Gradient enhanced gaussian process regression for constitutive modelling in finite strain hyperelasticity

Nathan Ellmer[†], Rogelio Ortigosa[†‡1], Jesús Martínez-Frutos[†‡], Antonio J. Gil[†]

† *Zienkiewicz Centre for Computational Engineering, Faculty of Science and Engineering, Swansea University, Bay Campus, SA1 8EN, United Kingdom*
‡ *Technical University of Cartagena, Campus Muralla del Mar, Cartagena, 30202 Murcia, Spain*

Abstract

This paper introduces a metamodelling technique that leverages gradient-enhanced Gaussian process regression (also known as gradient-enhanced Kriging), effectively emulating the response of diverse hyperelastic strain energy densities. The approach adopted incorporates principal invariants as inputs for the surrogate of the strain energy density. This integration enables the surrogate to inherently enforce fundamental physical constraints, such as material frame indifference and material symmetry, right from the outset. The proposed approach provides accurate interpolation for energy and the first Piola-Kirchhoff stress tensor (e.g. first order derivatives with respect to inputs). The paper presents three notable innovations. *Firstly*, it introduces the utilization of Gradient-Enhanced Kriging to approximate a diverse range of phenomenological models, encompassing numerous isotropic hyperelastic strain energies and a transversely isotropic potential. *Secondly*, this study marks the inaugural application of this technique for approximating the effective response of composite materials. This includes rank-one laminates, for which analytical solutions are feasible. However, it also encompasses more complex composite materials characterized by a Representative Volume Element (RVE) comprising an elastomeric matrix with a centered spherical inclusion. This extension opens the door for future application of this technique to various RVE types, facilitating efficient three-dimensional computational analyses at the macro-scale of such composite materials, significantly reducing computational time compared to FEM[2]. The *third innovation*, facilitated by the integration of these surrogate models into a 3D Finite Element computational framework, lies in the assessment of these models scenarios encompassing intricate cases of extreme twisting and more importantly, buckling instabilities in thin-walled structures, thereby highlighting both the practical applicability and robustness of the proposed approach.

Keywords: Kriging, Machine Learning, Constitutive Modelling, Hyperelasticity, Anisotropy

1. Introduction

Modern design relies upon the use of modelling to develop configurations which maximise the potential of the device for the required application. A challenge associated with modelling these materials is their often composite nature. An example is the use of layered lamination which enables one to enhance the overall properties of the device by combining multiple materials with different advantageous properties [1–3]. The effective properties of these type of composites can be obtained [4, 5] by making use of rank-n homogenisation theory [6], requiring a Newton-Raphson type procedure at the microscopic level. Other more challenging (from the computational standpoint) types of composite materials, such as an elastomeric matrix where spherical or ellipsoidal inclusions are incorporated, require computational homogenisation for the computation of their effective properties, which further escalates the computational demands.

Machine Learning (ML) techniques are increasingly being employed across many fields including that of constitutive modelling. Artificial Neural Networks (ANNs) have been utilised by Linka et al. [7] for the development of mechanical constitutive models and Klein et al. [8] in the context of electro-mechanically

---

[1]Corresponding author: rogelio.ortigosa@upct.es

coupled constitutive models making use of either physical (in the physical lab) or numerical/in-silico data. Gaussian Process Regression (GPR) (or Kriging from now on) is also gaining traction and has been applied by Aggarwal et al. [9] in the development of data-based constitutive models in soft tissue applications. This study is primarily centered on the examination of incompressible transversely isotropic models, which are articulated using a reduced set of two invariants. Significantly, a fundamental aspect of this research is the rigorous imposition of polyconvexity for the Kriging prediction from its inception. Previously, it had been applied [10] in order to approximate the behaviour of the Mooney-Rivlin model. An advantage of using Kriging over ANN approach is their inherent probabilistic nature which includes the ability to specify prior knowledge, produce a distribution over potential predictive functions, and the capability to directly calculate uncertainty over the prediction [11–13]. One can also control the degree to which Kriging interpolates between known points through specifying the noise used in the correlation function [14].

Kriging [15–17] involves predicting the values of a Gaussian random field based on observations collected at a finite set of observation points. This has become a popular method for a large range of applications, such as geostatistics [17], numerical code approximation [18], global optimization (Jones [19]) or machine learning (Rasmussen and Williams [11]). Utilising the properties of Gaussian distributions, Kriging defines a joint distribution in terms of the observed data and the data to be predicted. A covariance is defined which spatially correlates all of the data and can be used to weight the importance of observations on the predictions. The joint distribution can then be conditionalised on the observed data leaving a distribution describing the prediction. This is characterised by a mean and covariance which once obtained, enables one to sample the distribution and thereby make predictions [20]. This particular type of emulator has gained increasing popularity in the last decade due to its flexibility in capturing nonlinear functions and its ability to provide statistical information on the predicted output [19, 21]. The probabilistic response of the prediction has a two-fold benefit: (1) it provides probabilistic confidence intervals of the prediction, and (2) it can help defining infill sampling strategies to adaptively refine the metamodel depending on the region of interest.

This paper proposes a metamodelling technique based on a gradient-enhanced Kriging to emulate the response of various ground truth hyperelastic strain energy densities used to characterise the behaviour of soft/flexible materials. The primary goal for this approach is to develop an offline method capable of emulating the strain energy density (ground truth model). By incorporating principal invariants as inputs to the surrogate model, a notable advantage arises: the ability to enforce fundamental physical constraints, such as material frame indifference and material symmetry, right from the outset. One of the notable advantages of Gradient Kriging lies in its ability to interpolate precise values not only for energy but also for the first Piola-Kirchhoff stress tensor (e.g. derivatives of the strain energy with respect to its inputs). This capability ensures a higher level of accuracy in satisfying the governing equations. Furthermore, incorporating derivative information into the model allows for a significant reduction in the number of required sampling points while still achieving the desired accuracy at unobserved points.

A prominent advantage, in comparison to other machine learning methods presented in the literature like neural networks, is found in the inherent interpolatory nature of Kriging. This characteristic allows for the precise matching of the first Piola-Kirchhoff stress tensor at the sampling points, rather than relying on the minimization of an aggregated fitness function (typically used in neural networks).

The layout of this paper is as follows: Section 2 lays the foundation by introducing the fundamental components of nonlinear continuum mechanics, with a specific focus on consitutive modeling in finite strain hyperelasticity. Subsequently, in Section 3, we present a comprehensive self-contained overview of Gaussian Process Regression (GPR) and Kriging, providing the necessary background for our methodology. Section 4 carries out the calibration of the Kriging-based surrogate models by utilizing synthetic data derived from well-established ground truth strain energy densities, including Mooney-Rivlin, Gent, Arruday-Boyce, and others. Finally, in Section 5, we demonstrate the implementation of these surrogate models within a 3D Finite Element computational platform. Rigorous assessments are performed to evaluate the accuracy of these models in various challenging scenarios, comparing displacement and stress fields with their respective ground-truth analytical model counterparts. Notably, the examples encompass intricate cases involving extreme twisting and buckling instabilities in a thin-walled structure, showcasing the practical applicability and robustness of the proposed approach.

Notation: Throughout this paper, $\boldsymbol{A} : \boldsymbol{B} = A_{IJ}B_{IJ}$, $\forall \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{3\times3}$, and the use of repeated indices implies summation. The tensor product is denoted by $\otimes$ and the second order identity tensor by $\boldsymbol{I}$. The tensor cross product operation $\times$ between two artibrary second order tensor $\boldsymbol{A}$ and $\boldsymbol{B}$ entails $[\boldsymbol{A} \times \boldsymbol{B}]_{IJ} = \mathcal{E}_{IPQ}\mathcal{E}_{JRS}A_{PR}B_{QS}$. Furthermore, $\boldsymbol{\mathcal{E}}$ represents the third-order alternating tensor. The full and special orthogonal groups in $\mathbb{R}^3$ are represented as $\mathrm{O}(3) = \{\boldsymbol{A} \in \mathbb{R}^{3\times3}, |\, \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}\}$ and $\mathrm{SO}(3) = \{\boldsymbol{A} \in \mathbb{R}^{3\times3}, |\, \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}, \det\boldsymbol{A} = 1\}$, respectively and the set of invertible second order tensors with positive determinant is denoted by $\mathrm{GL}^+(3) = \{\boldsymbol{A} \in \mathbb{R}^{3\times3} |\, \det\boldsymbol{A} > 0\}$.

2. Finite strain elasticity

2.1. Differential governing equations in finite strain elasticity

Let $\mathcal{B}_0 \subset \mathbb{R}^3$ represent the undeformed or material configuration of an elastic solid. We assume the existence of an injective mapping $\boldsymbol{\phi} : \mathcal{B}_0 \to \mathbb{R}^3$ relating every material particle $\boldsymbol{X} \in \mathcal{B}_0$ with those in the deformed or spatial configuration $\mathcal{B} \subset \mathbb{R}^3$, namely $\boldsymbol{x} \in \mathcal{B}$, through the relationship $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{X})$ (see Figure 1). Associated with the mapping $\boldsymbol{\phi}$ it is possible to define its material gradient $\boldsymbol{F} \in \mathrm{GL}^+(3)$, and the cofactor $\boldsymbol{H}$ and determinant J of the latter as

$$\boldsymbol{F} = \partial_{\boldsymbol{X}}\boldsymbol{\phi}, \qquad J = \det\boldsymbol{F} = \frac{1}{6}\boldsymbol{F} : (\boldsymbol{F} \times \boldsymbol{F}), \qquad \boldsymbol{H} = \mathrm{Cof}\boldsymbol{F} = J\boldsymbol{F}^{-T} = \frac{1}{2}\boldsymbol{F} \times \boldsymbol{F}. \tag{1}$$
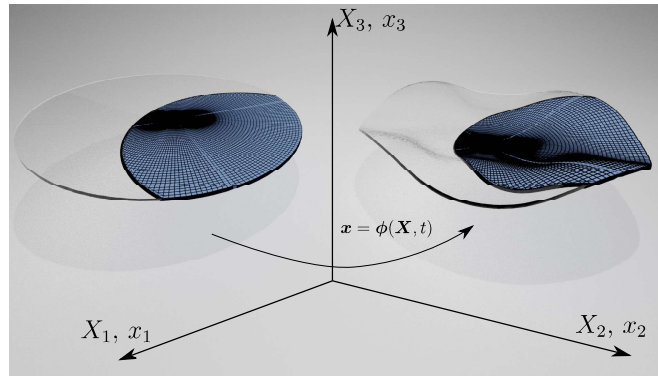

Figure 1: mapping of material quantities to the spatial quantities.

The response of the deformable solid $\mathcal{B}_0$ is governed by the following boundary value problem

$$\begin{aligned} \mathrm{DIV}\boldsymbol{P} + \boldsymbol{f}_0 &= \boldsymbol{0}, & &\text{in } \mathcal{B}_0 \\ \boldsymbol{\phi} &= \boldsymbol{\phi}^{\star}, & &\text{on } \partial_{\phi}\mathcal{B}_0 \\ \boldsymbol{P}\boldsymbol{N} &= \boldsymbol{t}_0, & &\text{in } \partial_t\mathcal{B}_0 \end{aligned} \tag{2}$$

where $\boldsymbol{P}$ represents the first Piola-Kirchhoff stress tensor, related to the deformation gradient tensor $\boldsymbol{F}$ through a suitable constitutive law (see Section 2.2). Furthermore, the undeformed boundary $\partial\mathcal{B}_0$ is divided into non-overlapping regions where Dirichlet and Neumann boundary conditions are imposed, according to $\partial\mathcal{B}_0 = \partial_{\phi}\mathcal{B}_0 \cup \partial_t\mathcal{B}_0$, with $\partial_{\phi}\mathcal{B}_0 \cap \partial_t\mathcal{B}_0 = \emptyset$. In addition, $\boldsymbol{f}_0$ represents the force per unit underformed volume $\mathcal{B}_0$, whilst $\boldsymbol{t}_0$ represents forces per unit under area acting upon $\partial_t\mathcal{B}_0$, being $\boldsymbol{N}$ the outward unit vector in $\boldsymbol{X} \in \partial_t\mathcal{B}_0$.

2.2. The strain energy density in hyperelasticity

In hyperelasticity, the constitutive model of the undeformed solid $\mathcal{B}_0$ is encapsulated in the strain energy density per unit underformed volume, denoted as

$$\Psi : \mathrm{GL}^+(3) \to \mathbb{R}, \qquad (\boldsymbol{F}) \mapsto \Psi(\boldsymbol{F}) \tag{3}$$

Differentiation of the strain energy density with respect to $\boldsymbol{F}$ yields the first Piola-Kirchhoff stress tensor $\boldsymbol{P}$ (2) as

$$\boldsymbol{P} = \partial_{\boldsymbol{F}}\Psi \tag{4}$$

3

The internal energy density $\Psi(\boldsymbol{F})$ must comply with the principle of objectivity or material frame indifference, i.e. invariance with respect to rotations $\boldsymbol{Q} \in \mathrm{SO}(3)$ of the spatial configuration, namely

$$\Psi(\boldsymbol{Q}\boldsymbol{F}) = \Psi(\boldsymbol{F}) \qquad \forall \, \boldsymbol{F} \in \mathrm{GL}^+(3), \, \boldsymbol{Q} \in \mathrm{SO}(3). \tag{5}$$

Furthermore, the strain energy density must adhere to the material symmetry group $\mathcal{G} \subseteq \mathrm{O}(3)$, which characterizes the isotropic or anisotropic nature of the underlying material. This can be mathematically stated as

$$\Psi(\boldsymbol{F}\boldsymbol{Q}) = \Psi(\boldsymbol{F}) \qquad \forall \, \boldsymbol{F} \in \mathrm{GL}^+(3), \, \boldsymbol{Q} \in \mathcal{G} \subseteq \mathrm{O}(3). \tag{6}$$

Crucially, the strain energy density and the first Piola-Kirchhoff stress tensor must vanish in the absence of deformations, namely

$$\Psi(\boldsymbol{F})|_{\boldsymbol{F}=\boldsymbol{I}} = 0, \qquad \partial_{\boldsymbol{F}}\Psi(\boldsymbol{F})|_{\boldsymbol{F}=\boldsymbol{I}} = \boldsymbol{0}. \tag{7}$$

The conditions (5), (6) and (7) can be considered as physical conditions. In addition to these, suitable mathematical conditions are also requested to be satisfied by the strain energy density function. In this regard, the strain energy density typically complies with some mathematical constraints related to the notion of convexity. These conditions will be described in the following Section.

2.3. Some notions on generalised convexity conditions

One of the simplest conditions is that of **convexity** of $\Psi(\boldsymbol{F})$, that is

$$\Psi(\lambda \boldsymbol{F}_1 + (1-\lambda)\boldsymbol{F}_2) \le \lambda\Psi(\boldsymbol{F}_1) + (1-\lambda)\Psi(\boldsymbol{F}_2); \quad \forall \, \boldsymbol{F}_1, \boldsymbol{F}_2 \in \mathrm{GL}^+(3); \quad \lambda \in [0,1], \tag{8}$$

which for functions with first order differentiability can be alternatively written as[2]

$$\Psi(\boldsymbol{F} + \delta\boldsymbol{F}) - \Psi(\boldsymbol{F}) - D\Psi(\boldsymbol{F})[\delta\boldsymbol{F}] \ge 0; \qquad \forall \, \boldsymbol{F} \in \mathrm{GL}^+(3), \delta\boldsymbol{F} \in \mathbb{R}^{3\times3}, \tag{9}$$

and for functions with second order differentiability as

$$D^2\Psi(\boldsymbol{F})[\delta\boldsymbol{F}; \delta\boldsymbol{F}] = \delta\boldsymbol{F} \bullet \boldsymbol{\mathcal{C}} \bullet \delta\boldsymbol{F} \ge 0; \qquad \forall \, \boldsymbol{F} \in \mathrm{GL}^+(3), \delta\boldsymbol{F} \in \mathbb{R}^{3\times3}, \tag{10}$$

which requires positive semi-definiteness of the fourth-order elasticity tensor $\boldsymbol{\mathcal{C}}$, defined as

$$\boldsymbol{\mathcal{C}} = \partial^2_{\boldsymbol{F}\boldsymbol{F}}\Psi. \tag{11}$$

However, convexity away from the origin (i.e. $\boldsymbol{F} \approx \boldsymbol{I}$) is not a suitable physical restriction as it precludes the realistic behavior of materials such as buckling [22]. An alternative mathematical restriction is that of quasiconvexity of $\Psi(\boldsymbol{F})$ [23]. Unfortunately, quasiconvexity is a nonlocal condition that is very difficult, even impossible, to be verified. A necessary restriction implied by quasiconvexity is that of **generalised rank-one convexity** of Ψ. A generalised rank-one convex energy density verifies

$$\Psi(\lambda \boldsymbol{F} + (1-\lambda)\tilde{\boldsymbol{F}}) \le \lambda\Psi(\boldsymbol{F}) + (1-\lambda)\Psi(\tilde{\boldsymbol{F}}); \quad \forall \, \boldsymbol{F} \in \mathrm{GL}^+(3); \quad \lambda \in [0,1], \tag{12}$$

and with $\tilde{\boldsymbol{F}} = \boldsymbol{F} + \delta\boldsymbol{F}$ and $\delta\boldsymbol{F} = \boldsymbol{u} \otimes \boldsymbol{V}$, with $\boldsymbol{u}, \boldsymbol{V}$ any arbitrary vectors. For the case of energies with first order differentiability, generalised rank-one convexity can alternatively be written as

$$\Psi(\boldsymbol{F} + \delta\boldsymbol{F}) - \Psi(\boldsymbol{F}) - D\Psi(\boldsymbol{F})[\delta\boldsymbol{F}] \ge 0; \qquad \delta\boldsymbol{F} = \boldsymbol{u} \otimes \boldsymbol{V}; \qquad \forall \, \boldsymbol{F} \in \mathrm{GL}^+(3), \boldsymbol{u}, \boldsymbol{V} \in \mathbb{R}^3, \tag{13}$$

and for energies with second order differentiability,

$$D^2\Psi(\boldsymbol{F})[\delta\boldsymbol{F}; \delta\boldsymbol{F}] = \delta\boldsymbol{F} \bullet \boldsymbol{\mathcal{C}} \bullet \delta\boldsymbol{F} \ge 0; \qquad \delta\boldsymbol{F} = \boldsymbol{u} \otimes \boldsymbol{V}; \qquad \forall \, \boldsymbol{F} \in \mathrm{GL}^+(3), \boldsymbol{u}, \boldsymbol{V} \in \mathbb{R}^3. \tag{14}$$

Condition (14) is referred to as the Legendre-Hadamard condition or **ellipticity** of Ψ, linked to the propagation of travelling plane wave within the material defined by a vector $\boldsymbol{V}$ and speed c. The

---

[2]$D\Psi(\boldsymbol{F})[\delta\boldsymbol{F}]$ represents the directional derivative of $\Psi(\boldsymbol{F})$, being $D\Psi(\boldsymbol{F})[\delta\boldsymbol{F}] = \partial_{\boldsymbol{F}}\Psi : \delta\boldsymbol{F}$

derivation of the above Legendre-Hadamard condition (14) has its roots in the study of the hyperbolicity (or stability in the quasi-static case) of the system of the generalised Cauchy-Maxwell equations [24] in order to ensure the existence of real wave speeds propagating throughout the domain. Crucially, the *ab initio* existence of real wave speeds for the specific system of governing equations (2) can be monitored via the study of the so-called acoustic tensor $\boldsymbol{Q}_{ac}$ defined as

$$[\boldsymbol{Q}_{ac}(\boldsymbol{F}, \boldsymbol{V})]_{ij} = [\boldsymbol{\mathcal{C}}(\boldsymbol{F})]_{iIjJ} V_I V_J \tag{15}$$

Specifically, the eigenvalues of the acoustic tensor $\boldsymbol{Q}_{ac}$ are proportional to the squared of the volumetric and shear wave speeds of the material. Hence, the above tensor $\boldsymbol{Q}_{ac}$ can be used as a suitable localisation measure for the onset of material instabilities by ensuring that the wave speeds are kept real throughout the entire deformation process. This can be achieved by ensuring that

$$\boldsymbol{u} \cdot \boldsymbol{Q}_{ac}(\boldsymbol{F}, \boldsymbol{V})\boldsymbol{u} \geq 0; \qquad \forall \boldsymbol{F} \in \mathrm{GL}^+(3), \boldsymbol{u}, \boldsymbol{V} \in \mathbb{R}^3 \tag{16}$$

2.4. Polyconvexity

A sufficient and local condition that complies with the rank-one condition in (12)-(14) is that of polyconvexity of Ψ. The strain energy density is defined as polyconvex [22] if there exists a convex and lower semicontinuous function $\mathbb{W} : \mathrm{GL}^+(3) \times \mathrm{GL}^+(3) \times \mathbb{R}^+ \to \mathbb{R} \cup \{+\infty\}$ (in general non-unique) defined as

$$\Psi(\boldsymbol{F}) = \mathbb{W}(\boldsymbol{\mathcal{U}}), \qquad \boldsymbol{\mathcal{U}} = (\boldsymbol{F}, \boldsymbol{H}, J). \tag{17}$$

Notice that convexity of $W(\boldsymbol{\mathcal{U}})$ implies that

$$\mathbb{W}(\lambda \boldsymbol{\mathcal{U}}_1 + (1-\lambda)\boldsymbol{\mathcal{U}}_2) \leq \lambda \mathbb{W}(\boldsymbol{\mathcal{U}}_1) + (1-\lambda)\mathbb{W}(\boldsymbol{\mathcal{U}}_2); \quad \forall \boldsymbol{\mathcal{U}}_1, \boldsymbol{\mathcal{U}}_2 \in \mathrm{GL}^+(3) \times \mathrm{GL}^+(3) \times \mathbb{R}^+; \quad \lambda \in [0,1]. \tag{18}$$

which for functions with sufficient differentiability, can be alternatively written as

$$\mathbb{W}(\boldsymbol{\mathcal{U}} + \delta\boldsymbol{\mathcal{U}}) - \mathbb{W}(\boldsymbol{\mathcal{U}}) - D\mathbb{W}(\boldsymbol{\mathcal{U}})[\delta\boldsymbol{\mathcal{U}}] \geq 0; \qquad \forall \boldsymbol{\mathcal{U}} \in \mathrm{GL}^+(3) \times \mathrm{GL}^+(3) \times \mathbb{R}^+, \delta\boldsymbol{\mathcal{U}} \in \mathbb{R}^{3\times3} \times \mathbb{R}^{3\times3} \times \mathbb{R}, \tag{19}$$

or even in terms of the Hessian operator $\partial^2_{\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{U}}}\mathbb{W}$ as

$$D^2\mathbb{W}(\boldsymbol{\mathcal{U}})[\delta\boldsymbol{\mathcal{U}}; \delta\boldsymbol{\mathcal{U}}] = \delta\boldsymbol{\mathcal{U}} \bullet \partial^2_{\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{U}}}\mathbb{W} \bullet \delta\boldsymbol{\mathcal{U}} \geq 0; \qquad \forall \boldsymbol{\mathcal{U}} \in \mathrm{GL}^+(3) \times \mathrm{GL}^+(3) \times \mathbb{R}^+, \delta\boldsymbol{\mathcal{U}} \in \mathbb{R}^{3\times3} \times \mathbb{R}^{3\times3} \times \mathbb{R}. \tag{20}$$

As presented by Ball in [22], polyconvexity in conjunction with suitable growth conditions, ensures the existence of minimisers in nonlinear elasticity.

