathcal{X} \leftarrow \mathcal{X} \cup (\mathcal{E}_{\mu_i}^p, E_{\mu_i}^p)$
 - 7 Construct an approximation function $\mathcal{E}_\mu^p \mapsto \text{Gpr}^p(\mathcal{E}_\mu^p)$ of the error E_μ^p using a Gaussian process regression and the data from $\mathcal{X}$
-

We recall that in model order reduction, the original hypothesis is the existence of a low-dimensional vector space where an acceptable approximation of the high-fidelity solution lies. The hypothesis is formalized under a rate of decrease for the Kolmogorov n -width with respect to the dimension of this vector space. The same hypothesis is made when using the Gappy-POD to reconstruct the dual quantities, which are expressed as a linear combination of constructed modes. For both the primal and dual quantities, the modes are computed by searching some low-rank structure of the high-fidelity data. The coefficients of the linear combination for reconstructing the primal quantities are given by the solution of the reduced Newton algorithm (2.13). After convergence, the residual is small, even in cases where the reduced order model exhibits large errors with respect to the high-fidelity reference: this residual gives no information on the distance between the reduced solution and the high-fidelity finite element space.

However, in the online phase of the ROM-Gappy-POD reconstruction (see [8, Algorithm 4]), the coefficients $\hat{p}_{\mu,k}$ (the accumulated plastic strain computed by the constitutive law solver during the online stage) contain information from the high-fidelity behavior law solver. Moreover, an overdetermined least-square is solved, which can provide a nonzero residual that implicitly contains this information from the high-fidelity behavior law solver: namely the distance between the prediction from the behavior law and the vector space spanned by the Gappy-POD modes (restricted to the reduced integration points): this is the term $\|Bz_\mu - \hat{p}_\mu\|_2$ in (3.10). Hence, the ability of the online variability to be expressed on the Gappy-POD modes is monitored through the behavior law solver on the reduced integration points. When the ROM is solved for an online variability not included in the offline variabilities, then the new physical solution cannot be correctly interpolated using the POD and Gappy-POD modes: hence, the ROM-Gappy-residual becomes large.

From Proposition 3.2, if $\|Bz_\mu - \hat{p}_\mu\|_2 = \|\tilde{p}_\mu - \hat{p}_\mu\|_2$ is large, then the global error $\|p_\mu - \tilde{p}_\mu\|_{L^2(\Omega)}$ and/or the error at the reduced integration points $\hat{x}_k$ is large, which makes $\|Bz_\mu - \hat{p}_\mu\|_2$ a good candidate for an enrichment criterion for the ROM. A limitation of the error indicator can occur if the online variability activates strong nonlinearities on areas containing no point from the reduced integration scheme, namely through the term $C_2 d(K, \text{Span}\{\psi_i^p\}_{1 \leq i \leq n^p})_{L^2(\Omega)}^2$ in (3.10).

We recall that the error indicator (3.12) is a regression of the function $\mathcal{E}_\mu^p \mapsto E_\mu^p$. In the online phase, we only need to evaluate $\mathcal{E}_\mu^p$ and do not require any estimation for the other terms and constants appearing in Propositions 3.1 and 3.2.

Equipped with an efficient error indicator, we are now able to assess the quality of the approximation made by the reduced order model in the online phase. If the error indicator is too large, an enrichment step occurs: the high-fidelity model is used to compute a new high-fidelity snapshot, which is used to update the POD and Gappy-POD basis, as well as the reduced integration schemes. Notice that for the enrichment steps to be computed, the displacement field and all the state variables of the previous time step need to be reconstructed on the complete mesh Ω to provide the high-fidelity solver with the correct material state. The workflow for the *online* stage with enrichment is presented in Fig. 3.2.

We refer to [8] for more details on this subject, and detailed numerical applications in nonlinear structural mechanics for this error indicator and its ability to enrich a ROM in the online stage.

Notice that another (noncertified) indicator in nonlinear solid mechanics with internal variables has been proposed in [1], aiming to approximate the dual norm of residuals in the same fashion as in the linear case described in Sect. 3.2. For such nonlinear case, rigorous error bounds are not obtained: a gappy-POD-based approximate representation of the stress tensor is used, and the inf-sup constant evaluation has been replaced by a normalization of the residual using the norm of the Riesz elements for the external loading.

