Unsupervised discovery of interpretable hyperelastic constitutive laws

Flaschel, Moritz; Kumar, Siddhant; De Lorenzis, Laura

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Unsupervised discovery of interpretable hyperelastic constitutive laws

Moritz Flaschel[a,1], Siddhant Kumar[a,b,1], Laura De Lorenzis[a,*]

[a] *Department of Mechanical and Process Engineering, ETH Zürich, 8092 Zürich, Switzerland*
[b] *Department of Materials Science and Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands*

Abstract

We propose a new approach for data-driven automated discovery of isotropic hyperelastic constitutive laws. The approach is *unsupervised*, i.e., it requires no stress data but only displacement and global force data, which are realistically available through mechanical testing and digital image correlation techniques; it delivers *interpretable* models, i.e., models that are embodied by parsimonious mathematical expressions discovered through sparse regression of a large catalogue of candidate functions; it is *one-shot*, i.e., discovery only needs one experiment — but can use more if available. The problem of unsupervised discovery is solved by enforcing equilibrium constraints in the bulk and at the loaded boundary of the domain. Sparsity of the solution is achieved by ℓ_p regularization combined with thresholding, which calls for a non-linear optimization scheme. The ensuing fully automated algorithm leverages physics-based constraints for the automatic determination of the penalty parameter in the regularization term. Using numerically generated data including artificial noise, we demonstrate the ability of the approach to accurately discover five hyperelastic models of different complexity. We also show that, if a "true" feature is missing in the function library, the proposed approach is able to surrogate it in such a way that the actual response is still accurately predicted.

1. Introduction

Data-driven approaches enabled by machine learning tools and facilitated by the large availability of data through modern experimental techniques are raising a rapidly growing interest in computational solid mechanics. A main focus of recent investigations has been the possibility to *bypass* or *surrogate* the development of material models, based on the recognition that, unlike the other components of a mechanical boundary value problem (conservation laws and essential constraints), the constitutive modeling step is of empirical, non-epistemic nature. As follows, we attempt a brief overview of the main available approaches for reversible material behavior (elasticity).

Bypassing material modeling altogether is the core idea of the (constitutive) model-free data-driven computing paradigm [1,2]. In these approaches, a boundary value problem is solved by retaining conservation laws and essential

---

constraints, and substituting the constitutive model with the direct use of data. The pioneering approach in [1] results in a data-driven solver which aims at assigning to each material point the state satisfying the conservation laws that is closest to the dataset. The first demonstrations focused on linear elasticity [3–5] but recent attempts extend it to geometrically nonlinear elasticity [3], general elasticity [4], elastodynamics [5], inelasticity [6], and fracture mechanics [7]. In [8], the phenomenological constitutive model is replaced with an experimental constitutive manifold, which is reconstructed from data using manifold learning methods. In a subsequent investigation focusing on elasticity [2], the data are used to identify a polynomial approximation of the strain energy density. Recently, this approach has been extended to a setting that fulfills the thermodynamics principles by construction [9]. More recent contributions pursue a hybrid approach where data are used to construct automatic corrections to existing hyperelastic and elastoplastic models [10,11].

A second stream of investigations has focused on the idea of *surrogating* constitutive models, e.g., through piecewise polynomial interpolation [12,13] or artificial neural networks (ANNs). The first use of ANNs to encode material models dates back to the 1990s [14]. The approach consists in replacing a constitutive relation with the training of an ANN based on data. The trained ANN encodes the relation between input and output variables (in the simplest case, strains and stresses, respectively) and can then be used as a black-box substitute of the constitutive equations. The method was further developed for a variety of applications [14–18]. Over alternative approaches to encode constitutive relations such as interpolations using piecewise linear functions, polynomial functions, or radial basis functions (RBF), ANN and especially deep ANN possess several advantages [19]: using appropriate regularization, they have good generalization properties to unseen data (albeit they show poor extrapolation beyond the interpolatory hull of the training data); they work well even for complex, highly inhomogeneous or anisotropic distribution of training points, i.e., dense along certain directions and sparse along others.

The above two categories of approaches are clearly *uninterpretable* from the standpoint of the material model, as they suppress it or encode it with black-box ANNs, respectively. This limits the insight that they can bring towards the physical understanding of the material behavior based on experimental observations. A further crucial aspect is that the vast majority of the approaches of both categories are rooted in *supervised learning*, i.e., they need data consisting of input–output pairs and thus must rely on stress data. Datasets may come from experimental measurements, or be generated computationally by simulations at the lower scales within multiscale approaches [20–22]. For mechanical tests, stress data are only obtainable in the simplest situations, e.g. uniaxial tensile or bending tests. The comprehensive observation of strain–stress relations relying on these tests is nearly impossible. Multiscale simulations can generate training datasets, but their computational cost is still unaffordable. Recognition of this issue is very recent in both the above categories of approaches. Within the data-driven computing paradigm, it motivated the development of the data-driven identification method [23,24], which formulates the inverse problem associated to the approach in [1]. By using only displacement field and load measurements, a set of admissible material strain–stress states is recovered with no assumption on the constitutive equation. Within the stream of research on surrogating constitutive models with ANNs, recent attempts to use only displacement and global force data have been performed in [19,25,26]. These approaches bear relation to the so-called physics-informed neural networks (PINN) [27]. The training of PINNs is performed with a cost function that, in addition to data, includes the governing equations, initial and boundary conditions. However, for the time being, these methods are limited to only very simple cases (constitutive models of known form with unknown parameters or unknown constitutive models but for one-dimensional cases) and very simple ordinary and partial differential equations (ODEs and PDEs) — e.g., common contemporary benchmarks include Burgers' equation, Darcy flow, and the advection–diffusion equation. In contrast, the "constitutive ingredient" (to be discovered) is much more complex and highly nonlinear in most solid mechanics PDEs, even in hyperelasticity.

A recent inspiring stream of research initiated by the physics community focuses on automated discovery of physical laws and governing equations [28]. These approaches are interpretable as they lead to closed-form expressions of the governing ODEs and PDEs for the investigated systems. A breakthrough was achieved in [29] by departing from a high-dimensional library of nonlinear candidate functions and using sparse regression to only select the dominant items, thus uncovering parsimonious governing ODEs. The approach has drawn tremendous attention [30–38]. An important extension has addressed the data-driven discovery of PDEs with time and spatial coordinates as variables [39,40]. Crucial improvements have been obtained by substituting numerical differentiation with automatic differentiation [41] and smoothing noisy data with neural networks [42–44]. These approaches do not limit in any way the form of the sought ODE or PDE, but only depart from the assumption of a possibly

very large catalogue of candidate component functions. Thus, apart from the choice of this catalogue, they do not leverage domain knowledge in the form of constraints stemming e.g. from conservation laws, thermodynamics, or symmetry requirements and seek the direct determination of the entire PDE. The consequences are two-fold: if directly applied to solid mechanics, these approaches may be more challenging than needed, and they may fail to comply with the epistemic requirements based on physics which amount to essential components of solid mechanics boundary value problems.

In this paper, inspired by the above ideas and tools, we propose a new approach which aims neither to bypass nor to surrogate, but rather to *automatically discover* constitutive material models, focusing on isotropic hyperelasticity in this first investigation. The approach delivers *interpretable* models, i.e., models that are embodied by parsimonious mathematical expressions based on sparse regression of a large catalogue of candidate functions, and is based on *unsupervised learning*, i.e., it takes as input only experimentally measurable data in the form of full-field displacements, as obtainable e.g., from digital image correlation (DIC) techniques, and global force data delivered by mechanical testing machines. Unsupervised discovery is achieved by enforcing the constraints stemming from physics, including balance of linear momentum and additional constraints for the elastic strain energy. Finally, using machine learning jargon, our approach is *one-shot*, meaning that discovery only needs one experiment — but can obviously use more if available. This is in contrast to conventional material model calibration techniques as well as supervised learning methods, that require several different experiments (e.g., uniaxial or biaxial tension, shearing, torsion, randomized strain paths, etc.).

2. Unsupervised discovery of hyperelastic constitutive laws

2.1. Problem setting

Consider a reference domain $\Omega \in \mathbb{R}^2$, with boundary $\partial\Omega$, subjected to a quasi-static mechanical test.[2] Dirichlet and Neumann boundary conditions are applied by the testing machine on $\partial\Omega_u \subseteq \partial\Omega$ and $\partial\Omega_t = \partial\Omega \backslash \partial\Omega_u$, respectively. Without loss of generality, we assume the experiments delivering the data to be conducted in displacement control, so that the Neumann boundary conditions are homogeneous and the forces applied by the testing machine to the specimen correspond to the reaction forces at the Dirichlet boundary.

Let $\mathcal{X} = \{\boldsymbol{X}^a \in \Omega : a = 1, \ldots, n_n\}$ denote a set of coordinates of n_n nodes in the reference configuration for which the displacements $\mathcal{U} = \{\boldsymbol{u}^a \in \mathbb{R}^2 : a = 1, \ldots, n_n\}$ are known. In the present two-dimensional setting, this discrete full-field displacement dataset is assumed to be known through DIC measurements. The extension to the three-dimensional setting would be conceptually straightforward and would require full-field displacement data e.g. from computed tomography combined with digital volume correlation. Additionally, the net reaction forces on portions of $\partial\Omega_u$ (corresponding to, e.g., each loaded side of the specimen) are assumed to be known and emulate the measurements from load cells.

Armed with these data, our objective is to discover the underlying constitutive law of the unknown material in one-shot (i.e., we only consider one loading experiment). We note that the experiment must be complex enough to activate a diverse range of deformation modes. We demonstrate in Section 3 that biaxial tension of a plate with a hole suffices for this purpose. For the scope of this initial work, our key assumptions are that the material is homogeneous, isotropic, and hyperelastic. Fig. 1 illustrates a step-by-step schematics of the proposed unsupervised algorithm, whereas a detailed description is presented in the following sections.