2.5. Invariant-based hyperelasticity

A simple manner to accommodate the principle of objectivity or material frame indifference and the requirement of material symmetry is through the dependence of the strain energy density function $\Psi(\boldsymbol{F})$ with respect to invariants of the right Cauchy-Green deformation gradient tensor $\boldsymbol{C} = \boldsymbol{F}^T\boldsymbol{F}$. Let $\mathbf{I} = \{I_1, I_2, \ldots, I_n\}$, represent the n objective invariants required to characterise a given material symmetry group $\mathcal{G}$. Then, it is possible to express the strain energy density $\Psi(\boldsymbol{F})$ equivalently as

$$\Psi(\boldsymbol{F}) = U(\mathbf{I}) \tag{21}$$

Application of the chain rule into equation (4) yields the first Piola-Kirchhoff stress tensor in terms of the derivatives of $U(\mathbf{I})$ as

$$\boldsymbol{P} = \sum_{i=1}^n \left(\partial_{I_i} U \right) \partial_{\boldsymbol{F}} I_i \tag{22}$$

Furthermore, application of the chain rule over equation (11) permits to obtain the elasticity tensor $\boldsymbol{\mathcal{C}}$ in terms of the derivatives of $U(\mathbf{I})$ as

$$\boldsymbol{\mathcal{C}} = \sum_{i=1}^n \sum_{j=1}^n \left(\partial^2_{I_i I_j} U \right) \partial_{\boldsymbol{F}} I_i \otimes \partial_{\boldsymbol{F}} I_j + \sum_{i=1}^n (\partial_{I_i} U) \partial^2_{\boldsymbol{F}\boldsymbol{F}} I_i. \tag{23}$$

2.5.1. Isotropic elasticity

For the case of isotropic elasticity, the invariants required to characterise this material symmetry group, and the first derivatives of the latter with respect to $\boldsymbol{F}$ (featuring in the definition of $\boldsymbol{P}$ in (22)) and second derivatives with respect to $\boldsymbol{F}$ (featuring in the definition of $\boldsymbol{\mathcal{C}}$ in (23)) are

$$
\begin{aligned}
& I_1 = \boldsymbol{F} : \boldsymbol{F} = \mathrm{tr}(\boldsymbol{C}), && I_2 = \boldsymbol{H} : \boldsymbol{H} = \mathrm{tr}(\mathrm{Cof}\boldsymbol{C}), && I_3 = J = (\det\boldsymbol{C})^{1/2}, \\
& \partial_{\boldsymbol{F}} I_1 = 2\boldsymbol{F}, && \partial_{\boldsymbol{F}} I_2 = 2\boldsymbol{H} \boldsymbol{\times} \boldsymbol{F}, && \partial_{\boldsymbol{F}} I_3 = \boldsymbol{H}, && (24) \\
& \partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_1 = 2\boldsymbol{\mathcal{I}}, && \partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_2 = 2\boldsymbol{F} \boldsymbol{\times} \boldsymbol{\mathcal{I}} \boldsymbol{\times} \boldsymbol{F} + 2\boldsymbol{\mathcal{I}} \boldsymbol{\times} \boldsymbol{H}, && \partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_3 = \boldsymbol{H} \otimes \boldsymbol{H} + \boldsymbol{\mathcal{I}} \boldsymbol{\times} \boldsymbol{F}
\end{aligned}
$$

with $\boldsymbol{\mathcal{I}}$ the fourth-order tensor defined as $[\boldsymbol{\mathcal{I}}]_{ijkl} = \delta_{ik}\delta_{jl}$, being δ_{ij} the ij^{th} component of the second order identity tensor, and with $[\boldsymbol{\mathcal{A}} \boldsymbol{\times} \boldsymbol{A}]_{iIjJ} = \mathcal{E}_{jkl}\mathcal{E}_{JKL}\mathcal{A}_{iIkK}A_{lL}$ and $[\boldsymbol{A} \boldsymbol{\times} \boldsymbol{\mathcal{A}}]_{iIjJ} = \mathcal{E}_{ikl}\mathcal{E}_{IKL}A_{kK}\mathcal{A}_{lLjJ}$, for $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{3\times3\times3\times3}$ and $\boldsymbol{A} \in \mathbb{R}^{3\times3}$. Inserting the expressions in (24) into (22) yield the following expression for the first Piola-Kirchhoff stress tensor $\boldsymbol{P}$

$$
\boldsymbol{P} = \left(\partial_{I_1} U\right) 2\boldsymbol{F} + \left(\partial_{I_2} U\right) 2\boldsymbol{H} \boldsymbol{\times} \boldsymbol{F} + \left(\partial_{I_3} U\right) \boldsymbol{H} \tag{25}
$$

Notice that for the elasticity tensor, all the terms featuring in its generic expression in equation (23), including the first and second derivatives of I_i with respect to $\boldsymbol{F}$ can be found in (24).

2.5.2. Transversely isotropic elasticity

For the case of transverse isotropy, where the preferred direction $\boldsymbol{N}$ is perpendicular to the plane of isotropy of the material, in addition to invariants $\{I_1, I_2, I_3\}$ in (24) and their respective first and second derivatives with respect to $\boldsymbol{F}$, two additional invariants are needed, found below

$$
\begin{aligned}
& I_4 = \boldsymbol{F}\boldsymbol{N} \cdot \boldsymbol{F}\boldsymbol{N} = \mathrm{tr}\left(\boldsymbol{C}\boldsymbol{N} \otimes \boldsymbol{N}\right), && I_5 = \boldsymbol{H}\boldsymbol{N} \cdot \boldsymbol{H}\boldsymbol{N} = \mathrm{tr}(\mathrm{Cof}\boldsymbol{C}\boldsymbol{N} \otimes \boldsymbol{N}), \\
& \partial_{\boldsymbol{F}} I_4 = 2\boldsymbol{F}\boldsymbol{N} \otimes \boldsymbol{N}, && \partial_{\boldsymbol{F}} I_5 = 2\left(\boldsymbol{H}\boldsymbol{N} \otimes \boldsymbol{N}\right) \boldsymbol{\times} \boldsymbol{F}, && (26) \\
& [\partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_4]_{iIjJ} = 2\delta_{ij}[\boldsymbol{N} \otimes \boldsymbol{N}]_{IJ}, && \partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_5 = \boldsymbol{F} \boldsymbol{\times} \partial_{\boldsymbol{H}\boldsymbol{H}}^2 I_5 \boldsymbol{\times} \boldsymbol{F} + \boldsymbol{\mathcal{I}} \boldsymbol{\times} \partial_{\boldsymbol{H}} I_5
\end{aligned}
$$

with $\partial_{\boldsymbol{H}\boldsymbol{H}}^2 I_5 = \partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_4$ and $\partial_{\boldsymbol{H}} I_5 = 2\left(\boldsymbol{H}\boldsymbol{N} \otimes \boldsymbol{N}\right)$. In this case, the first Piola Kirchhoff stress tensor $\boldsymbol{P}$ adopts the following expression

$$
\boldsymbol{P} = \left(\partial_{I_1} U\right) 2\boldsymbol{F} + \left(\partial_{I_2} U\right) 2\boldsymbol{H} \boldsymbol{\times} \boldsymbol{F} + \left(\partial_{I_3} U\right) \boldsymbol{H} + \left(\partial_{I_4} U\right) 2\boldsymbol{F}\boldsymbol{N} \otimes \boldsymbol{N} + \left(\partial_{I_5} U\right) 2\left(\boldsymbol{H}\boldsymbol{N} \otimes \boldsymbol{N}\right) \boldsymbol{\times} \boldsymbol{F} \tag{27}
$$

Notice that for the elasticity tensor $\boldsymbol{\mathcal{C}}$, all the terms featuring in its generic expression in equation (23), including the first and second derivatives of I_i with respect to $\boldsymbol{F}$ can be found in (24) and (26).

2.5.3. Orthotropy

For the case of orthotropy, there are three preferred directions $\{\boldsymbol{N}_1, \boldsymbol{N}_2, \boldsymbol{N}_3\}$ characterizing this material symmetry group, being the three of them unitary vectors mutually orthogonal between each other. Then, in addition to invariants $\{I_1, I_2, I_3\}$ in (24) and their respective first and second derivatives with respect to $\boldsymbol{F}$, three additional invariants are needed, found below

$$
\begin{aligned}
& I_4 = \boldsymbol{F}\boldsymbol{N}_1 \cdot \boldsymbol{F}\boldsymbol{N}_1, && I_5 = \boldsymbol{H}\boldsymbol{N}_1 \cdot \boldsymbol{H}\boldsymbol{N}_1, && I_6 = \boldsymbol{F}\boldsymbol{N}_2 \cdot \boldsymbol{F}\boldsymbol{N}_2, \\
& \partial_{\boldsymbol{F}} I_4 = 2\boldsymbol{F}\boldsymbol{N}_1 \otimes \boldsymbol{N}_1, && \partial_{\boldsymbol{F}} I_5 = 2\left(\boldsymbol{H}\boldsymbol{N}_1 \otimes \boldsymbol{N}_1\right) \boldsymbol{\times} \boldsymbol{F}, && \partial_{\boldsymbol{F}} I_6 = 2\boldsymbol{F}\boldsymbol{N}_2 \otimes \boldsymbol{N}_2 \\
& [\partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_4]_{iIjJ} = 2\delta_{ij}[\boldsymbol{N}_1 \otimes \boldsymbol{N}]_{IJ}, && \partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_5 = \boldsymbol{F} \boldsymbol{\times} \partial_{\boldsymbol{H}\boldsymbol{H}}^2 I_5 \boldsymbol{\times} \boldsymbol{F} + \boldsymbol{\mathcal{I}} \boldsymbol{\times} \partial_{\boldsymbol{H}} I_5, && [\partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_6]_{iIjJ} = 2\delta_{ij}[\boldsymbol{N}_2 \otimes \boldsymbol{N}_2]_{IJ}
\end{aligned}
$$
(28)

with $\partial_{\boldsymbol{H}\boldsymbol{H}}^2 I_5 = \partial_{\boldsymbol{F}\boldsymbol{F}}^2 I_4$ and $\partial_{\boldsymbol{H}} I_5 = 2\left(\boldsymbol{H}\boldsymbol{N}_1 \otimes \boldsymbol{N}_1\right)$. In this case, the first Piola Kirchhoff stress tensor $\boldsymbol{P}$ adopts the following expression

$$
\begin{aligned}
\boldsymbol{P} = {} & \left(\partial_{I_1} U\right) 2\boldsymbol{F} + \left(\partial_{I_2} U\right) 2\boldsymbol{H} \boldsymbol{\times} \boldsymbol{F} + \left(\partial_{I_3} U\right) \boldsymbol{H} \\
& + \left(\partial_{I_4} U\right) 2\boldsymbol{F}\boldsymbol{N}_1 \otimes \boldsymbol{N}_1 + \left(\partial_{I_5} U\right) 2\left(\boldsymbol{H}\boldsymbol{N}_1 \otimes \boldsymbol{N}_1\right) \boldsymbol{\times} \boldsymbol{F} + \left(\partial_{I_6} U\right) 2\boldsymbol{F}\boldsymbol{N}_2 \otimes \boldsymbol{N}_2.
\end{aligned}
\tag{29}
$$

Notice that for the elasticity tensor $\boldsymbol{\mathcal{C}}$, all the terms featuring in its generic expression in equation (23), including the first and second derivatives of I_i with respect to $\boldsymbol{F}$ can be found in (24) and (28).

Remark 1: While it may seem logical to incorporate the invariant $I_7 = \boldsymbol{H}\boldsymbol{N}_2 \cdot \boldsymbol{H}\boldsymbol{N}_2$ into an irreducible basis of invariants for orthotropic materials, this is not required. Only three anisotropic invariants are needed for a comprehensive characterization of the orthotropic material symmetry group. This aligns with the principles of defining irreducible bases of invariants using structured tensors, as elucidated by [25]. The structured tensor characterizing the orthotropic material symmetry group is as follows:

$$\boldsymbol{G} = \gamma_1 \boldsymbol{N}_1 \otimes \boldsymbol{N}_1 + \gamma_2 \boldsymbol{N}_2 \otimes \boldsymbol{N}_2 + \gamma_3 \boldsymbol{N}_3 \otimes \boldsymbol{N}_3 \tag{30}$$

with $\gamma_1, \gamma_2, \gamma_3, \in \mathbb{R}$. This tensor permits to define the two anisotropic invariants typically needed for this material symmetry group as

$$I_4 = \operatorname{tr}\left(\boldsymbol{F}^T \boldsymbol{F} \boldsymbol{G}\right); \qquad I_5 = \operatorname{tr}\left(\boldsymbol{H}^T \boldsymbol{H} \boldsymbol{G}\right) \tag{31}$$

Notice that the tensor $\boldsymbol{G}$ in (30) can be conveniently re-written as

$$\boldsymbol{G} = \gamma_3 \boldsymbol{I} + \alpha_1 \hat{\boldsymbol{G}} \tag{32}$$

with $\alpha_1 = (\gamma_1 - \gamma_3)$ and with the new structured tensor $\hat{\boldsymbol{G}}$ defined as

$$\hat{\boldsymbol{G}} = \left(\boldsymbol{N}_1 \otimes \boldsymbol{N}_1 + \beta \boldsymbol{N}_2 \otimes \boldsymbol{N}_2\right), \tag{33}$$

with $\beta = \frac{\alpha_2}{\alpha_1}$, and $\alpha_2 = (\gamma_2 - \gamma_3)$. Making use of (32) enables to re-write both I_4 and I_5 in (31) as

$$I_4 = \operatorname{tr}\left(\boldsymbol{F}^T \boldsymbol{F} \boldsymbol{G}\right) = \gamma_3 I_1 + \alpha_1 \hat{I}_4; \qquad \hat{I}_5 = \operatorname{tr}\left(\boldsymbol{H}^T \boldsymbol{H} \boldsymbol{G}\right) = \gamma_3 I_2 + \alpha_1 \hat{I}_5 \tag{34}$$

with

$$\hat{I}_4 = \operatorname{tr}\left(\boldsymbol{F}^T \boldsymbol{F} \hat{\boldsymbol{G}}\right); \qquad \hat{I}_5 = \operatorname{tr}\left(\boldsymbol{H}^T \boldsymbol{H} \hat{\boldsymbol{G}}\right) \tag{35}$$

Therefore, an irreducible basis of invariants comprises the set $\{I_1, I_2, I_3, \hat{I}_4, \hat{I}_5\}$, totaling five distinct invariants. Notably, both $\hat{I}_4$ and $\hat{I}_5$ depend on the structured tensor $\hat{\boldsymbol{G}}$ in (33), which, in turn, relies on the parameter β. Therefore, when including these five invariants and γ_1, the total count becomes six. This numeric correspondence precisely matches the count of invariants essential for the irreducible basis in (28), required for a comprehensive characterization of the orthotropic material symmetry group.

2.6. Application to composite materials

Section 2.5 considers the case of phenomenological strain energy densities, expressed in terms of principal invariants. However, in the context of composite materials, computing the effective or homogenized strain energy density, based on the individual phenomenological strain energy densities of each constituent, necessitates the use of a homogenization technique. In this section, we specifically focus on two distinctive type of composite materials. These are:

- Rank-one laminates. These laminates are composed of two constituents intercalated with the lamination orientation being perpendicular to a vector $\boldsymbol{N}$ (see Figure 2a).

- Representative Volume Element (RVE) comprising elastomeric matrix with spherical inclusion (see Figure 2b).
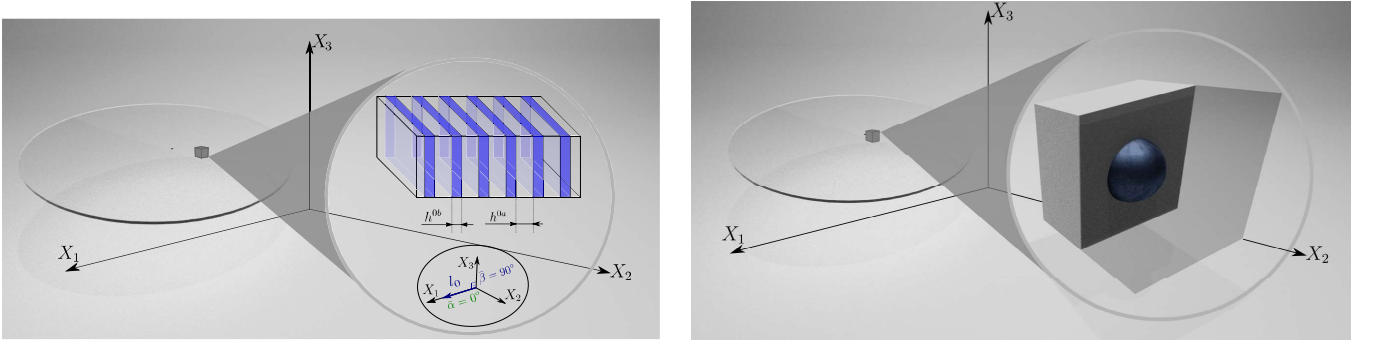
Figure 2: (a) Rank-one laminate composite material; (b) Composite material comprising an elastomer matrix endowed with spherical inclusions.

2.6.1. Rank-one laminate composite materials

For these composite materials, rank-n homogenisation theory [6] permits to express the macroscopic deformation gradient tensor $\boldsymbol{F}$ in terms of their microscopic counterparts $\boldsymbol{F}^a$ and $\boldsymbol{F}^b$, according to

$$\boldsymbol{F} = c^a \boldsymbol{F}^a + c^b \boldsymbol{F}^b; \qquad [\![\boldsymbol{F}]\!] \times \boldsymbol{N} = \boldsymbol{0}, \tag{36}$$

where the indices a and b differentiate the constituents and c^a and c^b denote their respective volume fractions, with $c^b = 1 - c^a$. A possible definition for $\boldsymbol{F}^a$ and $\boldsymbol{F}^b$ compatible with (36) is the following

$$\boldsymbol{F}^a\left(\boldsymbol{F}, \boldsymbol{\alpha}\right) = \boldsymbol{F} + c^b \boldsymbol{\alpha} \otimes \boldsymbol{N}; \qquad\qquad \boldsymbol{F}^b\left(\boldsymbol{F}, \boldsymbol{\alpha}\right) = \boldsymbol{F} - c^a \boldsymbol{\alpha} \otimes \boldsymbol{N}. \tag{37}$$

where $\boldsymbol{\alpha}$ represents the amplitude vector, which needs to be determined. This can be done by postulating the effective energy $\Psi(\boldsymbol{F})$ as

$$\Psi\left(\boldsymbol{F}\right) = \arg\min_{\boldsymbol{\alpha}}\{\hat{\Psi}\left(\boldsymbol{F}, \boldsymbol{\alpha}\right)\}; \qquad \hat{\Psi}\left(\boldsymbol{F}, \boldsymbol{\alpha}\right) = c^a \Psi^a\left(\boldsymbol{F}^a\left(\boldsymbol{F}, \boldsymbol{\alpha}\right)\right) + c^b \Psi^b\left(\boldsymbol{F}^b\left(\boldsymbol{F}, \boldsymbol{\alpha}\right)\right). \tag{38}$$

The stationary condition of $\hat{\Psi}$ with respect to $\boldsymbol{\alpha}$ yields

$$D\hat{\Psi}\left[\delta\boldsymbol{\alpha}\right] = c^a c^b \left(\boldsymbol{P}^a - \boldsymbol{P}^b\right) : \left(\delta\boldsymbol{\alpha} \otimes \boldsymbol{N}\right) = 0 \qquad \forall \delta\boldsymbol{\alpha} \qquad \implies \qquad [\![\boldsymbol{P}]\!]\boldsymbol{N} = \boldsymbol{0}. \tag{39}$$

which represent a nonlinear vector equation from which $\boldsymbol{\alpha}$ an be obtained. Finally, computation of $\boldsymbol{\alpha}$ permits to obtain the effective first Piola-Kirchhoff stress tensor, as

$$\boldsymbol{P} = c^a \boldsymbol{P}^a + c^b \boldsymbol{P}^b; \qquad \boldsymbol{P}^a = \partial_{\boldsymbol{F}^a} \Psi(\boldsymbol{F}^a\left(\boldsymbol{F}, \boldsymbol{\alpha}\right)), \qquad \boldsymbol{P}^b = \partial_{\boldsymbol{F}^b} \Psi(\boldsymbol{F}^b\left(\boldsymbol{F}, \boldsymbol{\alpha}\right)) \tag{40}$$

The specific form of the strain energy of each of the individual phases, namely $\Psi^a(\boldsymbol{F})$ and $\Psi^b(\boldsymbol{F})$ used in this work can be seen in Appendix A .

2.7. RVE with elastomeric matrix and spherical inclusion

In this case, the microscopic deformation gradient tensor $\boldsymbol{F}_\mu$ is defined in terms of the microscopic current position $\boldsymbol{x}_\mu$ as

$$\boldsymbol{F}_\mu = \frac{\partial \boldsymbol{x}_\mu}{\partial \boldsymbol{X}_\mu}. \tag{41}$$

being $\boldsymbol{X}_\mu$ the material counterpart of $\boldsymbol{x}_\mu$. In the context of periodic first-order homogenization [?], $\boldsymbol{x}_\mu$ can be decomposed as

$$\boldsymbol{x}_\mu = \boldsymbol{\alpha} + \boldsymbol{F}\boldsymbol{X}_\mu, \tag{42}$$

where $\boldsymbol{F}$ represents the macroscopic deformation gradient tensor and $\boldsymbol{\alpha} : \mathcal{B}_{0_\mu} \to \mathbb{R}^3$ represent the microscopic fluctuations. The latter are subjected to periodic boundary conditions of the form $[\![\boldsymbol{\alpha}]\!] = \boldsymbol{0}$, where $[\![(\cdot)]\!] = (\cdot)^+ - (\cdot)^-$ represents the jump of the field $(\cdot)$ across opposite boundaries of $\mathcal{B}_{0_\mu}$. The microscopic field $\boldsymbol{F}_\mu$ can be related with its macroscopic counterpart $\boldsymbol{F}$ as

$$\boldsymbol{F} = \frac{1}{V_\mu} \int_{\mathcal{B}_{0_\mu}} \boldsymbol{F}_\mu(\boldsymbol{X}_\mu) \, dV_\mu, \qquad V_\mu = \int_{\mathcal{B}_{0_\mu}} dV_\mu \tag{43}$$

8

We assume the existence of a strain energy density at both macro and microscales, denoted as $\Psi(\boldsymbol{F})$ and $\Psi_\mu(\boldsymbol{F}_\mu)$, respectively. Then, the effective or macroscopic strain energy density is postulated as

$$\underbrace{\Psi(\boldsymbol{F})V_\mu}_{\Pi(\boldsymbol{\alpha})} = \inf_{\boldsymbol{\alpha} \in \mathbb{V}} \left\{ \int_{\mathcal{B}_{0_\mu}} \Psi_\mu(\boldsymbol{F}_\mu)\, dV_\mu \right\}, \qquad \mathbb{V} = \left\{ \boldsymbol{\alpha} : (\boldsymbol{\alpha})_i \in H_0^1(\mathcal{B}_{0_\mu}),\ [\![\boldsymbol{\alpha}]\!] = \boldsymbol{0},\ \boldsymbol{\alpha} = \boldsymbol{0}\ \text{on}\ \partial_{\boldsymbol{\alpha}}\mathcal{B}_{0_\mu} \right\}$$

(44)

where $\partial_{\boldsymbol{\alpha}}\mathcal{B}_{0_\mu} \subset \partial\mathcal{B}_{0_\mu}$ is the region where zero Dirichlet boundary conditions are prescribed for $\boldsymbol{\alpha}$, preventing rigid body motions. The stationary conditions of the homogenized energy in (44) yields

$$D\Pi(\boldsymbol{\alpha})[\boldsymbol{v}_x] = \int_{\mathcal{B}_{0_\mu}} \boldsymbol{P}_\mu : \frac{\partial \boldsymbol{v}_x}{\partial \boldsymbol{X}_\mu}\, dV_\mu = 0,$$

(45)

which permit to obtain the micro-fluctuations $\boldsymbol{\alpha}$. In (45), $\boldsymbol{P}_\mu$ can be obtained through the following relationship

$$\boldsymbol{P}_\mu = \frac{\partial \Psi_\mu(\boldsymbol{F}_\mu)}{\partial \boldsymbol{F}_\mu},$$

(46)

where $\Psi_\mu(\boldsymbol{F}_\mu)$ adopts a different form for the matrix and the spherical inclusion (see Figure 2b). Their specificly, expressions can be found in Appendix A. Finally, use of the Hill-Mandel principle [?] permits to obtain the effective or homogenised first Piola-Kirchhoff stress tensor $\boldsymbol{P}$ as

$$\boldsymbol{P} = \frac{1}{V_\mu} \int_{\mathcal{B}_{0_\mu}} \boldsymbol{P}_\mu\, dV_\mu.$$

(47)

3. Kriging predictors

In the context of computer experiments, metamodelling or surrogate modelling involves substituting an expensive-to-evaluate model or simulator $U = \mathcal{M}(\mathbf{I})$ with a computationally efficient emulator $\hat{\mathcal{M}}(\mathbf{I})$. Both the simulator and emulator share the same input space $\mathcal{D}_I \subseteq \mathbb{R}^n$ and output space $\mathcal{D}_U \subseteq \mathbb{R}$. In our specific context, $\mathcal{M}$ represents the response of a ground truth strain energy density U, which is dependent on principal invariants $\mathbf{I}$ (as discussed in Section 2.5). This justifies replacing the input field, typically denoted as $\boldsymbol{x}$ in the literature, with $\mathbf{I}$, and the output, usually denoted as y, with U. Since the strain energy U is scalar in nature, the theoretical developments presented in this paper will solely focus on the case of scalar outputs.

