Model order reduction for large-scale structures with local nonlinearities

Zhenying Zhang*, Mengwu Guo, Jan S. Hesthaven

Chair of Computational Mathematics and Simulation Science,
École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Abstract

In solid mechanics, linear structures often exhibit (local) nonlinear behavior when close to failure. For instance, the elastic deformation of a structure becomes plastic after being deformed beyond recovery. To properly assess such problems in a real-life application, we need fast and multi-query evaluations of coupled linear and nonlinear structural systems, whose approximations are not straight forward and often computationally expensive. In this work, we propose a linear-nonlinear domain decomposition, where the two systems are coupled through the solutions on a prescribed linear-nonlinear interface. After necessary sensitivity analysis, e.g. for structures with a high dimensional parameter space, we adopt a non-intrusive method, e.g. Gaussian processes regression (GPR), to solve for the solution on the interface. We then utilize different model order reduction techniques to address the linear and nonlinear problems individually. To accelerate the approximation, we employ again the non-intrusive GPR for the nonlinearity, while intrusive model order reduction methods, e.g. the conventional reduced basis (RB) method or the static-condensation reduced-basis-element (SCRBE) method, are employed for the solution in the linear subdomain. The proposed method is applicable for problems with pre-determined linear-nonlinear domain decomposition. We provide several numerical examples to demonstrate the effectiveness of our method.

Keywords: Model order reduction, Reduced basis method, nonlinear structural analysis, Gaussian process regression, machine learning

1. Introduction

Benefiting from the rapid development of computational capabilities and simulation techniques, finite element methods (FEMs) [58, 59] have received extensive recognition as a tool for high-fidelity approximation of complex systems governed by partial differential equations. Nevertheless, the need for increasing resolution in simulations remains expensive for engineering applications. Hence, various model order reduction techniques have gained substantial attention for their capability to balance accuracy and efficiency.

During the last decades, rapid and reliable model order reduction techniques, e.g. the reduced basis (RB) method [21, 40, 41, 43], the proper generalized decomposition (PGD) [9, 10], and machine learning approaches [37, 38], have been developed to treat problems governed by parametrized partial differential equations. Such methods are designed to approximate high-dimensional finite element solutions through low dimensional surrogates in a *real-time* or *multi-query* context with an accuracy comparable to the finite element solution.

In this work, we focus on the *intrusive* RB method and the *non-intrusive* Gaussian processes regression (GPR) method. As the name suggests, *intrusive* methods, e.g., projection based methods, necessitate the modification of the deterministic model construction based on the intrinsic property of the underlying nonlinearity. Though *intrusive* models are capable of producing accurate results when proper methods

*Corresponding author.
Email addresses: zhengying.zhang@epfl.ch (Zhenying Zhang), mengwu.guo@epfl.ch (Mengwu Guo), Jan.Hesthaven@epfl.ch (Jan S. Hesthaven)

are applied, they may be computationally challenging and complex engineering application due to the modification needed for each individual problem. *Non-intrusive* methods, on the other hand, alleviate these complexities and are more generic for various types of nonlinearities. The RB method, carried out in an *offline-online* framework, is a widely acknowledged technique for model order reduction. One of its many merits is the low computational cost, which permits rapid numerical evaluation, where *a posteriori* error estimation provides quality control with respect to the original high fidelity solution. In the *offline* stage, a reduced space is constructed by the span of certain snapshots (finite element solutions at chosen parameters). The construction of this space is typically conducted through the Greedy algorithm [21, 40] or the proper orthogonal decomposition (POD) [33, 41]. The Greedy algorithm utilizes the RB intrinsic rigorous error estimator as the criteria to select a subset of solutions across the training parameters as the basis functions, while the POD approach adopts the singular value decomposition (SVD) to accommodate a large number of snapshots and truncate them, according to their singular values, to the desired amount of basis functions. Lastly, a (Petrov-)Galerkin projection is employed to reduce the affinely decomposable system and complete the *offline* stage.

In the *online* stage, the coefficients of the reduced basis functions are obtained by assembling and solving the reduced system at new parameter values. The RB method ensures that the size of the *online* problem is independent of the dimension of the original finite element system, thereby achieving major computational savings. For nonlinear and non-affine problems, this procedure is nonetheless not straightforward. To decouple and reduce such systems efficiently, the empirical interpolation method (EIM) [3] and its discrete derivative [7] have been proposed to restore the affine property of the underlying system. However, these methods are of an *intrusive* nature, which require the revision of existing codes depending on the nonlinearities, and are often less practical for complex problems. Hence, we focus on *non-intrusive* methods to treat the nonlinearities in this work.

Aside from lack of flexibility to efficiently treat nonlinear and non-affine problems, the traditional RB method is also restricted to a relatively small number of parameters as the *offline* cost, associated with a rich training set, increases drastically with the increase of parameter dimensions. To cope with such issues and to encourage industrial application, the Reduced-Basis-Element Method (RBEM) was first introduced in [35], and subsequently applied in [34, 36]. The Static-Condensation Reduced-Basis-Element (SCRBE) method [26, 27, 54] was developed to establish a component-based synthesis and model order reduction through a static condensation framework. The SCRBE method comprises the static condensation (SC) [57] method with the component mode synthesis (CMS) [11, 24, 25] method and the conventional RB method to reduce the number of parameters in each component and facilitate the efficient dimension reduction of component interiors and interfaces. Later, this method was extended to eliminate inactive modes on the ports [14–16] to ensure further model order reduction.

However, the SCRBE framework does not extend to nonlinear simulations. The intrinsic nature of the SC method, based on the Schur complement decomposition to eliminate the degree of freedom, confines this approach to linear problems. Nonetheless, the SCRBE method can still be implemented in systems with only local nonlinearities [2]. In this work, numerical examples of large-scale structures with local nonlinearities are considered in the context of component-based synthesis. However, we would like to point out that the method proposed in this work can be applied to any large-scale structure with local nonlinearity given a predetermined linear-nonlinear interface. In the framework of [2], a linear-nonlinear domain decomposition is assumed, prior to the system reformulation. A SCRBE approximation is then considered over the sublinear domain while a full finite element simulation is carried out in the nonlinear subdomain. The resulting hybrid linear-nonlinear formulation is coalesced through a constraint matrix to secure the consistency of solutions on the linear-nonlinear interface.

Driven by the rise of machine learning, *non-intrusive* methods have gained substantial attention. The GPR [18, 19] and the artificial neural networks (ANN) [22, 55] have been successfully applied to nonlinear problems for model order reduction, and reduced model error evaluation in [12]. A Gaussian process measures the similarity between sample points, using kernel functions, to predict the output values for new data points [42, 56]. Analogous to interpolation methods, the GPR is a regression based approach which maps the system parameters to the projection coefficients, and thus constructs the reduced solution from these coefficients and the chosen basis functions. This approach exploits the data-driven nature of machine learning techniques

and thoroughly decouples the *offline-online* stage through the input-output regression process.

In this work, we exploit both the speedup brought by the SCRBE method and the fully decoupled input-output mapping of the GPR for large structures with local nonlinearities. While we adopt the assumption of a pre-divided linear-nonlinear domain as in [2], we first solve for the solution at the linear-nonlinear interface, assisted by sensitivity analysis [45] where we analyze the uncertainty of the solution at the interface under the influences of the uncertainty in the parameters. Such analysis has been carried out by means of different sensitivity indices in the field of uncertainty quantification for decades. The variance based global sensitivity indices (Sobol' indices) [44, 51] and the derivative based global sensitivity measures (DGSM) [6, 28, 31] are widely used. The variance based sensitivity analysis has been adopted in [23] in the context of the RB method. In our work, we pursue the same analysis through derivative based measurements to reduce the number of parameters and obtain the solution on the interface. Equipped with this solution, the remaining system naturally separates into two parts: a linear problem that can be approximated efficiently by the *intrusive* SCRBE method and a nonlinear problem that can be treated using the *non-intrusive* GPR.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of the SCRBE methodology as well as the GPR method and the derivative based sensitivity analysis. Then our GPR-SCRBE approach is illustrated and the procedure is consequently specified in Section 3. In Section 4, we verify our approach on three examples from solid mechanics. Finally, we conclude our work in Section 5.

For clarity of the notation, italic symbols are adopted for functions, functionals and bilinear forms, such as the displacement u, linear functional f and bilinear form a; italic bold symbols are adopted for vectors and matrices in linear algebra, such as the coefficients of the displacement $\boldsymbol{u}$, the matrix of finite element basis functions $\boldsymbol{\mathcal{V}}_h$.

2. Preliminaries

In preparation for the introduction of the GPR-SCRBE method, we present the reduced basis method and its static condensation derivative, namely the SCRBE approximation, as well as the Gaussian processes regression and the derivative based sensitivity analysis in this section.

2.1. The RB method

The RB method is a well accepted model order reduction techniques. It was first applied to time independent elliptic problems, and subsequently extended to other classes of partial differential equations. The RB approximation is built upon a high fidelity finite element model of dimension $\mathcal{N}_h$, determined by the underlying mesh and the order of the polynomials used to approximate the solution. The method first constructs a parameter independent reduced basis space spanned either by a set of snapshots, defined by high fidelity finite element solutions at chosen parameter values, chosen through the Greedy algorithm [40, 43], or by the dominant modes of the snapshots obtained through a POD procedure [21, 41]. The former necessitates either an error estimator (the weak Greedy algorithm) or the true error computation at each training sample. The latter utilizes the singular value decomposition to accommodate the most significant modes in the resulting basis. For general nonlinear problems the effective and rigorous *a posteriori* error estimator is often not available. Consequently, the POD algorithm is adopted in this work to construct the reduced basis space for the nonlinear problem.

Let $\Omega \subset \mathbb{R}^d, d = 1, 2, 3$ be a bounded Lipschitz domain, and $\mathcal{D} \subset \mathbb{R}^p$ be a prescribed p-dimensional, compact parameter set. We consider a Hilbert space $\mathcal{V}$ with inner product $(\cdot, \cdot)_{\mathcal{V}}$ and associated norm $\|\cdot\|_{\mathcal{V}} = \sqrt{(\cdot, \cdot)_{\mathcal{V}}}$. The corresponding dual space is denoted by $\mathcal{V}'$. We consider a parametrized continuous, coercive bilinear form $a(\cdot, \cdot; \mu) : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ and a parametrized bounded linear functional $f(\cdot; \mu) \in \mathcal{V}'$. A typical problem looks for a solution $u(\mu) \in \mathcal{V}$ such that

$$a(u, v; \mu) = f(v; \mu), \quad \forall v \in \mathcal{V}. \tag{1}$$

We now introduce the finite element high fidelity solution $u_h(\mu) \in \mathcal{V}_h$ and its degree of freedom $\mathcal{N}_h$. The solution manifold can be expressed as $\mathcal{M}_h = \{u_h(\mu) : \mu \in \mathcal{D}\}$, and we select a set of N_s snapshots $\mathcal{M}_\theta = \{u_h(\mu_1), u_h(\mu_2), \cdots, u_h(\mu_{N_s})\}$. To represent those snapshots in a low rank space, a POD is employed

to extract N modes, associated with the biggest N singular values. The resulting orthonormal reduced space can be expressed as

$$\mathcal{V}_N = \text{span}\{\psi_1, \psi_2, \cdots, \psi_N\} \subset \mathcal{V}_h. \tag{2}$$

The Schmidt-Eckart-Young theorem [13, 41, 46] shows that $\mathcal{V}_N$ is the L^2-optimal basis of size N within the space spanned by $\mathcal{M}_\theta$. The approximation error with respect to the space $\mathcal{M}_\theta$ is bounded by the left out singular values $\sum_{i=N+1}^{N_s} \sigma_i^2$.

The solution in the reduced space $\mathcal{V}_N$ can be expressed as $u_N(\mu) = \sum_{i=1}^{N} \psi_i u_{Ni}$, so that (1) can be reinterpreted as a system of size N in the reduced space:

$$a(u_N, v_N; \mu) = f(v_N; \mu), \quad \forall v_N \in \mathcal{V}_N. \tag{3}$$

This procedure is a Galerkin projection of the finite element space onto the reduced space. For the optimality of this method and its associated *a priori* and *a posteriori* error estimation, we refer the reader to [21, 40, 41].

2.2. The SCRBE method

The SCRBE method has been successfully developed in [14, 16, 17, 26]. To briefly introduce this method, we define a component library consisting of n_{arch} archetype components and their associated physical domains $\hat{\Omega}_i$ and parameters $\hat{\mu}_i \in \hat{\mathcal{D}}_i \subset \mathbb{R}^{\hat{p}_i}, i = 1, \cdots, n_{\text{arch}}$. The boundary $\partial\hat{\Omega}_i$ of each archetype component is composed of n_i^γ disjoint local ports $\hat{\gamma}_{i,j}, j = 1, \cdots, n_i^\gamma$. Next, we introduce n_{inst} physical components with associated domain $\Omega_k = \mathcal{T}_k(\hat{\Omega}_{\pi(k)})$ and ports $\gamma_{k,j} = \mathcal{T}_k(\hat{\gamma}_{\pi(k,j)})$, $k = 1, \cdots, n_{\text{inst}}$, instantiated from the archetype library, where $\pi(k)$ maps the kth instantiation to its archetype component $\pi(k)$ in the library, $\pi(k, j)$ maps the jth port of instantiation k to the local port on its archetype component, and $\mathcal{T}_k : \hat{\gamma}_{\pi(k)} \to \Omega_k$ is a parametrized geometric mapping. We note that the kth instantiated component may connect to no more than $n_{\pi(k)}^\gamma$ other instantiations.

All n_{inst} instantiated components are then connected together to form the physical system $\Omega = \cup_{k=1}^{n_{\text{inst}}} \bar{\Omega}_k$. The resulting system parameter becomes $\mu = (\mu_1, \cdots, \mu_k) \in \mathcal{D} \subseteq \oplus_{k=1}^{n_{\text{inst}}} \hat{\mathcal{D}}_{\pi(k)}$. The concatenation of the local ports results in the reordering of all n_{glo}^γ global ports of the physical system. We note that the Dirichlet boundaries are exempted from the n_{glo}^γ ports. We define the connectivity of two local ports $\gamma_{k,j}$ and $\gamma_{k',j'}$ at the global port γ_l as $\rho_l = \{(k, j), (k', j')\}, l = 1, \cdots, n_{\text{glo}}^\gamma$. For ports on the global boundary, we have $\rho_l = \{(k, j)\}$ where $\gamma_{k,j}$ is the corresponding local port. We further define the port map $\pi_k(j) = l$ that maps a local port index j of a instantiation k to its global counterpart l.

We require conforming port spaces and denote the finite element dimension of global port l as $\mathcal{N}_l^\gamma = \mathcal{N}_{k,j}^\gamma = \mathcal{N}_{k',j'}^\gamma$ for all $\rho_l = \{(k, j), (k', j')\}$ or $\rho_l = \{(k, j)\}$, $l = 1, \cdots, n_{\text{glo}}^\gamma$. The total degrees of freedom on the global ports are $\mathcal{N}^\gamma = \sum_{l=1}^{n_{\text{glo}}^\gamma} \mathcal{N}_l^\gamma$. We further define a finite element space $\mathcal{V}_{h,\pi(k)}$, $k = 1, \cdots, n_{\text{inst}}$, of dimension $\mathcal{N}_{h,\pi(k)}$ on each instantiation, so that the elliptic problem (1) can be reformulated as finding $u_h(\mu) \in \mathcal{V}_h = \oplus_{k=1}^{n_{\text{inst}}} \mathcal{V}_{h,\pi(k)}$ such that

$$a(u_h(\mu), v; \mu) = \sum_{k=1}^{n_{\text{inst}}} a_{\pi(k)}(u_h(\mu)|_{\Omega_k}, v|_{\Omega_k}; \mu_k) = f(v; \mu) = \sum_{k=1}^{n_{\text{inst}}} f_{\pi(k)}(v|_{\Omega_k}; \mu_k), \quad \forall v \in \mathcal{V}_h. \tag{4}$$

2.2.1. Static condensation and reduced order approximation

The static condensation (SC) method eliminates degrees of freedom in the interior of each component or equally the bubble spaces $\mathcal{V}_{h,i}^0 = \{v \in \mathcal{V}_{h,i} : v|_{\hat{\gamma}_{i,j}} = 0, j = 1, \cdots, n_i^\gamma\} = \text{span}\{\phi_{i,1}, \cdots, \phi_{i,\mathcal{N}_i^0}\}$, where $\mathcal{N}_i^0$ is the interior degrees of freedom of archetype component i. The SC method expresses these as the degrees of freedom in the active terms that interact with other components, scilicet the part of the solution on the ports that lie in the port spaces $\mathcal{V}_{h,i,j}^\gamma = \text{span}\{\zeta_{i,j,1}, \cdots, \zeta_{i,j,\mathcal{N}_{i,j}^\gamma}\}$ for $i = 1, \cdots, n_{\text{arch}}$ and $j = 1, \cdots, n_i^\gamma$. Consequently, there are two elements of the model order reduction: i) the bubble reduction where we replace the finite element space inside each instantiation with a reduced basis space and ii) the port reduction which retains the first few dominant port modes.