2.2. Displacement field approximation via finite element mesh

To facilitate derivatives and integrals of the field quantities, we associate the grid at which the displacements are known to a finite element mesh with the nodal set $\mathcal{X}$. The displacement field is then approximated as

$$\boldsymbol{u}(\boldsymbol{X}) = \sum_{a=1}^{n_n} N^a(\boldsymbol{X})\,\boldsymbol{u}^a, \tag{1}$$

---

[2] While we consider a two-dimensional setting here motivated by DIC experiments, the subsequent derivations are also applicable to the three-dimensional case (which is not our focus here).
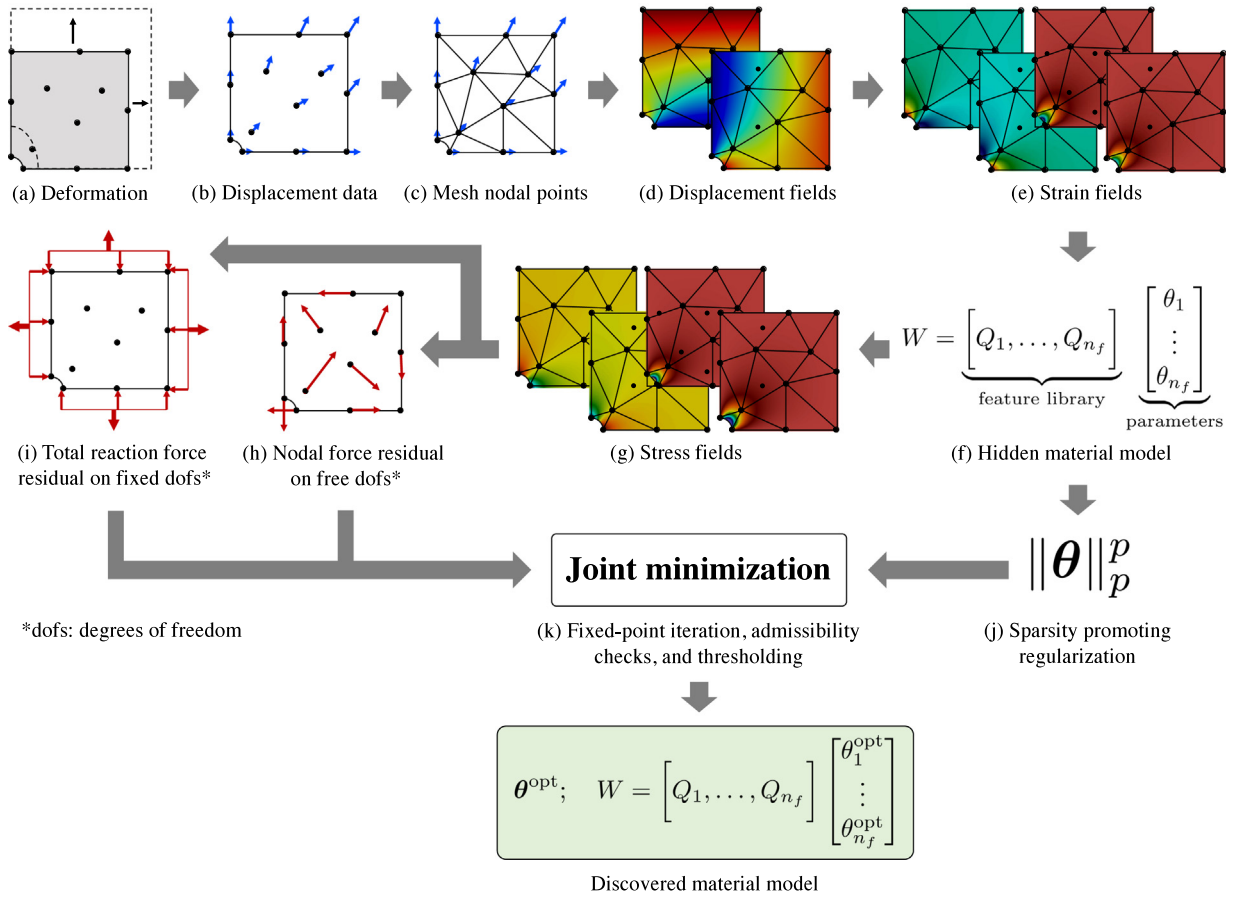
Fig. 1. Schematics of the unsupervised algorithm for discovering interpretable hyperelastic constitutive laws. Starting from a deformed body under loading (a), the point-wise displacement measurements (b) are interpolated by constructing a finite element mesh (c). The resulting continuous displacement field (d) is differentiated to obtain the strain field (e). The material strain energy density W is formulated as a linear combination of a large catalogue of nonlinear features $\boldsymbol{Q}$ and unknown material parameters $\boldsymbol{\theta}$ (f). The derivative of the strain energy density with respect to the strain field yields the stress field (g). A joint optimization problem (k) is formulated to find the unknown material parameters $\boldsymbol{\theta}$ such that the weak form of the linear momentum balance in the bulk material (h) and the reaction force balance on the Dirichlet boundaries (i) are satisfied. The joint optimization problem (k) also includes a sparsity-promoting regularization (j) to yield a parsimonious and interpretable material model (characterized by $\boldsymbol{\theta}^{\mathrm{opt}}$).

where $N^a : \Omega \to \mathbb{R}$ is the shape function associated with the node of reference coordinate $\boldsymbol{X}^a$. The deformation gradient is given by

$$\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{I} + \sum_{a=1}^{n_n} \boldsymbol{u}^a \otimes \nabla N^a(\boldsymbol{X}), \tag{2}$$

where ∇ denotes the gradient operator with respect to the reference coordinates.

2.3. Material model library

The objective is to find a constitutive law for the unknown material under consideration, e.g. a relation between the deformation gradient $\boldsymbol{F}$ and the first Piola–Kirchhoff stress tensor $\boldsymbol{P}$. Since the material is hyperelastic, our objective reduces to discovering the underlying strain energy density $W(\boldsymbol{F})$ from which the stress can be derived as $\boldsymbol{P} = \partial W/\partial \boldsymbol{F}$. Moreover, the objectivity requirement is a priori satisfied if we assume W to be a function of the right Cauchy–Green deformation tensor $\boldsymbol{C} = \boldsymbol{F}^T \boldsymbol{F}$. Finally, in the scope of isotropic materials, the strain energy

density can be further reduced to a function of the invariants of $\boldsymbol{C}$ as

$$W(\boldsymbol{C}) \equiv W(I_1(\boldsymbol{C}), I_2(\boldsymbol{C}), I_3(\boldsymbol{C})), \qquad \text{with} \quad I_1(\boldsymbol{C}) = \text{tr}(\boldsymbol{C}), \quad I_2(\boldsymbol{C}) = \frac{1}{2}[\text{tr}(\boldsymbol{C})^2 - \text{tr}(\boldsymbol{C}^2)], \quad I_3(\boldsymbol{C}) = \det(\boldsymbol{C}). \tag{3}$$

We consider a large feature library $\boldsymbol{Q} : \mathbb{R}^3 \rightarrow \mathbb{R}^{n_f}$ of n_f nonlinear functions of I_1, I_2, and I_3 whose linear combination yields the strain energy density as

$$W(I_1, I_2, I_3) = \boldsymbol{Q}^T(I_1, I_2, I_3)\boldsymbol{\theta}, \tag{4}$$

where $\boldsymbol{\theta} \in \mathbb{R}^{n_f}$ are the unknown material parameters to be estimated. In principle, any arbitrary strain energy density can be represented by (4) if the feature library is appropriately chosen. In the context of our work, we choose the following library

$$\boldsymbol{Q}(I_1, I_2, I_3) = \underbrace{\left[(\bar{I}_1 - 3)^i (\bar{I}_2 - 3)^{j-i} : j \in \{1, \dots, N\}, i \in \{0, \dots, j\}\right]^T}_{\text{Generalized Mooney–Rivlin features}}$$

$$\oplus \underbrace{\left[(J - 1)^{2k} : k \in \{1, \dots, M\}\right]^T}_{\text{Volumetric deformation features}} \oplus \underbrace{\left[\log\left(\bar{I}_2/3\right)\right]}_{\text{logarithmic feature}}, \tag{5}$$

where $\bar{I}_1 = J^{-2/3}I_1$, $\bar{I}_2 = J^{-4/3}I_2$, $J = \det(\boldsymbol{F}) = I_3^{1/2}$, and $\oplus$ denotes vector concatenation. The choice of N and M determines the total number of features n_f. Throughout this work, $N = M = 7$ is chosen, resulting in $n_f = 43$ features. The first set of features in (5) is motivated from polynomial hyperelasticity laws such as those described by generalized Mooney–Rivlin models for rubber-like materials [45–47]. The second set of features allows capturing volumetric deformation in compressible elasticity. The third term is motivated from logarithmic features encountered in models such as those proposed by Gent and Thomas [48]. We later exclude this term purposefully to test the generalization capability of our approach when the appropriate features are missing.

The components of the first Piola–Kirchhoff stress tensor are computed from (4) as

$$P_{ij} = \frac{\partial W(I_1, I_2, I_3)}{\partial F_{ij}} = \frac{\partial \boldsymbol{Q}^T(I_1, I_2, I_3)}{\partial F_{ij}}\boldsymbol{\theta}. \tag{6}$$

The derivatives of the feature library $\boldsymbol{Q}$ with respect to the deformation gradient can be calculated by applying the chain rule. Owing to the severe nonlinearities in $\boldsymbol{Q}$, manual differentiation is intractable. Therefore, we compute the derivatives via automatic differentiation (see [41] for a detailed review) that provides exact derivatives while bypassing the computational expense and accuracy issues of symbolic and numerical differentiation, respectively.

The above choice of the feature library is conducive to both physical and mathematical interpretability of the discovered models and, unlike in model-free and ANN encoding approaches, leverages decades of physical and phenomenological knowledge in modeling of hyperelastic materials, and in particular of rubber-like materials. For example, (5) naturally admits physics-based models, e.g., the Neo-Hookean [49] and Isihara [50] models which are derived via statistical homogenization of polymeric molecular chains. The model proposed by Arruda and Boyce [51], based on Langevin statistics of polymeric chains, does not have a closed form expression (due to the presence of the inverse Langevin function) and, in practice, is often approximated by a Taylor series with polynomial features which are already present in (5). The feature library also admits phenomenological models such as those proposed by Biderman [52], Haines and Wilson [53], Gent and Thomas [48], and many more. We refer the reader to the aforementioned references for physical interpretation of their respective features. The library can be easily expanded to include new features if needed, possibly based on the prior knowledge of general and/or specific features of the material behavior, e.g. when dealing with biological tissues rather than rubber-like materials. Moreover, physical constraints — e.g., material objectivity, material symmetries, stress-free reference configuration (i.e., $\boldsymbol{P}(\boldsymbol{F} = \boldsymbol{I}) = \boldsymbol{0}$), can be satisfied automatically. In the present case, objectivity and material symmetry (isotropy) are ensured a priori by (4), and a stress-free reference configuration is guaranteed by (5). Checking additional constraints [54] such as boundedness, coercivity, and convexity of the discovered strain energy density via (semi-)analytical methods is facilitated by the mathematically interpretable and parsimonious functional form. In general, leveraging such prior knowledge greatly reduces the solution space and makes the search for physically admissible models more tractable — particularly in the unsupervised setting, where the absence of stress data typically renders the search highly ill-posed.

2.4. Equilibrium constraints for unsupervised discovery

On the basis of the interpolated displacement data (Section 2.2) and the material model library (Section 2.3), the objective is now to find suitable material parameters $\boldsymbol{\theta}$, such that the physical constraints are satisfied, i.e., linear momentum balance[3] is fulfilled both in the bulk material and at the boundary. This is a crucial step, as the enforcement of these physical constraints acts as a substitute for the availability of stress data and thus allows for unsupervised discovery. Assuming negligible body forces, the weak formulation of linear momentum balance in the reference domain Ω under quasi-static conditions is given by

$$\int_{\Omega} \boldsymbol{P} : \nabla \boldsymbol{v} \, dV - \int_{\partial \Omega_t} \hat{\boldsymbol{t}} \cdot \boldsymbol{v} \, dS = 0, \quad \forall \quad \text{admissible } \boldsymbol{v}, \tag{7}$$

where $\hat{\boldsymbol{t}}$ is the surface traction acting on $\partial \Omega_t$ (zero in the present displacement-controlled setting, but non-zero for force-controlled experiments), and the test function $\boldsymbol{v}$ is admissible if it is sufficiently regular and vanishes on the Dirichlet boundary $\partial \Omega_u$. Note that we prefer the weak to the strong formulation of linear momentum balance because the double spatial derivatives required by the latter make it more sensitive to noise. This is also corroborated by the preferred use of the weak formulation in solving inverse problems based on full-field measurements, e.g., by applying the Virtual Field Method (see [55] for a review and [56] for a recent application to hyperelasticity). Such methods show methodical parallels to the approach presented in this paper, however, they rely on the assumption that the form of the constitutive law is known *a priori*.