In this paper, we make use of Kriging models [26, 27] which are also known as Kriging modelling. This particular type of emulator has gained increasing popularity in the last decade due to its flexibility in capturing nonlinear functions and its ability to provide statistical information on the predicted output [19, 21].

The construction of the emulator typically consists of three stages:

- The first stage involves the generation of a finite number, m, of observations

$$\mathscr{U} = \left\{ U^{(i)} = \mathcal{M}(\mathbf{I}^{(i)}),\ i = 1, \ldots, m \right\}$$

evaluated onto an experimental design $\mathscr{I} = \left\{ \mathbf{I}^{(i)}, \ldots, \mathbf{I}^{(m)} \right\}$. These observations are used in the process of searching the space of possible functions that emulate the behaviour of the simulator $\mathcal{M}$. This process is known as supervised learning in the field of statistical learning [11]. In this paper, the so-called Bayesian prediction methodology presented by Rasmussen and Williams [11] and Santner et al. [16] is adopted. In this approach a prior probability is assigned to functions based upon prior beliefs known around the model, thus promoting the most suitable functions. In practice it is difficult to know what type of functions are the most suitable ones, since the function $\mathcal{M}$ is a priori unknown. For this reason, the approach adopted in Kriging modelling consists in assuming an a priori structure for the emulator depending on a set of unknown parameters.

9

- At a second stage, optimization algorithms are used to find the optimal set of parameters that best fit the model to the observations. This stage is typically called model parameter estimation and metamodel training. All the technical steps required to carry out this stage are carefully described throughout Sections 3.1-3.3.

- Finally, the third stage consists of the validation of the surrogate model by estimating the accuracy via global and local indicators.

3.1. Kriging based prediction

Kriging modelling assumes that the output $U = \mathscr{M}(\mathbf{I})$[3] is a sample path of the Kriging to be characterised by

$$U = \boldsymbol{g}(\mathbf{I}) \cdot \boldsymbol{\beta} + Z(\mathbf{I}), \tag{48}$$

where $\boldsymbol{g}(\mathbf{I}) \cdot \boldsymbol{\beta}$ denotes the prior mean of the Kriging and corresponds to a linear regression model on a given functional basis $\{g_i, i = 1, \ldots, p\} \in \mathcal{L}_2(\mathcal{D}_I, \mathbb{R})$. The second term, $Z(\mathbf{I})$, denotes a zero mean Kriging with a constant variance σ_U^2 and a stationary autocovariance function as follows

$$C(\mathbf{I}, \mathbf{I}') = \sigma_U^2 \mathcal{R}(\mathbf{I}, \mathbf{I}', \boldsymbol{\theta}), \tag{49}$$

where $\mathcal{R}$ is a symmetric positive definite autocorrelation function, and $\boldsymbol{\theta}$, the vector of hyperparameters.

Remark 2: The stochastic process described by equation (48) represents the prior knowledge of the simulator. This prior embodies our beliefs regarding the types of functions we anticipate observing before any data is observed [11]. Consequently, the choice of the correlation function $\mathcal{R}$ should align with the known properties of the simulator, such as its derivability or periodicity. Various correlation functions are commonly used as examples, including the following:

$$\mathcal{R}(\boldsymbol{I}, \boldsymbol{I}', \boldsymbol{\theta}) = \begin{cases} \exp\left(\sum_{k=1}^{n} -\theta_k \left|I_k - I_k'\right|\right) & \text{Exponential kernel} \\[2ex] \exp\left(\sum_{k=1}^{n} -\theta_k \left|I_k - I_k'\right|^2\right) & \text{Gaussian kernel} \\[2ex] \left.\begin{array}{l} \prod_{k=1}^{n} 1 - 1.5\eta_k + 0.5\eta_k^3 \\[1ex] \eta_k = \min\{1, \theta_k \left|I_k - I_k'\right|\} \end{array}\right\} & \text{Spherical kernel} \\[3ex] \left.\begin{array}{l} \prod_{k=1}^{n} \chi_k(\eta_k) \\[1ex] \chi_k(\eta_k) = \begin{cases} 1 - 15\eta_k^2 + 30\eta_k^3, & 0 \le \eta_k \le 0.2 \\ 1.25(1 - \eta_k)^3, & 0.2 < \eta_k < 1 \\ 0, & \eta_k \ge 1 \end{cases} \\[3ex] \eta_k = \min\{1, \theta_k, \left|I_k - I_k'\right|\} \end{array}\right\} & \text{Spline kernel} \end{cases} \tag{50}$$

The construction of a Kriging model consist of the two-stage framework described in the upcoming subsections 3.2 and 3.3.

---

[3] Notice that we are making a slight abuse of notation. The symbol U now represents a probabilistic model, in contrast to the deterministic nature of the strain energy denoted as U in Section 2.

3.2. The conditional distribution of the prediction

The Bayesian prediction methodology assumes that observations gathered in the vector

$$\boldsymbol{U} = \left[U(\boldsymbol{I}^{(1)}), U(\boldsymbol{I}^{(2)}), \ldots, U(\boldsymbol{I}^{(m)}) \right]^T$$

along with any other unobserved value $U(\boldsymbol{I})$ are a realisation of a random vector distributed according to a joint parametric distribution. This section aims to obtain the random prediction for this unobserved quantity by exploiting this statistical dependency. The hypothesis of Gaussianity of $Z(\mathbf{I})$ on the right-hand side of equation (48) and the linearity of the regression model permits to conclude that the vector of observations $\boldsymbol{U}$ is also Gaussian, with

$$\boldsymbol{U} \sim \mathcal{N}(\boldsymbol{G}\boldsymbol{\beta}, \sigma_U^2 \boldsymbol{R}), \tag{51}$$

being $\boldsymbol{G}$ and $\boldsymbol{R}$ the regression and correlation matrices, defined as

$$G_{ij} = g_j(\mathbf{I}^{(i)}), \quad i = 1, \ldots, m, \ j = 1, \ldots, p, \tag{52}$$

and

$$R_{ij} = \mathcal{R}(\mathbf{I}^{(i)}, \mathbf{I}^{(j)}, \boldsymbol{\theta}) \quad i = 1, \ldots, m, \ j = 1, \ldots, m. \tag{53}$$

Similarly, a new random vector composed of the set of observations $\boldsymbol{U}$ and any other unobserved value $U(\boldsymbol{I})$ is jointly Gaussian distributed as

$$\left\{ \begin{matrix} \boldsymbol{U} \\ U(\boldsymbol{I}) \end{matrix} \right\} \sim \mathcal{N}\left(\left\{ \begin{matrix} \boldsymbol{G} \\ \boldsymbol{g}(\mathbf{I})^T \end{matrix} \right\} \boldsymbol{\beta}, \sigma_U^2 \begin{bmatrix} \boldsymbol{R} & \boldsymbol{r}(\mathbf{I}) \\ \boldsymbol{r}(\mathbf{I})^T & 1 \end{bmatrix} \right), \tag{54}$$

where $\boldsymbol{g}(\mathbf{I})$ is the vector of regressors evaluated at $\mathbf{I}$ and $\boldsymbol{r}(\mathbf{I})$ is the vector of cross-correlations between the observations and prediction given by

$$r_i(\mathbf{I}) = \mathcal{R}(\mathbf{I}^{(i)}, \mathbf{I}, \boldsymbol{\theta}) \quad i = 1, \ldots, m. \tag{55}$$

Assuming that the autocovariance function given by equation (49) is known, the conditional distribution of the prediction $\hat{U}(\boldsymbol{I}) = U(\boldsymbol{I})|\boldsymbol{U}$ is given by the best linear unbiased predictor (BLUP) theorem [16]. According to the BLUP the unobserved quantity of interest $U(\boldsymbol{I}) = \mathscr{M}(\mathbf{I})$ under the prior model in equation (48) is the Gaussian random variable $\hat{U}$ with mean

$$\mu_{\hat{U}}(\mathbf{I}) = \boldsymbol{g}(\mathbf{I}) \cdot \hat{\boldsymbol{\beta}} + \boldsymbol{r}(\mathbf{I}) \cdot \boldsymbol{R}^{-1}\left(\boldsymbol{U} - \boldsymbol{G}\hat{\boldsymbol{\beta}} \right), \tag{56}$$

and variance

$$\sigma_{\hat{U}}^2(\mathbf{I}) = \sigma_U^2 \left(1 - \boldsymbol{r}(\mathbf{I}) \cdot \boldsymbol{R}^{-1}\boldsymbol{r}(\mathbf{I}) + \boldsymbol{u}(\mathbf{I}) \cdot \left(\boldsymbol{G}^T \boldsymbol{R}^{-1} \boldsymbol{G} \right)^{-1} \boldsymbol{u}(\mathbf{I}) \right), \tag{57}$$

where

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{G}^T \boldsymbol{R}^{-1} \boldsymbol{G} \right)^{-1} \boldsymbol{G}^T \boldsymbol{R}^{-1} \boldsymbol{U}, \tag{58}$$

is the generalised least-squares estimate of the underlying regression problem, and

$$\boldsymbol{u}(\mathbf{I}) = \boldsymbol{G}^T \boldsymbol{R}^{-1} \boldsymbol{r}(\mathbf{I}) - \boldsymbol{g}(\mathbf{I}). \tag{59}$$

The derivations leading to equations (56)-(59), representing the mean value and variance of the BLUP (also denoted as universal Kriging predictor [16]), are considered standard procedures. For readers who are not familiar with these derivations, we recommend referring to the comprehensive treatment provided in Reference [28], which offers a detailed explanation of the underlying mathematical principles and methodologies.

Remark 3: As shown in Figure 3c the BLUP is an exact interpolator of the observations $\boldsymbol{U}$. To prove it, lets consider the prediction for the observation i

$$\begin{aligned}
\mu_{\hat{U}}(\mathbf{I}^{(i)}) &= \boldsymbol{g}(\mathbf{I}^{(i)}) \cdot \hat{\boldsymbol{\beta}} + \boldsymbol{r}(\mathbf{I}^{(i)}) \cdot \boldsymbol{R}^{-1}\left(\boldsymbol{U} - \boldsymbol{G}\hat{\boldsymbol{\beta}}\right) \\
&= \boldsymbol{g}(\mathbf{I}^{(i)}) \cdot \hat{\boldsymbol{\beta}} + \boldsymbol{R}^{-1}\boldsymbol{r}(\mathbf{I}^{(i)}) \cdot \left(\boldsymbol{U} - \boldsymbol{G}\hat{\boldsymbol{\beta}}\right)
\end{aligned} \tag{60}$$

Noticing that the vector $\boldsymbol{r}(\mathbf{I}^{(i)})$, from its definition in (55), is indeed the ith column of the correlation matrix, this allows $\boldsymbol{r}(\mathbf{I}^{(i)})$ to be written as $\boldsymbol{r}(\mathbf{I}^{(i)}) = \boldsymbol{R}e_i$, which permits that $\boldsymbol{R}^{-1}\boldsymbol{r}(\mathbf{I}^{(i)})$ in (60) can be expressed as

$$\boldsymbol{R}^{-1}\boldsymbol{r}(\mathbf{I}^{(i)}) = \boldsymbol{R}^{-1}\boldsymbol{R}e_i = e_i. \tag{61}$$

Substituting (61) into (60) allows $\mu_{\hat{U}}$ to be re-written as

$$\begin{aligned}
\mu_{\hat{U}}(\mathbf{I}^{(i)}) &= \boldsymbol{g}(\mathbf{I}^{(i)}) \cdot \hat{\boldsymbol{\beta}} + e_i \cdot \left(\boldsymbol{U} - \boldsymbol{G}\hat{\boldsymbol{\beta}}\right) \\
&= \boldsymbol{g}(\mathbf{I}^{(i)}) \cdot \hat{\boldsymbol{\beta}} + e_i \cdot \boldsymbol{U} - \boldsymbol{G}^T e_i \cdot \hat{\boldsymbol{\beta}}.
\end{aligned} \tag{62}$$

From the definition of the regression matrix $\boldsymbol{G}$ in (82) it is possible to express $\boldsymbol{G}^T e_i = \boldsymbol{g}(\mathbf{I}^{(i)})$, which permits $\mu_{\hat{U}}(\mathbf{I}^{(i)})$ to be given by

$$\mu_{\hat{U}}(\mathbf{I}^{(i)}) = \boldsymbol{g}(\mathbf{I}^{(i)}) \cdot \hat{\boldsymbol{\beta}} + e_i \cdot \boldsymbol{U} - \boldsymbol{g}(\mathbf{I}^{(i)}) \cdot \hat{\boldsymbol{\beta}} = U(\mathbf{I}^{(i)}). \tag{63}$$

3.3. Joint maximum likelihood estimation of the Kriging parameters

In the preceding sections, it has been assumed that the autocovariance function is known a priori. However, the type of correlation functions $\mathcal{R}(\mathbf{I}, \mathbf{I}', \boldsymbol{\theta})$, as well as the value of the variance σ_U^2 is generally unknown a priori. In this work, the type of the correlation function (see *Remark 1*) will be assumed a priori, and the determination of the hyperparameters $\boldsymbol{\theta}$ and the variance σ_U^2 will be obtained from the observation dataset following the *maximum likelihood estimation* (MLE) technique. The resulting predictor is known as the *empirical best linear unbiased predictors* (EBLUP) [16]. The Kringing parameter estimation entails the solution of the following minimisation problem

$$\left\{\boldsymbol{\beta}^*, \sigma_U^{2\,*}, \boldsymbol{\theta}^*\right\} = \arg\min_{\boldsymbol{\beta}, \sigma_U^2, \boldsymbol{\theta}} \mathscr{L}(\boldsymbol{U}|\boldsymbol{\beta}, \sigma_U^2, \boldsymbol{\theta}), \tag{64}$$

where $\mathscr{L}(\boldsymbol{U}|\boldsymbol{\beta}, \sigma_U^2, \boldsymbol{\theta})$ is the opposite log-likelihood of the observations $\boldsymbol{U}$ with respect to its multivariate normal distribution given by

$$\mathscr{L}(\boldsymbol{U}|\boldsymbol{\beta}, \sigma_U^2, \boldsymbol{\theta}) = \frac{1}{2\sigma_U^2}(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta})^T \boldsymbol{R}(\boldsymbol{\theta})^{-1}(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta}) + \frac{m}{2}log(2\pi) + \frac{m}{2}log(\sigma_U^2) + \frac{1}{2}log(|\boldsymbol{R}(\boldsymbol{\theta})|). \tag{65}$$

The MLE of $\boldsymbol{\beta}$ and σ_U^2 are obtained from the first order optimality conditions of $\mathscr{L}(\boldsymbol{U}|\boldsymbol{\beta}, \sigma_U^2, \boldsymbol{\theta})$, namely

$$\begin{cases}
\partial_{\boldsymbol{\beta}}\mathscr{L} = \dfrac{1}{\sigma_U^2}\boldsymbol{G}^T\boldsymbol{R}^{-1}\left(\boldsymbol{G}\boldsymbol{\beta} - \boldsymbol{U}\right) = 0; \\[2mm]
\partial_{\sigma_U^2}\mathscr{L} = \dfrac{1}{2\sigma_U^2}\left(m - \dfrac{\left(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta}\right) \cdot \boldsymbol{R}^{-1}\left(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta}\right)}{\sigma_U^2}\right) = 0;
\end{cases} \tag{66}$$

from which the following optimal values can be obtained

$$\begin{aligned}
\boldsymbol{\beta}^*(\boldsymbol{\theta}) &= \left(\boldsymbol{G}^T\boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{G}\right)^{-1}\boldsymbol{G}^T\left(\boldsymbol{R}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{U}; \\[2mm]
\sigma_U^{2\,*}(\boldsymbol{\theta}) &= \frac{1}{m}\left(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta}^*(\boldsymbol{\theta})\right) \cdot \left(\boldsymbol{R}(\boldsymbol{\theta})\right)^{-1} \cdot \left(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta}^*(\boldsymbol{\theta})\right).
\end{aligned} \tag{67}$$

These solutions are the so-called generalized least-squares estimates and were previously derived in (58). Notice in equation (67) that the dependence of both $\boldsymbol{\beta}^*$ and $\sigma_U^{2\,*}$ on the vector of hyperparameters $\boldsymbol{\theta}$, has been explicitly stated for the first time, which is a consequence of the definition itself of the correlation matrix $\boldsymbol{R}(\boldsymbol{\theta})$. This dependence will continue to be explicitly written in subsequent derivations.

Substituting the optimal values $\boldsymbol{\beta}^*(\boldsymbol{\theta})$ and $\sigma_U^{2}{}^*(\boldsymbol{\theta})$ into the log-likelihood function (65) enables it to be re-written as

$$\begin{aligned}
\mathscr{L}(\boldsymbol{U}|\boldsymbol{\beta}^*, \sigma_U^{2}{}^*, \boldsymbol{\theta}) &= \frac{m}{2} + \frac{m}{2}log(2\pi) + \frac{m}{2}log\left(\sigma_U^{2}{}^*(\boldsymbol{\theta})\right) + \frac{1}{2}log\left(|\boldsymbol{R}(\boldsymbol{\theta})|\right) \\
&= \frac{m}{2}log(\psi(\boldsymbol{\theta})) + \frac{m}{2}(log(2\pi) + 1),
\end{aligned}\tag{68}$$

where the reduced likelihood function has been introduced as

$$\psi(\boldsymbol{\theta}) = \sigma_U^{2}{}^*(\boldsymbol{\theta})|\boldsymbol{R}(\boldsymbol{\theta})|^{1/m}.\tag{69}$$

This entails that the minimisation problem in equation (64) is equivalent to

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \quad \psi(\boldsymbol{\theta}), \qquad s.t. \quad [\boldsymbol{\theta}]_i \geq 0 \quad i = \{1, 2, \cdots, n\}\tag{70}$$

Unfortunately, there is not an analytical solution for the optimal hyperparameters, $\boldsymbol{\theta} \in \mathbb{R}^n$, and instead, a numerical minimisation approach is typically advocated for. However, this is not devoid from numerical difficulties, stemming from the ill-conditioning of the correlation matrix $\boldsymbol{R}(\boldsymbol{\theta})$, existence of multiple local minima, etc. Various optimisation methods are available that address the problem described in (70), such as quasi-Newton methods [29], genetic algorithms [30] followed by a *dynamic hill climber* [31], or even the well-known function *fmincon* in MatLab [32]. In our work, we use of the algorithm *box-min* [33].

For the sake of clearness the process described above has been depicted graphically in Figure 3. To begin with, the only information known is that the simulator is a continuous and differentiable function. Considering this prior knowledge, Figure 3a shows 5 possible realisations of a Kriging with zero mean and quadratic exponential correlation providing the uniform shaded area. From the observations depicted in Figure 3b, a posterior model is obtained having optimised the parameters of the prior model (Figure 3d). As a result, Figure 3c shows the mean prediction and the shading represents variance of the emulator.

3.4. Gradient-enhanced Kriging based prediction

In addition to function observations, it is possible to make use of observations regarding derivatives of the output with respect to its inputs, with the aim of improving the accuracy of the predictor. The resulting model is known as Gradient Enhanced Kriging in the literature [34, 35], in contrast to the standard Kriging described through Sections 3.1-3.3. While the intuition behind the inclusion of derivative information to enhance the emulator's accuracy, using the same number of training points, may seem evident, we present a clear demonstration of this phenomenon in the simple one-dimensional example depicted in Figure 4. In this illustration, the ground truth model (an analytical one-dimensional function) exhibits significantly closer similarity to the behavior depicted by the gradient-enhanced Kriging (Figure 4(b)) than to the Kriging model that solely incorporates information from the function while excluding derivative information (Figure 4(b)). The comparison provides tangible evidence of the pronounced improvement achieved by incorporating derivatives in the emulator, underlining the significance of this approach for enhancing accuracy with the same number of training points.