We start with the introduction of the reduced bubble space, comprising basis functions obtained in the original bubble space $\mathcal{V}_h^0 = \oplus_{k=1}^{n_{\text{inst}}} \mathcal{V}_{h,\pi(k)}^0$ and the reduced port space constructed by truncated port modes in the original port space $\mathcal{V}_h^\gamma = \oplus_{k=1}^{n_{\text{inst}}} \oplus_{j=1}^{n_{\pi(k)}^\gamma} \mathcal{V}_{h,\pi(k),j}^\gamma$:

$$\begin{aligned}
\mathcal{V}_N^0 &= \oplus_{k=1}^{n_{\text{inst}}} \mathcal{V}_{N_{\pi(k)}}^0 = \oplus_{k=1}^{n_{\text{inst}}} \text{span}\{\psi_{\pi(k),1}, \cdots, \psi_{\pi(k),N_{\pi(k)}}\} \subset \mathcal{V}_h^0, \\
\mathcal{V}_M^\gamma &= \oplus_{k=1}^{n_{\text{inst}}} \oplus_{j=1}^{n_{\pi(k)}^\gamma} \mathcal{V}_{M_{\pi(k),j}}^\gamma = \oplus_{k=1}^{n_{\text{inst}}} \oplus_{j=1}^{n_{\pi(k)}^\gamma} \text{span}\{\chi_{\pi(k),j,1}, \cdots, \chi_{\pi(k),j,M_{\pi(k),j}}\} \subset \mathcal{V}_h^\gamma,
\end{aligned} \tag{5}$$

where $N = \sum_{k=1}^{n_{\text{inst}}} N_{\pi(k)}$ and $M = \sum_{k=1}^{n_{\text{inst}}} \sum_{j=1}^{n_{\pi(k)}^\gamma} M_{\pi(k),j}$, and $N_{\pi(k)}$ and $M_{\pi(k),j}$ indicate the number of reduced bubble functions associated with each instantiation and the number of reduced port functions associated with each local port, respectively.

We approximate the finite element solution $u_h(\mu) \in \mathcal{V}_h$ by $u_{N,M}(\mu) \in \mathcal{V}_{N,M} = \mathcal{V}_N^0 \oplus \mathcal{V}_M^\gamma$ consisting of two separate reduced spaces:

$$u_{N,M}(\mu) = u_{N,M}^0(\mu) + u_{N,M}^\gamma(\mu), \tag{6}$$

where $u_{N,M}^0(\mu) \in \mathcal{V}_N^0$ and $u_{N,M,k}^\gamma(\mu) \in \mathcal{V}_M^\gamma$. We note that the bubble solution $u_{N,M}^0(\mu) = \sum_{k=1}^{n_{\text{inst}}} u_{N,M,k}^0(\mu)$ can be recovered independently on each instantiation k by solving

$$a(u_{N,M,k}^0(\mu), v; \mu) = f(v; \mu), \quad \forall v \in \mathcal{V}_{N_{\pi(k)}}^0, \tag{7}$$

and $u_{N,M}^\gamma(\mu)$ is the solution of

$$a(u_{N,M}^\gamma(\mu), v; \mu) = f(v; \mu) - a(u_{N,M}^0(\mu), v; \mu), \quad \forall v \in \mathcal{V}_M^\gamma. \tag{8}$$

We note that $N_{\pi(k)} \ll \mathcal{N}_{\pi(k)}^0$ and the bubble solution $u_{N,M}^0(\mu) = \sum_{k=1}^{n_{\text{inst}}} u_{N,M,k}^0(\mu)$ can be recovered independently on each instantiation k through a Galerkin projection. The resulting equation to be solved constitutes a system of size $M \ll \mathcal{N}^\gamma$, thus saving a significant amount of computational effort. For the training of the two types of reduced spaces and more details on the model construction and model properties, we refer the reader to [14–17, 26, 49].

2.2.2. The hybrid-SCRBE method

We reiterate that we assume a prescribed linear-nonlinear domain decomposition, and in this section and onwards, we shall affix the subscripts "LIN" and "NLIN" to specify quantities that pertain to the linear and nonlinear subdomains, respectively. We define the linear subdomain where only linear operators act upon as $\Omega_{\text{LIN}}(\mu)$ and the nonlinear subdomain where the nonlinear operators are defined over as $\Omega_{\text{NLIN}}(\mu)$ such that

$$\bar{\Omega}_{\text{LIN}}(\mu) \cup \bar{\Omega}_{\text{NLIN}}(\mu) = \bar{\Omega}(\mu), \quad \Omega_{\text{LIN}}(\mu) \cap \Omega_{\text{NLIN}}(\mu) = \emptyset, \tag{9}$$

and the interface between them

$$\Gamma(\mu) = \bar{\Omega}_{\text{LIN}}(\mu) \cap \bar{\Omega}_{\text{NLIN}}(\mu). \tag{10}$$

We briefly summarize the method proposed in [2]. We first introduce the corresponding finite element spaces $\mathcal{V}_h^{\text{LIN}}(\mu) = \{v \in (H^1(\Omega_{\text{LIN}}(\mu)))^3 \,|\, v|_{\partial\Omega_{\text{LIN,D}}} = 0\}$ over $\Omega_{\text{LIN}}(\mu)$ and the space $\mathcal{V}_h^{\text{NLIN}}(\mu) = \{v \in (H^1(\Omega_{\text{NLIN}}(\mu)))^3 \,|\, v|_{\partial\Omega_{\text{NLIN,D}}} = 0\}$ over $\Omega_{\text{NLIN}}(\mu)$, where $\partial\Omega_{\text{LIN,D}}(\mu)$ and $\partial\Omega_{\text{NLIN,D}}(\mu)$ are Dirichlet boundary conditions on $\Omega_{\text{LIN}}(\mu)$ and $\Omega_{\text{NLIN}}(\mu)$, respectively. We have $\mathcal{V}_h = \{v \in (H^1(\Omega(\mu)))^3 \,|\, v_{\text{LIN}} \in \mathcal{V}_h^{\text{LIN}}(\mu), v_{\text{NLIN}} \in \mathcal{V}_h^{\text{NLIN}}(\mu)\}$. The operators can readily be defined as $a(\cdot,\cdot;\mu) : \mathcal{V}_h^{\text{LIN}} \times \mathcal{V}_h^{\text{LIN}} \to \mathbb{R}$, $b(\cdot,\cdot;\mu) : \mathcal{V}_h^{\text{NLIN}} \times \mathcal{V}_h^{\text{NLIN}} \to \mathbb{R}$, and $f(\cdot;\mu) : \mathcal{V}_h^{\text{LIN}} \to \mathbb{R}$. The elliptic problem (1) becomes: find $u_h(\mu) \in \mathcal{V}_h(\mu)$ such that

$$a(u_h^{\text{LIN}}(\mu), v^{\text{LIN}}; \mu) + b(u_h^{\text{NLIN}}(\mu), v^{\text{NLIN}}; \mu) = f(v^{\text{LIN}}; \mu), \quad \forall v \in \mathcal{V}_h(\mu). \tag{11}$$

We point out that the continuity condition on the linear-nonlinear interface Γ is weakly incorporated into (11) through the test function v which does not vanish on Γ.

In [2], this system is split into two parts: the linear model which approximates the solution $u_{N,M,h}^{\text{LIN}}(\mu)$

on $\Omega_{\mathrm{LIN}}(\mu)$ and can be estimated in the reduced space $\mathcal{V}_{N,M,h}^{\mathrm{LIN}} = \mathcal{V}_{N,M}$ by the SCRBE method. The nonlinear model solves for $u_{N,M,h}^{\mathrm{NLIN}}(\mu)$ on $\Omega_{\mathrm{NLIN}}(\mu)$ and can be treated by the FEM in its original dimension in $\mathcal{V}_{N,M,h}^{\mathrm{NLIN}} = \mathcal{V}_h^{\mathrm{NLIN}}$. These two models are coupled through the part of the shared solution on $\Gamma(\mu)$. To ensure the consistency of the solution on $\Gamma(\mu)$, the constraint $u_{N,M,h}^{\Gamma}(\mu) = u_{N,M,h}^{\mathrm{LIN}}(\mu)|_{\Gamma} = u_{N,M,h}^{\mathrm{NLIN}}(\mu)|_{\Gamma}$ is imposed, equally expressed in vector form $\boldsymbol{u}_{N,M,h}^{\Gamma_L}(\mu) = (\boldsymbol{\mathcal{V}}_L^{\Gamma})^T \boldsymbol{u}_{N,M,h}^{\Gamma_h}(\mu)$, where $\boldsymbol{u}_{N,M,h}^{\Gamma_L}(\mu)$ indicates the L coefficients of the port reduced functions on $\Gamma(\mu)$, $\boldsymbol{\mathcal{V}}_L^{\Gamma}$ is the matrix of the reduced port space on $\Gamma(\mu)$, comprising L basis vectors (with respect to the FE basis function), and $\boldsymbol{u}_{N,M,h}^{\Gamma_h}(\mu)$ indicates the coefficients of $u_{N,M,h}^{\Gamma}(\mu)$ in the FE basis. Hence, the solution on Γ can be expressed either as a vector of reduced port function coefficients or interpreted in terms of the finite element basis coefficients on the linear-nonlinear interface. The hybrid solution vector $\boldsymbol{u}_{N,M,h}(\mu)$ can be constructed as

$$\boldsymbol{u}_{N,M,h}(\mu) = \begin{bmatrix} \boldsymbol{u}_{N,M,h}^{\mathrm{NLIN}-\Gamma}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\mathrm{LIN}}(\mu) \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_{N,M,h}^{\mathrm{NLIN}-\Gamma}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\Gamma_L}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\mathrm{LIN}-\Gamma}(\mu) \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_{N,M,h}^{\mathrm{NLIN}-\Gamma}(\mu) \\ (\boldsymbol{\mathcal{V}}_L^{\Gamma})^T \boldsymbol{u}_{N,M,h}^{\Gamma_h}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\mathrm{LIN}-\Gamma}(\mu) \end{bmatrix}, \tag{12}$$

where $\boldsymbol{u}_{N,M,h}^{\mathrm{NLIN}-\Gamma}(\mu)$ and $\boldsymbol{u}_{N,M,h}^{\mathrm{LIN}-\Gamma}(\mu)$ represent $\boldsymbol{u}_{N,M,h}^{\mathrm{NLIN}}(\mu)$ and $\boldsymbol{u}_{N,M,h}^{\mathrm{LIN}}(\mu)$ with the part of coefficients on $\Gamma(\mu)$ removed, respectively. This leads to a constraint matrix $\boldsymbol{K}$ that facilitates the prolongation of the reduced subsystem and the FE subsystem such that

$$\boldsymbol{K}\boldsymbol{u}_{N,M,h}(\mu) = \boldsymbol{K} \begin{bmatrix} \boldsymbol{u}_{N,M,h}^{\mathrm{NLIN}-\Gamma}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\Gamma_L}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\mathrm{LIN}-\Gamma}(\mu) \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_{N,M,h}^{\mathrm{NLIN}-\Gamma}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\Gamma_h}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\Gamma_L}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\mathrm{LIN}-\Gamma}(\mu) \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_{N,M,h}^{\mathrm{NLIN}}(\mu) \\ \boldsymbol{u}_{N,M,h}^{\mathrm{LIN}}(\mu) \end{bmatrix}. \tag{13}$$

Thus, this coupled solution $u_{N,M,h}(\mu)$ is decomposed into two parts $u_{N,M,h}^{\mathrm{NLIN}}(\mu)$ and $u_{N,M,h}^{\mathrm{LIN}}(\mu)$ that express the solution over the linear and nonlinear subdomains, respectively. The resulting system consists of non-invasive blocks of the residual vector that represent the linear and nonlinear segments

$$\boldsymbol{K}^T \boldsymbol{R}^+(\boldsymbol{K}\boldsymbol{u}_{N,M,h}(\mu); \mu) = 0, \tag{14}$$

where $\boldsymbol{R}^+(\cdot;\mu)$ is the residual vector over the nonlinear subdomain with respect to $b(u_{N,M,h}^{\mathrm{NLIN}}(\mu), v_h^{\mathrm{NLIN}}; \mu)$ for all $v_h^{\mathrm{NLIN}} \in \mathcal{V}_{N,M,h}^{\mathrm{NLIN}}$ and over the linear subdomain with regard to $a(u_{N,M,h}^{\mathrm{LIN}}(\mu), v_{N,M,h}^{\mathrm{LIN}}; \mu) - f(v_{N,M,h}^{\mathrm{LIN}}; \mu)$ for all $v_{N,M,h}^{\mathrm{LIN}} \in \mathcal{V}_{N,M,h}^{\mathrm{LIN}}$, respectively. This formulation also leads to a non-intertwined Jacobian matrix, which can be solved by iterative methods, e.g. the Newton-Raphson method. For more information on this hybrid-SCRBE approach, we refer the reader to [2].

2.3. The GPR

A Gaussian process (GP) can be interpreted as a distribution over functions, comprising a collection of random variables, every finite subset of which has a multivariate normal distribution. In machine learning, a Gaussian process employs a kernel or covariance function to measure the similarity between the point of inference and the sampling points to determine the weights of regression [42]. It was shown in [39] that a fully connected single layer neural network with infinite many hidden units and Gaussian priors on weights and biases converges to a GP. Later, the convergence of infinitely wide deep neural networks to GPs was shown in [32]. In the context of machine learning and model order reduction, GPRs have been successfully applied to various problems and their error estimation in [12, 18, 19].

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i) : i = 1, 2, \cdots, M\}$ denote M observations, where $\mathbf{x}_i \in \mathbb{X} \subset \mathbb{R}^d$ are the d-dimensional inputs, with $\mathbb{X}$ being the input space, and $y_i \in \mathbb{R}$ are the corresponding outputs. A Gaussian process assumes that the input-output map follow an unknown regression function: $f : \mathbb{X} \to \mathbb{R}$, such that $y_i = f(\mathbf{x_i})$

or $y_i = f(\mathbf{x_i}) + \epsilon$ if corrupted by noise. In a GPR model, we first assume a prior on the unknown function f to be a GP, effected by noise fluctuations:

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), \kappa(\mathbf{x}, \mathbf{x}')), \quad y = f(\mathbf{x}) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_y^2), \tag{15}$$

where $m(\mathbf{x}) = \beta^T H(\mathbf{x})$ is the mean, $H(\mathbf{x})$ are the basis functions in $\mathbb{X}$, β^T are the corresponding coefficients, and $\kappa(\cdot, \cdot) : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ is the covariance function that estimates the resemblance of two inputs.

With many possible covariance functions, we briefly present one kernel that is used in this work, referred to as the automatic relevance determination-squared exponential (ARD-SE) covariance function:

$$\kappa(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{m=1}^{d}\frac{(x_m - x'_m)^2}{\sigma_m^2}\right). \tag{16}$$

This kernel takes the individual length scale for each input dimension into consideration, hence permitting a more flexible measurement.