Let $\mathcal{D} = \{(a, i) : a = 1, \ldots, n_n; i = 1, 2\}$ denote the set of all nodal degrees of freedom. $\mathcal{D}$ is further split into two subsets of free and fixed (via Dirichlet constraints) degrees of freedom: $\mathcal{D}^{\text{free}} \subseteq \mathcal{D}$ and $\mathcal{D}^{\text{fix}} = \mathcal{D} \backslash \mathcal{D}^{\text{free}}$, respectively. Using the same approximation basis as for $\boldsymbol{u}$ (Bubnov–Galerkin method), the test functions are approximated as

$$\boldsymbol{v}(\boldsymbol{X}) = \sum_{a=1}^{n_n} N^a(\boldsymbol{X}) \, \boldsymbol{v}^a, \quad \text{with} \quad v_i^a = 0 \text{ if } (a, i) \in \mathcal{D}^{\text{fix}}, \tag{8}$$

The weak form (7) then reduces to

$$\sum_{a=1}^{n_n} \boldsymbol{v}^a \cdot \left[\int_{\Omega} \boldsymbol{P} \nabla N^a \, dV - \int_{\partial \Omega_t} \hat{\boldsymbol{t}} N^a \, dS \right] = 0. \tag{9}$$

Substituting the unknown material model (6) and noting that (9) must hold for all admissible $\{\boldsymbol{v}^a : a = 1, \ldots, n_n\}$, we obtain the following componentwise force balance equations at all free degrees of freedom:

$$\int_{\Omega} \left(\frac{\partial \boldsymbol{Q}^T}{\partial F_{ij}} \boldsymbol{\theta} \right) \nabla_j N^a \, dV - \int_{\partial \Omega_t} \hat{t}_i N^a \, dS = 0, \quad \forall \quad (a, i) \in \mathcal{D}^{\text{free}}. \tag{10}$$

where we use the Einstein summation convention. The integrals are computed via numerical quadrature on the mesh. Under the assumption of material homogeneity, the material parameters $\boldsymbol{\theta}$ do not depend on the spatial coordinates and can be written outside the integrals. This results in a linear system of equations for $\boldsymbol{\theta}$ which can be vectorized and assembled into the form

$$\boldsymbol{A}^{\text{free}} \boldsymbol{\theta} = \boldsymbol{b}^{\text{free}}, \tag{11}$$

where $\boldsymbol{A}^{\text{free}} \in \mathbb{R}^{|\mathcal{D}^{\text{free}}| \times n_f}$ and $\boldsymbol{b}^{\text{free}} \in \mathbb{R}^{|\mathcal{D}^{\text{free}}|}$.

The material parameters $\boldsymbol{\theta}$ must also satisfy the force balance on the Dirichlet boundary $\partial \Omega_u$. For the degrees of freedom under Dirichlet constraints, the reaction force must equal the internal force from the material, i.e.,

$$\hat{r}_i^a = \int_{\Omega} \left(\frac{\partial \boldsymbol{Q}^T}{\partial F_{ij}} \boldsymbol{\theta} \right) \nabla_j N^a \, dV, \quad \forall \quad (a, i) \in \mathcal{D}^{\text{fix}}. \tag{12}$$

In general, the reaction force is not known at every fixed degree of freedom. Instead, only the sums of the reaction forces for subsets of the fixed degrees of freedom (each subset corresponding to one loaded side of the specimen and one coordinate direction) are known, emulating force measurements in an experiment. Let n_α be the number of these

---

[3] Note that the angular momentum balance is automatically fulfilled in the described setting.

subsets in $\partial \Omega_u$. For $\alpha = 1, \ldots, n_\alpha$, let $\mathcal{D}^{\text{fix},\alpha} \subseteq \mathcal{D}^{\text{fix}}$ be the set of degrees of freedom under Dirichlet constraints for which the sum of the reaction forces $\hat{R}^\alpha$ is known, such that $\cup_{\alpha=1}^{n_\alpha} \mathcal{D}^{\text{fix},\alpha} = \mathcal{D}^{\text{fix}}$ and $\mathcal{D}^{\text{fix},\alpha} \cap \mathcal{D}^{\text{fix},\beta} = \emptyset$ for $\beta \neq \alpha$. The force balance for each subset is given by

$$\hat{R}^\alpha = \sum_{(a,i) \in \mathcal{D}^{\text{fix},\alpha}} \hat{r}_i^a = \sum_{(a,i) \in \mathcal{D}^{\text{fix},\alpha}} \int_\Omega \left(\frac{\partial \boldsymbol{Q}^T}{\partial F_{ij}} \boldsymbol{\theta} \right) \nabla_j N^a \, dV, \quad \forall \quad \alpha = 1, \ldots, n_\alpha. \tag{13}$$

Note that each subset α only contains degrees of freedom corresponding to displacement components in one direction, so that the sum of the corresponding reaction force components is meaningful. Following the integration via numerical quadrature, this can be rearranged into additional n_α linear equations for $\boldsymbol{\theta}$ of the form

$$\boldsymbol{A}^{\text{fix}} \boldsymbol{\theta} = \boldsymbol{b}^{\text{fix}}, \tag{14}$$

where $\boldsymbol{A}^{\text{fix}} \in \mathbb{R}^{n_\alpha \times n_f}$ and $\boldsymbol{b}^{\text{fix}} \in \mathbb{R}^{n_\alpha}$.

The system of equations (11) and (14) are overdetermined and can be solved collectively in the least squares sense as

$$\boldsymbol{\theta}^{\text{opt}} = \arg \min_{\boldsymbol{\theta}} \left(\underbrace{\left\| \boldsymbol{A}^{\text{free}} \boldsymbol{\theta} - \boldsymbol{b}^{\text{free}} \right\|^2}_{\text{linear momentum balance}} + \lambda_r \underbrace{\left\| \boldsymbol{A}^{\text{fix}} \boldsymbol{\theta} - \boldsymbol{b}^{\text{fix}} \right\|^2}_{\text{reaction force balance}} \right), \tag{15}$$

with $\lambda_r > 0$ as a hyperparameter weighting the contribution of the reaction force balance term. This weighting coefficient is important to ensure that both boundary and interior force balance terms are contributing to a similar extent to the value of the objective function. In the chosen spatial discretization/mesh (discussed in Section 3), the number of boundary points is approximately two orders of magnitude smaller than the number of interior points. To ensure the similar influence of both sets of points in the optimization problem, $\lambda_r = 100$ was chosen to compensate for the fewer equations arising from the boundary points and kept constant throughout the numerical experiments. In our experience, small variations in λ_r do not have a significant influence on the solution to the minimization problem stated above. However, a violation of the balance equations at the boundary or in the interior is observed when choosing the order of magnitude of λ_r too small or too high, respectively. The above minimization problem is also equivalent to solving

$$\boldsymbol{\theta}^{\text{opt}} = \boldsymbol{A}^{\text{eqb}-1} \boldsymbol{b}^{\text{eqb}}, \quad \text{with} \quad \begin{aligned} \boldsymbol{A}^{\text{eqb}} &= \boldsymbol{A}^{\text{free}^T} \boldsymbol{A}^{\text{free}} + \lambda_r \boldsymbol{A}^{\text{fix}^T} \boldsymbol{A}^{\text{fix}}, \\ \boldsymbol{b}^{\text{eqb}} &= \boldsymbol{A}^{\text{free}^T} \boldsymbol{b}^{\text{free}} + \lambda_r \boldsymbol{A}^{\text{fix}^T} \boldsymbol{b}^{\text{fix}}. \end{aligned} \tag{16}$$

During an experiment, multiple full-field displacement datasets are usually available at different load steps and offer a richer information than from a single load step. Such information can be directly leveraged by additively combining the respective least square losses, i.e.,

$$\boldsymbol{\theta}^{\text{opt}} = \arg \min_{\boldsymbol{\theta}} \sum_{l=1}^{L} \left(\left\| \boldsymbol{A}^{\text{free},l} \boldsymbol{\theta} - \boldsymbol{b}^{\text{free},l} \right\|^2 + \lambda_r \left\| \boldsymbol{A}^{\text{fix},l} \boldsymbol{\theta} - \boldsymbol{b}^{\text{fix},l} \right\|^2 \right), \tag{17}$$

where the superscript l denotes the evaluation for load step $l = 1, \ldots, L$. For the sake of notational clarity, the subsequent derivations are only shown for a single load step but can be straightforwardly extended to the general case.

2.5. Parsimony and sparsity promotion

Despite the optimization in (15) being convex and apparently overdetermined in $\boldsymbol{\theta}$, the nature of the problem is significantly ill-posed due to the unsupervised nature of the learning problem — i.e., the lack of stress field measurements. Both the feature library $\boldsymbol{Q}$ in (5) and the coefficients $\boldsymbol{A}^{\text{eqb}}$ are highly nonlinear functions of the nodal displacements. Consequently, the condition number of $\boldsymbol{A}^{\text{eqb}}$ is large and the optimization problem (15) is highly sensitive to noise in displacement data measurements. Our numerical experiments (presented in Section 3.2) revealed that the optimization in (15) is not sufficient to identify the correct underlying material model. To address this challenge, we turn to additional sources of conceptual knowledge that can be leveraged to compensate for the lack of supervised data.

The optimization problem in (15) yields a dense solution, i.e., most entries in $\boldsymbol{\theta}$ are likely non-zero and many terms in the feature library (5) will be included in the material model. Previous works using supervised methods [45] have shown that material models with a large number of material parameters are less interpretable, poor at extrapolation to unseen strains, and more likely to be physically inadmissible. In this light, we seek material models that are parsimonious with as few features as possible to explain the displacement data.

To this end, we promote sparsity in $\boldsymbol{\theta}$ by introducing an ℓ_p regularization in (15) as

$$
\boldsymbol{\theta}^{\text{opt}} = \arg\min_{\boldsymbol{\theta}} \left(\left\| \boldsymbol{A}^{\text{free}}\boldsymbol{\theta} - \boldsymbol{b}^{\text{free}} \right\|^2 + \lambda_r \left\| \boldsymbol{A}^{\text{fix}}\boldsymbol{\theta} - \boldsymbol{b}^{\text{fix}} \right\|^2 + \lambda_p \|\boldsymbol{\theta}\|_p^p \right), \qquad \text{where} \qquad \|\boldsymbol{\theta}\|_p = \left(\sum_{i=1}^{n_f} |\theta_i|^p \right)^{1/p} .
\tag{18}
$$

The hyperparameters $\lambda_p > 0$ and $p \in [0, 1]$ control the degree of sparsity desired in $\boldsymbol{\theta}$ and thus the selection of features that must be activated in the feature library (5). The ℓ_0 regularization ($p = 0$) is equivalent to penalizing the number of activated features and thus promotes parsimony and interpretability. However, the optimization becomes a combinatorial subset selection problem which is computationally intractable for a large number of features. Traditionally, the ℓ_0 regularization is relaxed to the convex ℓ_1 regularization ($p = 1$) which promotes sparsity by zeroing out some coefficients while shrinking the rest. Also known as LASSO (least absolute shrinkage and selection operator) [57], ℓ_1 regularization offers a good compromise between computational complexity and model interpretability. Within [0, 1], a smaller value of p more aggressively promotes sparsity and interpretability at the cost of computational complexity (due to the higher degree of non-convexity in the penalty term) [58]. In our numerical experiments, we observed $p = 1/4$ to be a reasonable choice for the scope of our work.