In order to create a gradient enhanced predictor, the vector of observations is extended to include derivatives of the strain energy density U with respect to its input variables $\mathbf{I}$, such that

$$\boldsymbol{U} = \left[U^{(1)}, \ldots, U^{(m)}, \partial_{\mathbf{I}}U^{(1)}, \ldots, \partial_{\mathbf{I}}U^{(m)}\right]^T,\tag{71}$$

where

$$U^{(i)} = U(\boldsymbol{I}^{(i)}) \qquad \partial_{\mathbf{I}}U^{(i)} = \left[\partial_{I_1}U^{(i)}, \ldots, \partial_{I_n}U^{(i)}\right]^T.\tag{72}$$

To interpolate both the variable and its gradient at any unobserved location, the correlation matrix $\boldsymbol{R}$ is extended to take into account the correlation between the variable and its gradient as

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{UU} & \boldsymbol{R}_{UU'} \\ \boldsymbol{R}_{UU'}^T & \boldsymbol{R}_{U'U'} \end{bmatrix},\tag{73}$$
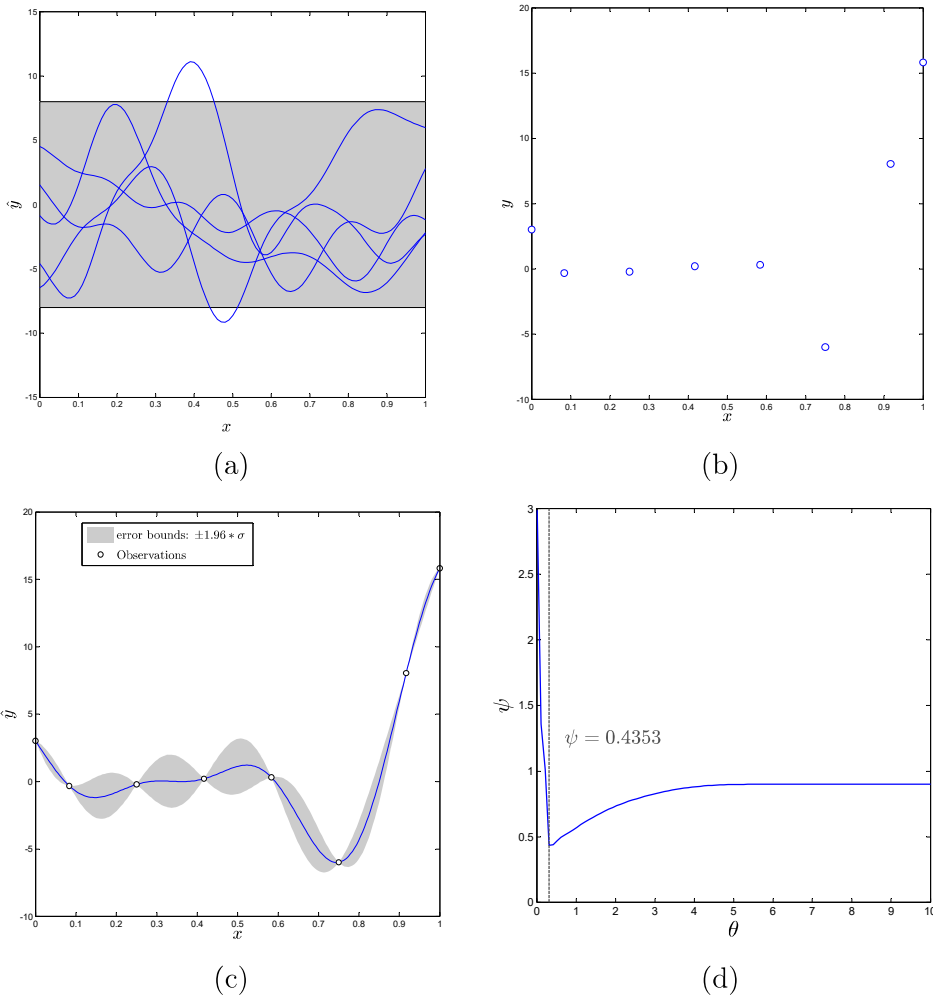
13

Figure 3: Stages involved in the construction of a metamodel for $\mathcal{M}(x) = (6x-2)^2 \cdot sin((6x-2)^2)$. (a) displays 5 realisations of the prior model; (b) demonstrates a selection of observations (*stage 1*); (c) portrays the posterior model having trained the parameters on the observations (*stage 2*); (d) presents the minimisation of the reduced likelihood function (*stage 2*).
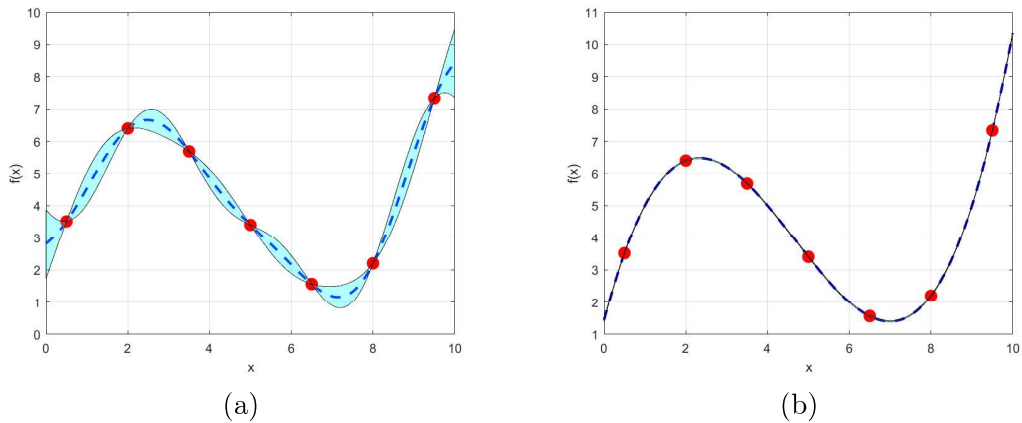

Figure 4: Comparison between trained Ordinary Kriging model, shown in (a), and trained Gradient Enhanced Kriging, shown in (b). Region of uncertainty shaded in blue is close to non-distinguishable in (b) promoting the advantages of using a gradient based approach for the same number of points.

where $\boldsymbol{R_{UU}}$ is the correlation matrix presented in (53) for the non-gradient case. $\boldsymbol{R_{UU'}}$ includes the partial derivatives of $\mathcal{R}$ according to

$$\boldsymbol{R_{UU'}} = \begin{bmatrix} \partial_{\mathbf{I}^{(1)}}\mathcal{R}(\mathbf{I}^{(1)},\mathbf{I}^{(1)},\boldsymbol{\theta}) & \dots & \partial_{\mathbf{I}^{(m)}}\mathcal{R}(\mathbf{I}^{(1)},\mathbf{I}^{(m)},\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \partial_{\mathbf{I}^{(1)}}\mathcal{R}(\mathbf{I}^{(m)},\mathbf{I}^{(1)},\boldsymbol{\theta}) & \dots & \partial_{\mathbf{I}^{(m)}}\mathcal{R}(\mathbf{I}^{(m)},\mathbf{I}^{(m)},\boldsymbol{\theta}) \end{bmatrix}, \tag{74}$$

given

$$\partial_{\mathbf{I}^{(j)}}\mathcal{R}(\mathbf{I}^{(i)},\mathbf{I}^{(j)},\boldsymbol{\theta}) = \left[\frac{\partial\mathcal{R}(\mathbf{I}^{(i)},\mathbf{I}^{(j)},\boldsymbol{\theta})}{\partial I_1^{(j)}}, \frac{\partial\mathcal{R}(\mathbf{I}^{(i)},\mathbf{I}^{(j)},\boldsymbol{\theta})}{\partial I_2^{(j)}}, \dots, \frac{\partial\mathcal{R}(\mathbf{I}^{(i)},\mathbf{I}^{(j)},\boldsymbol{\theta})}{\partial I_n^{(j)}} \right]^T. \tag{75}$$

The submatrix $\boldsymbol{R_{U'U'}}$ contains the second derivatives

$$\boldsymbol{R_{U'U'}} = \begin{bmatrix} \partial^2_{\mathbf{I}^{(1)}\mathbf{I}^{(1)}}\mathcal{R}(\mathbf{I}^{(1)},\mathbf{I}^{(1)},\boldsymbol{\theta}) & \dots & \partial^2_{\mathbf{I}^{(1)}\mathbf{I}^{(m)}}\mathcal{R}(\mathbf{I}^{(1)},\mathbf{I}^{(m)},\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \partial^2_{\mathbf{I}^{(m)}\mathbf{I}^{(1)}}\mathcal{R}(\mathbf{I}^{(m)},\mathbf{I}^{(1)},\boldsymbol{\theta}) & \dots & \partial^2_{\mathbf{I}^{(m)}\mathbf{I}^{(m)}}\mathcal{R}(\mathbf{I}^{(m)},\mathbf{I}^{(m)},\boldsymbol{\theta}) \end{bmatrix}, \tag{76}$$

where

$$\partial^2_{\mathbf{I}^{(i)}\mathbf{I}^{(j)}}\mathcal{R}(\mathbf{I}^{(i)},\mathbf{I}^{(j)},\boldsymbol{\theta}) = \begin{bmatrix} \partial^2_{\mathbf{I}^{(1)}\mathbf{I}^{(1)}}\mathcal{R}(\mathbf{I}^{(i)},\mathbf{I}^{(j)},\boldsymbol{\theta}) & \dots & \partial^2_{\mathbf{I}^{(1)}\mathbf{I}^{(m)}}\mathcal{R}(\mathbf{I}^{(i)},\mathbf{I}^{(j)},\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \partial^2_{\mathbf{I}^{(m)}\mathbf{I}^{(1)}}\mathcal{R}(\mathbf{I}^{(i)},\mathbf{I}^{(j)},\boldsymbol{\theta}) & \dots & \partial^2_{\mathbf{I}^{(m)}\mathbf{I}^{(m)}}\mathcal{R}(\mathbf{I}^{(i)},\mathbf{I}^{(j)},\boldsymbol{\theta}) \end{bmatrix}. \tag{77}$$

Similarly the vector of cross-correlations between the observations and the prediction is extended as follows

$$\boldsymbol{r}(\mathbf{I}) = \left[\mathcal{R}(\mathbf{I},\mathbf{I}^{(1)},\boldsymbol{\theta}), \dots, \mathcal{R}(\mathbf{I},\mathbf{I}^{(m)},\boldsymbol{\theta}), \partial_{\mathbf{I}^{(1)}}\mathcal{R}(\mathbf{I},\mathbf{I}^{(1)},\boldsymbol{\theta}), \dots, \partial_{\mathbf{I}^{(m)}}\mathcal{R}(\mathbf{I},\mathbf{I}^{(m)},\boldsymbol{\theta}) \right]^T. \tag{78}$$

Once these adaptations have been made, the new definitions for the various quantities can be substituted into the definitions detailed in Subsections 3.2 and 3.3. To start with, recall the mean prediction

$$\mu_{\hat{U}}(\mathbf{I}) = \boldsymbol{g}(\mathbf{I}) \cdot \hat{\boldsymbol{\beta}} + \boldsymbol{r}(\mathbf{I}) \cdot \boldsymbol{R}^{-1}\left(\boldsymbol{U} - \boldsymbol{G}\hat{\boldsymbol{\beta}} \right), \tag{79}$$

and the variance

$$\sigma^2_{\hat{U}}(\mathbf{I}) = \sigma^2_U\left(1 - \boldsymbol{r}(\mathbf{I}) \cdot \boldsymbol{R}^{-1}\boldsymbol{r}(\mathbf{I}) + \boldsymbol{u}(\mathbf{I}) \cdot \left(\boldsymbol{G}^T\boldsymbol{R}^{-1}\boldsymbol{G} \right)^{-1}\boldsymbol{u}(\mathbf{I}) \right), \tag{80}$$

with

$$\begin{aligned} \boldsymbol{\beta}^*(\boldsymbol{\theta}) &= \left(\boldsymbol{G}^T\boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{G} \right)^{-1}\boldsymbol{G}^T\left(\boldsymbol{R}(\boldsymbol{\theta}) \right)^{-1}\boldsymbol{U}; \\ \sigma^{2}_U{}^*(\boldsymbol{\theta}) &= \frac{1}{n(1+m)}\left(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta}^*(\boldsymbol{\theta}) \right) \cdot \left(\boldsymbol{R}(\boldsymbol{\theta}) \right)^{-1} \cdot \left(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta}^*(\boldsymbol{\theta}) \right); \end{aligned} \tag{81}$$

and

$$\boldsymbol{G}(\mathbf{I}) = \begin{bmatrix} \boldsymbol{G_U} \\ \boldsymbol{G_{U'}} \end{bmatrix}; \qquad \boldsymbol{G_U} = \begin{bmatrix} (\boldsymbol{g}(\mathbf{I}^{(1)}))^T \\ \vdots \\ (\boldsymbol{g}(\mathbf{I}^{(m)}))^T \end{bmatrix}; \qquad \boldsymbol{G_{U'}} = \begin{bmatrix} (\partial_{\mathbf{I}^{(1)}}\boldsymbol{g}(\mathbf{I}^{(1)}))^T \\ \vdots \\ (\partial_{\mathbf{I}^{(m)}}\boldsymbol{g}(\mathbf{I}^{(m)}))^T \end{bmatrix}, \tag{82}$$

where $\boldsymbol{g}(\boldsymbol{I}^{(i)}) = 1$ and hence $\partial_{\mathbf{I}^{(m)}}\boldsymbol{g}(\mathbf{I}^{(i)}) = \mathbf{0}_{1\times n}$. Finally, the optimal hyperparameters are achieved by minimizing the log-likelihood function.

$$\mathscr{L}(\boldsymbol{U}|\boldsymbol{\beta}^*,\sigma^{2}_U{}^*,\boldsymbol{\theta}) = \frac{m(1+n)}{2}log(\psi(\boldsymbol{\theta})) + \frac{m(1+n)}{2}(log(2\pi) + 1) \tag{83}$$

where the reduced likelihood function has been introduced as

$$\psi(\boldsymbol{\theta}) = \sigma^{2}_U{}^*(\boldsymbol{\theta})|\boldsymbol{R}(\boldsymbol{\theta})|^{\frac{1}{m(n+1)}}. \tag{84}$$

Remark 4: A potential aspect open to critique when applying the Kriging and Gradient Enhanced Kriging techniques, described in Sections 3.1 to 3.4, arises from the requisite inclusion of strain energy values (U) at each individual observation or training juncture. This necessity imparts notable constraints on their applicability to datasets originating from physical laboratory experiments, as opposed to those derived through in-silico or numerical means which, constitutes the focus of this paper. The challenge stems from the inherent difficulty in quantifying energy measurements.

Nevertheless, an advantageous facet of the Gradient Enhanced Kriging methodology lies in its adaptability to scenarios where only a singular observation point is utilized for strain energy (U), an instance realizable by selecting $\boldsymbol{F} = \boldsymbol{I}$, where U typically assumes a value of 0. Herein, derivative information at this specific point, coupled with derivatives at multiple other points, can be seamlessly incorporated. This tailored approach is delineated in greater detail within Section Appendix C.

3.5. Derivatives of strain energy density for Gradient Enhanced Kriging

As discussed in Section 3.4, the gradient-enhanced Kriging approach involves not only the strain energy density U but also its derivatives with respect to the input variables **I**. In cases where the material symmetry groups are isotropy or transverse isotropy and a principal invariant approach is employed (refer to Section 2.5), it becomes necessary to consider the derivatives of U with respect to either $\{I_1, I_2, I_3\}$ or $\{I_1, I_2, I_3, I_4, I_5\}$, respectively. In the case of analytical energies, such as those derived from a Mooney-Rivlin model, obtaining these derivatives is relatively straightforward. However, for more complex strain energy densities arising from composites through homogenisation techniques of varying complexity (e.g. rank-one laminates in Section 2.6.1), these derivatives may not be readily available. Instead, it may be necessary to derive these derivatives from the first Piola-Kirchhoff stress tensor $\boldsymbol{P}$. This can be achieved using standard linear algebra principles and will be specifically outlined for the two aforementioned cases, i.e., strain energies that are either isotropic or transversely isotropic.

In the case of isotropy, the expression for the first Piola-Kirchhoff stress tensor given in (25) permits to define $\boldsymbol{P}$ as a linear combination of three terms. This is conveniently re-written as

$$\boldsymbol{P} = \left(\partial_{I_1} U\right)\boldsymbol{V}_1 + \left(\partial_{I_2} U\right)\boldsymbol{V}_2 + \left(\partial_{I_3} U\right)\boldsymbol{V}_3 \tag{85}$$

where

$$\boldsymbol{V}_1 = 2\boldsymbol{F}; \qquad \boldsymbol{V}_2 = 2\boldsymbol{H} \times \boldsymbol{F}; \qquad \boldsymbol{V}_3 = \boldsymbol{H}, \tag{86}$$

can be understood as the linear independent vectors of a basis, whilst $\left(\partial_{I_i} U\right)$ represent the coordinates of $\boldsymbol{P}$ along the vectors $\boldsymbol{V}_i$, $(i = \{1, 2, 3\})$. As standard in basic courses of linear algebra, given $\boldsymbol{P}$, the coordinates $\left(\partial_{I_1} U\right)$ can be obtained through projection of the latter over the three vector of the basis, which yields the following linear system of equations

$$\begin{bmatrix} \boldsymbol{P} : \boldsymbol{V}_1 \\ \boldsymbol{P} : \boldsymbol{V}_2 \\ \boldsymbol{P} : \boldsymbol{V}_3 \end{bmatrix} = \boldsymbol{M}_{\text{iso}} \begin{bmatrix} \partial_{I_1} U \\ \partial_{I_2} U \\ \partial_{I_3} U \end{bmatrix}, \qquad [\boldsymbol{M}_{\text{iso}}]_{ij} = \boldsymbol{V}_i : \boldsymbol{V}_j, \quad i, j = \{1, 2, 3\} \tag{87}$$

From the algebraic system of equations in (87) it is possible to derive under which conditions it would be possible to obtain the solution of $\{\partial_{I_1} U, \partial_{I_2} U, \partial_{I_3} U\}$, which represent the components of $\boldsymbol{P}$ over the components of the basis $\boldsymbol{\mathcal{V}}$. Clearly, the above system of equations is solvable (i.e. $\{\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3\}$ are linearly independent vectors) when its determinant is different from 0. Without loss of generality, consideration of a diagonal deformation gradient tensor $\boldsymbol{F}$, written in terms of its principal values, namely $\boldsymbol{F} = \text{diag}[\lambda_1, \lambda_2, \lambda_3]$ permits to obtain the following expression for the determinant of $\boldsymbol{M}_{\text{iso}}$ as

$$\det \boldsymbol{M}_{\text{iso}} = 64(\lambda_1^2 - \lambda_2^2)^2(\lambda_1^2 - \lambda_3^2)^2(\lambda_2^2 - \lambda_3^2)^2 \tag{88}$$

Clearly, the three possible situations under which the above linear system of equations is not solvable corresponds when two of the principal stretches of $\boldsymbol{F}$ are identical or when the three of them are identical. The former can be mathematically stated as

$$\det \boldsymbol{M}_{\text{iso}} = 0 \iff \begin{cases} \lambda_i = \lambda_j, \ i \neq j, \ i, j = \{1, 2, 3\}, \\ \lambda_1 = \lambda_2 = \lambda_3 \end{cases} \tag{89}$$

16

For the case of transverse isotropy, the expression for the first Piola-Kirchhoff stress tensor in equation (27) can be conveniently written as

$$\boldsymbol{P} = \left(\partial_{I_1}U\right)\boldsymbol{V}_1 + \left(\partial_{I_2}U\right)\boldsymbol{V}_2 + \left(\partial_{I_3}U\right)\boldsymbol{V}_3 + \left(\partial_{I_4}U\right)\boldsymbol{V}_4 + \left(\partial_{I_5}U\right)\boldsymbol{V}_5 \tag{90}$$

where

$$\boldsymbol{V}_1 = 2\boldsymbol{F}, \qquad \boldsymbol{V}_2 = 2\boldsymbol{H} \times \boldsymbol{F}, \qquad \boldsymbol{V}_3 = \boldsymbol{H}, \qquad \boldsymbol{V}_4 = 2\boldsymbol{F}\boldsymbol{N} \otimes \boldsymbol{N}, \qquad \boldsymbol{V}_5 = 2\left(\boldsymbol{H}\boldsymbol{N} \otimes \boldsymbol{N}\right) \times \boldsymbol{F} \tag{91}$$

Projection of $\boldsymbol{P}$ in (90) over vectors $\boldsymbol{V}_i$, with $i = \{1, \ldots, 5\}$ in (91) yields the following system of algebraic equations

$$\begin{bmatrix} \boldsymbol{P} : \boldsymbol{V}_1 \\ \boldsymbol{P} : \boldsymbol{V}_2 \\ \boldsymbol{P} : \boldsymbol{V}_3 \\ \boldsymbol{P} : \boldsymbol{V}_4 \\ \boldsymbol{P} : \boldsymbol{V}_5 \end{bmatrix} = \boldsymbol{M}_{\text{ti}} \begin{bmatrix} \partial_{I_1}U \\ \partial_{I_2}U \\ \partial_{I_3}U \\ \partial_{I_4}U \\ \partial_{I_5}U \end{bmatrix}, \qquad [\boldsymbol{M}_{\text{ti}}]_{ij} = \boldsymbol{V}_i : \boldsymbol{V}_j, \quad i,j = \{1, \cdots, 5\} \tag{92}$$

Clearly, the above system of equations is solvable (i.e. $\boldsymbol{V}_i$ with $i = \{1, \ldots, 5\}$ are linearly independent vectors) when its determinant is different from 0. Consideration of a diagonal $\boldsymbol{F}$, written in terms of its principal stretches permits to obtain the following expression for the determinant of $\boldsymbol{M}_{\text{ti}}$ as

$$\det \boldsymbol{M}_{\text{ti}} = \det \boldsymbol{M}_{\text{iso}} c(\theta, \phi) g(\lambda_1, \lambda_2, \lambda_3, \theta, \phi) \tag{93}$$

where $(\theta, \phi) \in [0, 2\pi] \times [0, \pi]$ enable the spherical parametrisation of the preferred direction $\boldsymbol{N}$ according to

$$\boldsymbol{N} = \begin{bmatrix} \cos(\theta)\sin(\phi) & \sin(\theta)\sin(\phi) & \cos(\phi) \end{bmatrix} \tag{94}$$

Furthermore, in (93), the term $\det \boldsymbol{M}_{\text{iso}}$ can be found in equation (88), and the functions c and g are defined below

$$c(\theta, \phi) = 64\cos(\phi)^2\cos(\theta)^2\sin(\phi)^4\sin(\theta)^2$$
$$\begin{aligned} g(\lambda_1, \lambda_2, \lambda_3, \theta, \phi) = {} & \lambda_1^2\lambda_2^6 - 2\lambda_1^4\lambda_2^4 + \lambda_1^6\lambda_2^2 + \lambda_1^4\lambda_3^4 + \lambda_1^6\lambda_3^2 + \lambda_2^4\lambda_3^4 + \lambda_2^6\lambda_3^2 - 2\lambda_1^2\lambda_2^2\lambda_3^4 - \lambda_1^2\lambda_2^4\lambda_3^2 - \lambda_1^4\lambda_2^2\lambda_3^2 \\ & + \sin(\phi)^2\left(3\lambda_1^4\lambda_2^4 - \lambda_1^6\lambda_2^2 + \lambda_1^4\lambda_3^6 - \lambda_1^6\lambda_3^2 + \lambda_2^4\lambda_3^6 - 3\lambda_2^4\lambda_3^4 + \lambda_1^2\lambda_2^2\lambda_3^4 - \lambda_1^4\lambda_2^2\lambda_3^2\right) \\ & + \sin(\phi)^2\sin(\theta)^2\left(-\lambda_1^2\lambda_2^6 + \lambda_1^6\lambda_2^2 - 3\lambda_1^4\lambda_3^4 + \lambda_1^6\lambda_3^2 + 3\lambda_2^4\lambda_3^4 - \lambda_2^6\lambda_3^2 - \lambda_1^2\lambda_2^4\lambda_3^2 + \lambda_1^4\lambda_2^2\lambda_3^2\right) \end{aligned} \tag{95}$$

Therefore, the conditions that prevent the system of equations in (92) from being solved are

$$\det \boldsymbol{M}_{\text{ti}} = 0 \iff \begin{cases} \det \boldsymbol{M}_{\text{iso}} = 0 & \iff \begin{cases} \lambda_i = \lambda_j, \ i \neq j, \ i,j = \{1,2,3\}, \\ \lambda_1 = \lambda_2 = \lambda_3 \end{cases} \\ c(\theta, \phi) = 0 & \iff \lambda_i = \lambda_j, \ i \neq j, \ i,j = \{1,2,3\} \\ \begin{cases} \cos(\theta) = 0 \\ \sin(\theta) = 0 \\ \cos(\phi) = 0 \\ \sin(\phi) = 0 \end{cases} & \iff \boldsymbol{N} = \boldsymbol{V}_i, \ i = \{1,2,3\} \end{cases} \tag{96}$$

Remark 5: The ill-conditioning of the system of equations in (87) and (92) arises due to two primary factors: the presence of principal stretches of deformation with equal values (in both isotropy and transverse isotropic models) or the alignment of principal deformation directions with the preferred direction of transverse isotropy. To address this numerical issue, we propose a perturbation approach, wherein we introduce slight variations to the identical principal stretches and a misalignment of the coincident principal direction with the preferred direction. By doing so, we ensure the solvability of both (87) and (92).

3.5.1. Noise regularisation

Kriging fundamentally serves as an interpolation method. Nevertheless, it is widely recognized that in cases of substantial training data and the incorporation of derivative information into the training strategy (e.g., in the context of gradient-enhanced Kriging), the correlation matrix $\boldsymbol{R}$ defined in equation (73) can become ill-conditioned. To mitigate this issue, a customary practice is to introduce regularization by augmenting the correlation matrix with a diagonal matrix as follows: [20]

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{UU} + \varepsilon_1 \boldsymbol{I}_{m \times m} & \boldsymbol{R}_{UU'} \\ \boldsymbol{R}_{UU'}^T & \boldsymbol{R}_{U'U'} + \varepsilon_2 \boldsymbol{I}_{m \cdot n \times m \cdot n} \end{bmatrix}, \qquad \varepsilon_1, \varepsilon_2 \in \mathbb{R}^+ \tag{97}$$

While our paper primarily highlights the interpolation properties of this technique, we consistently employ sufficiently small values of ε_1 and ε_2 to mitigate potential challenges. It is noteworthy, as elucidated in *Remark 3*, that Kriging and its gradient counterpart can achieve interpolation when $\boldsymbol{R}$ remains unregularized, specifically for $\varepsilon_1 = \varepsilon_2 = 0$. However, when $\varepsilon_1 \neq 0$ and $\varepsilon_2 \neq 0$, and in the extreme scenario of both parameters assuming large values, Kriging transitions from an interpolation technique to a regression technique.

In this study, we exclusively analyze in-silico data that is intentionally devoid of noise. Consequently, it becomes feasible to employ regularization parameters of minimal magnitude, effectively causing Gradient-Enhanced Kriging to exhibit characteristics akin to an interpolation technique rather than a regression one. Nevertheless, for the sake of comprehensiveness, we undertake an investigation of the performance of Gradient-Enhanced Kriging in a regression context, particularly when confronted with severely ill-conditioned correlation matrices arising from noise-contaminated data. To elucidate this aspect, we employ two designated training samples, denoted as:

- **Unperturbed sample:** training sample devoid from noise in the output variables, including the values of the energy $\Psi(\boldsymbol{F}) = U(I_1, I_2, I_3)$ and its derivatives $\{\partial_{I_1} U, \partial_{I_2} U, \partial_{I_3} U\}$, where the ground-truth constitutive model from which these data have been generated in-silico corresponds with the Mooney-Rivlin model described in Appendix A.