Given M observations, a prior joint GP can be defined:

$$\mathbf{y}|\mathbf{X} \ \sim \ \mathcal{N}(m(\mathbf{x}), K_y)\,, \quad K_y = \kappa(\mathbf{X}, \mathbf{X}) + \sigma_y^2 \mathbf{I}_M, \tag{17}$$

where $\mathbf{y} = [y_1, y_2, \cdots, y_M]^{\mathrm{T}}$, $\mathbf{X} = [\ \mathbf{x}_1 \mid \mathbf{x}_2 \mid \ \cdots \ \mid \mathbf{x}_M\ ]$, and $\mathbf{I}_M$ is the M-dimensional unit matrix. To infer noise free output f^* at an unobserved point $\mathbf{x}^* \in \mathbb{X}$, the posterior distribution shall be drawn from

$$\begin{gathered} f^*|\mathbf{x}^*, \mathbf{X}, \mathbf{y} \ \sim \ \mathcal{N}(m^*(\mathbf{x}^*), K^*)\,, \\ m^*(\mathbf{x}^*) = m(\mathbf{x}^*) + \kappa(\mathbf{x}^*, \mathbf{X}) K_y^{-1}(\mathbf{y} - m(\mathbf{X}))\,, \quad K^* = \kappa(\mathbf{x}^*, \mathbf{x}^*) - \kappa(\mathbf{x}^*, \mathbf{X}) K_y^{-1} \kappa(\mathbf{X}, \mathbf{x}^*). \end{gathered} \tag{18}$$

The unknown hyperparameters $\theta = \{\sigma_f, \sigma_1, \cdots, \sigma_d, \sigma_y\}$ can be estimated by maximizing the marginal likelihood $p(\mathbf{y}|\mathbf{X}, \theta)$:

$$\begin{aligned} \theta_{\text{opt}} &= \arg\max_{\theta} \ \log p(\mathbf{y}|\mathbf{X}, \theta) \\ &= \arg\max_{\theta} \left\{ -\frac{1}{2}(\mathbf{y} - \beta^T H(\mathbf{X}))^T K_y^{-1}(\theta)(\mathbf{y} - \beta^T H(\mathbf{X})) - \frac{1}{2}\log|K_y(\theta)| - \frac{M}{2}\log(2\pi) \right\}. \end{aligned} \tag{19}$$

2.4. The DGSM

We reiterate that we deal with large-scale structures that permit high dimensional parameter spaces. However, the GPR often fails to learn a high dimensional multivariate problem, since the Euclidean length based inputs correlation becomes less informative as the input dimension increases, and the computational effort needed to learn one function grows exponentially [5, 53]. This is referred to as the curse of dimensionality [4]. In our work, instead of learning a nonlinear problem with high dimensional parameter inputs, we focus on methodologies that compress the input space while retaining parameters that bring significant uncertainty with respect to the quantity of interest. A common method to reduce the number of parameters is sensitivity analysis, which employs sensitivity indices to rank the importance of parameters.

Variance-based global sensitivity indices, e.g. Sobol' indices, necessitate a fairly large amount of model evaluations to acquire decent accuracy and convergence, and is computationally expensive for large scale engineering applications. Here, we introduce an alternative to the Sobol' indices, namely the derivative-based global sensitivity measures (DGSM), for the necessary sensitivity analysis to enable large-scale structural problems and allow high dimensional parameter spaces in a computational efficient manner. Albeit problem dependent, the computational effect for the evaluation of DGSMs is generally much lower than the corresponding cost for the Sobol' indices [30, 50].

Let l be a differentiable output function and $\theta = (\theta_1, \cdots, \theta_d)$ be the d-dimensional input defined in the d-dimensional unit hypercube. The partial derivative $\partial l / \partial \theta_i$ estimates the local variation of l with respect

to the local change of θ_i. This quantity shall be used here to construct the DGSM for $i = 1, \cdots, d$:

$$\nu_i = \int_{[0,1]^d} \left(\frac{\partial l}{\partial \theta_i}\right)^2 p(\theta) d\theta = \mathbb{E}\left[\left(\frac{\partial l}{\partial \theta_i}\right)^2\right], \tag{20}$$

where $p(\theta)$ is the probability density function. The element effect (EE) is adopted to evaluate $\partial l/\partial \theta_i$ [30, 52], expressed as a straightforward finite difference approach:

$$\frac{\partial l}{\partial \theta_i} = EE_i = \frac{l(\theta_1, \cdots, \theta_{i-1}, \theta_i + \Delta\theta_i, \theta_{i+1}, \cdots, \theta_d) - l(\theta)}{\Delta\theta_i}. \tag{21}$$

Though the estimation of DGSMs employs a Monte Carlo or Quasi Monte Carlo sampling method by averaging the evaluations of the partial derivatives, it requires fewer sampling points as compared with variance-based methods due to it's insensitiveness to the variance of the quantity of interests, where only the derivatives are employed. Hence, the computational effort of evaluating DGSMs is typically significantly lower than that of variance-based sensitivity indices [50].

We point out the connection between the DGSM indices ν_i and Sobol' indices [44, 51]

$$S_i^{tot} = \frac{\mathbb{E}_{\theta_{\sim i}}(\mathbb{V}_{\theta_i}(l|\theta_{\sim i}))}{\mathbb{V}(l)} = 1 - \frac{\mathbb{V}_{\theta_{\sim i}}(\mathbb{E}_{\theta_i}(l|\theta_{\sim i}))}{\mathbb{V}(l)}, \tag{22}$$

where $\mathbb{E}_{\theta_i}$ and $\mathbb{V}_{\theta_i}$ are the mean and variance, respectively, taken over θ_i, $\mathbb{V}(l)$ is the total variance of $l(\theta_1, \cdots, \theta_d)$, and $\theta_{\sim i}$ represents θ with ith component removed. It is shown in [29] that small DGSMs yield small total sensitivity indices such that

$$S_i^{tot} \leq \frac{C_i \nu_i}{\mathbb{V}(l)}, \tag{23}$$

where C_i is the Poincaré constant and its value depends on the probability distribution. Hence, parameters with low DGSMs are expected to have less significance on the corresponding output of interest, and they can be removed without reducing the accuracy of the global problem.

3. The GPR-SCRBE approximation

In this section, we present the GPR-SCRBE approach to tackle large scale problems in solid mechanics with local nonlinearities and separate parameter subspaces $\mathcal{D}_{\text{LIN}}$ and $\mathcal{D}_{\text{NLIN}}$, which are defined over Ω_{LIN} and Ω_{NLIN}, respectively. The parameter space can be rewritten as $\mathcal{D} = \{\mu \in \mathbb{R}^d \mid \mu^{\text{LIN}} \in \mathcal{D}_{\text{LIN}} \text{ and } \mu^{\text{NLIN}} \in \mathcal{D}_{\text{NLIN}}\}$. We note that μ^{NLIN} is considered of more importance for the treatment of the solution over the nonlinear subdomain than is μ^{LIN}. Hence, μ^{NLIN} are used as one part of the input to the nonlinear model. The contribution of μ^{LIN} can be attributed to the solution on the interface which serves as the other part of the input to the nonlinear model. We consider the problem (11) and use the approximation from the hybrid-SCRBE solver introduced in Sec.2.2.2 as the "truth". We notice that the essential step in a hybrid-SCRBE solve is the coupling of the linear estimation $u^{\text{LIN}}_{N,M,h}(\mu)$ and nonlinear solution $u^{\text{NLIN}}_{N,M,h}(\mu)$ through the linear-nonlinear interface $u^{\Gamma_L}_{N,M,h}(\mu)$, or equally $u^{\Gamma_h}_{N,M,h}(\mu)$. Since the nonlinear part resides in the high dimensional finite element space, the static condensation-based model order reduction from the hybrid-SCRBE solver can only reduce the linear subproblem efficiently. Fig. 3 shows a simplified decomposition of a domain and its associated solutions.

In order to enable a global reduction while utilizing the advanced computational acceleration provided by the SCRBE solver for linear systems simultaneously, we incorporate the sensitivity analysis with the GPR approach to decouple the physical system. Specifically, we employ the DGSMs with respect to $u^{\Gamma}_h(\mu)$ to reduce the number of parameters, retaining parameters introduce significant uncertainty on the behavior of $u^{\Gamma}_h(\mu)$, and are used to construct GPRs for its approximation. With the approximation on the linear-nonlinear interface, seeking an approximation of $u^{\text{LIN}-\Gamma}_h(\mu)$ corresponds to solving a linear SCRBE system. Taking the approximation of $u^{\Gamma}_h(\mu)$ as part of the inputs, another set of GPRs can be constructed to estimate

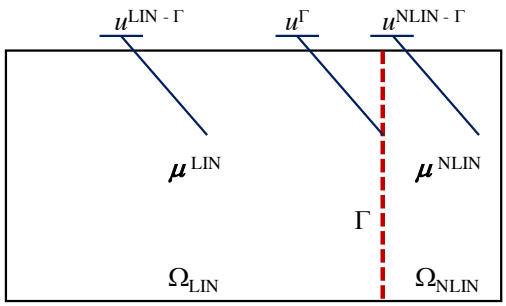


Figure 1: Decomposition of a domain into subdomains and their associated parameters and solutions.

$u_h^{\text{NLIN}-\Gamma}(\mu)$. For the sake of simplicity of the exposition, we drop the parameter dependence on Ω and Γ in this section.

For the preparation of the proposed approach and error analysis in this section, we decompose the hybrid-SCRBE solution $u_{N,M,h}(\mu)$ into the bubble reduction $u^0_{N,M,h}(\mu)$ and the port reduced solution $u^{\gamma}_{N,M,h}(\mu)$, and decompose $u^{\gamma}_{N,M,h}(\mu)$ further into $u^{\Gamma}_{N,M,h}(\mu) = u_{N,M,h}(\mu)|_{\Gamma}$ and $u^{\gamma-\Gamma}_{N,M,h}(\mu) = u_{N,M,h}(\mu)|_{\gamma-\Gamma}$, where $\gamma-\Gamma$ indicates all ports in the linear subdomain with the linear-nonlinear interface Γ removed. Similarly, the reduced space $\mathcal{V}^{\gamma}_{N,M,h}$ can be split into $\mathcal{V}^{\Gamma}_{N,M,h}$ over $\Gamma(\mu)$ and $\mathcal{V}^{\gamma-\Gamma}_{N,M,h}$ over $\Omega_{\text{LIN}}(\mu)\backslash\Gamma(\mu)$ to accommodate $u^{\Gamma}_{N,M,h}(\mu)$ and $u^{\gamma-\Gamma}_{N,M,h}(\mu)$, respectively. Assuming that $u^{\Gamma}_{N,M,h}(\mu)$ can be solved in advance as well, the linear part of the hybrid-SCRBE problem becomes

$$a(u^{\gamma-\Gamma}_{N,M,h}(\mu), v; \mu) = f(v;\mu) - a(u^0_{N,M,h}(\mu), v; \mu) - a(u^{\Gamma}_{N,M,h}(\mu), v; \mu), \quad \forall v \in \mathcal{V}^{\gamma-\Gamma}_{N,M,h}. \tag{24}$$

3.1. Methodology

We start with the special case where only one nonlinear subdomain and one linear-nonlinear interface are present, and we then generalize the method to the general setting where multiple nonlinear subdomains and interfaces coexist. We first carry out the sensitivity analysis over the interface. Since there may not be any output designed specifically for the linear-nonlinear interface, we integrate (21) and (23), and propose a modified version $\hat{\nu}_i$ as the DGSM for the ith parameter that does not require any output function:

$$\begin{aligned} \hat{\nu}_i &= \frac{\nu_i}{\sum_{i=1}^{d} \nu_i}, \\ \nu_i &= \mathbb{E}\left[\left(\frac{\|u^{\Gamma}_h(\mu_1, \cdots, \mu_{i-1}, \mu_i + \Delta\mu_i, \mu_{i+1}, \cdots, \mu_d) - u^{\Gamma}_h(\mu)\|}{\Delta\mu_i}\right)^2\right], \end{aligned} \tag{25}$$

where $\|\cdot\|$ denotes the L^2 norm. All parameters can be ranked according to their impact on $u^{\Gamma}_h(\mu)$. The first d_Γ parameters such that $\sum_{i=1}^{d_\Gamma} \hat{\nu}_i \geq r$ shall be retained. Here r is chosen by empirical judgment or engineering specification. We define this screening process as operator $S(\cdot) : \mathcal{D} \to \mathcal{D}_\Gamma$, and the d_Γ selected parameters comprise a reduced input domain $\mathcal{D}_\Gamma \subset \mathbb{R}^{d_\Gamma}$.

We then construct an orthonormal reduced basis space $\mathcal{V}^L_\Gamma = \text{span}\{\psi_1, \cdots, \psi_L\}$ from N_Γ snapshots $u^{\Gamma}_h(\mu_i), i = 1, \cdots, N_\Gamma$, by extracting the first L singular vectors. The reduced basis approximation on the interface is expressed as

$$u^{\Gamma}_L(\mu) = \sum_{i=1}^{L} u^{\Gamma}_{L,i}(\mu)\psi_i, \tag{26}$$

where $u^{\Gamma}_{L,i}$ is the individual coefficient which we can model through GPRs. For each basis coefficient i, $i = 1, \cdots, L$, we define $\pi^i_\Gamma : \mathcal{D}_\Gamma \to \mathbb{R}$ as the regression function that maps the parameters in $\mathcal{D}_\Gamma$ to the

ith coefficient of basis function ψ_i. To acquire the reduced solution on the interface, we estimate all L coefficients, which result in L Gaussian regression models.

For the ith GP model $\pi_\Gamma^i(\cdot)$, the training data set consists of n_{tr}^Γ training samples $\mathbf{x}_j^i = S(\mu_j) \in \mathcal{D}_\Gamma$, $y_j^i = \psi_i^T u_\Gamma^h(\mu_j)$, $j = 1, \cdots, n_{tr}^\Gamma$. We then define $\pi_\Gamma(\cdot) : \mathcal{D}_\Gamma \to \mathbb{R}^L$ as the collection of the L individual GPs such that

$$\boldsymbol{\pi}_\Gamma(S(\mu)) = [\pi_\Gamma^1(S(\mu)), \cdots, \pi_\Gamma^L(S(\mu))]^T. \tag{27}$$

The inferred RB approximation reads

$$\boldsymbol{u}_h^\Gamma(\mu) \approx \boldsymbol{u}_L^\Gamma(\mu) = \boldsymbol{\mathcal{V}}_\Gamma^L \boldsymbol{\pi}_\Gamma(S(\mu)) = \sum_{i=1}^{L} \pi_\Gamma^i(S(\mu)) \boldsymbol{\psi}_i. \tag{28}$$

To treat the local nonlinearities and construct a model order reduced approximation for $u_h^{\mathrm{NLIN}-\Gamma}(\mu)$, we take the nonlinear parameters μ^{NLIN} and the approximation on the linear-nonlinear interface $u_L^\Gamma(\mu)$ into consideration. Instead of the high dimensional $\mu \in \mathcal{D}$, we take $\mu^{\mathrm{NLIN}} \in \mathcal{D}_{\mathrm{NLIN}}$ and $u_L^\Gamma(\mu)$ as the input parameters to formulate a new set of GP models to fully reduce the nonlinear subsystem.

Let $\mathcal{V}_K^{\mathrm{NLIN}-\Gamma} = \mathrm{span}\{\phi_1, \cdots, \phi_K\}$ be an orthonormal reduced basis space of size K, constructed from the first K singular vectors of N_{NLIN} snapshots $u_h^{\mathrm{NLIN}-\Gamma}(\mu_i)$, $i = 1, \cdots, N_{\mathrm{NLIN}}$. The corresponding approximation is expressed as

$$u_K^{\mathrm{NLIN}-\Gamma}(\mu) = \sum_{i=1}^{K} u_{K,i}^{\mathrm{NLIN}-\Gamma}(\mu) \phi_i, \tag{29}$$

where $u_{K,i}^{\mathrm{NLIN}-\Gamma}$ is the individual coefficient of ϕ_i in $\mathcal{V}_K^{\mathrm{NLIN}-\Gamma}$. Similarly, we define $\pi_{\mathrm{NLIN}}^i : \mathcal{D}_{\mathrm{NLIN}} \times \mathbb{R}^L \to \mathbb{R}^K$ as the GP to approximate the ith coefficient of $u_K^{\mathrm{NLIN}-\Gamma}$. The required training data is a collection of n_{tr}^{NLIN} pairs $(\mathbf{x}_j^i, y_j^i)$, $j = 1, \cdots, n_{tr}^{\mathrm{NLIN}}$, where $\mathbf{x}_j^i = (\mu_j^{\mathrm{NLIN}}, u_\Gamma^L(\mu_j)) \in \mathcal{D}_{\mathrm{NLIN}} \times \mathbb{R}^L$ and $y_j^i = \phi_i^T u_h^{\mathrm{NLIN}-\Gamma}(\mu_j)$. The collection of projections π_{NLIN}^i forms the reduced coefficient vector

$$\boldsymbol{\pi}_{\mathrm{NLIN}}(\mu^{\mathrm{NLIN}}, u_L^\Gamma(\mu)) = [\pi_{\mathrm{NLIN}}^1(\mu^{\mathrm{NLIN}}, u_L^\Gamma(\mu)), \cdots, \pi_{\mathrm{NLIN}}^K(\mu^{\mathrm{NLIN}}, u_L^\Gamma(\mu))]^T. \tag{30}$$

The fully reduced estimation for the nonlinear subdomain reads

$$\boldsymbol{u}_h^{\mathrm{NLIN}-\Gamma}(\mu) \approx \boldsymbol{u}_K^{\mathrm{NLIN}-\Gamma}(\mu^{\mathrm{NLIN}}, \boldsymbol{u}_L^\Gamma(\mu)) = \boldsymbol{\mathcal{V}}_K^{\mathrm{NLIN}-\Gamma} \boldsymbol{\pi}_{\mathrm{NLIN}-\Gamma}(\mu^{\mathrm{NLIN}}, \boldsymbol{u}_L^\Gamma(\mu)). \tag{31}$$

To incorporate the two approximations in the linear-nonlinear coupled system, we take $u_{N,M,L,K}^\Gamma(\mu) := u_L^\Gamma(\mu)$ and $u_{N,M,L,K}^{\mathrm{NLIN}-\Gamma}(\mu) := u_K^{\mathrm{NLIN}-\Gamma}(\mu)$. Recall that N and M indicate the bubble and port reduction from the SCRBE approach, whereas L and K represent the two sets of GPRs on the linear-nonlinear interface and over the nonlinear subdomain, respectively. Analogous to the nonlinear subsystem, the linear subsystem can readily be solved as a function of μ^{LIN} and the linear-nonlinear interaction $u_{N,M,L,K}^\Gamma(\mu)$. We then form the global approximation $u_{N,M,L,K}(\mu)$ as

$$u_{N,M,L,K}(\mu) = u_{N,M,L,K}^\Gamma(\mu) + u_{N,M,L,K}^{\mathrm{NLIN}-\Gamma}(\mu^{\mathrm{NLIN}}, u_{N,M,L,K}^\Gamma(\mu)) + u_{N,M,L,K}^{\mathrm{LIN}-\Gamma}(\mu^{\mathrm{LIN}}, u_{N,M,L,K}^\Gamma(\mu)), \tag{32}$$

where two of the three terms, $u_{N,M,L,K}^\Gamma(\mu) \in \mathcal{V}_L^\Gamma$ and $u_{N,M,L,K}^{\mathrm{NLIN}-\Gamma}(\mu) \in \mathcal{V}_K^{\mathrm{NLIN}-\Gamma}$, have already been estimated through their associated GPRs. The remaining part $u_{N,M,L,K}^{\mathrm{LIN}-\Gamma}(\mu^{\mathrm{LIN}}, u_{N,M,L,K}^\Gamma(\mu))$ can be calculated by the SCRBE solver.