2.6. Numerical optimization strategy

2.6.1. Fixed-point iterative optimization

Due to the ℓ_p regularization with $0 < p < 1$, the objective function in (18) is non-convex and not continuously differentiable with respect to $\boldsymbol{\theta}$, thus, conventional gradient-based techniques such as gradient-descent and Newton (or Newton-like) methods do not perform well. The development of new algorithms for optimization problems including regularization terms of this type is an active research area — see, e.g., [59] and references therein for a brief review. Tibshirani [57] proposed a fixed-point iterative strategy for the ℓ_1-regularized problem by solving a weighted ℓ_2-regularized problem in each iteration. As follows, we generalize this idea to the ℓ_p-regularized problem.

The first optimality condition of the optimization problem (18) is given by

$$
2\boldsymbol{A}^{\text{eqb}}\boldsymbol{\theta} + p\lambda_p \left[\text{sgn}(\theta_1)|\theta_1|^{p-1}, \ldots, \text{sgn}(\theta_{n_f})|\theta_{n_f}|^{p-1} \right]^T = 2\boldsymbol{b}^{\text{eqb}},
\tag{19}
$$

where sgn denotes the sign or signum function. This is a nonlinear system of equations and calls for an iterative solution strategy. Let $\boldsymbol{\theta}^{(k)}$ be the guess solution at the kth iteration. Within the $(k + 1)$th iteration, (19) is modified as

$$
2\boldsymbol{A}^{\text{eqb}}\boldsymbol{\theta}^{(k+1)} + p\lambda_p \text{diag}\left(\left|\theta_1^{(k)}\right|^{p-2}, \ldots, \left|\theta_{n_f}^{(k)}\right|^{p-2} \right) \boldsymbol{\theta}^{(k+1)} = 2\boldsymbol{b}^{\text{eqb}},
\tag{20}
$$

where we made use of $\text{sgn}\left(\theta_i^{(k+1)}\right) \left|\theta_i^{(k+1)}\right| = \theta_i^{(k+1)}$. Therefore, (19) is reduced to a linear system of equations where the nonlinear terms are replaced by linear terms with weights based on the guess solution from the previous iteration, $\boldsymbol{\theta}^{(k)}$. This further yields the fixed-point iterative scheme given by

$$
\boldsymbol{\theta}^{(k+1)} \leftarrow \left[\boldsymbol{A}^{\text{eqb}} + \frac{p\lambda_p}{2} \text{diag}\left(\left|\theta_1^{(k)}\right|^{p-2}, \ldots, \left|\theta_{n_f}^{(k)}\right|^{p-2} \right) \right]^{-1} \boldsymbol{b}^{\text{eqb}}.
\tag{21}
$$

To avoid division by zero, whenever a coefficient $\left|\theta_i^{(k+1)}\right| < \epsilon_{\text{tol}}$ for some tolerance $\epsilon_{\text{tol}} \ll 1$, the corresponding ith feature is removed from the library $\boldsymbol{Q}$ and the fixed-point iteration scheme resumes with the remaining features only. Further iterations are carried out until the sequence $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(k+1)}, \ldots\}$ converges to a fixed point such that $\left\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\right\|_\infty < \epsilon_{\text{conv}}$ for a convergence tolerance $\epsilon_{\text{conv}} \ll 1$.

The objective function landscape becomes highly non-convex with the introduction of the ℓ_p regularization. Depending on the choice of the initial guess $\boldsymbol{\theta}^{(0)}$, the outlined fixed-point iteration scheme may converge to different
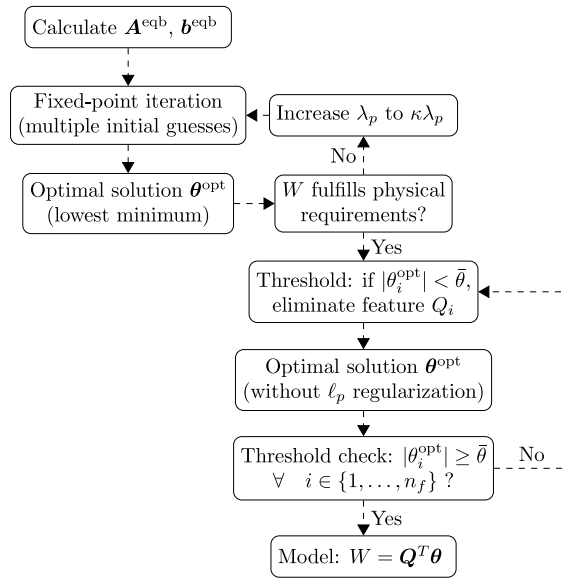
Fig. 2. Schematics of the numerical optimization strategy. Note that the value of n_f decreases during the iterative strategy as progressively more features are eliminated in the thresholding step.

local minima. For this reason, we perform multiple fixed-point iteration schemes in parallel, each with different randomly generated initial guesses. If the convergence criterion is not satisfied in a pre-specified maximum number of fixed-point iterations, the corresponding solution is discarded. From all the converged solutions, we finally choose the one which leads to the smallest value of the objective function (18).

2.6.2. Physical admissibility and the penalty parameter

As discussed in Section 2.5, the hyperparameters λ_p and p determine the degree of sparsity of the solution. Having selected $p = 1/4$ for our case (see Section 2.5), we now focus on the appropriate choice of λ_p (penalty parameter). If λ_p is too high, the resulting model will be too sparse and may not be rich enough to explain the displacement data. On the other hand, if λ_p is too low, the optimization problem (18) becomes highly ill-posed and may yield physically inadmissible models.

In order to correctly estimate λ_p, we adopt an iterative strategy with the fixed-point algorithm nested inside it. We start with a small initial value of $\lambda_p = \lambda_p^0 > 0$. If the resulting model (obtained via the fixed-point algorithm described in Section 2.6.1) is physically admissible (in the sense explained next), the model is accepted. Otherwise, λ_p is multiplicatively increased by a factor $\kappa > 1$, i.e., $\lambda_p = \kappa \lambda_p^0$. The process is repeated until a physically admissible model is obtained. By gradually increasing λ_p, we ensure that the resulting model is not too sparse or simple. Fig. 2 describes the schematics of the optimization procedure.

To verify the physical admissibility of a model, we perform a series of empirical checks. An hyperelastic strain energy density must satisfy $W(\boldsymbol{F}) \geq W(\boldsymbol{I}) = 0$ for all physically admissible $\boldsymbol{F}$. While it is computationally intractable to verify this condition for all $\boldsymbol{F}$, we can easily perform the check for the deformation gradients observed in the domain Ω. If the strain energy density is negative at any quadrature point in the finite element mesh, the material model is deemed inadmissible.

Additionally, consider the following deformation paths

$$\boldsymbol{F}^{\text{UT}}(\gamma) = \begin{bmatrix} 1 + \gamma & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{F}^{\text{UC}}(\gamma) = \begin{bmatrix} \frac{1}{1+\gamma} & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{F}^{\text{SS}}(\gamma) = \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix},$$

$$\boldsymbol{F}^{\text{BT}}(\gamma) = \begin{bmatrix} 1 + \gamma & 0 \\ 0 & 1 + \gamma \end{bmatrix}, \quad \boldsymbol{F}^{\text{BC}}(\gamma) = \begin{bmatrix} \frac{1}{1+\gamma} & 0 \\ 0 & \frac{1}{1+\gamma} \end{bmatrix}, \quad \boldsymbol{F}^{\text{PS}}(\gamma) = \begin{bmatrix} 1 + \gamma & 0 \\ 0 & \frac{1}{1+\gamma} \end{bmatrix}, \tag{22}$$

which describe uniaxial tension (UT), biaxial tension (BT), uniaxial compression (UC), biaxial compression (BC), simple shear (SS) and pure shear (PS). Along these deformation paths, the strain energy density must increase monotonically, i.e., the following condition must hold true for all paths

$$0 < W(\boldsymbol{F}(\gamma_i)) < W(\boldsymbol{F}(\gamma_j)) \quad \forall \quad \gamma_i < \gamma_j, \quad \gamma_i, \gamma_j \in (0, \infty). \tag{23}$$

For computational tractability, the above inequality is tested for a large number (n_γ) of samples of γ_i and γ_j from $(0, \gamma_{max})$ for some $\gamma_{max} \gg 1$. If the inequality is violated along any of the deformation paths in (22), the material model is rejected as unphysical.

While these checks are only empirical in nature, in our experience they proved sufficient to constrain the solution space and obtain satisfactory results, see Section 3.2. They also demonstrate that domain knowledge can be a very useful complement to a "blind" algorithm and help in determining the parameters of an otherwise difficult estimation. It is envisioned that a similar approach may involve the check of more rigorous mathematical properties of the strain energy density (e.g. polyconvexity or coercivity), which was not pursued in this work. Regardless of these checks, the mathematical interpretability and parsimony of the discovered models enable *a posteriori* verification of e.g., polyconvexity with (semi-)analytical techniques, which is not feasible with black-box approaches including neural networks and model-free methods.

2.6.3. Thresholding

As a result of the ℓ_p regularization, the optimal solution obtained from the fixed-point iterations and physical admissibility tests contains a large number of material parameters that are close or equal to zero. These "inactive features" can thus be permanently discarded from the initial library. This is performed through a simple thresholding operation analogous to the one used for supervised sparse regression by Brunton et al. [29], i.e., all parameters whose absolute value is below a threshold $\bar{\theta}$ are set to zero. Subsequently, we can deactivate the ℓ_p regularization and solve the unregularized optimization problem (15) for the remaining parameters. Note that the unregularized problem does not require an iterative technique and is solved directly using (16) . If any of the resulting non-zero parameters is again below the defined threshold $\bar{\theta}$, the thresholding procedure is repeated until convergence is reached (see Fig. 2 for schematics).

3. Numerical benchmarks

3.1. Data generation

For the scope of this work, we emulate DIC data with artificial data generated via the finite element method (FEM). We consider a hyperelastic square plate with a hole (as shown in Fig. 3) under displacement-controlled asymmetric biaxial tension with the loading parameter δ. Due to the two-fold symmetry, only a quadrant of the plate is considered with symmetry boundary conditions on the left and bottom boundaries. All lengths and displacements are normalized with respect to the side length of the quadrant. The plate is assumed to be in plane strain conditions.[4] The domain is meshed with linear triangular elements. For the prescribed boundary conditions (controlled via δ), the nodal displacements are recorded from the FEM solution at a total of L load steps. The total horizontal reaction force on the left and right boundaries, and the total vertical reaction force on the top and bottom boundaries are also recorded. We specifically choose this combination of geometry and loading, in-contrast to traditional bi-axial tension or torsion tests, because it leads to a strain field that is rich enough to solve the ill-posed problem of identifying the material model with no stress data and just one experiment.

The FEM simulations are performed with the following material models:

- NH2: Neo-Hookean solid with quadratic volumetric strain energy
- NH4: Neo-Hookean solid with biquadratic volumetric strain energy
- IH: Isihara solid [50] with quadratic deviatoric strain energy
- HW: Haines–Wilson solid [53] with cubic deviatoric strain energy
- GT: Gent–Thomas solid [48] with logarithmic deviatoric strain energy

---

[4] Note that, in plane strain, the right Cauchy–Green strain invariants are linearly dependent as $I_2 = (I_1 + I_3 - 1)$, thus the feature library can be written as a function of only two invariants, i.e., $\boldsymbol{Q}(I_1, I_3)$.
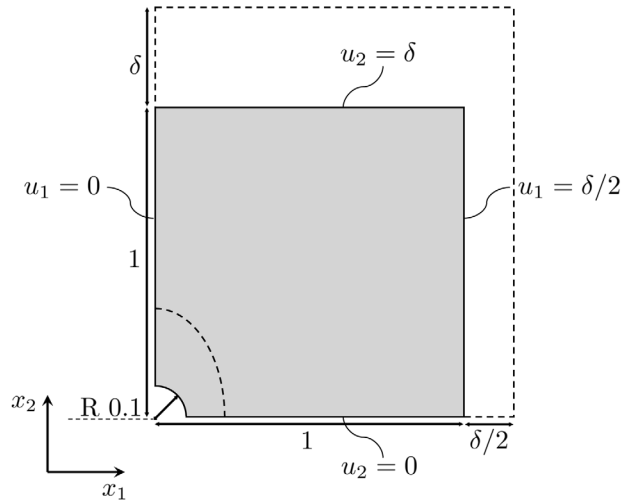
Fig. 3. Geometry and boundary conditions of a quadrant of a plate with a hole under displacement-controlled asymmetric biaxial tension.