- **Noisy sample:** this training sample has been obtained by perturbing the **deterministic sample** according to:

$$\widetilde{U} = U + \mathcal{N}(0, \sigma_U); \qquad \widetilde{\partial_{I_i} U} = \partial_{I_i} U + \mathcal{N}(0, \sigma_{\partial_{I_i} U}), \, i = \{1, 2, 3\} \tag{98}$$

with

$$\sigma_U = 0.2 \,\text{mean}(U); \qquad \sigma_{\partial_{I_i} U} = 0.2 \,\text{mean}(\partial_{I_i} U), \, i = \{1, 2, 3\} \tag{99}$$

In both datasets, Figure 5 illustrates the performance of Gradient-Enhanced Kriging utilizing an interpolation-oriented approach, characterized by the use of small ε_1 and ε_2 values. As anticipated, in the case of the **unperturbed sample** (see Figure 5a), Kriging impeccably reproduces the training data points (represented by circles). Conversely, in the **noisy sample**, Kriging strives to replicate the perturbed and irregular data to the greatest extent possible. Discrepancies observed at certain points, resulting in the loss of interpolation properties, can be attributed to the necessity of employing small ε_1 and ε_2 values. Notably, it is evident that the condition number of the matrix $\boldsymbol{R}$ experiences a substantial increase when dealing with the **noisy sample**, as illustrated in Figure 5b. This observation aligns with expectations and raises concerns regarding the predictive accuracy of Kriging between training points, potentially leading to undesired oscillations.

Alternatively, we have explored a regression-based methodology, as detailed in [20]. In this context, the regularization parameters $\{\varepsilon_1, \varepsilon_2\}$ are treated as supplementary hyperparameters. Consequently, both sets of hyperparameters, namely $\{\theta_1, \theta_2, \theta_3\}$ and $\{\varepsilon_1, \varepsilon_2\}$, are optimized through the minimization of the reduced likelihood function $\psi(\widetilde{\boldsymbol{\theta}})$

$$\widetilde{\boldsymbol{\theta}}^* = \arg\min_{\widetilde{\boldsymbol{\theta}}} \quad \psi(\widetilde{\boldsymbol{\theta}}), \qquad s.t. \quad [\widetilde{\boldsymbol{\theta}}]_i \geq 0 \quad i = \{1, 2, \cdots, 5\} \tag{100}$$
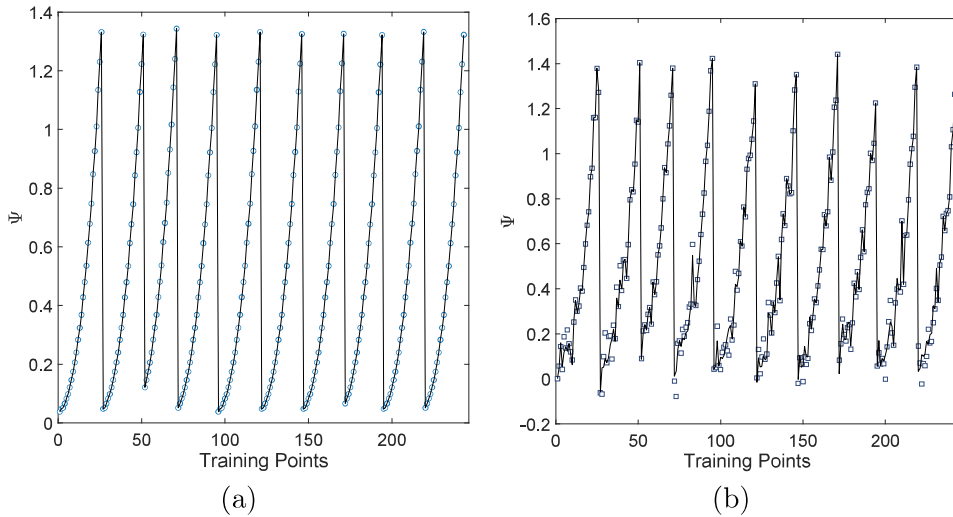
Figure 5: (a) ***Unperturbed sample***; (b) ***Noisy sample***. Performance of interpolation-based Gradient-Enhanced Kriging. Circles (a) or squares (b) represent the training points used. Continuous corresponds with the prediction of the Kriging.

where the augmented set of hyperparameters is defined as $\widetilde{\boldsymbol{\theta}} = \{\theta_1, \theta_2, \theta_3, \varepsilon_1, \varepsilon_2\}$. Applying this approach to only the the **noisy sample** yields the outcomes depicted in Figure 6. The values of $\{\varepsilon_1, \varepsilon_2\}$ are determined to strike a balance between the interpolation and regression properties of the Kriging response. Naturally, the response does not precisely match the noisy data, thereby avoiding the introduction of undesirable oscillations caused by data perturbations (see Figure 6a). Crucially, when removing the noise from the data, we can observe that the regression Kriging is capable of fitting these data (see Figure 6b).
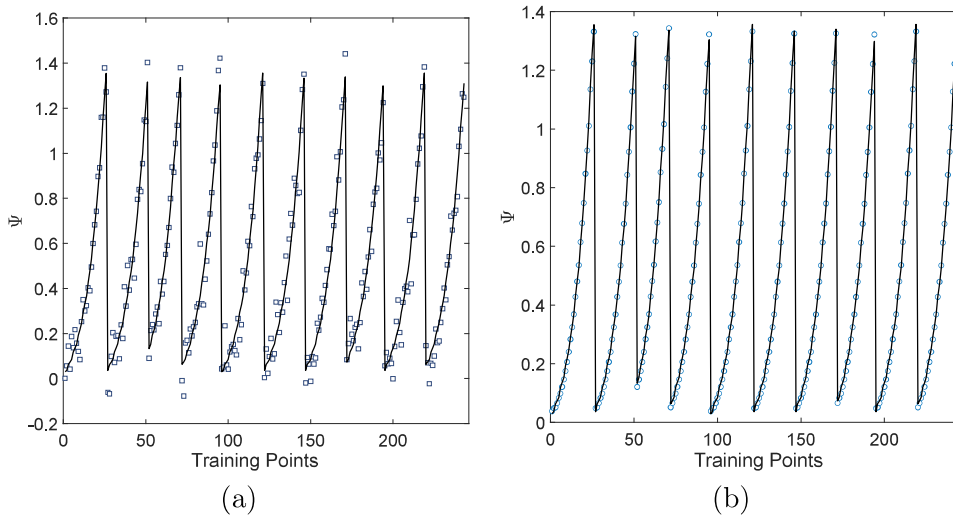
Figure 6: Performance of regression-based Gradient-Enhanced Kriging. (a) Squares represent the training points from the noisy sample; (b) Circles represent the values of the **noisy sample** devoid from noise, showing that the Kriging model is capable of reproducing the behaviour of the unperturbed ground truth model.

4. Calibration of Kriging and Gradient Enhanced Kriging metamodels

4.1. Design of Experiments

In this section, we present the procedure used for generating synthetic data, utilizing a diverse set of ground truth constitutive models. The strain energy densities and material parameters for these models can be found in Appendix A. The models employed encompass the following: Mooney-Rivlin, quadratic Mooney-Rivlin, Gent, Yeoh, Arruda-Boyce, a transversely isotropic model, a rank-one laminate composite, and a computationally homogenised RVE comprising elastomer matrix with centered spherical

inclusion. Table 4.1 shows all the ground truth models considered, and in parallel, we indicate the material symmetry group of its corresponding Kriging or Gradient-Enhanced Kriging (since all data in this paper has been generated in-silico, we advocate for the Gradient-Enhanced Kriging approach described in Section 3.4, encompassing energy and derivative information at all training points) counterpart.

Ground truth model	Kriging-based counterpart
Mooney-Rivlin	Isotropic model, $U = U(I_1, I_2, I_3)$
Gent model	Isotropic model, $U = U(I_1, I_2, I_3)$
Yeoh model	Isotropic model, $U = U(I_1, I_2, I_3)$
Arruda-Boyce model	Isotropic model, $U = U(I_1, I_2, I_3)$
Transversely Isotropic (TI) model	TI model, $U = U(I_1, I_2, I_3, I_4, I_5)$
Rank-one laminate composite	TI model, $U = U(I_1, I_2, I_3, I_4, I_5)$
RVE	Orthotropic model, $U = U(I_1, I_2, I_3, I_4, I_5, I_6)$

To generate the data, we follow the methodology described in [36]. The deformation gradient tensor $\boldsymbol{F}$ is parametrized using a carefully selected sample of deviatoric directions, amplitudes, and Jacobians J (i.e., the determinant of $\boldsymbol{F}$). The process of generating the sampling points for deviatoric directions, amplitudes, and Jacobians is outlined in Algorithm 1.

Algorithm 1 Pseudo-code for sample generation

1: Set the number of amplitudes, directions and determinants: $\{n_{\boldsymbol{t}}, n_{\boldsymbol{X}}, n_{\boldsymbol{J}}\}$;
2: Initialise the vector of amplitudes and determinants: $\boldsymbol{t} = [0, ..., 1.7]_{n_{\boldsymbol{t}} \times 1}$; $\boldsymbol{J} = [0.8, ..., 1.2]_{n_{\boldsymbol{J}} \times 1}$;
3: Initialise a vector of Latin Hypercube Sampled angles: $\boldsymbol{\phi}_1 = [0, 2\pi]_{n_{\boldsymbol{X}} \times 1}$;
4: Initialise three vectors of Latin Hypercube Sampled angles: $\boldsymbol{\phi}_{2,...,4} = [0, \pi]_{n_{\boldsymbol{X}} \times 1}$;
5: Construct the directions, $\boldsymbol{X}$, using an extended Spherical parametrisation in $\mathbb{R}^5$ - detailed in (101);
6: Evaluate the deformation gradient tensors, $\boldsymbol{F}$, parametrised in terms of deviatoric directions $\boldsymbol{X}$, amplitudes t and determinants J - detailed in Algorithm 2;
7: Evaluate the invariants, energy and stress - detailed in (24), Section Appendix A, (25) respectively.

Regarding the deviatoric directions, we have constructed them using a spherical parametrization in $\mathbb{R}^5$. Specifically, we represent these directions in terms of four relevant angular measures $\{\phi_1, \phi_2, \phi_3, \phi_4\}$ within this 5-dimensional space, as follows:

$$\boldsymbol{X}^i = \begin{bmatrix} \cos \phi_1^i \\ \sin \phi_1^i \cos \phi_2^i \\ \sin \phi_1^i \sin \phi_2^i \cos \phi_3^i \\ \sin \phi_1^i \sin \phi_2^i \sin \phi_3^i \cos \phi_4^i \\ \sin \phi_1^i \sin \phi_2^i \sin \phi_3^i \sin \phi_4^i \end{bmatrix}; \qquad 1 \le i \le n_{\boldsymbol{X}} \tag{101}$$

Once the sample is generated following Algorithm 1, the reconstruction of the deformation gradient tensor becomes possible at each of the sampling points. This reconstruction process is demonstrated in Algorithm 2, where $\boldsymbol{\Psi}$ represents the basis for symmetric and traceless tensors (refer to Appendix Appendix B for details on $\boldsymbol{\Psi}$).

4.2. Calibration and Validation

The synthetic data, generated as described in Section 4.1, serves to calibrate both Kriging and Gradient Enhanced Kriging surrogate models, following the principles outlined in Section 3. To assess the accuracy of the Kriging and Gradient Enhanced Kriging model at non-observation points (different from those used for calibration), a set of evaluation points can be generated using the same procedure as described in Section 4.1. These evaluation points are not part of the calibration data but are used to test the performance of the surrogate model. One commonly used metric in our work to determine

Algorithm 2 Pseudo-code for construction of the set of deformation gradient tensors

---

1: **for** $i = 1 : n_{\boldsymbol{X}}$ **do**
2:     **for** $j = 1 : n_J$ **do**
3:         **for** $k = 1 : n_{\boldsymbol{t}}$ **do**
4:             $\boldsymbol{F} = J_j^{1/3} \exp\left(t_k \left[\sum_{l=1}^{5} X_l^i \boldsymbol{\Psi}_l \right] \right);$
5:         **end for**
6:     **end for**
7: **end for**

---

the accuracy is the relative error in the first Piola-Kirchhoff stress tensor at all evaluation points (non-observation/calibration points). This relative error is denoted as $E_{\boldsymbol{P}}$ and is defined as follows

$$E_{\boldsymbol{P}} = \frac{\sum_{i=1}^{n} \| \boldsymbol{P}^{An^i} - \boldsymbol{P}^{Kr^i} \|}{\sum_{i=1}^{n} \| \boldsymbol{P}^{An^i} \|}; \qquad n = 10,000 \tag{102}$$

where $\|\boldsymbol{A}\|$ denotes the Frobenius norm of $\boldsymbol{A}$, n is the number of experiments, $\boldsymbol{P}^{An^i}$ and $\boldsymbol{P}^{Kr^i}$ represent the analytical and Kriging-predicted first Piola-Kirchhoff stress tensors, respectively.

It is important to highlight that for validation purposes, the authors typically employ a validation set comprising $10,000$ data points. This approach ensures that the sampled set is sufficiently dense to assess whether the substantially smaller calibration set is capable of producing an accurate model. The validation process enables a comprehensive evaluation of the surrogate model's performance across a wide range of data points and helps validate its reliability and generalizability.

The calibration and validation process will be conducted for a comprehensive battery of constitutive models, each exhibiting a certain degree of nonlinearity. The models included in this study are as follows: (a) Mooney-Rivlin model; (b) Arruda-Boyce; (c) Gent; (d) Quadratic Mooney-Rivlin; (e) Yeoh; (f) Transversely isotropic; (g) Rank-one laminate composite; (h) RVE with spherical inclusion. The specific expressions for the strain energy densities of each model can be found in Appendix A, along with the values of the various material parameters involved. For each of these models, we have generated **2 training samples**, each comprising a total number of training points of $N = \{20, 60\}$. For each training set, we have calibrated both the Kriging and Gradient Enhanced Kriging models for all 7 ground truth models. The results, in terms of the mean squared error for first Piola-Kirchhoff stress tensor $\boldsymbol{P}$, denoted as $R^2(\boldsymbol{P})$ and the values of $E_{\boldsymbol{P}}$ in (102) and $\hat{E}_{\boldsymbol{P}}$, the latest defined as

$$\hat{E}_{\boldsymbol{P}} = \max \left(\frac{\| \boldsymbol{P}^{An^i} - \boldsymbol{P}^{Kr^i} \|}{\| \boldsymbol{P}^{An^i} \|}, \right) \quad i = \{1, \cdots, n = 10,000\}, \tag{103}$$

are presented in Tables 1 (for $N = 20$ training points) and 2 (for $N = 60$ training points).

The results from the analysis (see Tables 1 and 2), show the performance of both Kriging and Gradient Enhanced Kriging. In both tables, the $R^2(\boldsymbol{P})$ achieved using both approaches are remarkably high (close to 1), indicating excellent accuracy in predicting the first Piola-Kirchhoff stress tensor. However, when considering the two alternative metrics, namely $E_{\boldsymbol{P}}$ and $\hat{E}_{\boldsymbol{P}}$, these show a considerably superior accuracy of the Gradient Enhanced Kriging, typically obtaining one order of magnitude smaller values with respect to the Kriging counterpart. For instance, most values of $\hat{E}_{\boldsymbol{P}}$ for $N = 20$ indicate that the worst relative error in the validation sample are above approximately 10%, with even 25% for the Yeoh model, whereas this value is considerably reduced by the gradient Kriging, except for the Arruda-Boyce model, which seems similarly challenging for both approaches. The most challenging case corresponds with that where the ground truth model is the RVE, due to the larger number inputs (i.e. 6 invariants) and to the high contrast between the matrix and the inclusion ($f_m = 10$, see Table A.10). This is only shown for the case of $N = 60$ training points (Table 2), where it can be seen that the worst relative error for the Kriging is above 30%, whereas its gradient counterpart attains a value below 2%.

This observations highlight the clear advantage of employing the Gradient Enhanced technique, as it enables accurate predictions of the first Piola-Kirchhoff stress tensor even with an extremely small

Table 1: $R^2(\boldsymbol{P})$, $E_{\boldsymbol{P}}$ and $\hat{E}_{\boldsymbol{P}}$ for all seven models for number of training points $N = 20$, for both Kriging and Gradient Enhanced Kriging. $\{n_t, n_{\boldsymbol{X}}, n_J\} = \{2, 5, 2\}$

Ground truth model:	Ordinary Kriging			Gradient Kriging		
	$R^2\,(\|\boldsymbol{P}\|)$	$E_{\boldsymbol{P}}$	$\hat{E}_{\boldsymbol{P}}$	$R^2\,(\|\boldsymbol{P}\|)$	$E_{\boldsymbol{P}}$	$\hat{E}_{\boldsymbol{P}}$
Arruda-Boyce	0.9973	2.21×10^{-2}	1.19×10^{-1}	0.9956	1.17×10^{-2}	1.36×10^{-1}
Gent	0.9920	4.82×10^{-2}	2.10×10^{-1}	0.9987	7.03×10^{-3}	7.36×10^{-2}
Yeoh	0.9997	2.39×10^{-2}	2.51×10^{-1}	1.0000	3.56×10^{-3}	1.42×10^{-2}
Mooney-Rivlin	0.9978	3.90×10^{-2}	1.01×10^{-1}	1.0000	3.17×10^{-3}	1.72×10^{-2}
Quadratic Mooney-Rivlin	0.9991	3.18×10^{-2}	7.96×10^{-2}	1.0000	1.69×10^{-3}	6.04×10^{-3}
Transversely isotropic	0.9959	5.72×10^{-2}	3.52×10^{-1}	0.9997	5.84×10^{-3}	4.27×10^{-2}
Rank-one laminate	0.9698	1.33×10^{-1}	2.11×10^{-1}	0.9995	1.25×10^{-2}	6.46×10^{-2}

Table 2: $R^2(\boldsymbol{P})$, $E_{\boldsymbol{P}}$ and $\hat{E}_{\boldsymbol{P}}$ for all seven models for number of training points $N = 60$, for both Kriging and Gradient Enhanced Kriging. $\{n_t, n_{\boldsymbol{X}}, n_J\} = \{5, 6, 2\}$

Ground truth model:	Ordinary Kriging			Gradient Kriging		
	$R^2\,(\|\boldsymbol{P}\|)$	$E_{\boldsymbol{P}}$	$\hat{E}_{\boldsymbol{P}}$	$R^2\,(\|\boldsymbol{P}\|)$	$E_{\boldsymbol{P}}$	$\hat{E}_{\boldsymbol{P}}$
Arruda-Boyce	0.9993	8.75×10^{-3}	8.54×10^{-3}	1.0000	1.84×10^{-4}	9.89×10^{-4}
Gent	0.9998	5.77×10^{-3}	4.11×10^{-2}	1.0000	9.65×10^{-5}	9.89×10^{-4}
Yeoh	1.0000	7.26×10^{-3}	1.25×10^{-1}	1.0000	3.42×10^{-4}	1.97×10^{-3}
Mooney-Rivlin	0.9999	5.04×10^{-3}	1.74×10^{-2}	1.0000	2.23×10^{-4}	8.91×10^{-4}
Quadratic Mooney-Rivlin	0.9999	9.16×10^{-3}	3.29×10^{-2}	1.0000	4.24×10^{-4}	3.22×10^{-3}
Transversely isotropic	0.9999	7.71×10^{-3}	6.06×10^{-2}	1.0000	9.93×10^{-4}	2.19×10^{-2}
Rank-one laminate	0.9998	9.79×10^{-3}	5.15×10^{-2}	1.0000	3.20×10^{-3}	1.21×10^{-2}
RVE	0.9952	3.12×10^{-2}	3.22×10^{-1}	0.9999	6.62×10^{-3}	1.82×10^{-2}

number of training points. This makes Gradient Enhanced Kriging a very convenient and efficient choice compared to its standard Kriging counterpart.

Lastly, for the sake of completeness, Figure 7 provides a visual representation of the selected training points (marked in red) within the space defined by the invariants I_1 and I_2. This plot encompasses the two transversely isotropic calibrated Gradient Enhanced Kriging constitutive models, namely those whose ground truth counterpart are the transversely isotropic model and the rank-one laminate model, showcasing the relative error in the first Piola-Kirchhoff stress tensor across the evaluation points. The figure offers valuable insights into how the surrogate models perform in different regions of the parameter space and allows us to observe the accuracy distribution over the entire range of invariants.

Remark 6: Kriging and its gradient-enhanced variant inherently enable the development of infill strategies based on metamodel epistemic uncertainty. Such strategies involve the selection of new training points guided by the variance of the surrogate model $\sigma_{\hat{U}}^2(\mathbf{I})$ in equation (80). This selection process identifies a point in the n-dimensional space by $\{I_1, I_2, \ldots, I_n\}$ where $\sigma_{\hat{U}}^2(\mathbf{I})$ reaches its maximum value.

However, it is crucial to note that the deformation gradient tensor $\boldsymbol{F}$ is parameterized using amplitudes t, angles $\{\phi_1, \phi_2, \phi_3, \phi_4,\}$ and by the Jacobian J (see Algorithm 2). The lower and upper values of these values defined the 6-dimensional hpercube $[t_{min}, t_{max}] \times [\phi_{1_{min}}, \phi_{1_{max}}] \times [\phi_{2_{min}}, \phi_{2_{max}}] \times \cdots \times [\phi_{4_{min}}, \phi_{4_{max}}] \times [J_{min}, J_{max}]$. This hypercube is what we refer to as the parametrization space. When
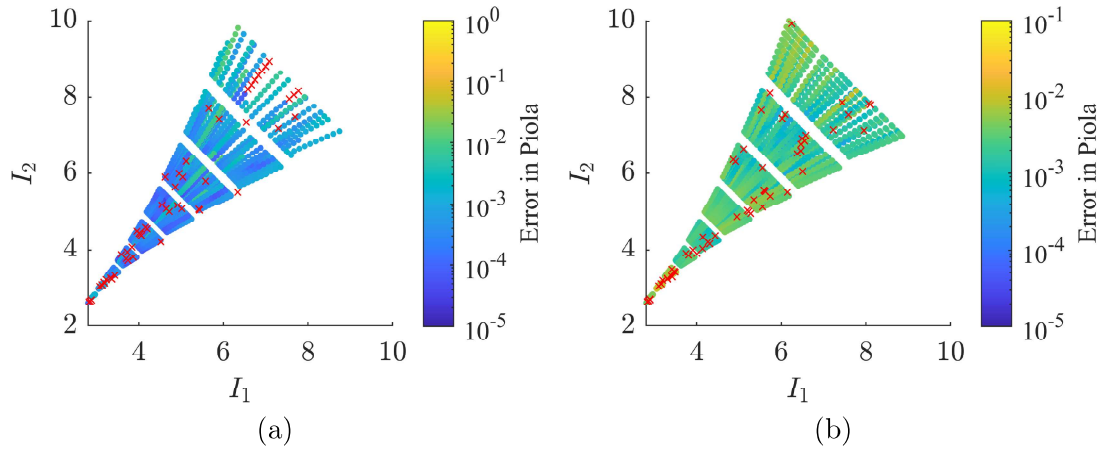
Figure 7: (a-b) Distribution of validation points across the first two invariants. Red crosses represent the location of the calibration/training points and the colour bar represents the distribution of error in $E_{\boldsymbol{P}}$. (a) represents the transversely isotropic and (b) the rank-one laminate models.

a substantial number of points are distributed within this parametrization space, their corresponding $\boldsymbol{F}$ values can be computed, and subsequently, the n invariants (dependent on the material symmetry group considered) can be derived. This process transforms the parametrization space into its counterpart in the design space. However, it is important to highlight that the boundaries of the design space form a polygon in $\mathbb{R}^n$ that may exhibit highly irregular shapes, in contrast to the n-cube shape of the parametrization space. For instance, the projection of the design space onto the plane defined by invariants $I_1 - I_2$, as shown in Figure 7, demonstrates the irregular nature of this projection.