We recall that in the SCRBE approach, $u_{N,M}(\mu)$ is split into $u_{N,M}^0(\mu)$ and $u_{N,M}^\gamma(\mu)$, where $u_{N,M}^0(\mu) \in \mathcal{V}_N^0$ represents the reduced bubble approximations that can be solved individually in advance on each components and $u_{N,M}^\gamma(\mu) \in \mathcal{V}_M^\gamma$ reflects the reduced port approximations that reside on each boundary of the linear

subsystem (including the linear-nonlinear interface Γ). Similarly, we have

$$\begin{aligned} u^{\mathrm{LIN}}_{N,M,L,K}(\mu, u^{\Gamma}_{N,M,L,K}(\mu)) &= u^{\Gamma}_{N,M,L,K}(\mu) + u^{\mathrm{LIN}-\Gamma}_{N,M,L,K}(\mu^{\mathrm{LIN}}, u^{\Gamma}_{N,M,L,K}(\mu)) \\ &= u^{0}_{N,M,L,K}(\mu^{\mathrm{LIN}}) + u^{\gamma}_{N,M,L,K}(\mu). \end{aligned} \tag{33}$$

We then propose to split the port approximation $u^{\gamma}_{N,M,L,K}(\mu)$ as

$$u^{\mathrm{LIN}}_{N,M,L,K}(\mu, u^{\Gamma}_{N,M,L,K}(\mu)) = u^{0}_{N,M,L,K}(\mu^{\mathrm{LIN}}) + u^{\Gamma}_{N,M,L,K}(\mu) + u^{\gamma-\Gamma}_{N,M,L,K}(\mu^{\mathrm{LIN}}, u^{\Gamma}_{N,M,L,K}(\mu)), \tag{34}$$

where $u^{0}_{N,M,L,K}(\mu^{\mathrm{LIN}}) \in \mathcal{V}^{0}_{N,M,L,K} = \mathcal{V}^{0}_{N}$, $u^{\Gamma}_{N,M,L,K}(\mu) \in \mathcal{V}^{\Gamma}_{N,M,L,K} = \mathcal{V}^{\Gamma}_{L}$, $u^{\gamma-\Gamma}_{N,M,L,K}(\mu^{\mathrm{LIN}}, u^{\Gamma}_{N,M,L,K}(\mu)) \in \mathcal{V}^{\gamma-\Gamma}_{N,M,L,K} = \mathcal{V}^{\gamma-\Gamma}_{M}$ is the port approximation without counting the linear-nonlinear interface Γ, and $\mathcal{V}^{\gamma-\Gamma}_{N,M,L,K} \subset \mathcal{V}^{\gamma}_{N,M,L,K} = \mathcal{V}^{\gamma}_{M}$ indicates the reduced space $\mathcal{V}^{\gamma}_{N,M,L,K}$ with the expression on Γ removed. Given that $u^{\Gamma}_{N,M,L,K}(\mu)$ is obtained through GPRs and $u^{0}_{N,M,L,K}(\mu^{\mathrm{LIN}})$ can be approximated individually on each component, analogous to the formulation (8), we then solve for $u^{\gamma-\Gamma}_{N,M,L,K}(\mu) \in \mathcal{V}^{\gamma-\Gamma}_{N,M,L,K}$ such that for all $v \in \mathcal{V}^{\gamma}_{N,M,L,K}$

$$a(u^{\gamma-\Gamma}_{N,M,L,K}(\mu^{\mathrm{LIN}}, u^{\Gamma}_{N,M,L,K}(\mu)), v; \mu) = f(v;\mu) - a(u^{0}_{N,M,L,K}(\mu^{\mathrm{LIN}}), v; \mu) - a(u^{\Gamma}_{N,M,L,K}(\mu), v; \mu). \tag{35}$$

In a divide-and-conquer manner, the global RB approximation $u_{N,M,L,K}(\mu)$ is thus separated into four segments

$$\begin{aligned} u_{N,M,L,K}(\mu) = & u^{\Gamma}_{N,M,L,K}(\mu) + u^{0}_{N,M,L,K}(\mu^{\mathrm{LIN}}) \\ & + u^{\mathrm{NLIN}-\Gamma}_{N,M,L,K}(\mu^{\mathrm{NLIN}}, u^{\Gamma}_{N,M,L,K}(\mu)) + u^{\gamma-\Gamma}_{N,M,L,K}(\mu^{\mathrm{LIN}}, u^{\Gamma}_{N,M,L,K}(\mu)). \end{aligned} \tag{36}$$

The first and second terms can be approximated independently with complexity $\mathcal{O}(Ln^{tr}_{\Gamma})$ and $\mathcal{O}(N^3)$. Hereafter, the third and fourth terms can be estimated individually by utilizing the first two results at the computational cost of $\mathcal{O}(Kn^{tr}_{\mathrm{NLIN}})$ and $\mathcal{O}(M^3)$, respectively. We note that compared to the hybrid-SCRBE approach with online computational cost being proportional to the FE degrees of freedom over the nonlinear subdomain and the number of iterations needed for convergence, the computational saving of the proposed method is huge, as the online evaluation of each GPR necessitates only the calculation of one function value, given pre-trained hyper-parameters. The main cost of the proposed method is dominated by the linear part, i.e. the treatment of the solution over the linear subdomain, given the solution on the interface. This cost is $\mathcal{O}(N^3)$ as denoted by the second term in (36) with N being the number of active port modes of the linear subsystem. However, we point out that the offline training cost of the proposed approach is presumably higher than the hybrid-SCRBE approach. The computational cost is $\mathcal{O}(n^3_{tr})$, where n_{tr} indicates the number of training samples, to invert the matrix $K_y(\theta)$ at each iteration of the training process, as shown in (19).

Now, we extend this method to a general setting where several nonlinear components and multiple linear-nonlinear interfaces are present. Let Ω^i_{NLIN}, $i = 1, \cdots, n_{\mathrm{NLIN}}$, be n_{NLIN} subdomains of Ω, and $\Gamma_{i,j}$, $j = 1, \cdots, n^i_{\Gamma}$, be n^i_{Γ} linear-nonlinear interfaces of Ω^i_{NLIN}, where

$$\bar{\Omega} = \bar{\Omega}_{\mathrm{LIN}} \cup_{i=1}^{n_{\mathrm{NLIN}}} \bar{\Omega}^i_{\mathrm{NLIN}} \quad \text{and} \quad \bar{\Omega}_{\mathrm{LIN}} \cap \bar{\Omega}^i_{\mathrm{NLIN}} = \emptyset, \quad \forall i = 1, \cdots, n_{\mathrm{NLIN}}. \tag{37}$$

Analogous to (36), we decompose the global RB approximation $u_{N,M,L,K}(\mu)$ into segments on the interfaces and the interior of the nonlinear components

$$\begin{aligned} u_{N,M,L,K}(\mu) = & u^{0}_{N,M,L,K}(\mu^{\mathrm{LIN}}) + u^{\gamma-\Gamma}_{N,M,L,K}(\mu^{\mathrm{LIN}}, u^{\Gamma,1,1}_{N,M,L,K}(\mu), \cdots, u^{\Gamma,n_{\mathrm{NLIN}},n^{n_{\mathrm{NLIN}}}_{\Gamma}}_{N,M,L,K}(\mu)) \\ & + \sum_{i=1}^{n_{\mathrm{NLIN}}} \sum_{j=1}^{n^i_{\Gamma}} u^{\Gamma,i,j}_{N,M,L,K}(\mu) + \sum_{i=1}^{n_{\mathrm{NLIN}}} u^{\mathrm{NLIN}-\Gamma,i}_{N,M,L,K}(\mu^{\mathrm{NLIN}}, u^{\Gamma,i,1}_{N,M,L,K}(\mu), \cdots, u^{\Gamma,i,n^i_{\Gamma}}_{N,M,L,K}(\mu)). \end{aligned} \tag{38}$$

We note that a total number of $\sum_{i=1}^{n_{\text{NLIN}}} n_\Gamma^i$ GPRs are required to approximate $u^{\Gamma,i,j}_{N,M,L,K}(\mu)$ and n_{NLIN} GPRs are needed to solve for $u^{\text{NLIN}-\Gamma,i}_{N,M,L,K}(\mu^{\text{NLIN}}, u^{\Gamma,i,1}_{N,M,L,K}(\mu), \cdots, u^{\Gamma,i,n_\Gamma^i}_{N,M,L,K}(\mu))$. Consequently, we reformulate (35) to solve for $u^{\gamma-\Gamma}_{N,M,L,K}(\mu^{\text{LIN}}, u^{\Gamma}_{N,M,L,K}(\mu)) \in \mathcal{V}^{\gamma-\Gamma}_{N,M,L,K}$, such that for all $v \in \mathcal{V}^{\gamma}_{N,M,L,K}$

$$\begin{aligned} a(u^{\gamma-\Gamma}_{N,M,L,K}(\mu^{\text{LIN}}, u^{\Gamma}_{N,M,L,K}(\mu)), v; \mu) =& f(v;\mu) - a(u^0_{N,M,L,K}(\mu^{\text{LIN}}), v; \mu) \\ & - a(\sum_{i=1}^{n_{\text{NLIN}}} \sum_{j=1}^{n_\Gamma^i} u^{\Gamma,i,j}_{N,M,L,K}(\mu), v; \mu). \end{aligned} \tag{39}$$

3.2. Error analysis

We reiterate that we use the hybrid-SCRBE approximation as our truth reference. Hence, we compare our solution $u_{N,M,L,K}(\mu)$ with the hybrid-SCRBE solution $u_{N,M,h}(\mu)$ for error analysis. Precisely, we compare the solution of each component individually. For a nonlinear component i, $i = 1, \cdots, n_{\text{NLIN}}$, we take the n_Γ^i neighboring linear components that share the linear-nonlinear interfaces j, $j = 1, \cdots, n_\Gamma^i$, with the nonlinear component i for the error analysis. For the interior of the nonlinear component and its interfaces, we define

$$\begin{aligned} e^{\Gamma}_{i,j}(\mu) &= \frac{\|u^{\Gamma,i,j}_{N,M,L,K}(\mu) - u^{\Gamma,i,j}_{N,M,h}(\mu)\|}{\|u^{\Gamma,i,j}_{N,M,h}(\mu)\|}, \\ e^{\text{NLIN}-\Gamma}_{i}(\mu) &= \frac{\|u^{\text{NLIN}-\Gamma,i}_{N,M,L,K}(\mu^{\text{NLIN}}, u^{\Gamma,i,1}_{N,M,L,K}(\mu), \cdots, u^{\Gamma,i,n_\Gamma^i}_{N,M,L,K}(\mu)) - u^{\text{NLIN}-\Gamma,i}_{N,M,h}(\mu)\|}{\|u^{\text{NLIN}-\Gamma,i}_{N,M,h}(\mu)\|}. \end{aligned} \tag{40}$$

To facilitate the error analysis in Sec. 4, we define the POD solutions $u^i_{\text{POD}}(\mu)$ as the FE solution $u^i_h(\mu)$ of the nonlinear component i projected onto the reduced spaces, such that the vector form can be defined as

$$\begin{aligned} \boldsymbol{u}^{\Gamma,i,j}_{\text{POD}}(\mu) &= (\boldsymbol{\mathcal{V}}^{\Gamma,i,j}_{N,M,L,K})^T \boldsymbol{u}^{\Gamma,i,j}_{N,M,h}(\mu), \\ \boldsymbol{u}^{\text{NLIN}-\Gamma,i}_{\text{POD}}(\mu) &= (\boldsymbol{\mathcal{V}}^{\text{NLIN}-\Gamma,i}_{N,M,L,K})^T \boldsymbol{u}^{\text{NLIN}-\Gamma,i}_{N,M,h}(\mu), \end{aligned} \tag{41}$$

where $\boldsymbol{\mathcal{V}}^{\Gamma,i,j}_{N,M,L,K}$ and $\boldsymbol{\mathcal{V}}^{\text{NLIN}-\Gamma,i}_{N,M,L,K}$ are the matrices of the RB basis coefficients over the linear-nonlinear interfaces and nonlinear subdomain interiors, respectively. We note that in the nonlinear setting, projected POD solutions are often more accurate than solutions obtained by solving a reduced nonlinear problem using either intrusive or non-intrusive approaches in the same space, due to the limitation of available data and the intrinsic nonlinear behavior [7, 41]. This could potentially lead to the stagnation of the error convergence. For the comparison of errors, we define the relative POD errors as

$$\begin{aligned} e^{\Gamma}_{\text{POD},i,j}(\mu) &= \frac{\|u^{\Gamma,i,j}_{\text{POD}}(\mu) - u^{\Gamma,i,j}_{N,M,h}(\mu)\|}{\|u^{\Gamma,i,j}_{N,M,h}(\mu)\|}, \\ e^{\text{NLIN}-\Gamma}_{\text{POD},i}(\mu) &= \frac{\|u^{\text{NLIN}-\Gamma,i}_{\text{POD}}(\mu) - u^{\text{NLIN}-\Gamma,i}_{N,M,h}(\mu)\|}{\|u^{\text{NLIN}-\Gamma,i}_{N,M,h}(\mu)\|}. \end{aligned} \tag{42}$$

As discussed in Sec. 2.1, $u^{\Gamma,i,j}_{\text{POD}}(\mu)$ and $u^{\text{NLIN}-\Gamma,i}_{\text{POD}}(\mu)$ are the best approximation that can be obtained in $\mathcal{V}^{\Gamma,i,j}_{N,M,L,K}$ and $\mathcal{V}^{\text{NLIN}-\Gamma,i}_{N,M,L,K}$. Therefore, $e^{\Gamma}_{\text{POD},i,j}(\mu)$ and $e^{\text{NLIN}-\Gamma}_{\text{POD},i}(\mu)$ shall be viewed as the lower bounds of $e^{\Gamma}_{i,j}(\mu)$ and $e^{\text{NLIN}-\Gamma}_{i}(\mu)$, respectively.