These models serve as benchmarks, as they are to be recovered through the exclusive use of nodal displacements and reaction force data using our approach. The exact strain energy densities are listed in Table 1.

In contrast to the synthetically generated data, real DIC data are unavoidably affected by noise in the measured displacement field. To emulate real data, we add artificial noise to the synthetic displacement data coming from the FEM simulations. As the noise in DIC measurements depends on the pixel accuracy of the imaging device, it would be unrealistic to assume the noise level to be proportional to the displacement magnitude. Instead, it is more realistic to assume the same absolute noise level (also referred to as noise floor) at every degree of freedom and for every load step, independently of the corresponding magnitude of displacement. We therefore add artificial noise to the displacement data such that

$$u_i^{a,l} = u_i^{\text{fem},a,l} + u_i^{\text{noise},a,l}, \quad u_i^{\text{noise},a,l} \sim \mathcal{N}(0, \sigma) \quad \forall \quad (a, i) \in \mathcal{D}, \ l \in \{1, \dots, L\}, \tag{24}$$

where $u_i^{\text{fem},a,l}$ is the ith component of the displacement at node a at load step l obtained from the FEM, and $u_i^{\text{noise},a,l}$ denotes the noise added to it — taken as a Gaussian noise with zero mean and standard deviation $\sigma > 0$. A realistic estimate for σ depends on the specimen size and the pixel accuracy of the measurement device. The noise floor typically ranges within 0.1–1% of the pixel size (see, e.g., [60]) resulting in a reported strain measurement error of less than 10^{-4} relative to the domain length scale [61]. We hence consider two noise levels $\sigma \in \{10^{-4}, 10^{-3}\}$ (normalized with the plate quadrant length), which are (assuming sufficiently large specimen size and deformation) representative of respectively low and high noise levels encountered in modern DIC experiments.

Displacement measurements from DIC are often denoised by assuming spatial smoothness in the data. In previous works on data driven discovery of governing laws, denoising has been performed either as a preprocessing step or simultaneously with the optimization problem [42–44]. Here, we preprocess the noisy displacements by denoising using kernel ridge regression (KRR). The noisy displacements in $\mathcal{U}$ are spatially interpolated by a ridge estimator kernelized by an RBF. The interpolation reduces the noise and the interpolated displacements are used as inputs to our algorithm. In Appendix A, we briefly review KRR in the context of our work; for more details, the reader is referred to Saunders et al. [62]. We specifically chose KRR as the denoiser because other methods are either computationally expensive (e.g., Gaussian process regression), only limited to rectangular domains with structured discretizations (e.g., Gaussian filters), prone to overfitting (e.g., neural networks), or prone to underfitting (e.g., polynomial interpolation).

3.2. Results

All subsequent results are based on the parameters (including the algorithm hyperparameters) presented in Appendix B, which were identified upon a careful testing phase. The strain energy densities discovered by the

Table 1

Strain energy density of the (true) hidden and discovered material models for different noise levels σ.

Benchmarks		Strain energy density (W)
NH2	Truth	$0.5000(\bar{I}_1 - 3) + 1.5000(J - 1)^2$
	$\sigma = 0$	$0.5000(\bar{I}_1 - 3) + 1.5000(J - 1)^2$
	$\sigma = 10^{-4}$	$0.4995(\bar{I}_1 - 3) + 1.4998(J - 1)^2$
	$\sigma = 10^{-3}$	$0.4936(\bar{I}_1 - 3) + 1.4986(J - 1)^2$
NH4	Truth	$0.5000(\bar{I}_1 - 3) + 1.5000(J - 1)^4$
	$\sigma = 0$	$0.5000(\bar{I}_1 - 3) + 1.5000(J - 1)^4$
	$\sigma = 10^{-4}$	$0.5005(\bar{I}_1 - 3) + 1.4973(J - 1)^4$
	$\sigma = 10^{-3}$	$0.4848(\bar{I}_1 - 3) + 1.4728(J - 1)^4$
IH	Truth	$0.5000(\bar{I}_1 - 3) + 1.0000(\bar{I}_2 - 3) + 1.0000(\bar{I}_1 - 3)^2 + 1.5000(J - 1)^2$
	$\sigma = 0$	$0.5000(\bar{I}_1 - 3) + 1.0000(\bar{I}_2 - 3) + 1.0000(\bar{I}_1 - 3)^2 + 1.5000(J - 1)^2$
	$\sigma = 10^{-4}$	$0.5306(\bar{I}_1 - 3) + 0.9576(\bar{I}_2 - 3) + 0.9917(\bar{I}_1 - 3)^2 + 1.5041(J - 1)^2$
	$\sigma = 10^{-3}$	$1.6323(\bar{I}_2 - 3) + 1.5546(\bar{I}_1 - 3)(\bar{I}_2 - 3) + 0.0304(\bar{I}_1 - 3)^2(\bar{I}_2 - 3)^3 + 1.4498(J - 1)^2$
HW	Truth	$0.5000(\bar{I}_1 - 3) + 1.0000(\bar{I}_2 - 3) + 0.7000(\bar{I}_1 - 3)(\bar{I}_2 - 3) + 0.2000(\bar{I}_1 - 3)^3 + 1.5000(J - 1)^2$
	$\sigma = 0$	$0.5000(\bar{I}_1 - 3) + 1.0000(\bar{I}_2 - 3) + 0.7000(\bar{I}_1 - 3)(\bar{I}_2 - 3) + 0.2000(\bar{I}_1 - 3)^3 + 1.5000(J - 1)^2$
	$\sigma = 10^{-4}$	$0.5853(\bar{I}_1 - 3) + 0.9101(\bar{I}_2 - 3) + 0.6475(\bar{I}_1 - 3)(\bar{I}_2 - 3) + 0.2089(\bar{I}_1 - 3)^3 + 1.5050(J - 1)^2$
	$\sigma = 10^{-3}$	$1.3600(\bar{I}_1 - 3) + 1.4284(\bar{I}_2 - 3)^3 + 1.5469(J - 1)^2$
GT	Truth	$0.5000(\bar{I}_1 - 3) + 1.5000(J - 1)^2 + 1.0000 \log(\bar{I}_2/3)$
	$\sigma = 0$	$0.5000(\bar{I}_1 - 3) + 1.5000(J - 1)^2 + 0.9999 \log(\bar{I}_2/3)$
	$\sigma = 10^{-4}$	$0.4978(\bar{I}_1 - 3) + 1.5002(J - 1)^2 + 1.0095 \log(\bar{I}_2/3)$
	$\sigma = 10^{-3}$	$0.3578(\bar{I}_1 - 3) + 0.4695(\bar{I}_2 - 3) + 1.5152(J - 1)^2$
GT*	Truth	$0.5000(\bar{I}_1 - 3) + 1.5000(J - 1)^2 + 1.0000 \log(\bar{I}_2/3)$
	$\sigma = 0$	$0.4105(\bar{I}_1 - 3) + 0.3783(\bar{I}_2 - 3) + 1.5040(J - 1)^2$
	$\sigma = 10^{-4}$	$0.3974(\bar{I}_1 - 3) + 0.4011(\bar{I}_2 - 3) + 1.5045(J - 1)^2$
	$\sigma = 10^{-3}$	$0.3578(\bar{I}_1 - 3) + 0.4695(\bar{I}_2 - 3) + 1.5152(J - 1)^2$

GT* denotes the case where the logarithmic feature is excluded from the feature library $\boldsymbol{Q}$.

proposed unsupervised algorithm from the data generated with the five chosen material models are reported in Table 1. A comparison of the strain energy densities of the true models and the discovered models along three of the deformation paths in (22) – in particular, uniaxial tension (UT), simple shear (SS) and pure shear (PS) – is presented in Fig. 4. Fig. 5 further shows such a comparison for selected stress components, which can be deduced from the strain energy density by differentiating (6). As follows, these results are discussed for the three different noise levels (σ) artificially superposed to the displacement measurements.

In the case of displacements without noise ($\sigma = 0$), all the material models are exactly discovered with negligible errors in the magnitude of the material parameters. In the low noise case ($\sigma = 10^{-4}$), the material models are again correctly discovered (i.e., all the correct features are identified). There are small differences in the material parameters, which however do not affect the accuracy of the strain energy density. This conclusion is corroborated by the negligible difference in the plots of the strain energy densities and stress components along the chosen deformation paths in Figs. 4 and 5.

In the high noise case ($\sigma = 10^{-3}$), the discovered models show good agreement with the hidden models for the two Neo-Hookean solids (NH2 and NH4). However, the unsupervised algorithm fails to correctly identify the model features for the Isihara (IH), Haines–Wilson (HW) and Gent–Thomas (GT) solids. Despite the difference in the functional form, the discovered model for the GT solid shows good agreement with the true model in Figs. 4 and 5. However, the discovered models for the IH and HW solids deviate from the true models — particularly at large pure shear deformations. Hence, in these cases, they only serve as good approximations in the small deformation regime. This also reveals the limitations in the extrapolatory capability of our approach in the case of high noise levels. Specifically, data with higher noise levels require larger displacements and loadings to improve the generalizability of the discovered models to larger deformations. This is also expected because the higher-order and nonlinear features are more dominant at larger deformations only. Moreover, the chosen experiment does not feature very large levels of pure shear deformations, which explains why the discovered model for higher noise deviates from
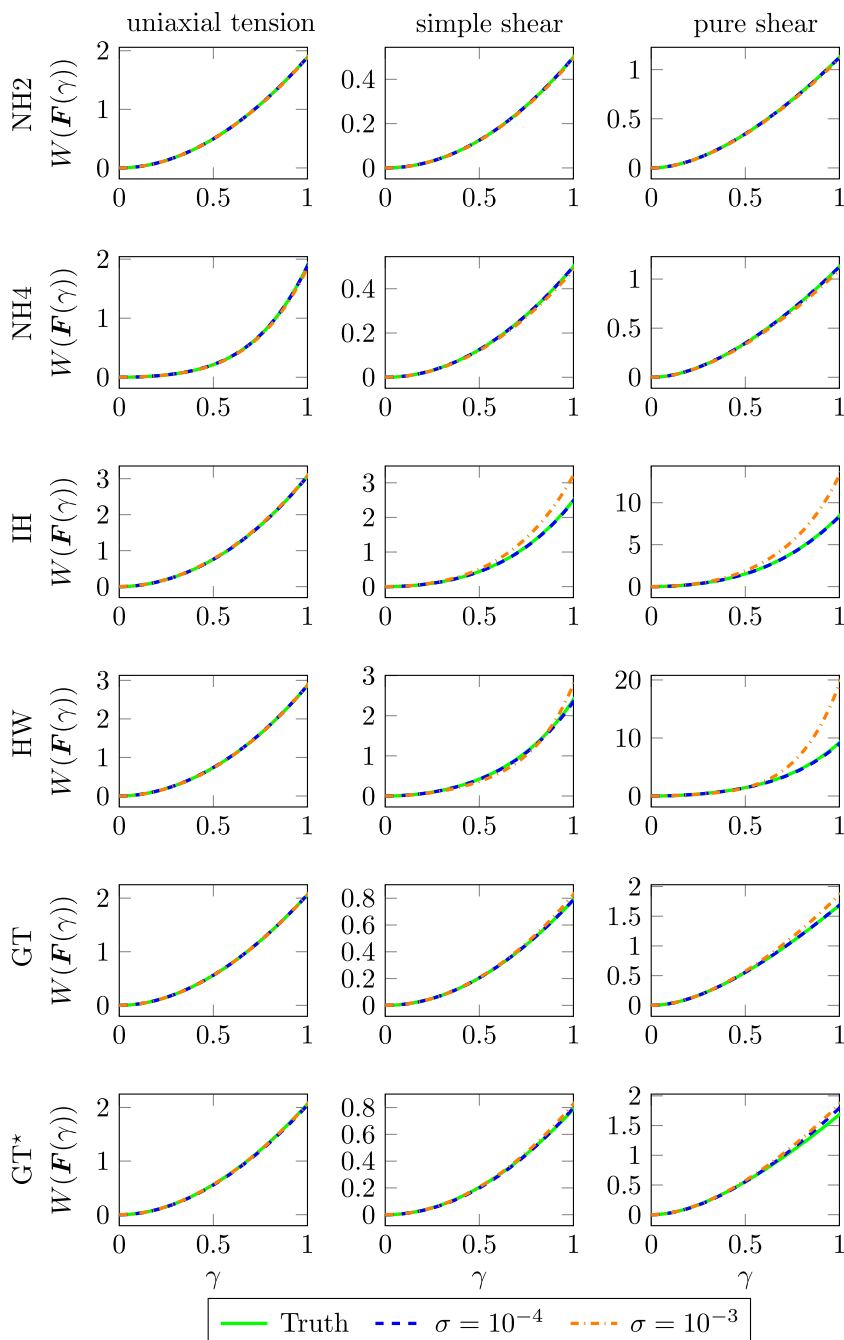
Fig. 4. Strain energy density $W(\boldsymbol{F}(\gamma))$ along different deformation paths (see (22)) of the (true) hidden and discovered material models for different noise levels σ. The discovered models for the $\sigma = 0$ case are omitted as they are exactly equal to the hidden models. GT* denotes the case when the logarithmic feature is excluded from the feature library $\boldsymbol{Q}$.