Consequently, when seeking the point with the maximum epistemic uncertainty (i.e., maximum value of $\sigma_{\hat{U}}^2(\mathbf{I})$), it must fall within the boundaries of the design space to avoid extrapolation. This requirement holds true for various machine learning techniques, including neural networks. However, identifying such a point within an irregular polygon in $\mathbb{R}^n$ is a non-trivial task and has not been pursued in this study. Given that our data is generated in-silico, we propose an infill strategy based on selecting a point within the design space that results in the maximum relative error between the in-silico data and its predicted Kriging counterpart.

23

5. Numerical three-dimensional examples

The analysis conducted in Section 4 provides substantial evidence supporting the superiority of gradient enhanced Kriging over its energy-only counterpart, which lacks the consideration of first derivatives information. As a result of these promising findings, we intend to leverage these advanced gradient enhanced metamodels to evaluate their accuracy and performance in the context of challenging three-dimensional Finite Element simulations. This strategic approach seeks to explore the potential of enhanced predictive capabilities and further validate the applicability of gradient enhanced Kriging for addressing complex engineering problems with superior accuracy and efficiency.

5.1. Simple uniform deformation example

Before delving into complex 3D Finite Element simulations, we examine the case of uniform deformations, recognizing its potential to offer valuable and insightful information. In such cases, instabilities may arise, potentially leading to loss of ellipticity. In order to examine the instabilities present in the ground truth and Kriging models, we will generate them by introducing an electromechanical term to the strain energy density as

$$\Psi_{\text{total}}(\boldsymbol{F}, \boldsymbol{D}_0) = \Psi(\boldsymbol{F}) + p\,(J-1) + \frac{1}{2\varepsilon J} II_{\boldsymbol{FD}_0}; \qquad II_{\boldsymbol{FD}_0} = \boldsymbol{FD}_0 \cdot \boldsymbol{FD}_0, \tag{104}$$

where $\Psi(\boldsymbol{F})$ denotes the strain energy of the material (whether it is the ground truth or gradient Kriging model). The Lagrange multiplier p is included to ensure the incompressibility constraint, while the third term represents the electromechanical contribution. The material electric field $\boldsymbol{E}_0$ is the conjugate of $\boldsymbol{D}_0$, and can be obtained as

$$\boldsymbol{E}_0 = \partial_{\boldsymbol{D}_0} \Psi = \frac{1}{\varepsilon J} \boldsymbol{F}^T \boldsymbol{F} \boldsymbol{D}_0, \tag{105}$$

which permits, given $\boldsymbol{E}_0$, to obtain an explicit expression for $\boldsymbol{D}_0$, namely

$$\boldsymbol{D}_0 = \varepsilon J \boldsymbol{F}^{-1} \boldsymbol{F}^{-T} \boldsymbol{E}_0. \tag{106}$$
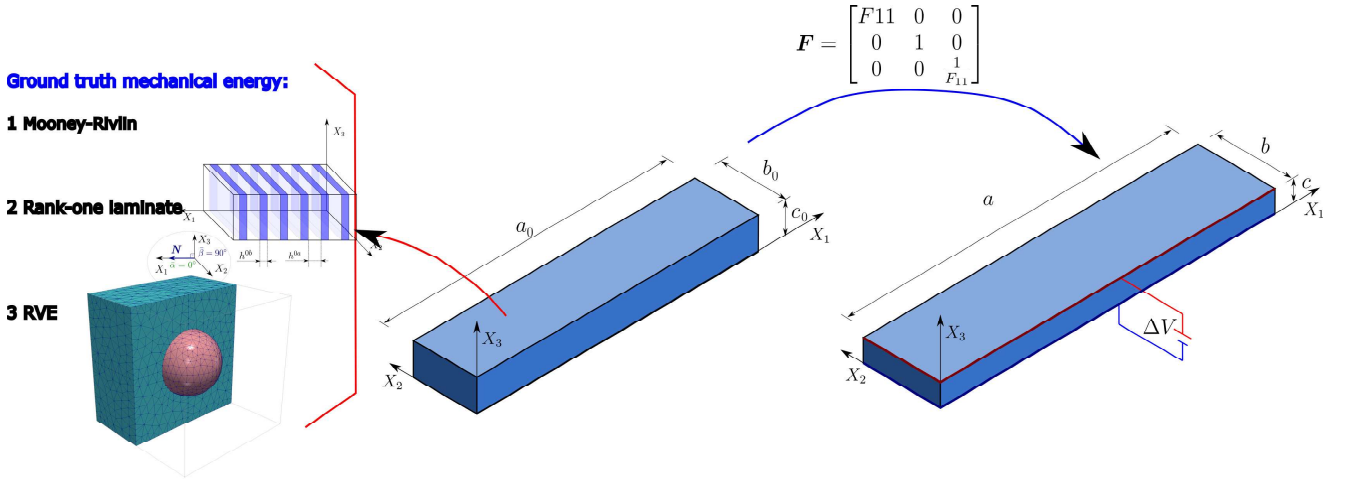
Figure 8: Uniform deformation induced in prism of dimensions $[a_0, b_0, c_0]$ after applying voltage gradient $\Delta V = -\frac{(\boldsymbol{E}_0)_3}{c_0}$, yielding uniformly prism with dimensions $[a, b_0, c]$. Mechanical energies used for the ground truth models are: Mooney-Rivlin, rank-one laminate and RVE (see Appendix A). Induced deformation through addition of the coupled term $\frac{1}{2\varepsilon J} II_{\boldsymbol{FD}_0}$ into the strain energy in (104).

We impose the following specific form of the deformation gradient tensor $\boldsymbol{F}$ and on $\boldsymbol{E}_0$, i.e.

$$\boldsymbol{F} = \begin{bmatrix} F_{11} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & F_{33} \end{bmatrix}; \qquad \boldsymbol{E}_0 = \begin{bmatrix} 0 \\ 0 \\ E_0 \end{bmatrix} \tag{107}$$

For any value of E_0, $\{F_{11}, F_{33}, p\}$ can be obtained from the stationary conditions of Ψ_{total} in (104), i.e

$$D\Psi_{\text{total}}[\delta F_{11}] = \boldsymbol{P} : D\boldsymbol{F}[\delta F_{11}] = P_{11}\delta F_{11} = 0;$$
$$D\Psi_{\text{total}}[\delta F_{33}] = \boldsymbol{P} : D\boldsymbol{F}[\delta F_{33}] = P_{33}\delta F_{33} = 0; \tag{108}$$
$$D\Psi_{\text{total}}[\delta p] = (J - 1)\delta p = 0$$

with the first Piola-Kirchhoff stress tensor $\boldsymbol{P}$ defined as

$$\boldsymbol{P} = \partial_{\boldsymbol{F}}\Psi(\boldsymbol{F}) + p\boldsymbol{H} + \frac{1}{\varepsilon J}\left(\boldsymbol{F}\boldsymbol{D}_0 \otimes \boldsymbol{D}_0 - \frac{1}{2J}II_{\boldsymbol{F}\boldsymbol{D}_0}\boldsymbol{H}\right) \tag{109}$$

Therefore, given E_0 in (107), the following nonlinear permit to determine $\{F_{11}, F_{33}, \lambda\}$, i.e.

$$\mathcal{R}_{F_{11}} = P_{11} = 0; \qquad \mathcal{R}_{F_{33}} = P_{33} = 0; \qquad \mathcal{R}_p = J - 1 = 0; \tag{110}$$

The objective is to analyse, for this specific uniform deformation case, the behaviour of the Gradient-Enhanced Kriging model when the following ground truth models are consider: (a) Mooney-Rivlin; (b) rank-one laminate composie; (c) composite material characterised by RVE with elastomeric matrix and spherical inclusion.

5.1.1. Mooney-Rivlin ground truth model

We consider a Mooney-Rivlin model (whose analytical expression is in equation (A.1)) and rank-one laminate model (see Section 2.6.1 and equation (A.7)) for the mechanical contribution of the energy, namely $\Psi(\boldsymbol{F})$. For the ground truth Mooney-Rivlin model, we choose the material parameters given in Table A.3.

Figure $9a - b$ illustrates the equilibrium path (obtained by means of an arc-length technique) for both the Mooney-Rivlin ground truth model and the gradient-enhanced Kriging calibrated model. Remarkably, these paths are practically indistinguishable, demonstrating strong similarity between the two models. Furthermore, Figure $9a - b$ highlights the specific regions where convexity is lost in both models with respect to F_{11}, F_{33}.

For the ground truth Mooney-Rivlin model, which incorporates the electromechanical term defined in (104), it is well-established that it maintains its ellipticity, as indicated by equation (16). This characteristic is reflected in Figure 9_c, which reveals the absence of regions where the acoustic tensor loses its positive definiteness for all possible directions $\boldsymbol{V}$ parametrized spherically. Similarly, the gradient-enhanced model depicted in Figure 9_d exhibits a comparable region of convexity loss. Notably, it also satisfies the ellipticity condition within the explored region shown in Figure 9.

5.1.2. Rank-one laminate composite ground truth model

For the rank-one laminate model, we use the material parameters in Table A.9, selecting the mechanical contrast f_m as $f_m = 10$, and introducing a slight modification in the lamination orientation $\boldsymbol{N}$, which is now selected to be

$$\boldsymbol{N} = \begin{bmatrix} \cos\theta\sin\psi & \sin\theta\sin\psi & \cos\psi \end{bmatrix}; \qquad \theta = 0 \qquad \psi = \frac{3.4\pi}{8} \tag{111}$$

Figure $10a - b$ depicts the equilibrium path obtained through an arc-length technique for both the rank-one laminate ground truth model and the gradient-enhanced Kriging calibrated model. Once again, these paths exhibit remarkable similarity, highlighting the strong resemblance between the two models. Moreover, Figure $10c - d$ showcases the specific regions where both models experience a loss of convexity with respect to F_{11}, F_{33}, as well as regions where the ellipticity condition is violated. Notably, these regions exhibit striking similarities between the two models. A noteworthy observation is that the region of loss of ellipticity in the gradient-enhanced Kriging model extends to a higher value of F_{11} compared to the ground truth model.

On the equilibrium path depicted in Figure $10c - d$, three points are highlighted. At each point, the minors of the acoustic tensor are monitored for both models. Figure 11 indeed demonstrates the remarkable similarity in the minor of the acoustic tensor for all directions $\boldsymbol{V}$ spherically parameterized
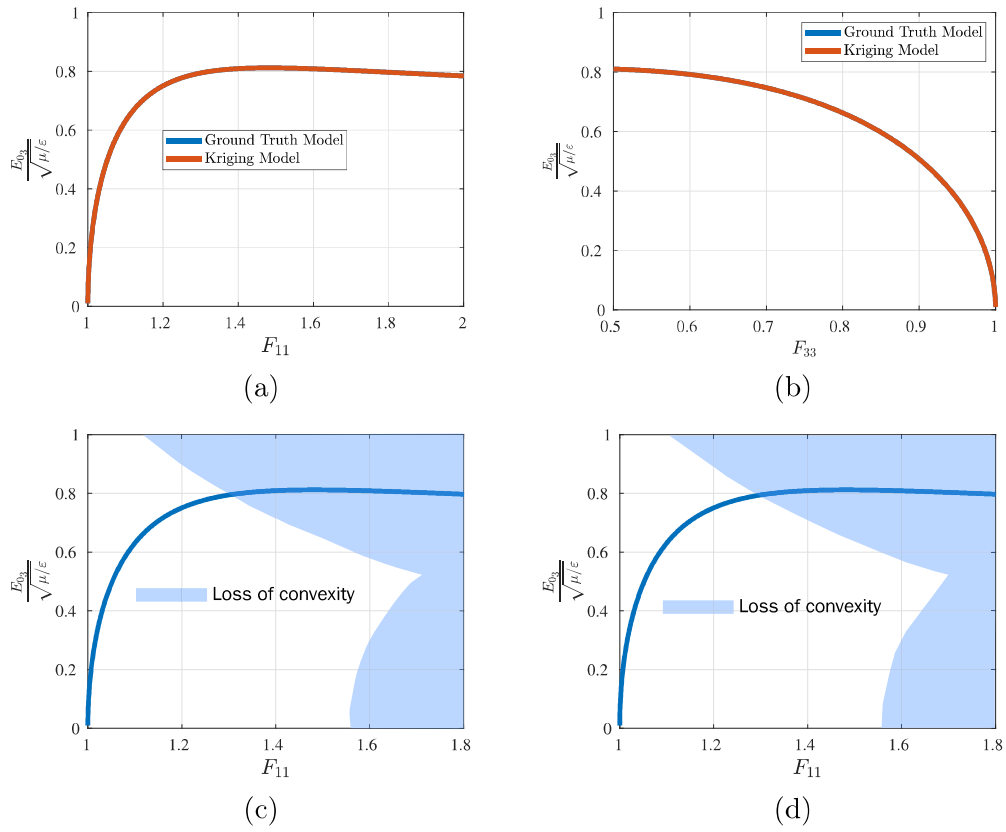
Figure 9: (a) and (b) represent equilibrium path for $\Psi(\boldsymbol{F})$ in (104) representing the Mooney-Rivlin ground truth model and the gradient enhanced Kriging calibrated model, respectively. (c) and (d) represent the regions Mooney-Rivlin ground truth and its gradient enhanced Kriging counterpart, respectively, lose convexity with respect to $\{F_{11}, F_{33}\}$.
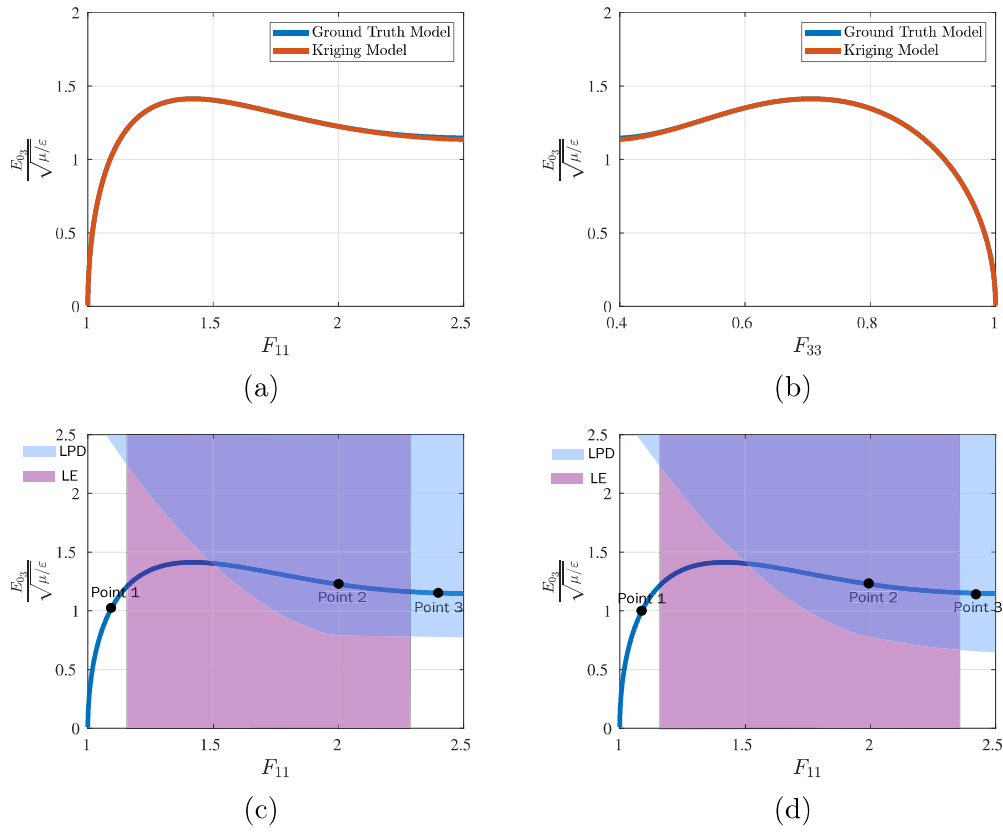
Figure 10: (a) and (b) represent equilibrium path for $\Psi(\boldsymbol{F})$ in (104) representing the rank-one laminate ground truth model and the gradient enhanced Kriging calibrated model. (c) and (d) represent the regions where the ground truth model and its gradient enhanced counterpart lose convexity with respect to $\{F_{11}, F_{33}\}$ and ellipticity.

according to (111), where θ, ψ correspond to the longest and shortest sides of the rectangular grid utilized in each plot of Figure 11.

Lastly, Figure 12 visually represents the shear modulus $\widetilde{\mu}$ associated with Ψ_{total} for both models, which is defined as

$$\widetilde{\mu} = \partial^2_{\boldsymbol{F}\boldsymbol{F}}\Psi_{\text{tot}} : \boldsymbol{e}(\theta,\psi) \otimes \boldsymbol{e}(\theta,\psi) \otimes \boldsymbol{e}(\theta,\psi) \otimes \boldsymbol{e}(\theta,\psi),\tag{112}$$

where $\boldsymbol{e}(\theta,\psi)$ is spherically parameterized in terms of θ and ψ according to (111), yielding a reasonable similarity between both models at the three points considered.
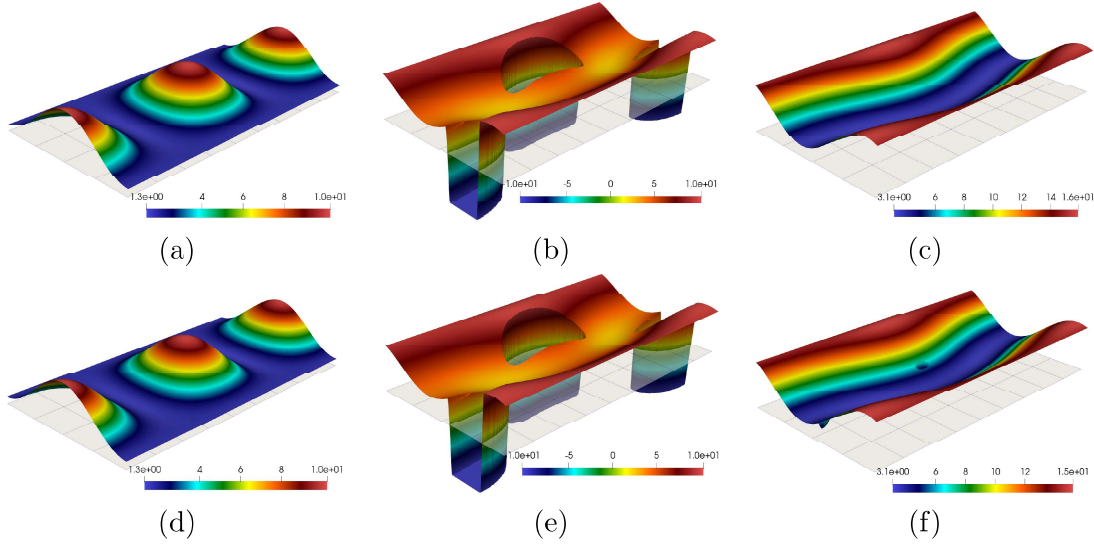
Figure 11: Spherical parametrisation of the least of the minors of the acoustic tensor at: (a) Point 1 in Figure 10c (ground truth model); (d) Point 1 in Figure 10d (Gradient Enhanced Kriging); (b) Point 1 in Figure 10c (ground truth model); (e) Point 1 in Figure 10d (Gradient Enhanced Kriging); (c) Point 1 in Figure 10c (ground truth model); (f) Point 1 in Figure 10d (Gradient Enhanced Kriging).
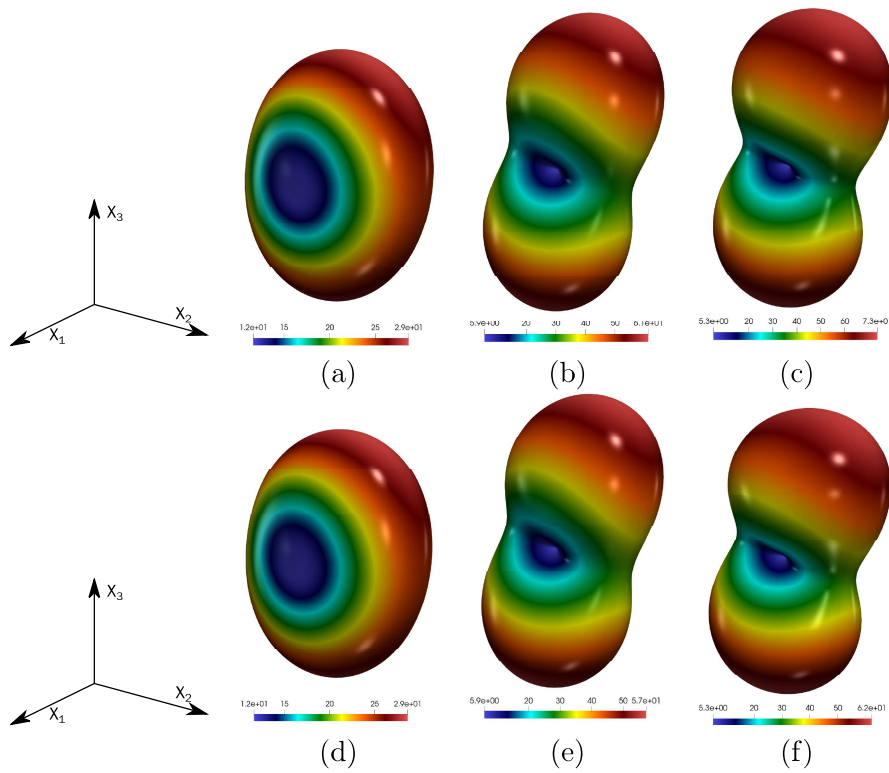
Figure 12: Spatial parametrisation of the shear modulus $\widetilde{\mu}$ in (112) at: (a) Point 1 in Figure 10c (ground truth model); (d) Point 1 in Figure 10d (Gradient Enhanced Kriging); (b) Point 1 in Figure 10c (ground truth model); (e) Point 1 in Figure 10d (Gradient Enhanced Kriging); (c) Point 1 in Figure 10c (ground truth model); (f) Point 1 in Figure 10d (Gradient Enhanced Kriging).

5.1.3. RVE with elastomer matrix and spherical inclusion

Now we consider the case of a composite material characterised by the RVE shown in Figure 13, comprising an elastomeric hexahedral matrix with an centered spherical inclusion. The homogenisation principles which permit to obtain the effective properties (i.e. energy U and first Piola-Kirchhoff stress tensor $\boldsymbol{P}$) are shown in Section 2.7. The specific form of the strain energies for both matrix and inclusion, as well as their respective material parameters, can be found in Appendix A.

As already indicated, the Gradient-Enhanced Kriging counterpart is an orthotropic material characterised by the prefered directions $\boldsymbol{N}_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$, $\boldsymbol{N}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$ and $\boldsymbol{N}_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$. Notice that $\boldsymbol{N}_3$ is not needed in order to define the 6 required irreducible basis of invariants associated to this material symmetry group, as indicated in Section 2.5.3.