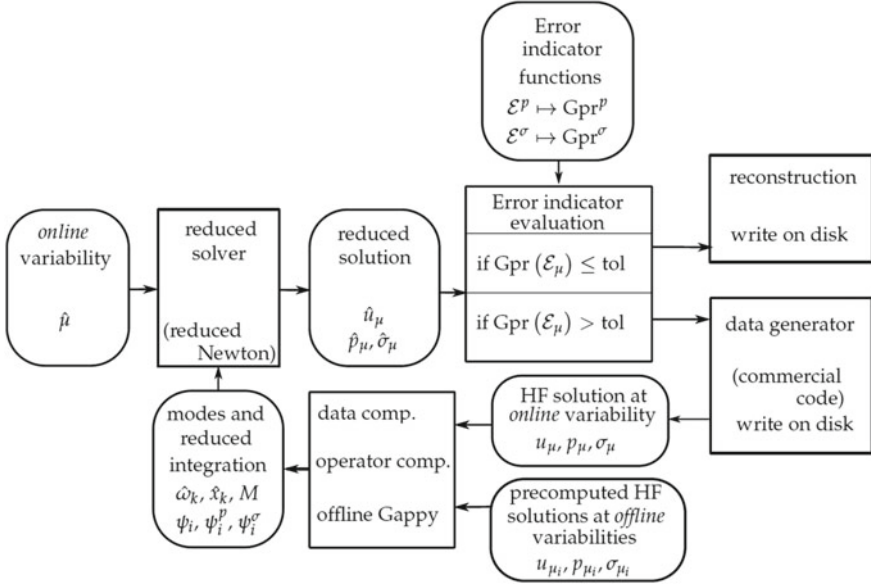


Fig. 3.2 Workflow for the online stage with enrichment [8]

3.4 In Computational Fluid Dynamics

In the following section, we present a priori error estimates due to the POD-Galerkin approximation applied to fluid dynamics equations in particular and parabolic nonlinear PDEs in general. It is a theoretical result on the convergence of the POD-Galerkin reduced order model towards the high-fidelity semi-discretized equations in the sense of the spatial variable. The solution of these semi-discretized equations is denoted $\tilde{u}^h$ over a time interval $[0, T]$ such that $h = \frac{1}{M}$ and M is the cardinal of a Hilbert basis that is capable of generating the high-fidelity semi-discretized solution at some specified M time instants. This orthonormal basis can be obtained in an a posteriori fashion thanks to the snapshots POD method. It is denoted by $(\psi_i)_{i=1, \dots, M}$.

The following convergence result includes furthermore a discussion on the stability of the Galerkin projection technique applied to parabolic PDEs. This result has been developed and published in the following papers: [2–5].

In the literature, we can find many works around the problem of defining convergent and a priori upper bounds for the POD-Galerkin reduced models for parabolic equations. It is a subject of great interest, if we can quantify efficiently the error of a reduced solution $\hat{u}$ obtained with an approximation technique of the corresponding high fidelity semi-discretized solution $\tilde{u}^h$. This problem can be seen as a theoretical confidence interval around a training point of an approximation model (details on some convergence results of the literature should be added). Let us denote by Ω the open and bounded domain of the spatial variable, such that $\Omega \in \mathbb{R}^d$, where

$d = 1$ or 2 . Let us consider the parabolic PDEs for which the weak formulation in the space of the solution u^h spanned by the POD basis of cardinal M , is described as follows:

- b is a trilinear form defined over $[H^1(\Omega)]^d \times [H^1(\Omega)]^d \times [H^1(\Omega)]^d$,
- a dissipating term defined as a bi-linear and coercive form a over $[H^1(\Omega)]^d \times [H^1(\Omega)]^d$,
- a linear form F_t defined over $[L^2(\Omega)]^d$,
- β is the coercivity constant of the bilinear form a ,
- C_a and C_b are respectively the norms of a and b in the space $[L^2(\Omega)]^d$,
- $K = \|u^h\|_{L^\infty(0,T;[L^2(\Omega)]^d)}$.