For the error analysis of the adjacent linear components, we note that there is no error in the approximation of $u^0_{N,M,L,K}(\mu^{\text{LIN}})$ with respect to $u^0_{N,M,h}(\mu)$, which is the RB approximation obtained from the hybrid-SCRBE solver, because both of them are solved individually on the interior of each component by the same procedure. Since the hybrid-SCRBE provides only RB solutions $u^{\gamma-\Gamma}_{N,M,h}$ for the linear components, the POD solutions and, consequently, the POD errors, cannot be estimated. Hence, we define the linear

error as

$$e_{i,j}^{\gamma-\Gamma}(\mu) = \frac{\|u_{N,M,L,K}^{\gamma-\Gamma,i,j}(\mu^{\mathrm{LIN}}, u_{N,M,L,K}^{\Gamma,i,j}(\mu)) - u_{N,M,h}^{\gamma-\Gamma,i,j}(\mu)\|}{\|u_{N,M,h}^{\gamma-\Gamma,i,j}(\mu)\|}, \tag{43}$$

where $u_{N,M,L,K}^{\gamma-\Gamma,i,j}(\mu^{\mathrm{LIN}}, u_{N,M,L,K}^{\Gamma,i,j}(\mu))$ is the solution of the adjacent linear component j of nonlinear component i. We further point out that the test space $\mathcal{V}_{N,M,L,K}^{\gamma-\Gamma}$ of (35) is equivalent to the test space $\mathcal{V}_{N,M,h}^{\gamma-\Gamma}$ in the hybrid-SCRBE solver. This results in a global bound of $\|u_{N,M,L,K}^{\gamma-\Gamma}(\mu) - u_{N,M,h}^{\gamma-\Gamma}(\mu)\|$ in terms of $\|u_{N,M,L,K}^{\Gamma}(\mu) - u_{N,M,h}^{\Gamma}(\mu)\|$, such that

$$\|u_{N,M,L,K}^{\gamma-\Gamma}(\mu) - u_{N,M,h}^{\gamma-\Gamma}(\mu)\| \leq C(\mu)\|u_{N,M,L,K}^{\Gamma}(\mu) - u_{N,M,h}^{\Gamma}(\mu)\|, \tag{44}$$

where $C(\mu)$ depends only on μ.

Proof. For simplicity, we refer to the parameter dependence $(\mu^{\mathrm{LIN}}, u_{N,M,L,K}^{\Gamma}(\mu))$ as (μ). Recall that the two approximations $u_{N,M,L,K}^{\gamma-\Gamma}(\mu)$ and $u_{N,M,h}^{\gamma-\Gamma}(\mu)$ are obtained from

$$\begin{aligned} a(u_{N,M,L,K}^{\gamma-\Gamma}(\mu), v; \mu) &= f(v;\mu) - a(u_{N,M,L,K}^{0}(\mu), v; \mu) - a(u_{N,M,L,K}^{\Gamma}(\mu), v; \mu) \quad \forall v \in \mathcal{V}_{N,M,L,K}^{\gamma-\Gamma}, \\ a(u_{N,M,h}^{\gamma-\Gamma}(\mu), v; \mu) &= f(v;\mu) - a(u_{N,M,h}^{0}(\mu), v; \mu) - a(u_{N,M,h}^{\Gamma}(\mu), v; \mu) \quad \forall v \in \mathcal{V}_{N,M,h}^{\gamma-\Gamma}. \end{aligned} \tag{45}$$

Since $u_{N,M,L,K}^{0}(\mu) = u_{N,M,h}^{0}(\mu)$ and $\mathcal{V}_{N,M,L,K}^{\gamma-\Gamma}$ is equivalent to $\mathcal{V}_{N,M,h}^{\gamma-\Gamma}$, we have

$$a(u_{N,M,L,K}^{\gamma-\Gamma}(\mu) - u_{N,M,h}^{\gamma-\Gamma}(\mu), v; \mu) = a(u_{N,M,h}^{\Gamma}(\mu) - u_{N,M,L,K}^{\Gamma}(\mu), v; \mu)$$

By definition, the bilinear form a is coercive and continuous, we can then define the coercivity and continuity constants with respect to $\|\cdot\|$ as

$$\begin{aligned} &\alpha(\mu)\|v\|^2 \leq a(v, v; \mu) \quad \forall v \in \mathcal{V}_h, \\ &a(v, w; \mu) \leq \gamma(\mu)\|v\|\|w\| \quad \forall v, w \in \mathcal{V}_h. \end{aligned} \tag{46}$$

Applying the coercivity and continuity constant to the equation above, we have

$$\begin{aligned} \alpha(\mu)\|u_{N,M,L,K}^{\gamma-\Gamma}(\mu) - u_{N,M,h}^{\gamma-\Gamma}(\mu)\|^2 \leq& a(u_{N,M,L,K}^{\gamma-\Gamma}(\mu) - u_{N,M,h}^{\gamma-\Gamma}(\mu), u_{N,M,L,K}^{\gamma-\Gamma}(\mu) - u_{N,M,h}^{\gamma-\Gamma}(\mu); \mu) \\ =& a(u_{N,M,L,K}^{\Gamma}(\mu) - u_{N,M,h}^{\Gamma}(\mu), u_{N,M,L,K}^{\gamma-\Gamma}(\mu) - u_{N,M,h}^{\gamma-\Gamma}(\mu); \mu) \\ \leq& \gamma(\mu)\|u_{N,M,L,K}^{\Gamma}(\mu) - u_{N,M,h}^{\Gamma}(\mu)\|\|u_{N,M,L,K}^{\gamma-\Gamma}(\mu) - u_{N,M,h}^{\gamma-\Gamma}(\mu)\|, \end{aligned}$$

so that

$$\|u_{N,M,L,K}^{\gamma-\Gamma}(\mu) - u_{N,M,h}^{\gamma-\Gamma}(\mu)\| \leq \frac{\gamma(\mu)}{\alpha(\mu)}\|u_{N,M,L,K}^{\Gamma}(\mu) - u_{N,M,h}^{\Gamma}(\mu)\|.$$

Let $C(\mu) = \frac{\gamma(\mu)}{\alpha(\mu)}$ and this completes the proof. □

We note that in active learning [47] or Greedy training [40], such a posteriori error estimation can potentially identify the insufficiently rich parameter subdomain, where the sample set employed in the sensitivity analysis is not comprehensive enough to capture the derivative behaviors in that region. In such a case, corrections to the previous model can be made on the fly by including samples that maximizes the a posteriori error indicator.

4. Numerical results

We consider three dimensional elasto-plastic problems with local linear isotropic hardening. For an elasto-plastic body with small deformation, the definition of the Cauchy strain tensor $\boldsymbol{\varepsilon}$ and the equation of

equilibrium are given as follows

$$\begin{aligned} \boldsymbol{\varepsilon} &= \frac{1}{2}\left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^{\mathrm{T}}\right], \\ \operatorname{div} \boldsymbol{\sigma} + \boldsymbol{b} &= 0, \end{aligned} \tag{47}$$

where $\boldsymbol{u}$ is the displacement field, $\boldsymbol{\sigma}$ is the Cauchy stress tensor and $\boldsymbol{b}$ is a body force. In the theory of classical rate-dependent plasticity, e.g. [8, 20, 48], it is assumed that the strain tensor $\boldsymbol{\varepsilon}$ can be decomposed into an elastic part and a plastic component, denoted by $\boldsymbol{\varepsilon}^e$ and $\boldsymbol{\varepsilon}^p$, respectively, such that $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^e + \boldsymbol{\varepsilon}^p$. The stress response is only related to the elastic strain $\boldsymbol{\varepsilon}^e$. Considering the linear isotropic elasticity, Hooke's law yields

$$\boldsymbol{\sigma} = \boldsymbol{C} : \boldsymbol{\varepsilon}^e = \boldsymbol{C} : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^p), \tag{48}$$

where the stiffness tensor is defined as

$$\boldsymbol{C} = \frac{1}{E}[(1+\nu)\boldsymbol{I} - \nu \mathbf{1} \otimes \mathbf{1}]. \tag{49}$$

Here E is the Young's modulus, ν is the Poisson's ratio, $\boldsymbol{I}$ is the fourth-order identity tensor and $\mathbf{1}$ is the second-order identity tensor.

Next, we define a material internal variable $\boldsymbol{q} \in \mathbb{R}^m$ and a *yield function* $g : \mathbb{R}^{3\times 3} \times \mathbb{R}^m \to \mathbb{R}$. This function describes the occurrence and development of the plasticity. When $g(\boldsymbol{\sigma}, \boldsymbol{q}) < 0$, the state $(\boldsymbol{\sigma}, \boldsymbol{q})$ remains inside the elastic domain. It moves to and remains on the *yield surface* only when $g = 0$. For the associative hardening considered in this work, the *flow rule* is given by

$$\dot{\boldsymbol{\varepsilon}}^p = \gamma \partial_{\boldsymbol{\sigma}} g(\boldsymbol{\sigma}, \boldsymbol{q}). \tag{50}$$

To insure this inequality constraint of g, a nonnegative function γ, referred to as the *consistency parameter*, is introduced. Hence, the inequality constraint conforms to the *Kuhn–Tucker complementarity conditions*

$$\gamma \geq 0, \ g(\boldsymbol{\sigma}, \boldsymbol{q}) \leq 0, \ \text{and } \gamma g(\boldsymbol{\sigma}, \boldsymbol{q}) = 0, \tag{51}$$

and the *consistency requirement*

$$\gamma \dot{g}(\boldsymbol{\sigma}, \boldsymbol{q}) = 0. \tag{52}$$

Evidently, $\gamma = 0$ holds for any elastic state $g < 0$. On the other hand, when $g = 0$, $\dot{g} < 0$ ($\gamma = 0$) is referred to as the *elastic unloading*. Meanwhile, $\dot{g} = 0$ along with $\gamma = 0$ is called *neutral loading*, and $\dot{g} = 0$ with $\gamma > 0$ is termed *plastic loading*.

We further assume that the hardening depends only on the total plastic deformation, quantified by the effective plastic strain ε_p, i.e. $\boldsymbol{q} = \boldsymbol{q}(\varepsilon_p)$. This scalar ε_p is defined as

$$\varepsilon_p = C|\boldsymbol{\varepsilon}^p|, \tag{53}$$

where C is a positive constant and can be determined via the uniaxial test of a given material.

In the plastic or neutral loading stage, the consistency condition $\dot{g} = 0$ yields

$$\gamma = \frac{1}{h}\partial_{\boldsymbol{\sigma}} g : \dot{\boldsymbol{\sigma}} = \frac{\partial_{\boldsymbol{\sigma}} g : \boldsymbol{C} : \dot{\boldsymbol{\varepsilon}}}{\partial_{\boldsymbol{\sigma}} g : \boldsymbol{C} : \partial_{\boldsymbol{\sigma}} g + h}, \quad \text{with } h := -C|\partial_{\boldsymbol{\sigma}} g| \left(\partial_{\boldsymbol{q}} g \cdot \partial_{\varepsilon_p} \boldsymbol{q}\right). \tag{54}$$

Thus we obtain the expression of the stress rate $\dot{\boldsymbol{\sigma}}$ in terms of the total strain rate $\dot{\boldsymbol{\varepsilon}}$ as

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{C}^{ep} : \dot{\boldsymbol{\varepsilon}}, \tag{55}$$

where $\boldsymbol{C}^{ep}$ is the elasto-plastic stiffness tensor defined by

$$\boldsymbol{C}^{ep} = \boldsymbol{C} - \operatorname{sgn}(\gamma)\ \frac{\boldsymbol{C} : (\partial_{\boldsymbol{\sigma}} g \otimes \partial_{\boldsymbol{\sigma}} g) : \boldsymbol{C}}{\partial_{\boldsymbol{\sigma}} g : \boldsymbol{C} : \partial_{\boldsymbol{\sigma}} g + h}\,, \tag{56}$$

with sgn denoting the sign function.

In the J_2 flow theory, the yield function, often referred to as the *von Mises yield criterion*, is given by

$$g(\boldsymbol{\sigma}, (\boldsymbol{\alpha}, \sigma_Y)) = J_2(\boldsymbol{\sigma} - \boldsymbol{\alpha}) - \sigma_Y^2/3\,, \tag{57}$$

where $\boldsymbol{\alpha}$ is a set of internal variables representing the center of the von Mises yield surface, σ_Y is the von Mises flow stress, and $J_2(\boldsymbol{\tau}) = |\boldsymbol{\tau}|^2/2 - \operatorname{tr}[\boldsymbol{\tau}]^2/6$ denotes the second deviatoric stress invariant. This J_2-plasticity model is adopted in this work and we consider the case of linear isotropic hardening, i.e. $\partial_{\varepsilon_p}\sigma_Y$ is a positive constant and $\boldsymbol{\alpha} = \boldsymbol{0}$.

In this work, the FE solver from `Akselos`[1] is employed as the reference solver in the first numerical example, while the hybrid-SCRBE solver from `Akselos` is used as reference in the second and third numerical example. We assume that the hybrid-SCRBE solutions are accurate enough for engineering applications, so that we can use the hybrid-SCRBE solution as the reference, or truth solution, to validate our approach. In all numerical examples, the `MATLAB` function `RegressionGP.fit` is used to train the GPR models and construct predictions.

4.1. Numerical example: steel beams

The first example consists of two components of connecting steel beams as shown in Fig. 4.1. We assume that the component on the left is plastic by setting a low yield stress to this component, and the component on the right is elastic by applying a very high yield stress. A homogeneous Dirichlet boundary condition is applied on the plastic side of the beams, and boundary on the elastic side of the beams is assumed free. The degrees of freedom of the full model are $64,785$ in the original finite element space. We introduce two parameters: one nonlinear parameter μ_1, ranging from 250 to 280 MPa, to indicate the plasticity yield stress of the nonlinear component, and one linear parameter μ_2, ranging from 1×10^7 to 1.2×10^7 N/m^3, to reflect the body force exerted on the linear component. We note that the Young's modulus and the Poisson's ratio of both components are set to 200 GPa and 0.3, respectively. The tangent modulus of the linear isotropic hardening is set to 0.3 GPa, the yield stress of the linear component is set to 5×10^5 MPa so that plasticity does not occur, and the body force of the nonlinear component is set to 1×10^7 N/m^3. To construct the model, we randomly generate 500 uniformly distributed sample points in the parameter domain as the training set, and another 500 samples as the testing set.

We note that since there are only two parameters, sensitivity analysis is not necessary. We also point out that due to the relatively small number of degrees of freedom, we are able to solve this model in the high fidelity finite element space. Hence instead of the hybrid-SCRBE solver, we employ the FE solver and utilize these high fidelity solutions as training samples, testing sets, and truth references. In addition, we construct the reduction model in the traditional RB sense as described in Sec. 2.1 over the linear component without static condensation such that

$$\begin{aligned} u_{N,L,K}(\mu) &= u^{\mathrm{LIN}-\Gamma}_{N,L,K}(\mu) + u^{\Gamma}_{N,L,K}(\mu) + u^{\mathrm{NLIN}-\Gamma}_{N,L,K}(\mu), \\ a(u^{\mathrm{LIN}-\Gamma}_{N,L,K}(\mu), v; \mu) &= f(v;\mu) - a(u^{\Gamma}_{N,L,K}(\mu), v; \mu), \quad \forall v \in \mathcal{V}^{\mathrm{LIN}-\Gamma}_{N,L,K}\,, \end{aligned} \tag{58}$$

where N indicates the traditional model order reduction over the linear subdomain, In this case, $u_{N,M,L,K}(\mu)$ and $\mathcal{V}^{\mathrm{LIN}-\Gamma}_{N,M,L,K}$ in Sec. 3.1 reduce to $u_{N,L,K}(\mu)$ and $\mathcal{V}^{\mathrm{LIN}-\Gamma}_{N,L,K}$, respectively.