the true model especially along this deformation path. Using experiments with more complex strain fields,[5] or using more than one experiment, is expected to be beneficial in cases where high noise levels are to be expected.

---

[5] See [55] and references therein for a variety of possible experimental setups.
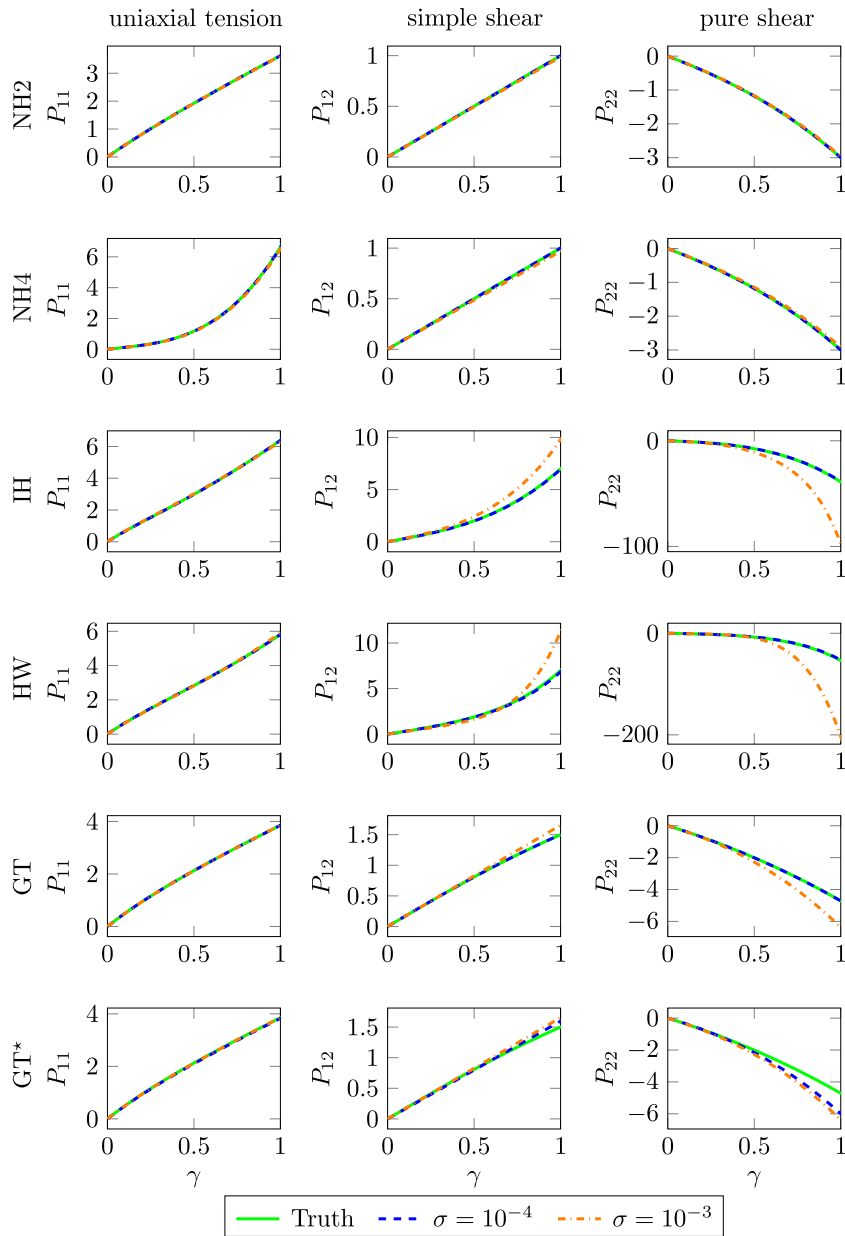
Fig. 5. Stress components $P_{11}(\boldsymbol{F}^{\mathrm{UT}}(\gamma))$ along uniaxial tension deformation path, $P_{12}(\boldsymbol{F}^{\mathrm{SS}}(\gamma))$ along simple shear deformation path and $P_{22}(\boldsymbol{F}^{\mathrm{PS}}(\gamma))$ along pure shear deformation path (see (22)) of the (true) hidden and discovered material models for different noise levels σ. The discovered models for the $\sigma = 0$ case are omitted as they are exactly equal to the hidden models. GT* denotes the case when the logarithmic feature is excluded from the feature library $\boldsymbol{Q}$.

The question arises whether the unsupervised algorithm generalizes well when the feature library cannot represent the hidden material model due to the lack of the corresponding feature(s). To this end, we purposefully exclude the logarithmic feature $\log(\bar{I}_2/3)$ from the feature library $\boldsymbol{Q}$ ($n_f = 42$ in this case). Interestingly, in this situation the discovered models for the GT solid, denoted by GT* (Table 1 and Figs. 4 and 5), surrogate the logarithmic feature by its linearization, i.e., by the feature ($\bar{I}_2 - 3$). This is also observed for the high noise level ($\sigma = 10^{-3}$), for which it also occurs when the logarithmic feature is included in the feature library. The strain energy density and stress response is nevertheless reproduced quite accurately (Figs. 4 and 5). These results suggest that the proposed

approach is able not only to identify but also to surrogate model features, depending on the availability of such features in the catalogue of candidate functions, in such a way that the predicted response of the model is accurately reproduced.

4. Conclusions and outlook

We developed a new unsupervised sparse regression approach for automated discovery of constitutive models for hyperelastic material behavior. The approach requires only measurable (displacement and global force) data, fulfills many important physics constraints a priori, and delivers interpretable models from a potentially very large library of candidate functions by enforcing the satisfaction of linear momentum balance in the interior and at the boundary of the domain. In order to discover parsimonious models with as little non-zero terms as possible, we adopt a sparsity promoting technique based on the classical ℓ_p regularization and, unlike in the previous literature, we determine the corresponding penalty parameter automatically through the enforcement of additional physics constraints. The devised algorithm combines this regularization with thresholding in a fully automatic fashion. In our numerical experiments, where the datasets were obtained by finite element simulations with or without added artificial noise, the approach was shown to be able to correctly identify five chosen hyperelastic models out of a large catalogue of functions based on only one loading experiment (i.e., in machine learning jargon, in one shot), with an accuracy depending on the magnitude of the noise and on the complexity of the model. For no or low noise, the models were identified perfectly or to high accuracy. For high noise, very accurate results were still obtained for the simplest models whereas the most complex ones were surrogated by others with reduced accuracy. In such cases, it is envisioned that loading experiments with more complex strain states and larger deformation levels as well as the availability of datasets from different experiments will be of help. Finally, we also showed that, if a "true" feature is missing in the library of candidate functions, the proposed approach is able to surrogate it in such a way that the actual response is still accurately predicted.

We believe that the proposed approach has the potential for many future developments, including the employment on experimental data[6] in the two- and three-dimensional settings and the extension to the more complex cases of non-homogeneous, anisotropic, and irreversible material behavior.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Code and data availability

The codes and data generated during the current study are available from the authors upon reasonable request.

Appendix A. Denoising via kernel ridge regression (KRR)

Here, we provide a brief review of KRR in the context of this work. For more details, the reader is referred to Saunders et al. [62].

Let $\boldsymbol{y} = \left(y^1, y^2, \dots, y^{n_n} \right)^T \in \mathbb{R}^{n_n}$ denote a vector of noisy scalar measurements, where y^a is the measurement at the node of reference coordinate $\boldsymbol{X}^a \in \mathcal{X}$. We assume an ansatz for the denoised field $\hat{y} : \Omega \to \mathbb{R}$ as

$$\hat{y}(\boldsymbol{X}) = \boldsymbol{\varphi}(\boldsymbol{X})^T \boldsymbol{\beta}, \tag{A.1}$$

where $\boldsymbol{\varphi} : \Omega \to \mathbb{R}^d$ denotes a highly nonlinear map to the d-dimensional feature space and $\boldsymbol{\beta} \in \mathbb{R}^d$ performs a linear combination of the rows of $\boldsymbol{\varphi}$. Here, d is assumed to be much higher than the number of measurements, i.e., $d \gg n_n$, and can even be infinite.

The objective is to determine the coefficients $\boldsymbol{\beta}$ such that the differences between noisy and denoised measurements are minimized subject to ℓ_2 regularization (also known as ridge regularization), i.e.,

$$\boldsymbol{\beta}^{\text{opt}} = \arg \min_{\boldsymbol{\beta}} \| \boldsymbol{\Phi} \boldsymbol{\beta} - \boldsymbol{y} \|^2 + \xi \| \boldsymbol{\beta} \|^2, \qquad \text{with} \qquad \boldsymbol{\Phi} = \left[\boldsymbol{\varphi}(\boldsymbol{X}^1), \boldsymbol{\varphi}(\boldsymbol{X}^2), \dots, \boldsymbol{\varphi}(\boldsymbol{X}^{n_n}) \right]^T \in \mathbb{R}^{n_n \times d}. \tag{A.2}$$

---

[6] Depending on the experimental setting, the algorithmic framework may need to be reformulated considering the plane stress assumption.