The following result is proved:

Theorem 3.1 *If $n \ll M$ is the dimension of the truncated POD-Galerkin reduced solution $\hat{u}$ that represent the training point (solution) u^h , then we can derive the following a priori upper bound of the $[L^2(\Omega)]^d$ – error between $\hat{u}(t)$ and $\tilde{u}^h(t)$, $\forall t \in [0, T]$:*

$$\|(\tilde{u}^h - \hat{u})(t)\|_{[L^2(\Omega)]^d}^2 \leq f_1(N). \quad (3.13)$$

Where $f_1(n)$ is the remainder of the sum of a convergent series; $f_1(n)$ converges to 0 when N converges to M . More precisely, $f_1(n)$ is a function of the remainder of the sum of the POD eigenvalues $(\lambda_k)_{k=1,\dots,M}$ obtained from the Snapshots POD applied to M temporal snapshots of u^h : it is expressed into two different fashions: Either

$$f_1(n) = \left(1 + 2C_b K + \frac{C_a^2}{\epsilon}\right) T \sum_{k=n+1}^M \lambda_k, \quad (3.14)$$

if $(\epsilon - 2\beta + 6C_b K) \leq 0$, for a strictly positive real number ϵ , or

$$f_1(n) = \left(T + \left(2C_b K + \frac{C_a^2}{\epsilon}\right) T \exp(T[2C_a K + K^2(1 + C_a^2)])\right) \sum_{k=n+1}^M \lambda_k, \quad (3.15)$$

if $(\epsilon - 2\beta + 6C_b K) > 0$, for all strictly positive real numbers ϵ .

The mathematical proof of Theorem 3.1 is based on the properties of the forms a and b , the application of the Young inequality and the Gronwall lemma. For the details of the proof, refer to [4].

Remark 3.1 The result of Theorem 3.1 is applicable in particular for the 1D Burgers equation and the 2D unsteady and incompressible Navier-Stokes equations, by remarking the following:

- When $d = 1$, $H^1(\Omega) \subset L^\infty(\Omega)$: there exists $C_\infty^1 \in \mathbb{R}^{+*}$ such that $\forall v \in H^1(\Omega)$, $\|v\|_{L^\infty(\Omega)} \leq C_\infty^1 \|\nabla v\|_{L^2(\Omega)}$.

- When $d = 2$ or $d = 3$, $H^1(\Omega) \subset L^4(\Omega)$: there exists $C_4^1 \in \mathbb{R}^{+*}$ such that $\forall v \in H^1(\Omega)$, $\|v\|_{L^4(\Omega)} \leq C_4^1 \|\nabla v\|_{L^2(\Omega)}$.
- When $d = 2$: there exists $C \in \mathbb{R}^{+*}$ such that:

$$\|v\|_{L^4(\Omega)} \leq C \|v\|_{L^2(\Omega)}^{1/2} \|\nabla v\|_{L^2(\Omega)}^{1/2}.$$

- If we denote by S the square matrix of dimension M defined by: $S_{i,j} = \langle \psi_i, \psi_j \rangle_{[H^1(\Omega)]^d}$, then $\forall v$ in $\mathcal{V}^h$ the solutions space spanned by the complete POD basis ψ we have the following inequality: $\|v\|_{[H^1(\Omega)]^d} \leq \sqrt{\|S\|} \|v\|_{[L^2(\Omega)]^d}$. Where $\|S\|$ is the spectral norm of the matrix S . For details, refer to [14].

Remark 3.2 A result of stability with respect to time of the POD-Galerkin reduced model for nonlinear and dissipated parabolic PDEs can be derived from result 3.1. If μ denotes only the viscosity constant in this particular case (without any loss of generality), then the POD-Galerkin reduced model is stable with respect to time when μ satisfies the following inequality:

$$\mu \geq \frac{6C_b K + \epsilon}{2\beta},$$

where ϵ is a strictly positive real number.