Figure 13 illustrates the equilibrium path for both the effective response of the RVE ground truth model and the orthotropic gradient-enhanced Kriging calibrated model up to a stretch of $F_{11} = 1.9$. Remarkably, these paths are practically indistinguishable, demonstrating strong similarity between the two models. For completeness, we show the value of the micro–fluctuations $(\boldsymbol{\alpha})_2$ in the undeformed configuration of the RVE for three different points in the equilibrium path of the computationally homogenised ground truth model.
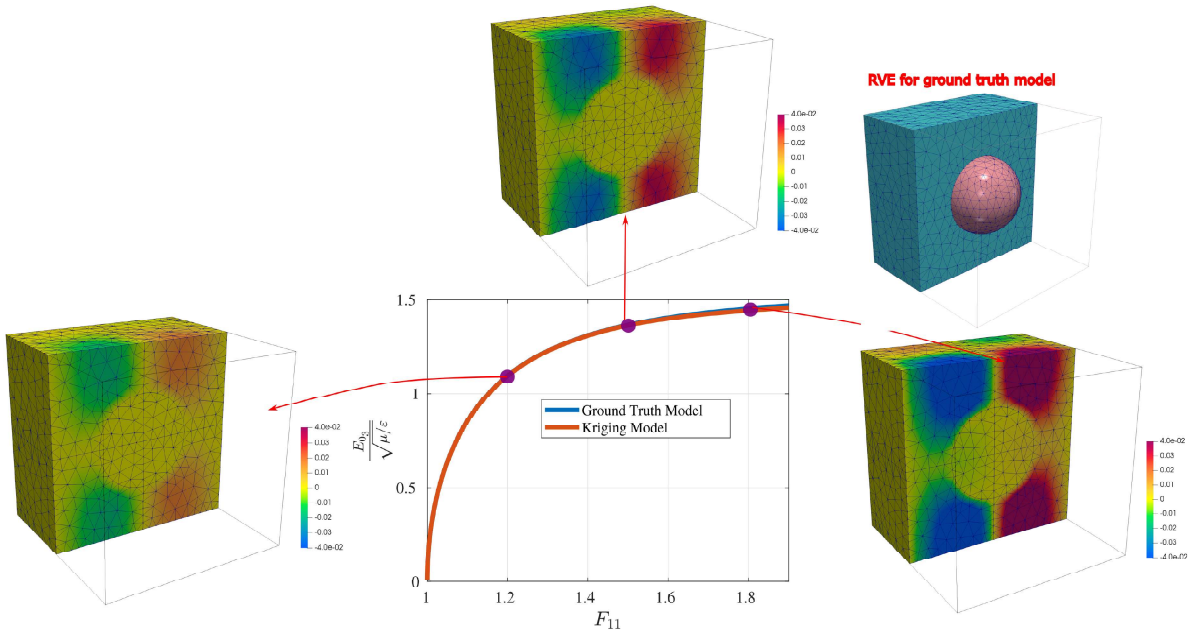
Figure 13: (a) and (b) represent equilibrium path for $\Psi(\boldsymbol{F})$ in (104) representing the Mooney-Rivlin ground truth model and the gradient enhanced Kriging calibrated model, respectively. The parameter μ has been taken as that from the matrix, namely $\mu = \frac{1}{2}(\mu_1^m + \mu_2^m)$. Contour plot of micro-fluctuations $(\boldsymbol{\alpha})_2$ in the undeformed configuration of the RVE for three different points in the equilibrium path of the computationally homogenised ground truth model.

5.2. Cantilever Beam Example

As indicated in the abstract and introductory section of this paper, a primary goal is to integrate gradient enhanced Kriging models into an in-house Finite Element computational framework. This integration aims to substantiate the accuracy and efficacy of these metamodels by rigorously comparing their predictions against the Finite Element solution provided by their respective ground truth counterparts. Specifically, the evaluation will encompass complex and demanding scenarios such as bending, torsion, and wrinkles, offering a robust assessment of the metamodels' performance under challenging conditions.

The first of these Finite Element examples is represented by the cantilever beam depicted in Figure 14. The geometry and boundary conditions for this example are described precisely on Figure 14. With regards to the Finite Element discretisation used, tri-quadratic $Q2$ Finite Elements have been used for the interpolation of the displacement field.
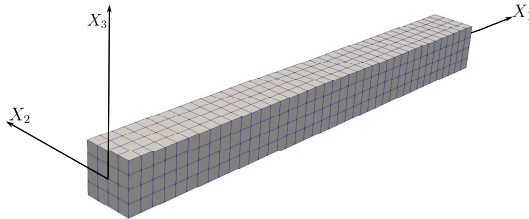
Figure 14: Cantilever beam example: beam with squared section 0.1×0.1 (m^2) and length 1 m. Beam completely fixed at $X_1 = 0$ and applied surface force acting at $X_1 = 1$ along the X_3 direction with value $0.018 Nm^{-2}$.

In this example, we consider a Mooney-Rivlin model (whose analytical expression is in equation (A.1)) and rank-one laminate model (see Section 2.6.1 and equation (A.7)) for the mechanical contribution of the energy, namely $\Psi(\boldsymbol{F})$. For the ground truth Mooney-Rivlin model, we choose the material parameters given in Table A.3. For the rank-one laminate model, we use the material parameters in Table A.9, selecting the mechanical contrast f_m as $f_m = 10$.

Upon application of the surface force, the induced deformation for both ground truth and gradient enhanced Kriging models can be seen in Figure 15. The similarity between the ground truth and its gradient enhanced Kriging counterpart is remarkable for both Figures 15a and 15b, where the ground truth models correspond with the Mooney-Rivlin and rank-one laminate models, respectively. Crucially, the similarity also reflects in the countour plot of the hydrostatic pressure $p = \frac{1}{3}\text{tr}(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma} = J^{-1}\boldsymbol{P}\boldsymbol{F}^T$ denotes the Cauchy stress tensor. These consistent similarities validate the accuracy and reliability of the gradient enhanced Kriging models, highlighting their potential for effectively capturing the intricate behavior of the underlying physical systems
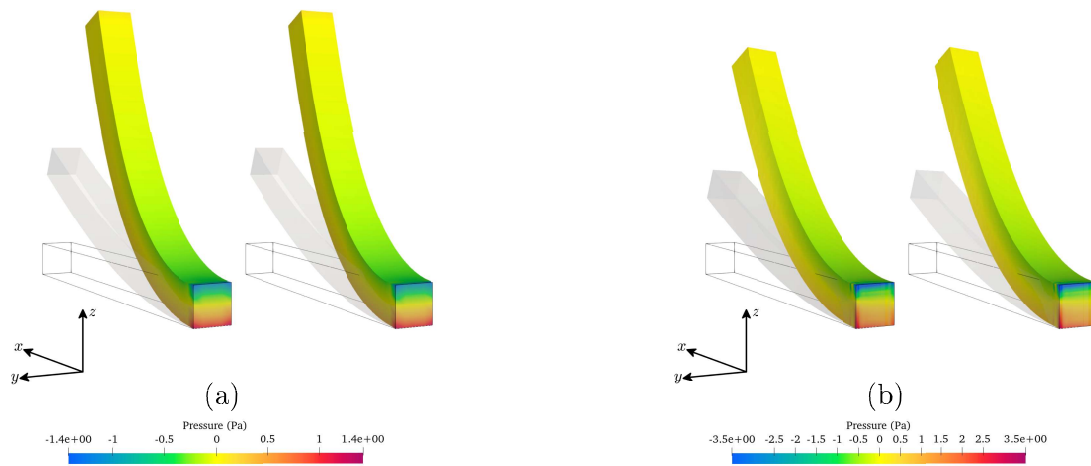
Figure 15: Cantilever beam bending using (a) Mooney-Rivlin (b) a rank-one laminate constitutive model. Note the left and right beams represents the analytical and gradient enhanced Kriging based solutions, respectively. The axis $\{x, y, z\}$ correspond to $\{X_1, X_2, X_3\}$ in Figure 14.

5.3. Column Twisting Example

The next example considers a column with unit area in the xy plane and with length 30 in the Z direction. Tri-quadratic $Q2$ Finite Elements have been employed in a Finite Element discretisation comprising 8 elements in the x-, y- directions and 240 elements in the z-direction. The base of the column ($Z = 0$) is completely fixed, whereas an initial angular velocity $\boldsymbol{\Omega}_0$ is considered, defined as

$$\boldsymbol{\Omega}_0 = \frac{\Omega_0 z}{30} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T; \qquad \Omega_0 = 100 \text{ rad/s} \tag{113}$$

The simulation has been carried out using a dynamic setting, by introducing a fictitious density of $\rho = 0.01$ (kg/mm^3) and a mass-proportional Rayleigh coefficient of value 100 (1/s), by employing a leap frog time integrator, with time step $\Delta t = 3 \times 10^{-5}$ (s). We again consider a Mooney-Rivlin model (whose analytical expression is in equation (A.1)) and rank-one laminate model (see Section 2.6.1 and equation (A.7)) for the mechanical contribution of the energy, namely $\Psi(\boldsymbol{F})$. For the ground truth Mooney-Rivlin model, we choose the material parameters given in Table A.3. For the rank-one laminate model, we use the material parameters in Table A.9, selecting the mechanical contrast f_m as $f_m = 10$.

The results shown in Figures 16 and 17, for various time step of the Finite Element simulation are aligned with those reported in Figure 15, namely, not only the displacements but also the contour plot distribution of the hydrostatic pressure display an almost indistinguishable difference between the ground truth models (Mooney-Rivlin in Figure 16 and rank-one laminate in Figure 17) and their corresponding gradient-enhanced Kriging counterparts. This further reinforces the accuracy and reliability of the gradient-enhanced Kriging models, substantiating their ability to faithfully capture the intricate behavior of the physical systems under investigation.
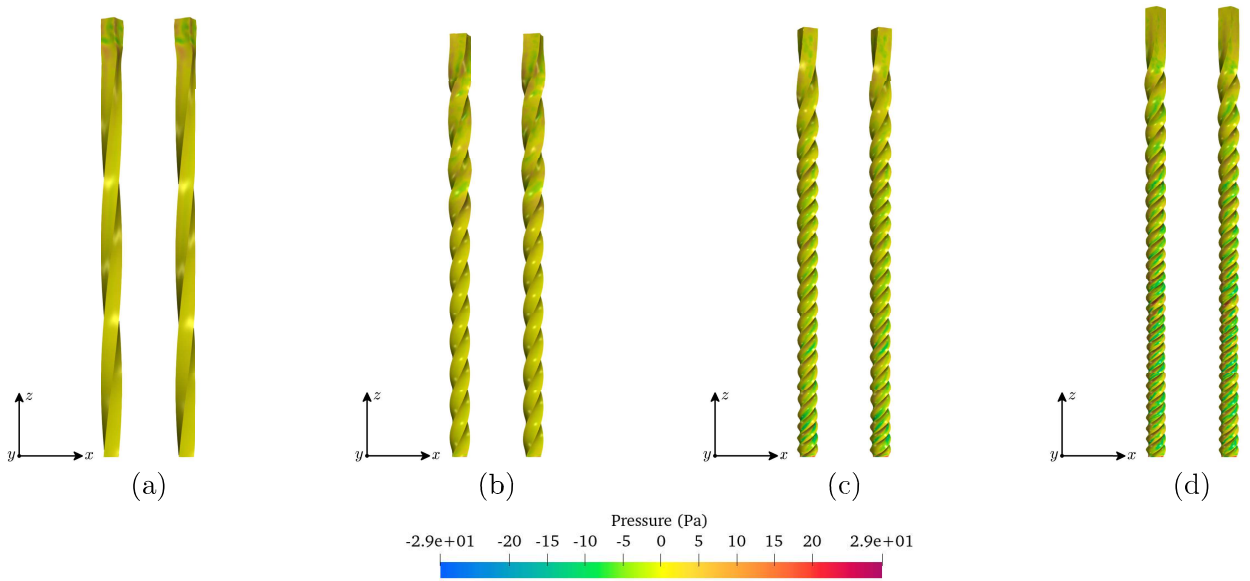

Figure 16: contour pressure distribution for time steps: (a) $3,500$; (b) $15,000$; (c) $35,000$; (d) $60,000$. The left half of each diagram represents the Finite Element solution by the Mooney-Rivlin model, and the right half represents the solution by the gradient enhanced Kriging counterpart.
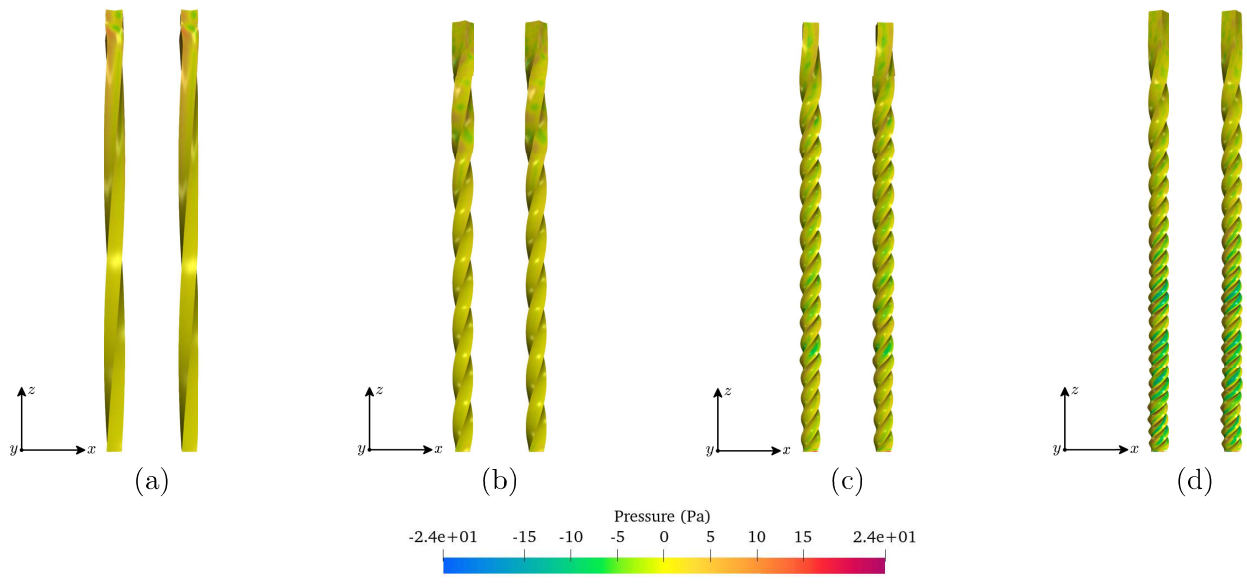
Figure 17: contour pressure distribution for time steps: (a) $2,500$; (b) $10,000$; (c) $25,000$; (d) $40,000$. The left half of each diagram represents the Finite Element solution by the rank-one laminate model, and the right half represents the solution by the gradient enhanced Kriging counterpart.

5.4. Thin Walled Wrinkling

For that we consider the following example with geometry and boundary conditions shown in Figure 18, which represents a thin cylinder of thickness 0.03 (m) and outer radious 3 (m), with a length of 4 (m). The cylinder is clampled at its base (i.e. $X_3 = 0$) and it is subjected to an initial velocity profile $\boldsymbol{v}|_{t=0}$ defined as

$$\boldsymbol{v}|_{t=0} = \begin{bmatrix} 0 & 0 & v_0 \end{bmatrix}^T ; \qquad v_0 = 80\,(\mathrm{m/s}) \tag{114}$$
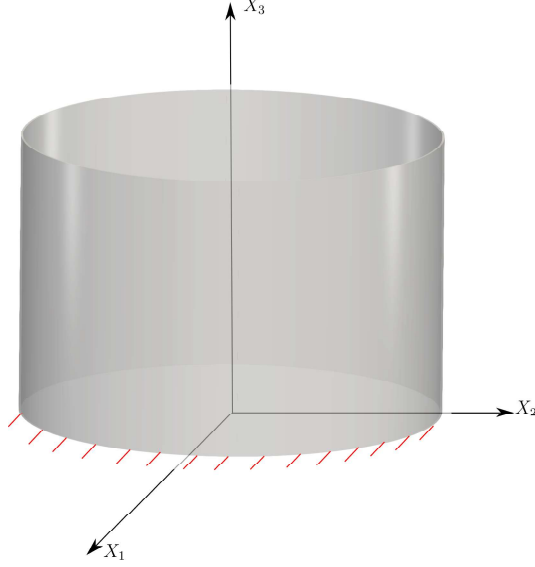
Figure 18: Thin walled wrinkling: Geometry and boundary conditions.

Regarding the Finite Element discretisation considered, $Q2$ (tri-quadratic) hexahedral elements have been used, in a mesh comprising $300 \times 480 \times 2$ elements in X_3 direction, circumferential direction, and across thickness, respectively. The simulation has been carried out using a dynamic setting, by introducing a fictitious density of $\rho = 0.01$ (kg/mm^3) and a mass-proportional Rayleigh coefficient of value 100 (1/s), by employing a leap frog time integrator, with a time step $\Delta t = 1e \times 10^{-5}$ (s). With the aim of emulating a nearly incompressible behaviour, the following volumetric term has been added to the strain energy of the model

$$\frac{\lambda}{2}(J-1)^2; \qquad \lambda = 100, \tag{115}$$

where λ plays the role of a penalty term. This has been done for the ground truth model (isotropic and rank-one laminate) and for their isotropic and transversely isotropic gradient enhanced Kriging counterparts.

In Figure 19, the results obtained from both the ground truth Mooney-Rivlin model and its gradient enhanced Kriging counterpart demonstrate a striking concurrence in the observed wrinkling pattern throughout the various depicted time steps. This alignment is not only limited to the wrinkling patterns but also extends to the contour plot distribution of the hydrostatic pressure p. Similarly, in Figure 20, a comparable level of excellence is evident in the comparison between the ground truth rank-one laminate model and its gradient enhanced transversely isotropic Kriging counterpart. Once again, this manifests in the agreement observed in the wrinkling patterns and stress distribution, further affirming the robustness and accuracy of the gradient enhanced Kriging models in capturing intricate behavior across different material models.
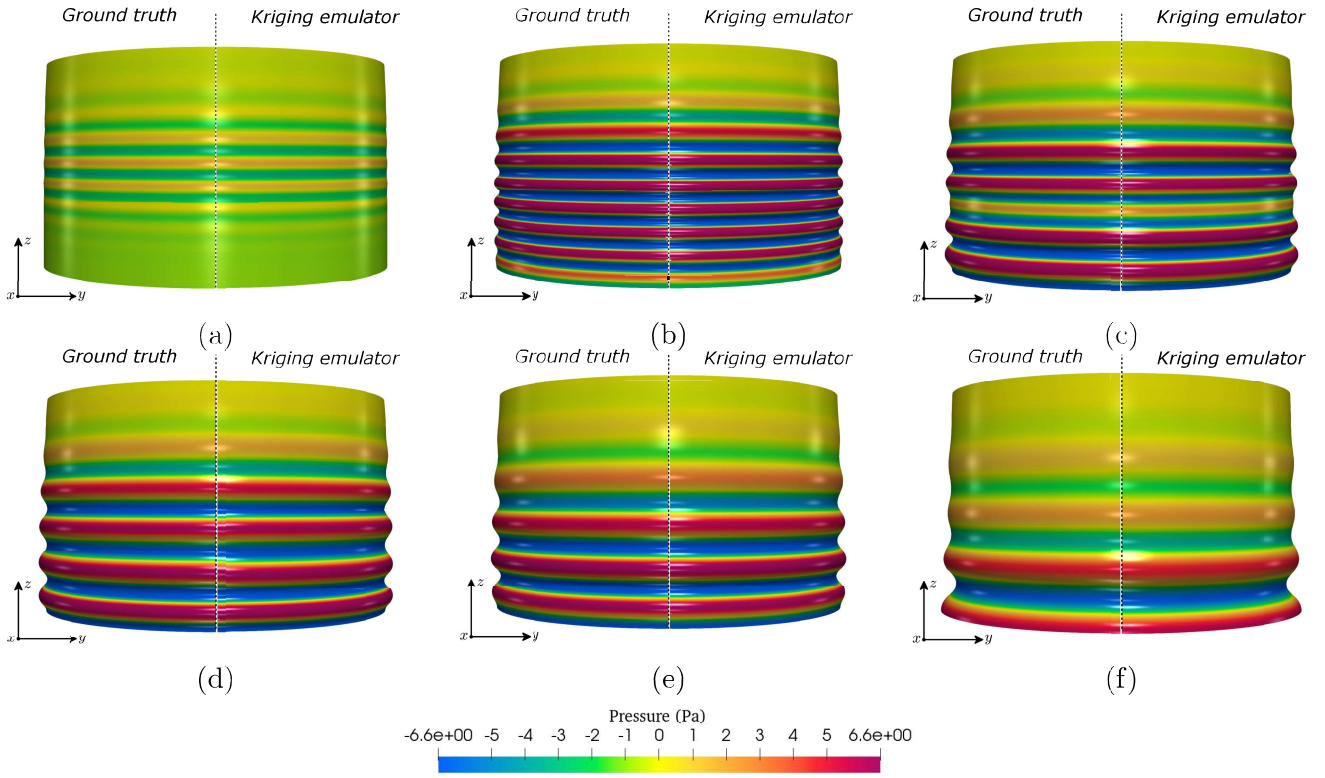
Figure 19: Contour plot distribution of hydrostatic pressurefor various time steps: (a) $2,300$; (b) $3,200$; (c) $6,200$; (d) $7,500$; (e) $11,000$; (f) $22,000$. The left half of each diagram represents the Finite Element solution by the Mooney-Rivlin model, and the right half represents the solution by the gradient enhanced Kriging counterpart.
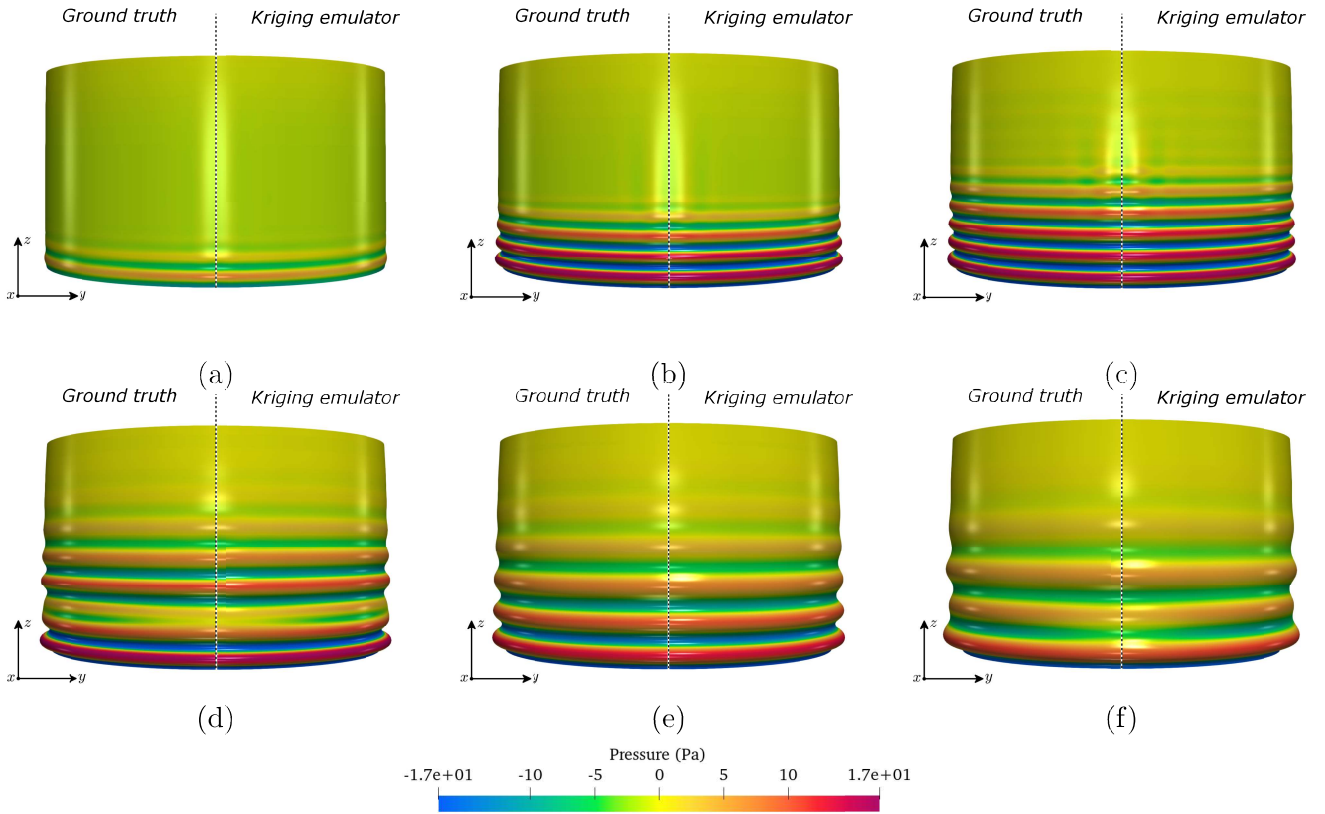

Figure 20: Contour plot distribution of hydrostatic pressurefor various time steps: (a) 550; (b) 900; (c) $1,300$; (d) $3,050$; (e) $5,000$; (f) $9,750$. The left half of each diagram represents the Finite Element solution by the rank-one laminate model, and the right half represents the solution by the gradient enhanced Kriging counterpart.