Here, $\xi > 0$ controls the strength of the ℓ_2 regularization. The first optimality condition of (A.2) yields

$$\boldsymbol{\beta}^{\mathrm{opt}} = \underbrace{\left(\xi \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)^{-1}}_{d \times d} \boldsymbol{\Phi}^T \boldsymbol{y}. \tag{A.3}$$

This involves inverting a $d \times d$ matrix, which is computationally intractable if d is infinite.

To make (A.3) tractable, we perform the following manipulation,

$$\boldsymbol{\beta}^{\mathrm{opt}} = \left(\xi \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^T \underbrace{\left(\xi \boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right)\left(\xi \boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right)^{-1}}_{=\boldsymbol{I}} \boldsymbol{y} = \left(\xi \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)^{-1}$$
$$\times \left(\xi \boldsymbol{\Phi}^T + \boldsymbol{\Phi}^T \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right)\left(\xi \boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right)^{-1} \boldsymbol{y}. \tag{A.4}$$

This is rearranged into

$$\boldsymbol{\beta}^{\mathrm{opt}} = \left(\xi \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)^{-1} \left(\xi \boldsymbol{I} + \boldsymbol{\Phi}^T \boldsymbol{\Phi}\right) \boldsymbol{\Phi}^T \left(\xi \boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right)^{-1} \boldsymbol{y}, \tag{A.5}$$

which further simplifies to

$$\boldsymbol{\beta}^{\mathrm{opt}} = \boldsymbol{\Phi}^T \underbrace{\left(\xi \boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right)^{-1}}_{n_n \times n_n} \boldsymbol{y}. \tag{A.6}$$

Notably, (A.3) reduces to inverting a finite-dimensional $n_n \times n_n$ matrix which is computationally tractable in contrast to a $d \times d$ matrix (recall, $d \gg n_n$).

Substituting (A.6) into the ansatz (A.1), the denoised field is obtained as

$$\hat{y}(\boldsymbol{X}) = \boldsymbol{\varphi}(\boldsymbol{X})^T \boldsymbol{\beta}^{\mathrm{opt}} = \boldsymbol{\varphi}(\boldsymbol{X})^T \boldsymbol{\Phi}^T \left(\xi \boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right)^{-1} \boldsymbol{y}. \tag{A.7}$$

We then use the *kernel trick* [63]. For $\boldsymbol{X}, \tilde{\boldsymbol{X}} \in \Omega$, let

$$\mathcal{K}(\boldsymbol{X}, \tilde{\boldsymbol{X}}) = \boldsymbol{\varphi}(\boldsymbol{X})^T \boldsymbol{\varphi}(\tilde{\boldsymbol{X}}), \tag{A.8}$$

be a positive definite kernel defining the inner product in the feature space φ. The denoised field is then reduced to

$$\hat{y}(\boldsymbol{X}) = \boldsymbol{k}(\boldsymbol{X})^T \left(\xi \boldsymbol{I} + \boldsymbol{K}\right)^{-1} \boldsymbol{y},$$
$$\text{with} \quad k_a(\boldsymbol{X}) = \mathcal{K}(\boldsymbol{X}, \boldsymbol{X}^a) \quad \text{and} \quad K_{ab} = \mathcal{K}(\boldsymbol{X}^a, \boldsymbol{X}^b), \quad \forall \quad a, b \in \{1, \dots, n_n\}. \tag{A.9}$$

The trick is that we only need to define a valid kernel $\mathcal{K}$ (must satisfy the requirements of an inner product) to solve for the denoised field, without actually ever defining the (possibly infinitely-dimensional) feature space φ. To this end, we choose the RBF kernel defined as

$$\mathcal{K}(\boldsymbol{X}, \tilde{\boldsymbol{X}}) = \exp\left(\frac{-\|\boldsymbol{X} - \tilde{\boldsymbol{X}}\|^2}{2\chi^2}\right), \tag{A.10}$$

where $\chi > 0$ controls the length scale of the kernel. Since KRR does not assume a parametric ansatz (φ remains undefined in (A.1)), it is a *non-parametric* regression.

In the context of this work, the displacement components are denoised separately by setting $\boldsymbol{y} = (u_1^1, u_1^2, \dots, u_1^{n_n})^T$ and $\boldsymbol{y} = (u_2^1, u_2^2, \dots, u_2^{n_n})^T$ in (A.9). The only hyperparameters are the regularization strength (ξ) and kernel length scale (χ). These hyperparameters are optimized by minimizing the error $\sum_{a=1}^{n_n} \|\hat{y}(\boldsymbol{X}^a) - y^a\|^2$ via random search.

We note that KRR is equivalent to estimating the posterior mean in Gaussian process regression [64,65]. However, the latter also has additional complexity as it provides uncertainties and posterior sampling for the predictions $\hat{\boldsymbol{y}}$ (unlike the deterministic framework of KRR). Consequently, despite the asymptotic computational complexity of both regressions being $O(n_n^3)$, we observed that KRR is significantly faster for moderately-sized problem ($n_n \sim 10^4 - 10^5$) in the context of this work.

Appendix B. Numerical protocol and computational costs

Table A.2 lists the default set of parameters and hyperparameters used in the data generation and the proposed algorithm. For the NH2 and NH4 models, displacement data from $L = 4$ load steps are considered. For the rest of

Table A.2

Default parameters and hyperparameters for the data generation and the proposed algorithm.

Parameter	Notation	Value
Number of nodes in mesh	n_n	63,601
Number of reaction force constraints	n_α	4
Number of load steps	L	NH2, NH4: 4
		IH, HW, GT: 8
Loading parameter	δ	$\{0.1 \times l : l = 1, \ldots, L\}$
Number of features	n_f	$\{42, 43\}$
Coefficient for reaction force balance	λ_r	100
ℓ_p regularization	p	$1/4$
Initial coefficient for ℓ_p regularization	λ_p^0	0.01
Multiplicative factor for increasing λ_p	κ	5
Number of parallel fixed-point iterations	–	200
Maximum number of fixed-point iteration steps	–	200
Number of samples for admissibility checks	n_γ	75
Maximum loading parameter for admissibility checks	$\gamma_{\max}$	10^9
Zero tolerance for θ during fixed-point iteration	ϵ_{tol}	10^{-6}
Convergence tolerance for θ during fixed-point iteration	ϵ_{conv}	10^{-3}
Threshold for θ after fixed-point iteration	$\bar{\theta}$	0.01
Displacement noise standard deviation	σ	$\{0, 10^{-4}, 10^{-3}\}$

the models, i.e., IH, HW, and GT, with higher order terms in the strain energy density, the number of considered load steps is increased to $L = 8$ to allow for similar maximum strain levels as in the simpler models.

For the chosen number of nodes (n_n) in the spatial discretization, the computational time needed to execute the algorithm depicted in Fig. 2 with one initial guess for the fixed-point iterations is of the order of ten minutes for an average modern processor. We highlight that the processes of fixed-point iterations with different initial guesses can be executed in an embarrassingly parallel fashion on a distributed cluster.

References

[1] T. Kirchdoerfer, M. Ortiz, Data-driven computational mechanics, Comput. Methods Appl. Mech. Engrg. 304 (2016) 81–101, http://dx.doi.org/10.1016/j.cma.2016.02.001, URL: https://linkinghub.elsevier.com/retrieve/pii/S0045782516300238.

[2] R. Ibañez, D. Borzacchiello, J.V. Aguado, E. Abisset-Chavanne, E. Cueto, P. Ladeveze, F. Chinesta, Data-driven non-linear elasticity: constitutive manifold construction and problem discretization, Comput. Mech. 60 (2017) 813–826, http://dx.doi.org/10.1007/s00466-017-1440-1.

[3] L.T.K. Nguyen, M.A. Keip, A data-driven approach to nonlinear elasticity, Comput. Struct. 194 (2018) 97–115, http://dx.doi.org/10.1016/j.compstruc.2017.07.031, URL: https://linkinghub.elsevier.com/retrieve/pii/S0045794917301311.

[4] S. Conti, S. Müller, M. Ortiz, Data-driven problems in elasticity, Arch. Ration. Mech. Anal. 229 (2018) 79–123, http://dx.doi.org/10.1007/s00205-017-1214-0.

[5] T. Kirchdoerfer, M. Ortiz, Data-driven computing in dynamics: Data-driven computing in dynamics, Internat. J. Numer. Methods Engrg. 113 (2018) 1697–1710, http://dx.doi.org/10.1002/nme.5716.

[6] R. Eggersmann, T. Kirchdoerfer, S. Reese, L. Stainier, M. Ortiz, Model-free data-driven inelasticity, Comput. Methods Appl. Mech. Engrg. 350 (2019) 81–99, http://dx.doi.org/10.1016/j.cma.2019.02.016, URL: http://arxiv.org/abs/1808.10859, arXiv: 1808.10859.

[7] P. Carrara, L. De Lorenzis, L. Stainier, M. Ortiz, Data-driven fracture mechanics, 2020, arXiv:2006.03133 [cs] URL: http://arxiv.org/abs/2006.03133. arXiv: 2006.03133.

[8] R. Ibañez, E. Abisset-Chavanne, J.V. Aguado, D. Gonzalez, E. Cueto, F. Chinesta, A manifold learning approach to data-driven computational elasticity and inelasticity, Arch. Comput. Methods Eng. 25 (2018) 47–57, http://dx.doi.org/10.1007/s11831-016-9197-9.

[9] D. González, F. Chinesta, E. Cueto, Thermodynamically consistent data-driven computational mechanics, Contin. Mech. Thermodyn. 31 (2019) 239–253, http://dx.doi.org/10.1007/s00161-018-0677-z.

[10] R. Ibáñez, E. Abisset-Chavanne, D. González, J.L. Duval, E. Cueto, F. Chinesta, Hybrid constitutive modeling: data-driven learning of corrections to plasticity models, Int. J. Mater. Form. 12 (2019) 717–725, http://dx.doi.org/10.1007/s12289-018-1448-x.

[11] D. González, F. Chinesta, E. Cueto, Learning corrections for hyperelastic models from data, Front. Mater. 6 (2019) 14, http://dx.doi.org/10.3389/fmats.2019.00014, URL: https://www.frontiersin.org/article/10.3389/fmats.2019.00014/full.

[12] T. Sussman, K.J. Bathe, A model of incompressible isotropic hyperelastic material behavior using spline interpolations of tension-compression test data, Commun. Numer. Methods. Eng. 25 (2009) 53–63, http://dx.doi.org/10.1002/cnm.1105.

[13] J. Crespo, M. Latorre, F.J. Montáns, WYPIWYG hyperelasticity for isotropic, compressible materials, Comput. Mech. 59 (2017) 73–92, http://dx.doi.org/10.1007/s00466-016-1335-6.

[14] J. Ghaboussi, J.H. Garrett, X. Wu, Knowledge-based modeling of material behavior with neural networks, J. Eng. Mech. 117 (1991) 132–153, http://dx.doi.org/10.1061/(ASCE)0733-9399(1991)117:1(132), URL: http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9399%281991%29117%3A1%28132%29.

[15] M. Lefik, B. Schrefler, Artificial neural network as an incremental non-linear constitutive model for a finite element code, Comput. Methods Appl. Mech. Engrg. 192 (2003) 3265–3283, http://dx.doi.org/10.1016/S0045-7825(03)00350-5, URL: https://linkinghub.elsevier.com/retrieve/pii/S0045782503003505.

[16] Y. Shen, K. Chandrashekhara, W. Breig, L. Oliver, Finite element analysis of V-ribbed belts using neural network based hyperelastic material model, Int. J. Non-Linear Mech. 40 (2005) 875–890, http://dx.doi.org/10.1016/j.ijnonlinmec.2004.10.005, URL: https://linkinghub.elsevier.com/retrieve/pii/S0020746204001635.