Remark 3.3 More particularly, in the case of linear and dissipated parabolic PDEs, the stability condition of the POD-Galerkin reduced model becomes equivalent to

$\mu \geq \frac{\epsilon}{2c_p}$. The error of the POD-Galerkin reduced model with respect to the high fidelity training point can be estimated exactly as in the Céa lemma applied for elliptic and sesquilinear PDEs: if $\mu \geq 1$ then,

$$\|(u^h - \hat{u})(t)\|_{[L^2(\Omega)]^d}^2 \leq \left(1 + \frac{C_a^2}{2\beta}\right) T \sum_{n=N+1}^M \lambda_n.$$

Based on the same methodology, we propose in what follows an a priori error estimate and a convergence result for POD-Galerkin reduced model parameter wise. In other words, we show that when the parameters change with respect to the training ones, a confidence interval is obtained around the new test solution. The width of this confidence interval converges to the truncation error of the ROM at the training parameters. This parametric convergence result is formulated as follows:

Theorem 3.2 *Let us denote by $u_{\mu_0}^h$ a training solution associated with a training parameter μ_0 . So we suppose we have only one training point for the POD-Galerkin reduced model, without any loss of generality. If ψ^{μ_0} denotes the Hilbert basis obtained from the POD applied to the high fidelity training solution and $\hat{u}_{\mu, \mu_0}$ denotes the truncated POD-Galerkin reduced solution that approaches the test point (solution) u_{μ}^h in the reduced POD space of dimension n spanned by ψ^{μ_0} , then:*

$$\left\| (u_\mu^h - \widehat{u}_{\mu, \mu_0})(t) \right\|^2 \leq f_1^{\mu_0}(n) (1 + \|\mu - \mu_0\|^{2\alpha}) + f_2^{\mu_0}(n) \|\mu - \mu_0\|^\alpha, \quad (3.16)$$

where $f_2^{\mu_0}(n)$ is the remainder of the sum of a convergent series; $f_2^{\mu_0}(n)$ converges to 0 when n converges to M . More precisely $f_2^{\mu_0}(n)$ is a function of the remainder of the sum of the orthogonal projection coefficients of the characteristic function 1_{Ω_x} such that Ω_x tends to Ω when x tends to $\partial\Omega$: it is expressed as follows:

$$f_2^{\mu_0}(n) = \left(\| (u_\mu^h - u_{\mu_0}^h)(0) \|^{-4} + \frac{B}{A} (\exp(-2At) - 1) \right)^{-1/2} C_a^2 \| \nabla u_{\mu_0}^h \|^2 \sum_{k=n+1}^M \langle 1_{\Omega_x}, \psi_k^{\mu_0} \rangle^2, \quad (3.17)$$

where A and B are strictly positive constants of which the detailed expressions are given in [2].

The proof of the above theorem is published in [2]. It is based on the properties of the two forms a and b , the application of the Young inequality and the resolution of a nonlinear ordinary differential inequality of Riccati type.

3.5 A Note on Accuracy of a Posteriori Error Bounds and Round-Off Errors

In this section, we explain why the online-efficient error bound (3.7) may be sensitive to round-off errors.

In computers, real numbers are represented by a finite number of bits, called floating-point representation. Current architectures are optimized for the IEEE 754 double-precision binary floating-point format. Let x and y be real numbers. When computing the operation $x + y$, the result returned by the computer can be different from its theoretical value. Whenever the difference is substantial, a loss of significance occurs. A well-known case of loss of significance is when x and y are almost opposite numbers. Suppose that $x = -y$. We denote by $\text{maxfl}(x + y)$ the result that the computer returns when the maximal accumulation of round-off errors occurs when computing the summation. There holds

$$|\text{maxfl}(x + y)| \approx 2\epsilon|x|, \quad (3.18)$$

where ϵ is called the machine precision. In double precision, $\epsilon = 5 \times 10^{-16}$ (see [12, Sect. 1.2]).

When implementing an algorithm, one should ensure that each step is free of such a loss of significance. In some cases, simply changing the order of the operations can prevent these situations. As an illustration, consider $x = 1$, $y = 1 + 10^{-7}$, and the operation $x^2 - 2xy + y^2$. This is a sum of terms where the first intermediate result in the sum is 14 orders larger than the result. Therefore, a loss of significance is expected. The relative error of this computation is about 8×10^{-4} . Computing $(x - y)^2$, which is the factorization of the considered operation, leads to a relative

error of about 10^{-9} . Thus, the terms of the sum are only 7 orders larger than the results, leading to a less catastrophic loss of significance. In this specific case, the remedy consists in carrying out the sum before the multiplication. In our projection-based ROM context, the evaluation of the formula $\mathcal{E}_2$ suffers from such a loss of significance, as we now explain.