We show first in Fig. 3 three RB coefficients from the training set on the linear-nonlinear interface and three RB coefficients over the nonlinear subdomain with regard to the two corresponding parameter values. We observe that as the index of the basis function increases, the coefficient values become less smooth and harder to predict, which may be an indicator for a denser training set for higher dimensional coefficients, or decreasing accuracy for fixed number of training data. Next, we show the predictive results for the RB

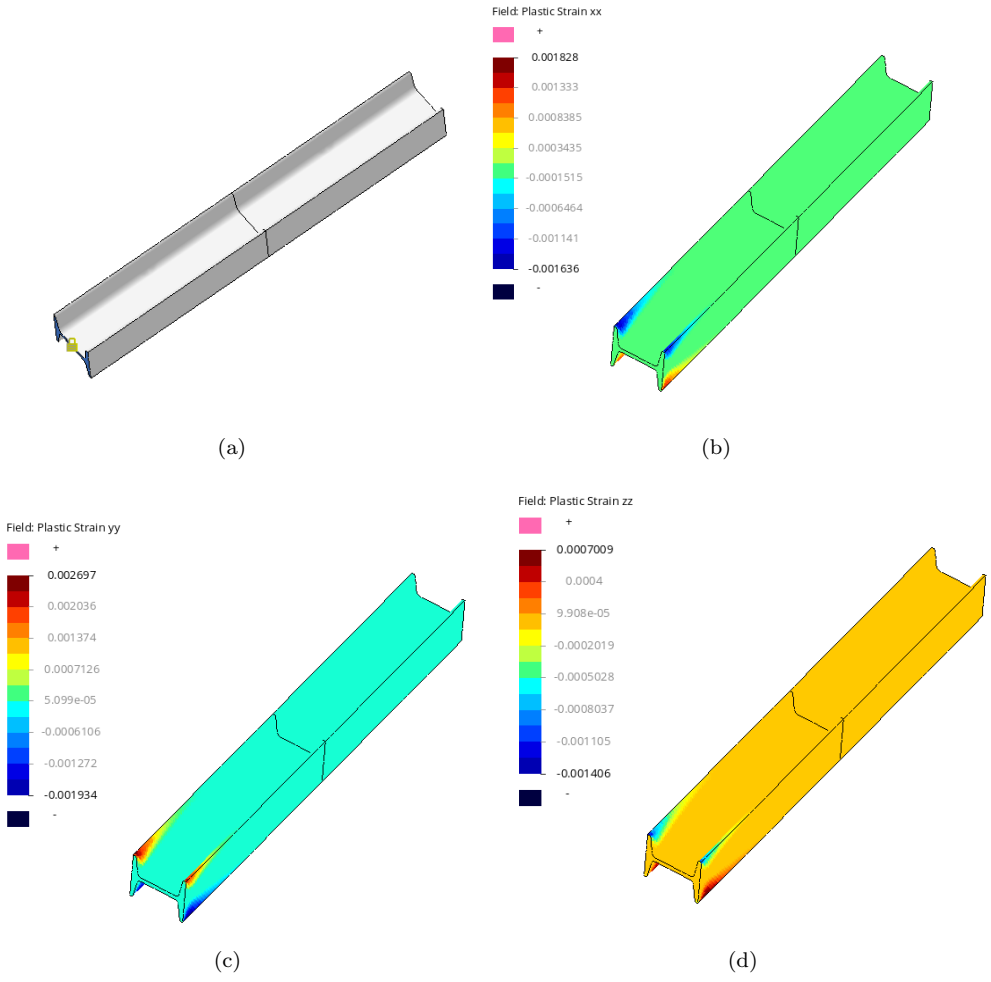


Figure 2: Steel beams model: (a) model visualization - the component on the left is treated as a plastic model, the component on the right is linear elastic; (b)-(d) example of results (plastic strain ϵ_{xx}, ϵ_{yy}, and ϵ_{zz}) at $\mu_1 = 250$ MPa and $\mu_2 = 1 \times 10^7$ N/m^3.

coefficients both on the linear-nonlinear interface and over the nonlinear domain with a 95% confidence level in Fig. 4. We notice that the confidence range enlarges as the index of the RB basis function increases, hence resulting in a larger amount of uncertainty.

Lastly, we present the convergence of the model which is constructed from 500 randomly generated training samples. We validate it against solutions at another 500 parameter values, and show the relative errors and their corresponding POD errors in Fig. 5. We note that as discussed in Sec. 3.2, the relative error over the linear subdomain is proportional to the relative error on the linear-nonlinear interface, as a result of the solution on the linear-nonlinear interface being considered as an external source acted on the linear subsystem. Similarly, we use the solution on the interface as part of the inputs to the nonlinear model, and observe that the error on the interface again reflects the error over the nonlinear subdomain. We further point out that the plateau in the convergence is due to the decreasing smoothness of the POD mode coefficients as the POD mode index number increases, making it more and more difficult for the GPR to achieve the same predictive accuracy, as shown in Fig. 3. Consequently, more and more data is required to guarantee a continuous decrease of errors for an increasing number of POD modes. We show in Fig. 5(c) and (d) the convergence results on the linear-nonlinear interface and over the nonlinear domain obtained from different number of training samples. The observation confirms our conjecture. We show that the stagnation in Fig. 5 confirms the observation in Fig. 3 and Fig. 4 that the predictive accuracy of the GPR decreases as the index number increases due to the lacking of training data. We note that compared to the nonlinear FE solver of $64,785$ unknowns, we get up to 10^5 speedup using the GPR solver. This number is expected to grow larger as the degrees of freedom of the original problem increases.

4.2. Numerical example: chair

The second example is a chair that consists of 41 components of 13 identical steel beams (8 on the bottom, 4 in the middle, 1 on the top) and 28 other components as shown in Fig. 6. The component in yellow is assumed to be plastic through a low yield stress, and all other components are elastic by applying a high yield stress. Homogeneous Neumann boundary conditions are applied at the bottom of the chair. The degrees of freedom of the full model are $1,779,975$ in the original finite element space. We set one out of the 13 steel beam components to be nonlinear with 4 parameters: the yield stress $\mu_{\text{NLIN},1} \in [30, 32]$ MPa, the body force $\mu_{\text{NLIN},2} \in [5, 6] \times 10^7$ N/m^3, the Young's modulus $\mu_{\text{NLIN},3} \in [200, 220]$ GPa, and the Poisson's ratio $\mu_{\text{NLIN},4} \in [0.28, 0.3]$. For the remaining $i = 1, \cdots, 12$ linear steel beam components, we define three parameters for each component: the body force $\mu_{\text{LIN},3i-2} \in [5, 6] \times 10^7$ N/m^3, the Young's modulus $\mu_{\text{LIN},3i-1} \in [200, 220]$ GPa, and the Poisson's ratio $\mu_{\text{LIN},3i} \in [0.28, 0.3]$. We note that the tangent modulus of the linear isotropic hardening is set to 0.3 GPa, the body force of all other components is set to 5×10^7 N/m^3, the Young's modulus and the Poisson's ratio of all linear components are set to 200 GPa and 0.3, respectively. In total, we have 40 parameters for the model.

We construct the model using 500 randomly generated and uniformly distributed parameter points as the training set, and another 500 such samples as the testing set. We note that since there are two linear-nonlinear interfaces on the nonlinear component, we denote the interface on top as Γ_1 and the interface on bottom as Γ_2. We show sensitivity results in fig. 7(a). In this analysis we employ 20 parameter samples and we observe that the most important parameters for both interfaces are the nonlinear and linear parameters of the components that are close to these interfaces. We further notice that after sorting the parameters according to the significance of their sensitivity indices, the first 24 parameters capture a majority of the model uncertainty on the linear-nonlinear interfaces. However, we show later that for engineering accuracy, a small number of parameters suffices. In Fig. 7(b), we show the first 20 singular values of each solution part. The singular values over the nonlinear subdomain inevitably decay slower than the ones on the interfaces. It can be deduced that the error over the nonlinear subdomain will dominate and may potentially serve as the error indicator for the whole system.

Lastly, we present the convergence results of 500 randomly generated testing parameters, the mean relative errors and their corresponding mean relative POD errors, shown in Fig. 7(c)-(f). In Fig. 7(c)-(d), our observation confirms the expectation that the first 24 out of the 40 parameters reflect the system behavior well on the linear-nonlinear interfaces. To reach an accuracy level of 10^{-3}, 20 parameters needs to be incorporated in the model. Including more parameters as inputs does not improve the predictive results

significantly. In Fig. 7(e) and (f), the relative errors over both the nonlinear and the linear subdomains show similar trends, but the error in the nonlinear subdomain is higher, which confirms the nontrivial fact that the nonlinear error dominates. We note that the plateaus in convergence is caused by the inaccurate prediction for POD modes with higher index numbers, as in the previous numerical example. Increasing the amount of training data potentially improves the quality of the GPRs. We note that we get around 38 times speedup when compared to the hybrid solve under the same condition. Since the speedup of the GPR-SCRBE approach is governed by the linear SCRBE solver, the speedup can be viewed as the computational saving of the linear SCRBE solver when compared to the nonlinear hybrid-SCRBE solver.

4.3. Numerical example: structural building

The third example is a three story structural building that consists of 408 components, among which 120 are horizontal and vertical steel beams and the rest 288 are other components, e.g. connectors and adapters. As shown in Fig. 8(a)-(d), the two components in yellow are treated as plastic components with a low yield stress, while the material behavior of all other components is elastic by setting a significantly higher yield stress. Among the two nonlinear components, the one on the left is indexed component 1 and the one on the right is nonlinear component 2. Homogeneous Neumann boundary conditions are applied on the bottom of the structural building. There are more than 15 million degrees of freedom in the original finite element model. We set $i = 1, 2$ out of the 120 steel beam components to be nonlinear with 2 parameters: the yield stress $\mu_{NLIN,2i-1} \in [6, 7] \times 10^3$ MPa and the body force $\mu_{NLIN,2i} \in [8, 9] \times 10^8$ N/m^3. For the remaining $i = 1, \cdots, 118$ materially elastic linear steel beam components, we assign the body force $\mu_{LIN,i} \in [8, 9] \times 10^8$ N/m^3 as the parameter for each component. We note that the Young's modulus and the Poisson's ratio are set to 200 GPa and 0.3, respectively. The yield stress of the linear component is set to 5×10^5 MPa so that plasticity shall not occur, and the body force of all non-steel beam components is set to 8×10^8 N/m^3. In total, we have 122 parameters for the whole model.

We construct the GPR-SCRBE model using 500 randomly generated and uniformly distributed points in the parameter domain as the training set, and another 500 such samples are employed as the testing set. We first show the result of the sensitivity analysis in Fig. 8(e), which is obtained from 20 parameter samples as discussed in Sec. 3.1. We notice that the most important parameters are the yield stress and the body forces of components located close to the linear-nonlinear interface. We observe that after sorting the parameters according to their sensitivity indices, the first 12 parameters capture the majority of the model behavior on the linear-nonlinear interface of each component. With the first 20 significant parameters, almost all characteristics of the solutions on the interface can be well represented. In Fig. 8(f), we demonstrate the first 20 singular values of each solution parts. Evidently, the singular values over the nonlinear subdomain decay slower than on the interfaces. We conjecture that the error over the nonlinear subdomain dominates and may serve as an error indicator of the whole system.

We then present the convergence of the solution at 500 randomly generated testing parameters, the mean relative errors and their corresponding mean relative POD errors, shown in Fig. 9(a)-(f). In Fig. 9(a)-(b), our observation confirms our conjecture that the first 12 out of 122 most significant parameters describe the solution on the linear-nonlinear interface well, with an average relative error below 10^{-2}. Afterwards, adding more parameters does not improve the predictive results remarkably. In Fig. 9(c)-(f), the relative errors over the nonlinear and linear subdomains show similar behavior, but the magnitude of the nonlinear errors is higher. We reiterate the data-driven nature of GPRs, which results in the stagnation of the error convergence.

Lastly, we point out that we get around 22 times speedup with respect to the hybrid solve under the same condition. The speedup of the GPR-SCRBE approach is governed by the linear SCRBE solver of cost $\mathcal{O}(M^3)$, since it is almost free to evaluate the GP regressions at chosen parameter values. The computational cost does not increase visibly even if we empoly a few more basis functions for the GP regressions, so that the speedup of GPR-SCRE solvers with different numbers of GP basis functions stays the same.

5. Conclusions

A hybrid GPR and SCRBE approach is proposed to enable model order reduction of large-scale structures with local nonlinearities. In our framework, a prescribed linear-nonlinear domain division is prerequisite and an RB space is constructed for the linear-nonlinear interface. Rather than the conventional Galerkin approach, the GPR is used to carry out the reduced approximation for each basis coefficient to allow a full decoupling of the offline and online stages. Equipped with the approximation on the interface, the system is fully decoupled into one linear and one nonlinear subsystem. The linear subsystem can be treated by a SCRBE solver which ensures a high model order reduction and a controllable accuracy. The nonlinear subsystem is treated by GPRs, where, instead of the full parameter space, only nonlinear parameters and the RB solution at the linear-nonlinear interface are taken as model inputs. Our method is validated against three numerical examples of increasing complexity, and is shown to be an effective tool for the solution of large-scale structures with local nonlinearities and high dimensional parameter domains.

We reiterate that all numerical results in this work are obtained in the context of component synthesis based SCRBE method. However, the proposed method can be applied to any problem with a pre-determined linear-nonlinear interface, i.e. thermal fin problems with local nonlinear heat transfer or fluid structure interactions. With the validation of numerical results, we conclude that the proposed method provides the capability that fit well in industrial-scale engineering applications in a multi-query and real-time context.

Acknowledgment

This work was supported by the Swiss Commission for Technology and Innovation (CTI) under Grant No. 25964.2 PFIW-IW, and is a collaboration with Akselos S.A.. We are grateful that they provided us with their software for research purpose. We would also like to thank our colleague Jonas Ballani from Akselos S.A. for fruitful discussions.

References

[1] Akselos software. `https://akselos.com/`.

[2] J. Ballani, D. Huynh, D. Knezevic, L. Nguyen, and A. Patera. A component-based hybrid reduced basis/finite element method for solid mechanics with local nonlinearities. *Computer Methods in Applied Mechanics and Engineering*, 329:498–531, 2018.

[3] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004.

[4] R. Bellman. Dynamic programming and lagrange multipliers. *Proceedings of the National Academy of Sciences*, 42(10):767–769, 1956.

[5] Y. Bengio, O. Delalleau, and N. L. Roux. The curse of highly variable functions for local kernel machines. In *Advances in neural information processing systems*, pages 107–114, 2006.

[6] F. Campolongo, J. Cariboni, and A. Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental modelling & software*, 22(10):1509–1518, 2007.

[7] S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.

[8] W.-F. Chen and D.-J. Han. *Plasticity for structural engineers*. J. Ross Publishing, 2007.

[9] F. Chinesta and P. Ladevéze. *Separated Representations and PGD-Based Model Reduction*. Springer, 2014.

[10] F. Chinesta, P. Ladeveze, and E. Cueto. A short review on model order reduction based on proper generalized decomposition. *Archives of Computational Methods in Engineering*, 18:395–404, 2011.

[11] R. Craig and M. Bampton. Coupling of substructures for dynamic analyses. *AIAA journal*, 6(7):1313–1319, 1968.

[12] M. Drohmann and K. Carlberg. The romes method for statistical modeling of reduced-order-model error. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):116–145, 2015.

[13] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[14] J. Eftang, D. Huynh, D. Knezevic, E. Ronquist, and A. Patera. Adaptive port reduction in static condensation. *IFAC Proceedings Volumes*, 45(2):695–699, 2012.

[15] J. L. Eftang and A. T. Patera. Port reduction in component-based static condensation for parametrized problems: Approximation and a posteriori error estimation. *IJNME*, 96(5):269–302, 2013.

[16] J. L. Eftang and A. T. Patera. Port reduction in parametrized component static condensation: approximation and a posteriori error estimation. *International Journal for Numerical Methods in Engineering*, 96(5):269–302, 2013.