6. Concluding Remarks

This paper has advocated for a metamodeling technique based on gradient-enhanced Gaussian process regression, allowing for the emulation of diverse hyperelastic strain energy densities. By incorporating principal invariants as inputs for the surrogate of the strain energy density, this approach inherently enforces fundamental physical constraints, such as material frame indifference and material symmetry, right from the beginning.

The proposed method has demonstrated high accuracy in interpolating both the energy and the first Piola-Kirchhoff stress tensor, effectively capturing first-order derivatives with respect to the inputs. To validate its performance, the surrogate models have been implemented within a 3D Finite Element computational platform, and their accuracy has been extensively assessed in various challenging scenarios. Comparisons have been made between displacement and stress fields obtained from the surrogate models and those derived from the ground-truth analytical models. Remarkably, the proposed approach showcased its practical applicability and robustness through intricate cases of extreme twisting and buckling instabilities in a thin-walled structure. The integration of derivative information in the surrogate models resulted in exceptional accuracy, leading to a striking resemblance between the surrogate models and their corresponding ground-truth counterparts, particularly concerning their respective second derivatives. This remarkable similarity ensures that both types of models, surrogate and ground truth, exhibited comparable behavior in terms of their positive definiteness (e.g., convexity/concatenivity) and ellipticity properties. The achieved accuracy and fidelity of the surrogate models enable reliable predictions of the mechanical behavior of the materials under study, making this approach a promising tool in constitutive model in more complex scenarious such as those involving multi-physics interactions, i.e. nonlinear electro or magneto mechanics.

Appendix A. Constitutive Models

The strain energy density for the Mooney-Rivlin model is:

Appendix A.1. Mooney-Rivlin

$$U\left(I_1, I_2, I_3\right) = \frac{\mu_1}{2}\left(I_1 - 3\right) + \frac{\mu_2}{2}\left(I_2 - 3\right) - \left(\mu_1 + 2\mu_2\right)\ln\left(I_3\right) + \frac{\lambda}{2}\left(I_3 - 1\right)^2 \qquad (A.1)$$

and the material parameters used are

Table A.3: Material parameters used with the Mooney-Rivlin model.

Parameter:	μ_1	μ_2	λ
Value:	0.5	0.5	5

Appendix A.2. Quadratic Mooney-Rivlin

The strain energy density for "quadratic" Mooney-Rivlin model is:

$$U\left(I_1, I_2, I_3\right) = \frac{\mu_1}{2}\left(I_1\right)^2 + \frac{\mu_2}{2}\left(I_2\right)^2 - 6\left(\mu_1 + 2\mu_2\right)\ln\left(I_3\right) + \frac{\lambda}{2}\left(I_3 - 1\right)^2 \qquad (A.2)$$

and the material parameters used are

Table A.4: Material parameters used with the Quadratic Mooney-Rivlin model.

Parameter:	μ_1	μ_2	λ
Value:	0.5	0.5	5

Appendix A.3. Gent

The strain energy density for Gent model is:

$$U\left(I_1, I_3\right) = -\frac{\mu}{2}J_m\ln\left(1 - \frac{I_1 - 3}{J_m}\right) - \mu\ln\left(I_3\right) + \frac{\lambda}{2}\left(I_3 - 1\right)^2 \qquad (A.3)$$

and the specific values for the material parameters used are

Table A.5: Material parameters used with the Gent model.

Parameter:	μ	J_m	λ
Value:	1	19	5

Appendix A.4. Yeoh

The strain energy density for Yeoh model is:

$$U\left(I_1, I_3\right) = C_{10}\left(I_1 - 3\right) + C_{20}\left(I_1 - 3\right)^2 + C_{30}\left(I_1 - 3\right)^3 - 2C_{10}\ln\left(I_3\right) + \frac{\lambda}{2}\left(I_3 - 1\right)^2 \qquad (A.4)$$

and the material parameters used are:

Parameter:	C_{10}	C_{20}	C_{30}	λ
Value:	1	1	1	5

Appendix A.5. Arruda-Boyce

The strain energy density for Arruda-Boyce model is:

$$U(I_1, I_3) = a_1 \left(\beta(I_1) \lambda_c(I_1) - a_2 \ln \left(\frac{\sinh(\beta(I_1))}{\beta(I_1)} \right) \right) - A \ln(I_3) + \frac{1}{2} \lambda(I_3 - 1)^2 \tag{A.5}$$

where

$$\lambda_c(I_1) = \sqrt{\frac{1}{3}} \sqrt{I_1}; \qquad \mathcal{L}^{-1}(x) = \frac{3x - x^3}{1 - x^2}; \qquad \beta(I_1) = \mathcal{L}^{-1}\left(\frac{\lambda_c(I_1)}{a_2} \right)$$

The material parameters used in the model are:

Table A.7: Material parameters used with the Arruda-Boyce model.

Parameter:	a_1	a_2	λ
Value:	2.1899	$\sqrt{6}$	4.9159

Appendix A.6. Transversely Isotropic

The strain energy of the polyconvex transversely isotropic model is:

$$U(I_1, I_2, I_3, I_4, I_5) = \frac{\mu_1}{2}(I_1 - 3) + \frac{\mu_2}{2}(I_2 - 3) - (\mu_1 + 2\mu_2 + \mu_3)\ln(I_3) \tag{A.6}$$

$$+ \frac{\lambda}{2}(I_3 - 1)^2 + \frac{\mu_3}{2\alpha}(I_4)^\alpha + \frac{\mu_3}{2\beta}(I_5)^\beta + \frac{1}{2}\left(\frac{1}{2\alpha}\mu_3 + \frac{1}{2\beta}\mu_3 \right)$$

The material parameters used for this model are:

Table A.8: Material parameters used with the Transversely Isotropic model.

Parameter:	μ_1	μ_2	μ_3	λ	α	β	$\boldsymbol{N}$
Value:	0.5	0.5	7.5	5	2	2	$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$

Appendix A.7. Rank-One Laminate

We consider Mooney-Rivlin strain energy densities for the individual phases a and b within this composite (refer to Section 2.6.1), namely

$$U^a(I_1^a, I_2^a, I_3^a) = \frac{1}{2}\mu_1^a(I_1^a - 3) + \frac{1}{2}\mu_2^a(I_2^a - 3) - (\mu_1^a + 2\mu_2^a)\ln(I_3^a) + \frac{1}{2}\lambda^a(I_3^a - 1)^2$$

$$U^b\left(I_1^b, I_2^b, I_3^b\right) = \frac{1}{2}\mu_1^b\left(I_1^b - 3\right) + \frac{1}{2}\mu_2^b\left(I_2^b - 3\right) - \left(\mu_1^b + 2\mu_2^b\right)\ln\left(I_3^b\right) + \frac{1}{2}\lambda^b\left(I_3^b - 1\right)^2$$

$$\tag{A.7}$$

being the effective strain energy $\Psi(\boldsymbol{F})$

$$\Psi(\boldsymbol{F}) = \arg\min_{\boldsymbol{\alpha}}\{\hat{\Psi}(\boldsymbol{F}, \boldsymbol{\alpha})\}; \qquad \hat{\Psi}(\boldsymbol{F}, \boldsymbol{\alpha}) = c^a \Psi^a\left(\boldsymbol{F}^a(\boldsymbol{F}, \boldsymbol{\alpha})\right) + c^b \Psi^b\left(\boldsymbol{F}^b(\boldsymbol{F}, \boldsymbol{\alpha})\right), \tag{A.8}$$

with

$$\Psi^a\left(\mathbf{F}^a\left(\mathbf{F}, \boldsymbol{\alpha}\right)\right) = U^a\left(I_1^a, I_2^a, I_3^a\right); \qquad \Psi^b\left(\mathbf{F}^b\left(\mathbf{F}, \boldsymbol{\alpha}\right)\right) = U^b\left(I_1^b, I_2^b, I_3^b\right) \tag{A.9}$$

where $\{I_1^a, I_2^a, I_3^a\}$ and $\{I_1^b, I_2^b, I_3^b\}$ represent the principal invariants of $\mathbf{F}^a$ and $\mathbf{F}^b$, related to the macroscopic deformation gradient tensor $\mathbf{F}$ through equation (37). The material parameters used for this composite material are found below:

Table A.9: Material parameters used with the Rank One Laminate model.

Parameter:	μ_1^a	μ_2^a	λ^a	α	β	f_m	c
Value:	0.5	0.5	5	$\pi/4$	$\pi/4$	$\{1.1, 2, 10, 100\}$	0.6

where f_m represents the contrast in material properties, namely

$$f_m = \frac{\mu_1^b}{\mu_1^a} = \frac{\mu_2^b}{\mu_2^a} = \frac{\lambda^b}{\lambda^a} \tag{A.10}$$

Appendix A.8. RVE with spherical inclusions

The RVE is divided into the region associated to the matrix $\mathbf{X}_\mu \in \mathcal{B}_{0_\mu}^m$ and the inclusion $\mathbf{X}_\mu \in \mathcal{B}_{0_\mu}^i$, such that $\mathcal{B}_{0_\mu} = \mathcal{B}_{0_\mu}^m \cup \mathcal{B}_{0_\mu}^i$, and $\mathcal{B}_{0_\mu}^m \cap \mathcal{B}_{0_\mu}^i = \emptyset$. At each region, we define the energy density $U_\mu(\mathbf{X}_\mu, I_{1_\mu}, I_{2_\mu}, I_{3_\mu}) = \Psi_\mu(\mathbf{X}_\mu, \mathbf{F}_\mu)$ according to

$$\Psi_\mu(\mathbf{X}_\mu, \mathbf{F}_\mu) = U_\mu(\mathbf{X}_\mu, I_{1_\mu}, I_{2_\mu}, I_{3_\mu}) = \begin{cases} U_\mu^m(I_{1_\mu}, I_{2_\mu}, I_{3_\mu}) & \mathbf{X}_\mu \in \mathcal{B}_{0_\mu}^m \\ U_\mu^i(I_{1_\mu}, I_{2_\mu}, I_{3_\mu}) & \mathbf{X}_\mu \in \mathcal{B}_{0_\mu}^i \end{cases} \tag{A.11}$$

We consider Mooney-Rivlin strain energy densities of both matrix and inclusion, namely

$$U_\mu^m(I_{1_\mu}, I_{2_\mu}, I_{3_\mu}) = \frac{1}{2}\mu_1^m\left(I_{1_\mu}^m - 3\right) + \frac{1}{2}\mu_2^m\left(I_{2_\mu}^m - 3\right) - (\mu_1^m + 2\mu_2^m)\ln\left(I_{3_\mu}^m\right) + \frac{1}{2}\lambda^m\left(I_{3_\mu}^m - 1\right)^2$$
$$U_\mu^i(I_{1_\mu}, I_{2_\mu}, I_{3_\mu}) = \frac{1}{2}\mu_1^i\left(I_{1_\mu}^i - 3\right) + \frac{1}{2}\mu_2^i\left(I_{2_\mu}^i - 3\right) - \left(\mu_1^i + 2\mu_2^i\right)\ln\left(I_{3_\mu}^i\right) + \frac{1}{2}\lambda^i\left(I_{3_\mu}^i - 1\right)^2$$

$$\tag{A.12}$$

The material parameters used for this composite material are found below:

Table A.10: Material parameters used for the RVE.

Parameter:	μ_1^m	μ_2^m	λ^m	f_m	R
Value:	0.5	0.5	5	10	0.5

where R represents the diameter of the centered sphere within the hexahedral RVE and f_m represents the contrast in material properties, namely

$$f_m = \frac{\mu_1^i}{\mu_1^m} = \frac{\mu_2^i}{\mu_2^m} = \frac{\lambda^i}{\lambda^m} \tag{A.13}$$

Appendix B. Basis for Symmetric Traceless Second Order Tensors

$$\Psi_1 = \sqrt{\frac{1}{6}} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad \Psi_2 = \sqrt{\frac{1}{2}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad \Psi_3 = \sqrt{\frac{1}{2}} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{B.1}$$

$$\Psi_4 = \sqrt{\frac{1}{2}} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \Psi_5 = \sqrt{\frac{1}{2}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{B.2}$$

Appendix C. Gradient-enhanced Gaussian-process based prediction using a single observation point in the strain energy

The gradient enhanced Kriging approach, described in Section 3.4, can be particularised to the case when there is only one observation point for energy, whilst still retaining n observation points for the derivatives of the energy with respect to invariants. In order to particularise this approach to this scenario, the vector of observations $\boldsymbol{U}$ is now defined as

$$\boldsymbol{U} = \left[U^{(1)} \partial_{\mathbf{I}} U^{(1)}, \ldots, \partial_{\mathbf{I}} U^{(m)} \right]^T, \tag{C.1}$$

where

$$U^{(1)} = U(\boldsymbol{I}^{(1)}) \qquad \partial_{\mathbf{I}} U^{(i)} = \left[\partial_{I_1} U^{(i)}, \ldots, \partial_{I_n} U^{(i)} \right]^T. \tag{C.2}$$

The correlation matrix $\boldsymbol{R}$ is similar to that in (73), with the same block structure, namely

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{UU} & \boldsymbol{R}_{UU'} \\ \boldsymbol{R}_{UU'}^T & \boldsymbol{R}_{U'U'} \end{bmatrix}, \tag{C.3}$$

with $\boldsymbol{R}_{UU} = \mathcal{R}(\mathbf{I}^1, \mathbf{I}^1, \boldsymbol{\theta})$. $\boldsymbol{R}_{UU'}$ includes the partial derivatives of $\mathcal{R}$ according to

$$\boldsymbol{R}_{UU'} = \left[\partial_{\mathbf{I}^{(1)}} \mathcal{R}(\mathbf{I}^{(1)}, \mathbf{I}^{(1)}, \boldsymbol{\theta}) \quad \ldots \quad \partial_{\mathbf{I}^{(m)}} \mathcal{R}(\mathbf{I}^{(1)}, \mathbf{I}^{(m)}, \boldsymbol{\theta}) \right], \tag{C.4}$$

given

$$\partial_{\mathbf{I}^{(j)}} \mathcal{R}(\mathbf{I}^{(i)}, \mathbf{I}^{(j)}, \boldsymbol{\theta}) = \left[\frac{\partial \mathcal{R}(\mathbf{I}^{(i)}, \mathbf{I}^{(j)}, \boldsymbol{\theta})}{\partial I_1^{(j)}}, \frac{\partial \mathcal{R}(\mathbf{I}^{(i)}, \mathbf{I}^{(j)}, \boldsymbol{\theta})}{\partial I_2^{(j)}}, \ldots, \frac{\partial \mathcal{R}(\mathbf{I}^{(i)}, \mathbf{I}^{(j)}, \boldsymbol{\theta})}{\partial I_n^{(j)}} \right]^T. \tag{C.5}$$

The submatrix $\boldsymbol{R}_{U'U'}$ exactly the same as that in (76)-(77). Similarly the vector of cross-correlations between the observations and the prediction is extended as follows

$$\boldsymbol{r}(\mathbf{I}) = \left[\mathcal{R}(\mathbf{I}, \mathbf{I}^{(1)}, \boldsymbol{\theta}), \partial_{\mathbf{I}^{(1)}} \mathcal{R}(\mathbf{I}, \mathbf{I}^{(1)}, \boldsymbol{\theta}), \ldots, \partial_{\mathbf{I}^{(m)}} \mathcal{R}(\mathbf{I}, \mathbf{I}^{(m)}, \boldsymbol{\theta}) \right]^T. \tag{C.6}$$

Once these adaptations have been made, the new definitions for the various quantities can be substituted into the definitions detailed in Subsections 3.2 and 3.3. To start with, recall the mean prediction

$$\mu_{\hat{U}}(\mathbf{I}) = \boldsymbol{g}(\mathbf{I}) \cdot \hat{\boldsymbol{\beta}} + \boldsymbol{r}(\mathbf{I}) \cdot \boldsymbol{R}^{-1} \left(\boldsymbol{U} - \boldsymbol{G}\hat{\boldsymbol{\beta}} \right), \tag{C.7}$$

and the variance

$$\sigma_{\hat{U}}^2(\mathbf{I}) = \sigma_U^2 \left(1 - \boldsymbol{r}(\mathbf{I}) \cdot \boldsymbol{R}^{-1} \boldsymbol{r}(\mathbf{I}) + \boldsymbol{u}(\mathbf{I}) \cdot \left(\boldsymbol{G}^T \boldsymbol{R}^{-1} \boldsymbol{G} \right)^{-1} \boldsymbol{u}(\mathbf{I}) \right), \tag{C.8}$$

with

$$\boldsymbol{\beta}^*(\boldsymbol{\theta}) = \left(\boldsymbol{G}^T \boldsymbol{R}(\boldsymbol{\theta})^{-1} \boldsymbol{G} \right)^{-1} \boldsymbol{G}^T \left(\boldsymbol{R}(\boldsymbol{\theta}) \right)^{-1} \boldsymbol{U};$$

$$\sigma_U^{2 \, *}(\boldsymbol{\theta}) = \frac{1}{1 + nm} \left(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta}^*(\boldsymbol{\theta}) \right) \cdot \left(\boldsymbol{R}(\boldsymbol{\theta}) \right)^{-1} \cdot \left(\boldsymbol{U} - \boldsymbol{G}\boldsymbol{\beta}^*(\boldsymbol{\theta}) \right); \tag{C.9}$$

and

$$\boldsymbol{G}(\mathbf{I}) = \begin{bmatrix} \boldsymbol{G_U} \\ \boldsymbol{G_{U'}} \end{bmatrix}; \qquad \boldsymbol{G_U} = \boldsymbol{g}(\mathbf{I}^{(1)}); \qquad \boldsymbol{G_{U'}} = \begin{bmatrix} \left(\partial_{\mathbf{I}^{(1)}} \boldsymbol{g}(\mathbf{I}^{(1)})\right)^T \\ \vdots \\ \left(\partial_{\mathbf{I}^{(m)}} \boldsymbol{g}(\mathbf{I}^{(m)})\right)^T \end{bmatrix}. \tag{C.10}$$

[1] K. Bertoldi, M. Gei, Instabilities in multilayered soft dielectrics, JMPS 59 (2011) 18–42. doi:https://doi.org/10.1016/j.jmps.2010.10.001.

[2] M. Gei, R. Springhetti, E. Bortot, Performance of soft dielectric laminated composites, Smart Materials and Structures 22 (2013) 1–8. doi:https://doi.org/10.1088/0964-1726/22/10/104014.

[3] L. Tian, L. Tevet-Deree, G. deBotton, K. Bhattacharya, Dielectric elastomer composites, JMPS 60 (2012) 181–198. doi:https://doi.org/10.1016/j.jmps.2011.08.005.

[4] F. Marín, J. Martínez-Frutos, R. Ortigosa, A. J. Gil, A convex multi-variable based computational framework for multilayered electro-active polymers, CMAME 374 (2021) 1–42. doi:https://doi.org/10.1016/j.cma.2020.113567.

[5] F. Marín, R. Ortigosa, J. Martínez-Frutos, A. J. Gil, Viscoelastic up-scaling rank-one effects in in-silico modelling of electro-active polymers, CMAME 389 (2022) 1–44. doi:https://doi.org/10.1016/j.cma.2021.114358.

[6] G. deBotton, L. Tevet-Deree, E. A. Socolsky, Electroactive heterogeneous polymers: Analysis and applications to laminated composites, Mechanics of Advanced Materials and Structures 14 (2007) 13–22. doi:https://doi.org/10.1080/15376490600864372.

[7] K. Linka, M. Hillgärtner, K. P. Abdolazizi, R. C. Aydin, M. Itskov, C. J. Cryon, Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning, Journal of Computational Physics 429 (2021) 1–17. doi:https://doi.org/10.1016/j.jcp.2020.110010.

[8] D. K. Klein, R. Ortigosa, J. Martínez-Frutos, O. Weeger, Finite electro-elasticity with physics-augmented neural networks, CMAME 400 (2022) 1–33. doi:https://doi.org/10.1016/j.cma.2022.115501.

[9] A. Aggarwal, B. S. Jensen, S. Pant, C. Lee, Strain energy density as a gaussian process and its utilization in stochastic finite element analysis: Application to planar soft tissues, CMAME 404 (2023) 1–27. doi:https://doi.org/10.1016/j.cma.2022.115812.

[10] A. L. Frankel, R. E. Jones, L. P. Swiler, Tensor basis gaussian process models of hyperelastic materials, Journal of Machines Learning for Modelling and Computing 1 (2020) 1–17.

[11] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning, MIT Press, Cambridge, Massachusetts, 2006.

[12] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006.

[13] J. I. Marden, Multivariate Statistics: Old School, CreateSpace Independent Publishing, 2015.

[14] J. N. Fuhg, M. Marino, N. Bouklas, Local approximate gaussian process regression for data-driven constitutive models: development and comparison with neural networks, CMAME 388 (2022) 1–23. doi:https://doi.org/10.1016/j.cma.2021.114217.

[15] D. G. Krige, A statistical approach to some basic mine valuation problems on the witwatersrand, Journal of the Southern African Institute of Mining and Metallurgy 52 (6) (1951) 119–139.

[16] T. Santner, B. Williams, W. Notz, The design and analysis of computer experiments, Springer series in Statistics. Springer, 2003.

[17] G. Matheron, Trait de gostatistique applique, Editions Technip, 1962.

[18] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, Statistical Science 4 (4) (1989) 409–423.

[19] D. Jones, A taxonomy of global optimization methods based on response surfaces, Journal of Global Optimization (21) (2001) 345–383.

[20] J. Ollar, C. Mortished, R. Jones, J. Sienz, V. Toropov, Gradient based hyper-parameter optimisation for well conditioned kriging metamodels, Structural and Multidisciplinary Optimization 55 (2017) 2029–2044. doi:https://doi.org/10.1007/s00158-016-1626-8.

[21] J. Martínez-Frutos, D. Herrero, Kriging-based infill sampling criterion for constraint handling in multi-objective optimization, Journal of Global Optimization (64) (2016) 97–115.

[22] J. M. Ball, Convexity conditions and existence theorems in nonlinear elasticity, Archive for Rational Mechanics and Analysis 63 (4) (1976) 337–403.

[23] J. M. Ball, Geometry, Mechanics and Dynamics, Springer, 2002, Ch. Some open problems in Elasticity, pp. 3–59.

[24] R. Ortigosa, A. Gil, A new framework for large strain electromechanics based on convex multi-variable strain energies: Conservation laws, hyperbolicity and extension to electro-magneto-mechanics, Computer Methods in Applied Mechanics and Engineering 309 (2016) 202–242. doi:https://doi.org/10.1016/j.cma.2016.05.019.

[25] J. Schröder, P. Neff, V. Ebbing, Anisotropic polyconvex energies on the basis of crystallographic motived structural tensors, Journal of Mechanics and Physics of Solids 56 (2008) 3486–3506.

[26] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, Statistical Science (4) (1989) 409–435.

[27] J. Kleijnen, Kriging metamodeling in simulation: a review, Eur. J. Oper. Res. (192) (2009) 707–716.

[28] V. Dubourg, Adaptive surrogate models for reliability analysis and reliability-based design optimization, Ph.D. thesis, Universite Blaise Pascal - Clermont II (2011).

[29] J.Nocedal, S. Wright, Numerical Optimization, Springer, New York, 2006.

[30] D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Professional, 1989.

[31] H. Li, Y. Tian, Dynamic hill climber: A new optimizer for continuous optimization problems, Entropy 20 (4) (2018) 275.

[32] MathWorks, Matlab optimization toolbox, https://www.mathworks.com/help/optim/ug/fmincon.html (Accedido en 2023).

[33] X. Wu, Y. Yuan, et al., Boxmin: A fast and accurate derivative-free algorithm for black-box optimization, IEEE Transactions on Cybernetics 50 (2) (2020) 503–515.

[34] Z. Han, S. Görtz, R. Zimmermann, Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function, Aerosp Sci Technol (25) (2013) 177–289.

[35] L. Laurent, R. L. Riche, B. Soulier, P.-A. Boucard, An overview of gradient-enhanced metamodels with applications, Archives of Computational Methods in Engineering 26 (2019) 61–106.

[36] O. Kunc, F. Fritzen, Finite strain homogenization using a reduced basis and efficient sampling, Mathematical and Computational Applications 24 (2019) 1–28. doi:https://doi.org/10.3390/mca24020056.