We investigate the influence of round-off errors when computing the error bounds (3.5) and (3.7) for respectively $\mathcal{E}_1(\mu)$ and $\mathcal{E}_2(\mu)$. As observed in the previous paragraph, the computation of a polynomial using a factorized form is more accurate than using the developed form, in particular at points close to its roots. Here, $(\tilde{\beta}_\mu \mathcal{E}_2(\mu))^2$ is a multivariate polynomial of degree 2 in $\hat{x}_\mu$ computed in a developed form, whereas the scalar product $(G_\mu u_\mu, G_\mu u_\mu)_{\mathcal{V}}$ used in the computation of $\mathcal{E}_1(\mu)$ is not developed. The following holds (see [9, Proposition 2.2.1] for the proof)

Proposition 3.3 *Let $\mu \in \mathcal{P}$ and let $\max\text{fl}(\tilde{\beta}_\mu \mathcal{E}_k(\mu))$, $k = 1, 2$, denote the evaluation of $\tilde{\beta}_\mu \mathcal{E}_k(\mu)$ when the maximum accumulation of round-off errors occurs. There holds*

$$\begin{aligned} \max\text{fl}(\tilde{\beta}_\mu \mathcal{E}_1(\mu)) &\geq 2\delta\epsilon, \\ \max\text{fl}(\tilde{\beta}_\mu \mathcal{E}_2(\mu)) &\geq 2\delta\sqrt{\epsilon}, \end{aligned} \quad (3.19)$$

where $\delta = \|G_{00}\|_{\mathcal{V}}$ and ϵ is the machine precision.

From this proposition, we notice that the online-efficient formula $\mathcal{E}_2(\mu)$ suffers from an important loss of significance.

We present below an error estimator proposed in [9] that enjoys both accuracy and online-efficiency. Let $\sigma := 1 + 2dn + (dn)^2$. For a given $\mu \in \mathcal{P}_{\text{trial}}$ and the resulting $\hat{u}_\mu \in \text{Span}\{\psi_1, \dots, \psi_n\}$ solving the reduced problem (3.2), we define $\hat{X}(\mu) \in \mathbb{C}^\sigma$ as the vector with components $(1, \hat{x}_{\mu_I}, \hat{x}_{\mu_I}^*, \hat{x}_{\mu_I}^* \hat{x}_{\mu_J})$, where $\hat{x}_{\mu_I} = \alpha_k^\mu \gamma_i^\mu$ (we recall that γ_i^μ are the coefficients of the reduced solution in the reduced basis, see (3.3), and α_k^μ the coefficients of the affine decomposition of a_μ in (3.4)), with $1 \leq I, J \leq dn$ (with $I = i + n(k-1)$ such that $1 \leq i \leq n, 1 \leq k \leq d$, and with $J = j + n(l-1)$ such that $1 \leq j \leq n, 1 \leq l \leq d$). We can write the right-hand side of (3.7) as a linear form in $\hat{X}(\mu)$ as follows:

$$\delta^2 + 2\text{Re}(s^t \hat{x}_\mu) + \hat{x}_\mu^{*t} S \hat{x}_\mu = \sum_{p=1}^{\sigma} t_p \hat{X}_p(\mu), \quad (3.20)$$

where t_p is independent of μ (as δ, s , and S are independent of μ) and $\hat{X}_p(\mu)$ is the p -th component of $\hat{X}(\mu)$.

Consider the function of two variables $(p, \mu) \mapsto \hat{X}_p(\mu)$, for all $p \in \{1, \dots, \sigma\}$ and all $\mu \in \mathcal{P}$. We look for an approximation of this function in the form

$$\forall \mu \in \mathcal{P}, \forall p \in \{1, \dots, \sigma\}, \hat{X}_p(\mu) \approx \sum_{r=1}^{\hat{\sigma}} \lambda_r^{\hat{\sigma}}(\mu) \hat{X}_p(\mu_r), \quad (3.21)$$

for a certain parameter $\hat{\sigma} \leq \sigma$. The Empirical Interpolation Method (EIM) provides a numerical procedure to construct this approximation and to choose the interpolation points (see [6, 15]), which leads to the following formula for computing the error bound

$$\mathcal{E}_3(\mu) := \tilde{\beta}_\mu^{-1} \left(\sum_{r=1}^{\hat{\sigma}} \lambda_r^{\hat{\sigma}}(\mu) V_r \right)^{\frac{1}{2}}, \quad (3.22)$$

where $V_r = \|G_{\mu_r} \hat{u}_{\mu_r}\|_{\mathcal{V}}^2$, and where $\lambda^{\hat{\sigma}}(\mu)$ and μ_r are provided by EIM, see [9, Sect. 3.2] for all the details of this derivation. There holds (see [9, Proposition 3.2.1]):