[17] J. L. Eftang and A. T. Patera. A port-reduced static condensation reduced basis element method for large component-synthesized structures: approximation and a posteriori error estimation. *Advanced Modeling and Simulation in Engineering Sciences*, 1(1):3, 2014.
[18] M. Guo and J. S. Hesthaven. Data-driven reduced order modeling for time-dependent problema. *To appear in Computer Methods in Applied Mechanics and Engineering*, 2018.
[19] M. Guo and J. S. Hesthaven. Reduced order modeling for nonlinear structural analysis using gaussian process regression. *Computer Methods in Applied Mechanics and Engineering*, 341:807–826, 2018.
[20] W. Han and B. D. Reddy. *Plasticity: mathematical theory and numerical analysis*, volume 9. Springer Science & Business Media, 2012.
[21] J. S. Hesthaven, G. Rozza, and B. Stamm. *Certified reduced basis methods for parametrized partial differential equations*. Springer, 2016.
[22] J. S. Hesthaven and S. Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. Technical report, 2018.
[23] J. S. Hesthaven and S. Zhang. On the use of anova expansions in reduced basis methods for parametric partial differential equations. *Journal of Scientific Computing*, 69(1):292–313, 2016.
[24] U. L. Hetmaniuk and R. B. Lehoucq. A special finite element method based on component mode synthesis. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(3):401–420, 2010.
[25] W. C. Hurty. Dynamic analysis of structural systems using component modes. *AIAA journal*, 3(4):678–685, 1965.
[26] D. B. P. Huynh, D. J. Knezevic, and A. T. Patera. A static condensation reduced basis element method: approximation and a posteriori error estimation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(1):213–251, 2013.
[27] D. B. P. Huynh, D. J. Knezevic, and A. T. Patera. A static condensation reduced basis element method: Complex problems. *Computer Methods in Applied Mechanics and Engineering*, 259:197–216, 2013.
[28] S. Kucherenko et al. A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices. *Computer Physics Communications*, 181(7):1212–1217, 2010.
[29] S. Kucherenko and B. Iooss. Derivative-based global sensitivity measures. *Handbook of uncertainty quantification*, pages 1241–1263, 2017.
[30] S. Kucherenko, M. Rodriguez-Fernandez, C. Pantelides, and N. Shah. Monte carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering & System Safety*, 94(7):1135–1148, 2009.
[31] S. Kucherenko and I. Sobol. Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79(10):3009–3017, 2009.
[32] J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, and J. Sohl-Dickstein. Deep neural networks as gaussian processes. *arXiv preprint arXiv:1711.00165*, 2017.
[33] Y. Liang, H. Lee, S. Lim, W. Lin, K. Lee, and C. Wu. Proper orthogonal decomposition and its applications part i: Theory. *Journal of Sound and Vibration*, 252(3):527–544, 2002.
[34] A. E. Løvgren, Y. Maday, and E. M. Rønquist. A reduced basis element method for the steady stokes problem. *ESAIM: Mathematical Modelling and Numerical Analysis*, 40(3):529–552, 2006.
[35] Y. Maday and E. M. Rønquist. A reduced-basis element method. *Journal of scientific computing*, 17(1-4):447–459, 2002.
[36] Y. Maday and E. M. Ronquist. The reduced basis element method: application to a thermal fin problem. *SIAM Journal on Scientific Computing*, 26(1):240–258, 2004.
[37] K. S. Mohamed. *Machine learning for model order reduction*. Springer, 2018.
[38] A. S. Z. Moosavi, R. Stefanescu, and A. Sandu. Efficient construction of local parametric reduced order models using machine learning techniques. *CoRR*, abs/1511.02909, 2015.
[39] R. M. Neal. *Bayesian learning for neural networks*, volume 118. Springer, 1995.
[40] A. T. Patera and G. Rozza. *Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations*. Copyright MIT 2007, MIT Pappalardo Graduate Monographs in Mechanical Engineering, http://www.augustine.mit.edu, 2007.
[41] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced basis methods for partial differential equations: an introduction*, volume 92. Springer, 2015.
[42] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*. MIT press Cambridge, 2006.
[43] G. Rozza, D. B. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):1, 2007.
[44] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010.
[45] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.
[46] E. Schmidt. Zur theorie der linearen und nichtlinearen integralgleichungen. i. teil: Entwicklung willkürlicher funktionen nach systemen vorgeschriebener. *Mathematische Annalen*, 63:433–476, 1907.
[47] B. Settles. Active learning literature survey. Technical report, University of Wisconsin, Madison, 2010.
[48] J. C. Simo and T. J. Hughes. *Computational Inelasticity*. Springer Science & Business Media, 2006.
[49] K. Smetana and A. T. Patera. Optimal local approximation spaces for component-based static condensation procedures. *SIAM Journal on Scientific Computing*, 38(5):A3318–A3356, 2016.
[50] I. Sobol and S. Kucherenko. Derivative based global sensitivity measures. *Procedia-Social and Behavioral Sciences*, 2(6):7745–7746, 2010.
[51] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics*

and computers in simulation, 55(1-3):271–280, 2001.

[52] B. Sudret and C. V. Mai. Computing derivative-based global sensitivity measures using polynomial chaos expansions. *Reliability Engineering & System Safety*, 134:241–250, 2015.

[53] R. Tripathy, I. Bilionis, and M. Gonzalez. Gaussian processes with built-in dimensionality reduction: Applications to high-dimensional uncertainty propagation. *Journal of Computational Physics*, 321:191–223, 2016.

[54] S. Vallaghé and A. T. Patera. The static condensation reduced basis element method for a mixed-mean conjugate heat exchanger model. *SIAM Journal on Scientific Computing*, 36(3):B294–B320, 2014.

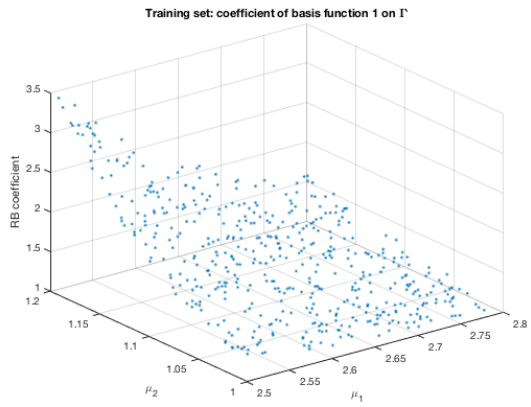
[55] Q. Wang, J. S. Hesthaven, and D. Ray. Non-intrusive reduced order modeling of unsteady flows using artificial neural networks with application to a combustion problem. *Journal of computational physics*, 2018.

[56] C. K. Williams and C. E. Rasmussen. Gaussian processes for regression. In *Advances in neural information processing systems*, pages 514–520, 1996.

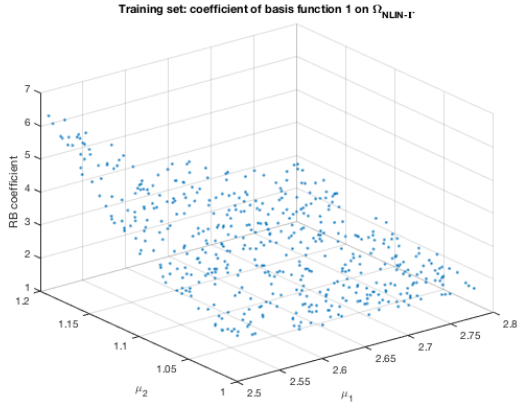
[57] E. L. Wilson. The static condensation algorithm. *International Journal for Numerical Methods in Engineering*, 8(1):198–203, 1974.

[58] O. C. Zienkiewicz, R. L. Taylor, and D. D. Fox. *The Finite Element Method for Solid and Structural Mechanics, 7th edition*. Elsevier, 2014.

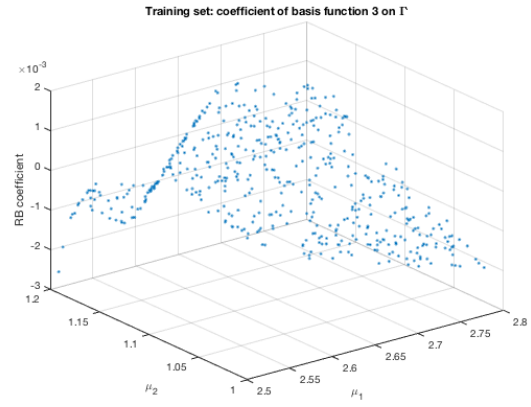
[59] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu. *The Finite Element Method: Its Basis and Fundamentals, 7th edition*. Elsevier, 2013.



(a)

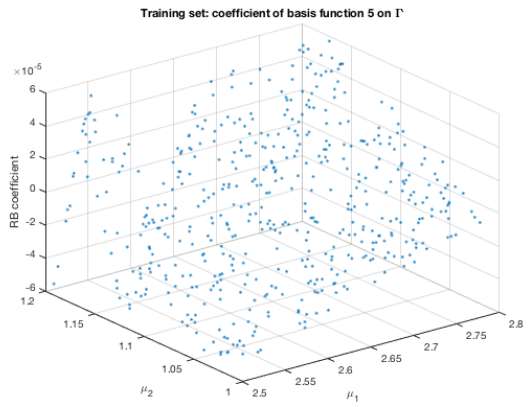


(b)

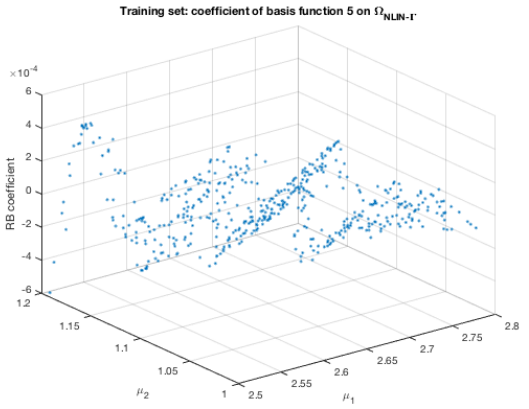


(c)

(d)

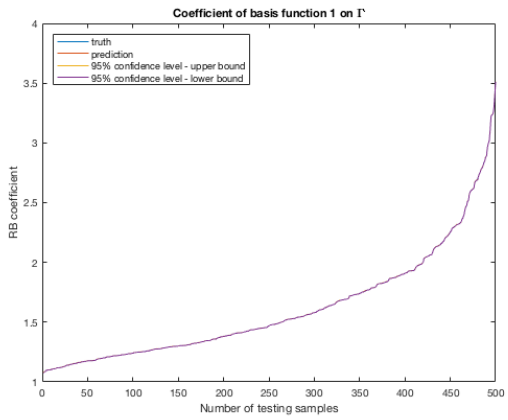


(e)

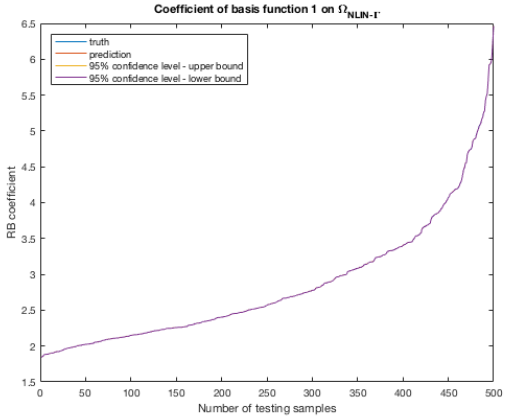


(f)

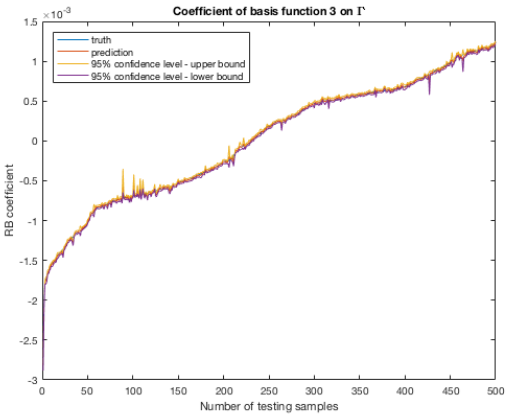
Figure 3: Steel beams model: visualization of the training set versus the RB coefficient values. (a)(c)(e) the first, third and fifth RB coefficients of the training set on the linear-nonlinear interface; (b)(d)(f) the first, third and fifth RB coefficients of the training set over the nonlinear subdomain.



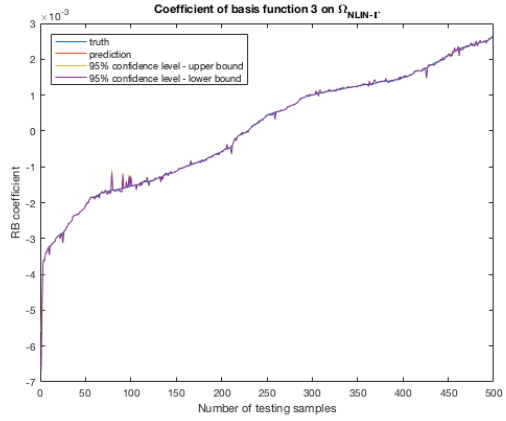
(a)



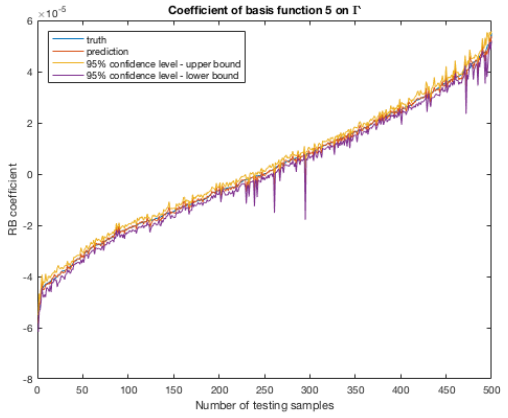
(b)



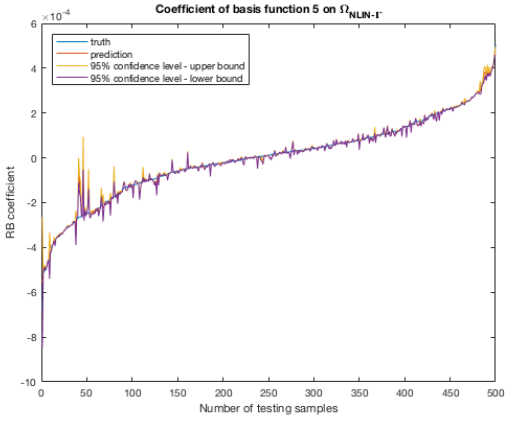
(c)



(d)



(e)



(f)

Figure 4: Steel beams model: predictive results for RB coefficients. (a)(c)(e) predictive results of the first, third and fifth RB coefficients on the linear-nonlinear interface; (b)(d)(f) predictive results of the first, third and fifth RB coefficients over the nonlinear subdomain.

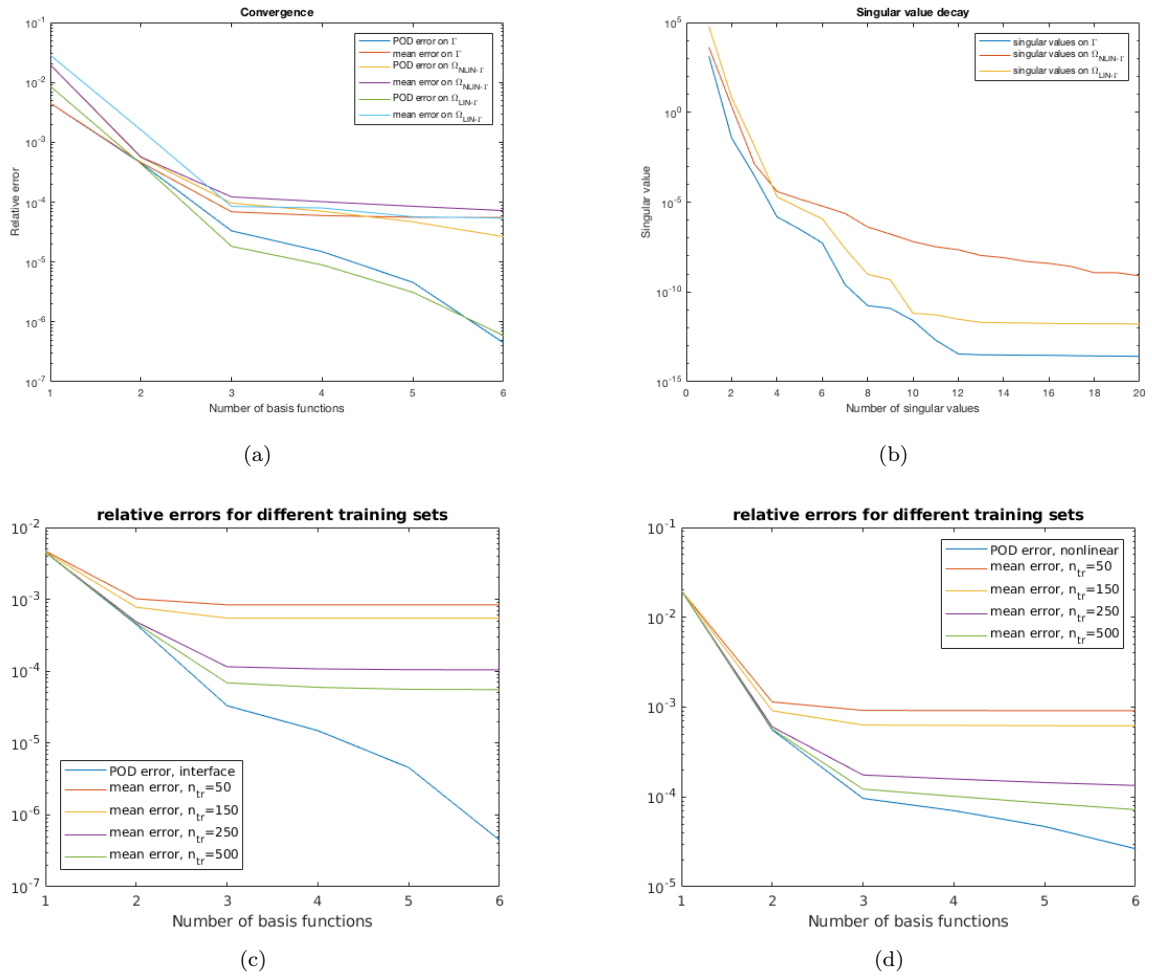


Figure 5: Steel beams model: (a) convergence result of approximation errors, as well as POD errors, as defined in (42) and (43); (b) the first 20 singular values of training samples; (c)-(d) convergence result of approximation errors on the interface and over the nonlinear subdomain when different numbers of training samples are employed, respectively.