[17] M. Mozaffar, R. Bostanabad, W. Chen, K. Ehmann, J. Cao, M.A. Bessa, Deep learning predicts path-dependent plasticity, Proc. Natl. Acad. Sci. 116 (2019) 26414–26420, http://dx.doi.org/10.1073/pnas.1911815116, URL: https://www.pnas.org/content/116/52/26414, arXiv:https://www.pnas.org/content/116/52/26414.full.pdf.

[18] S. Kumar, S. Tan, L. Zheng, D.M. Kochmann, Inverse-designed spinodoid metamaterials, npj Comput. Mater. 6 (2020) 73, http://dx.doi.org/10.1038/s41524-020-0341-6.

[19] D.Z. Huang, K. Xu, C. Farhat, E. Darve, Learning constitutive relations from indirect observations using deep neural networks, J. Comput. Phys. 416 (2020) 109491, http://dx.doi.org/10.1016/j.jcp.2020.109491, URL: https://linkinghub.elsevier.com/retrieve/pii/S0021999120302655.

[20] Z. Hashin, Analysis of composite materials—A survey, J. Appl. Mech. 50 (1983) 481–505, http://dx.doi.org/10.1115/1.3167081.

[21] Z. Yuan, J. Fish, Toward realization of computational homogenization in practice, Internat. J. Numer. Methods Engrg. 73 (2008) 361–380, http://dx.doi.org/10.1002/nme.2074.

[22] K. Wang, W. Sun, A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning, Comput. Methods Appl. Mech. Engrg. 334 (2018) 337–380, http://dx.doi.org/10.1016/j.cma.2018.01.036, URL: https://linkinghub.elsevier.com/retrieve/pii/S0045782518300380.

[23] A. Leygue, M. Coret, J. Réthoré, L. Stainier, E. Verron, Data-based derivation of material response, Comput. Methods Appl. Mech. Engrg. 331 (2018) 184–196, http://dx.doi.org/10.1016/j.cma.2017.11.013, URL: https://linkinghub.elsevier.com/retrieve/pii/S0045782517307156.

[24] M. Dalémat, Measuring stress field without constitutive equation, Mech. Mater. 9 (2019).

[25] A.M. Tartakovsky, C.O. Marrero, P. Perdikaris, G.D. Tartakovsky, D. Barajas-Solano, Learning parameters and constitutive relationships with physics informed deep neural networks, 2018, arXiv:1808.03398 [physics] URL: http://arxiv.org/abs/1808.03398, arXiv: 1808.03398.

[26] E. Haghighat, M. Raissi, A. Moure, H. Gomez, R. Juanes, A deep learning framework for solution and discovery in solid mechanics, 2020, arXiv:2003.02751 [cs, stat] URL: http://arxiv.org/abs/2003.02751. arXiv: 2003.02751.

[27] M. Raissi, P. Perdikaris, G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707, http://dx.doi.org/10.1016/j.jcp.2018.10.045, URL: https://linkinghub.elsevier.com/retrieve/pii/S0021999118307125.

[28] M. Schmidt, H. Lipson, Distilling free-form natural laws from experimental data, Science 324 (2009) 81–85, http://dx.doi.org/10.1126/science.1165893.

[29] S.L. Brunton, J.L. Proctor, J.N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. 113 (2016) 3932–3937, http://dx.doi.org/10.1073/pnas.1517384113.

[30] J.C. Loiseau, S.L. Brunton, Constrained sparse Galerkin regression, 2016, arXiv:1611.03271 [physics] URL: http://arxiv.org/abs/1611.03271, arXiv: 1611.03271.

[31] S. Zhang, G. Lin, Robust data-driven discovery of governing physical laws with error bars, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 474 (2018) 20180305, http://dx.doi.org/10.1098/rspa.2018.0305.

[32] K. Champion, B. Lusch, J.N. Kutz, S.L. Brunton, Data-driven discovery of coordinates and governing equations, Proc. Natl. Acad. Sci. 116 (2019) 22445–22451, http://dx.doi.org/10.1073/pnas.1906995116.

[33] M. Hoffmann, C. Fröhner, F. Noé, Reactive SINDy: Discovering governing reactions from concentration data, J. Chem. Phys. 150 (2019) 025101, http://dx.doi.org/10.1063/1.5066099.

[34] Z. Lai, S. Nagarajaiah, Sparse structural system identification method for nonlinear dynamic systems with hysteresis/inelastic behavior, Mech. Syst. Signal Process. 117 (2019) 813–842, http://dx.doi.org/10.1016/j.ymssp.2018.08.033, URL: https://linkinghub.elsevier.com/retrieve/pii/S0888327018305727.

[35] K. Wu, Numerical aspects for approximating governing equations using data, J. Comput. Phys. 22 (2019).

[36] Z. Huang, Y. Tian, C. Li, G. Lin, L. Wu, Y. Wang, H. Jiang, Data-driven automated discovery of variational laws hidden in physical systems, J. Mech. Phys. Solids 137 (2020) 103871, http://dx.doi.org/10.1016/j.jmps.2020.103871, URL: https://linkinghub.elsevier.com/retrieve/pii/S0022509619306246.

[37] H. Schaeffer, G. Tran, R. Ward, L. Zhang, Extracting structured dynamical systems using sparse optimization with very few samples, Multiscale Model. Simul. 18 (2020) 1435–1461, http://dx.doi.org/10.1137/18M1194730.

[38] Z. Wang, J. Estrada, E. Arruda, K. Garikipati, Discovery of deformation mechanisms and constitutive response of soft material surrogates of biological tissue by full-field characterization and data-driven variational system identification, 2020, http://dx.doi.org/10.1101/2020.10.13.337964, bioRxiv.

[39] S.H. Rudy, S.L. Brunton, J.L. Proctor, J.N. Kutz, Data-driven discovery of partial differential equations, Sci. Adv. 3 (2017) e1602614, http://dx.doi.org/10.1126/sciadv.1602614.

[40] H. Schaeffer, Learning partial differential equations via data discovery and sparse optimization, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 473 (2017) 20160446, http://dx.doi.org/10.1098/rspa.2016.0446.

[41] A.G. Baydin, B.A. Pearlmutter, A.A. Radul, J.M. Siskind, Automatic differentiation in machine learning: a survey, J. Mach. Learn. Res. 18 (2017) 43.

[42] J. Berg, K. Nyström, Data-driven discovery of PDEs in complex datasets, J. Comput. Phys. 384 (2019) 239–252, http://dx.doi.org/10.1016/j.jcp.2019.01.036, URL: http://arxiv.org/abs/1808.10788, arXiv: 1808.10788.

[43] G.J. Both, S. Choudhury, P. Sens, R. Kusters, DeepMoD: Deep learning for model discovery in noisy data, 2019, arXiv:1904.09406 [physics, q-bio, stat] URL: http://arxiv.org/abs/1904.09406, arXiv: 1904.09406.

[44] Z. Chen, Y. Liu, H. Sun, Deep learning of physical laws from scarce data, 2020, arXiv:2005.03448 [physics, stat] URL: http://arxiv.org/abs/2005.03448. arXiv: 2005.03448.

[45] S. Hartmann, Numerical studies on the identification of the material parameters of Rivlin's hyperelasticity using tension-torsion tests, Acta Mech. 148 (2001) 129–155, http://dx.doi.org/10.1007/BF01183674.

[46] G. Marckmann, E. Verron, Comparison of hyperelastic models for rubber-like materials, Rubber Chem. Technol. 79 (2006) 835–858, http://dx.doi.org/10.5254/1.3547969, URL: https://meridian.allenpress.com/rct/article/79/5/835/93139/Comparison-of-Hyperelastic-Models-for-RubberLike.

[47] A.F. Bower, Applied Mechanics of Solids, CRC press, 2009.

[48] A.N. Gent, A.G. Thomas, Forms for the stored (strain) energy function for vulcanized rubber, J. Polym. Sci. 28 (1958) 625–628, http://dx.doi.org/10.1002/pol.1958.1202811814.

[49] L.R.G. Treloar, The elasticity of a network of long-chain molecules - ii, Trans. Faraday Soc. 39 (1943) 241–246, http://dx.doi.org/10.1039/TF9433900241.

[50] A. Isihara, N. Hashitsume, M. Tatibana, Statistical theory of rubber like elasticity. iv. (two dimensional stretching), J. Chem. Phys. 19 (1951) 1508–1512, http://dx.doi.org/10.1063/1.1748111.

[51] E.M. Arruda, M.C. Boyce, A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials, J. Mech. Phys. Solids 41 (1993) 389–412, http://dx.doi.org/10.1016/0022-5096(93)90013-6, URL: https://www.sciencedirect.com/science/article/pii/0022509693900136.

[52] V. Biderman, Calculations of rubber parts (in russian), Raschet. Prochnost 40 (1958).

[53] D. Haines, W. Wilson, Strain–energy density function for rubberlike materials, J. Mech. Phys. Solids 27 (1979) 345–360, http://dx.doi.org/10.1016/0022-5096(79)90034-6, URL: http://www.sciencedirect.com/science/article/pii/0022509679900346.

[54] J.M. Ball, Convexity conditions and existence theorems in nonlinear elasticity, Arch. Ration. Mech. Anal. 63 (1976) 337–403.

[55] M. Grédiac, F. Pierron, S. Avril, E. Toussaint, The virtual fields method for extracting constitutive parameters from full-field measurements: a review, Strain 42 (2008) 233–253, http://dx.doi.org/10.1111/j.1475-1305.2006.tb01504.x.

[56] A. Tayeb, J.B. Le Cam, E. Robin, F. Canévet, M. Grédiac, E. Toussaint, X. Balandraud, Sensitivity based virtual fields for identifying hyperelastic constitutive parameters, in: 11th European Conference on Constitutive Models for Rubber, CRC Press/Balkema, Nantes, France, 2019, pp. 163–168, http://dx.doi.org/10.1201/9780429324710-29, URL: https://hal-univ-rennes1.archives-ouvertes.fr/hal-02499976.

[57] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1996) 267–288, http://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[58] l. E. Frank, J.H. Friedman, A statistical view of some chemometrics regression tools, Technometrics 35 (1993) 109–135, http://dx.doi.org/10.1080/00401706.1993.10485033.

[59] R. Wang, N. Xiu, S. Zhou, Fast Newton method for sparse logistic regression, 2019, arXiv:1901.02768 [math] URL: http://arxiv.org/abs/1901.02768, arXiv: 1901.02768.

[60] H. Schreier, J.J. Orteu, M.A. Sutton, Image Correlation for Shape, Motion and Deformation Measurements, Springer US, Boston, MA, 2009, http://dx.doi.org/10.1007/978-0-387-78747-3.

[61] M. Bonnet, A. Constantinescu, Inverse problems in elasticity, Inverse Problems 21 (2005) R1–R50, http://dx.doi.org/10.1088/0266-5611/21/2/R01.

[62] C. Saunders, A. Gammerman, V. Vovk, Ridge regression learning algorithm in dual variables, in: In Proceedings of the 15th International Conference on Machine Learning, Morgan Kaufmann, 1998, pp. 515–521.

[63] S. Theodoridis, K. Koutroumbas, Pattern Recognition, fourth ed., Academic Press, Burlington, Mass, 2009.

[64] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, Cambridge, Mass, 2006.

[65] G. Gundersen, 2020, URL: https://gregorygundersen.com/blog/2020/01/06/kernel-gp-regression/.