Proposition 3.4 *The computation of the formula $\mathcal{E}_3$ is well defined, and this formula is online-efficient.*

Besides, $\mathcal{E}_3$ involving a linear combination of accurately computed scalar products (see (3.5)), it is not subject to the loss of significance encountered in $\mathcal{E}_2$.

We refer to [9] for more details on the notion of validity of a formula to compute an error bound, additional variants for accurate and efficient error bounds (including one featuring a stabilized EIM), as well as numerical illustrations for a one-dimensional linear diffusion problem and and three-dimensional acoustic scattering problem. The error bound $\mathcal{E}_3$ can be of particular interest in the following situations: (i) when the stability constant of the original problem is very small (this is the case in many practical problems), (ii) when very accurate solutions are needed, (iii) when considering a nonlinear problem (for which, in some cases, no error bound is possible until a very tight tolerance is reached, see [18]).

References

1. E. Agouzal, J-P. Argaud, M. Bergmann, G. Ferté, T. Taddei, A projection-based reduced-order model for parametric quasi-static nonlinear mechanics using an open-source industrial code (2022)
2. N. Akkari, A. Hamdouni, L. Erwan, M. Jazar, On the sensitivity of the pod technique for a parameterized quasi-nonlinear parabolic equation. *Adv. Model. Simul. Eng. Sci.* **1**, 14, 08 (2014)
3. N. Akkari, A. Hamdouni, M. Jazar, Mathematical and numerical results on the parametric sensitivity of a rom-pod of the burgers equation. *Eur. J. Comput. Mech.* **23**(1–2), 78–95 (2014)
4. N. Akkari, A. Hamdouni, M. Jazar, Mathematical and numerical results on the sensitivity of the pod approximation relative to the burgers equation. *Appl. Math. Comput.* **247**, 951–961 (2014)
5. N. Akkari, A. Hamdouni, E. Liberge, M. Jazar, A mathematical and numerical study of the sensitivity of a reduced order model by pod (rom-pod), for a 2d incompressible fluid flow. *J. Comput. Appl. Math.* **270**, 522–530 (2014), in *Fourth International Conference on Finite Element Methods in Engineering and Sciences (FEMTEC 2013)*
6. 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 Mathématiques* **339**(9), 666–672 (2004)

7. F. Casenave, Accurate a posteriori error evaluation in the reduced basis method. *Comptes Rendus Mathématique* **350**(9–10), 539–542 (2012)
8. 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)
9. F. Casenave, A. Ern, T. Lelièvre, Accurate and online-efficient evaluation of the a posteriori error bound in the reduced basis method. *ESAIM Math. Model. Numer. Anal.* **48**(1), 207–229 (2014)
10. Y. Chen, J.S. Hesthaven, Y. Maday, J. Rodríguez, Improved successive constraint method based a posteriori error estimate for reduced basis approximation of 2d Maxwell's problem. *ESAIM Math. Model. Numer. Anal.* **43**(6), 1099–1116, 8 (2009)
11. R. Everson, L. Sirovich, Karhunen-Loève procedure for gappy data. *J. Opt. Soc. Am. A* **12**(8), 1657–1664 (1995)
12. D. Goldberg, What every computer scientist should know about floating point arithmetic. *ACM Comput. Surv.* **23**(1), 5–48 (1991)
13. D.B.P. Huynh, G. Rozza, S. Sen, A.T. Patera, A successive constraint linear optimization method for lower bounds of parametric coercivity and inf-sup stability constants. *Comptes Rendus Mathématique* **345**(8), 473–478 (2007)
14. K. Kunisch, S. Volkwein, Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische mathematik* **90**(1), 117–148 (2001)
15. 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)
16. 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)
17. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
18. M. Yano, A space-time Petrov-Galerkin certified reduced basis method: application to the Boussinesq equations. *SIAM J. Sci. Comput.* **36**(1), A232–A266 (2014)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