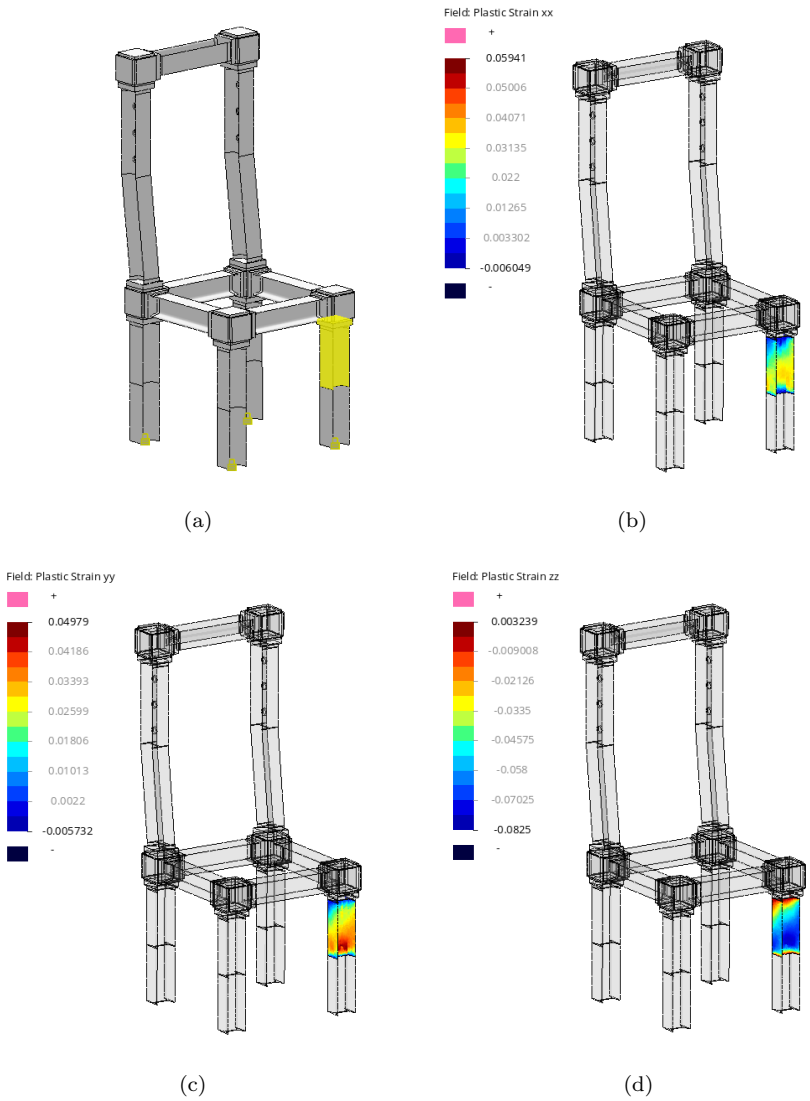


Figure 6: Chair model: (a) model visualization - the component in yellow is treated as a plastic model, the remaining components are linear elastic; (b)-(d) example of results (plastic strain ϵ_{xx}, ϵ_{yy}, and ϵ_{zz}) at $\mu_{\text{NLIN},1} = 30$ MPa, $\mu_{\text{NLIN},2} = 5 \times 10^7$ N/m^3, $\mu_{\text{NLIN},3} = 200$ GPa, $\mu_{\text{NLIN},4} = 0.28$, and $\mu_{\text{LIN},3i-2} = 5 \times 10^7$ N/m^3, $\mu_{\text{LIN},3i-1} = 200$ GPa, $\mu_{\text{LIN},3i} = 0.28$ for $i = 1, \cdots, 12$.

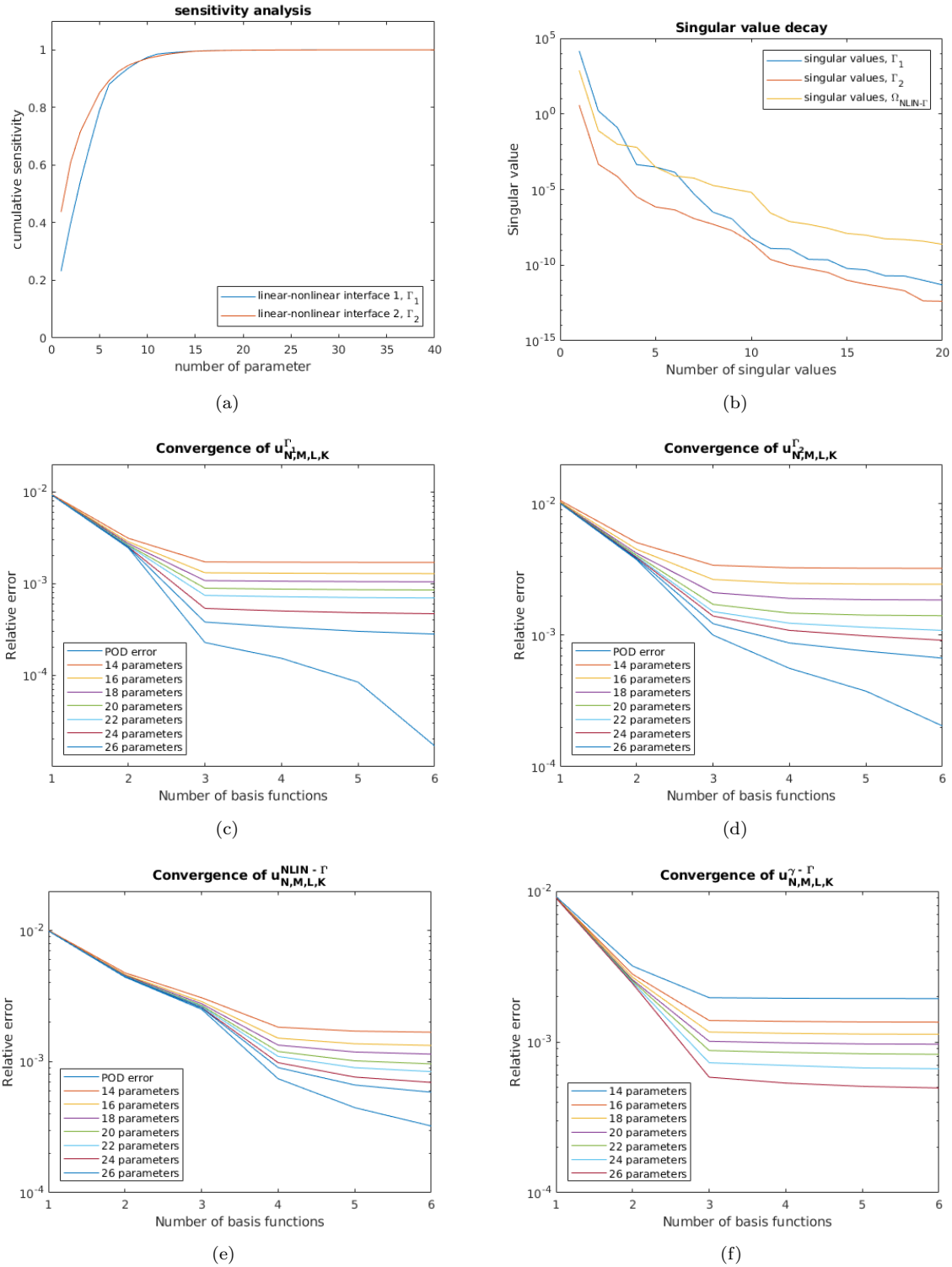
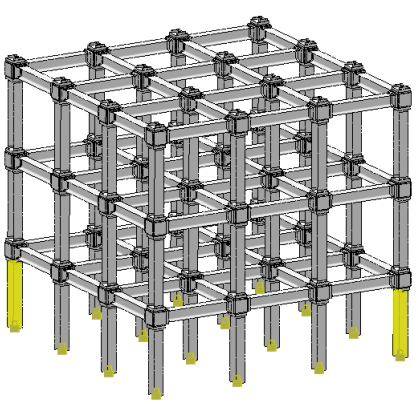
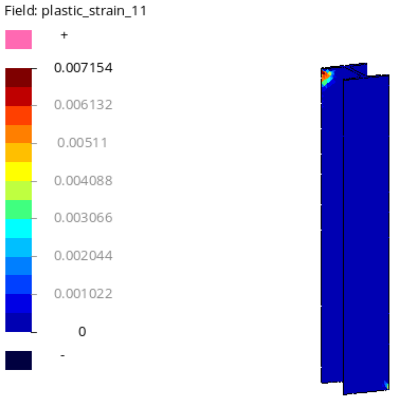


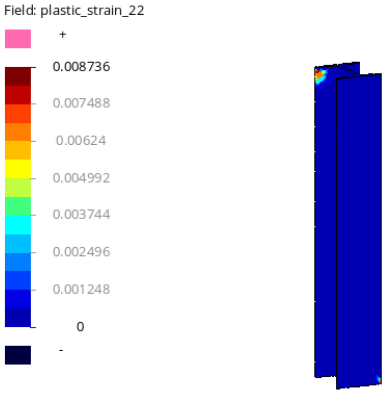
Figure 7: Chair model: (a) sensitivity analysis; (b) singular value decay; (c)-(f) convergence result of approximation errors, as well as POD errors, as defined in (42) and (43).



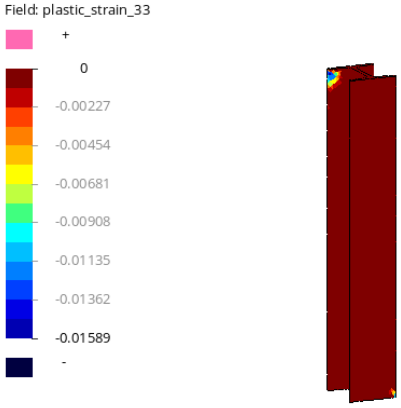
(a)



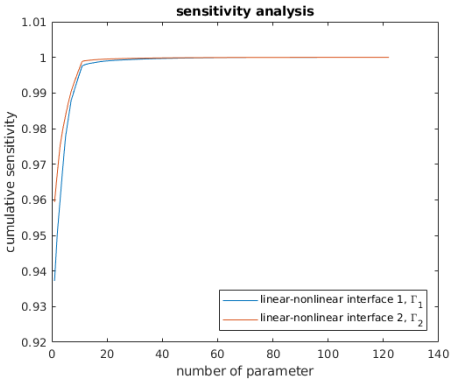
(b)



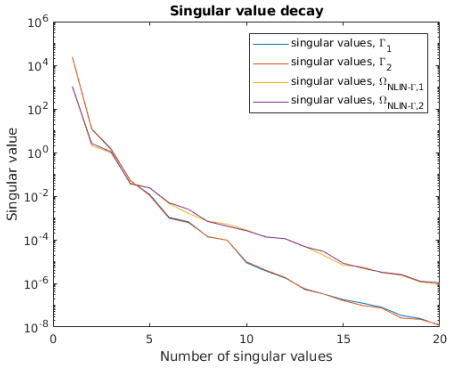
(c)



(d)

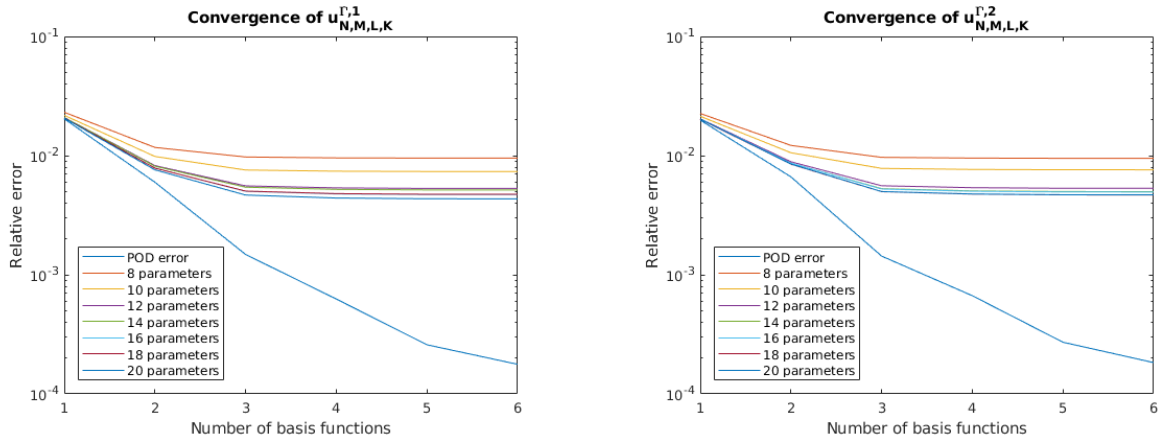


(e)



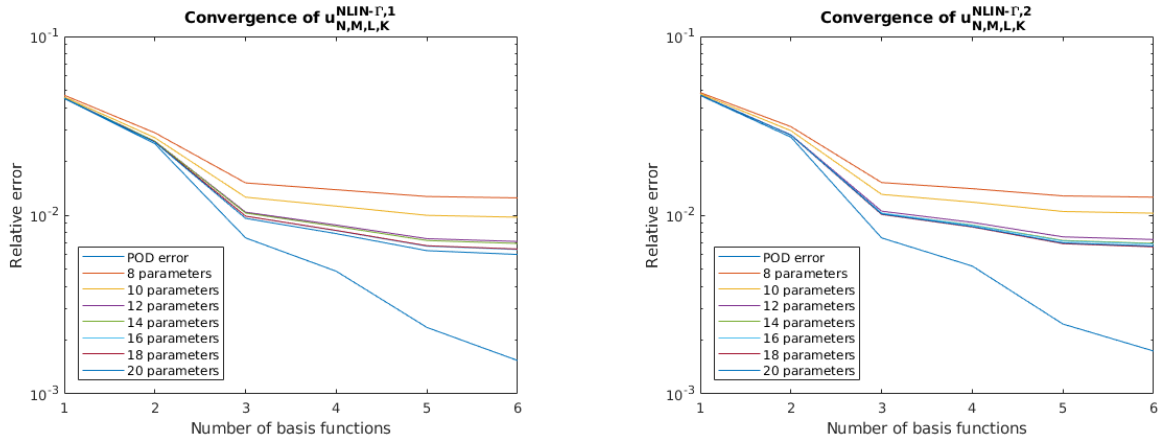
(f)

Figure 8: Structural building: (a) model visualization - the two components in yellow are nonlinear components, the one on the left is component 1 and the one on the right is component 2; (b)-(d) example of results (plastic strain ϵ_{xx}, ϵ_{yy}, and ϵ_{zz}) of nonlinear component 1 at $\mu_{NLIN,2i-1} = 6 \times 10^3$ MPa, $\mu_{NLIN,2i} = 8 \times 10^8$ N/m^3, $i = 1, 2$, and $\mu_{LIN,i} = 8 \times 10^5$ N/m^3, $i = 1, \cdots, 118$; (e) sensitivity analysis; (f) singular value decay.



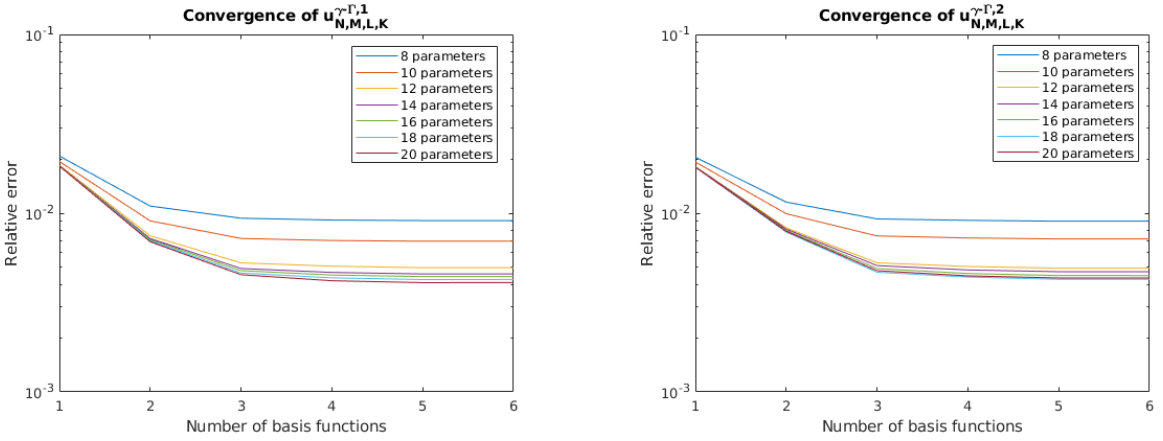
(a)

(b)



(c)

(d)



(e)

(f)

Figure 9: Structural building: (a) sensitivity analysis; (b) singular value decay; (c)-(f) convergence result of approximation errors, as well as POD errors, as defined in (42) and (43).