NN-mCRE: a modified Constitutive Relation Error framework for unsupervised learning of nonlinear state laws with physics-augmented Neural Networks

Antoine Benady, Emmanuel Baranger, Ludovic Chamoin

NN-mCRE: A modified Constitutive Relation Error framework for unsupervised learning of nonlinear state laws with physics-augmented Neural Networks

Antoine Benady
Université Paris-Saclay, CentraleSupélec
ENS Paris-Saclay, CNRS,
LMPS - Laboratoire de Mécanique Paris-Saclay,
4, Avenue des Sciences, 91190, Gif-sur-Yvette, France
antoine.benady@ens-paris-saclay.fr

Emmanuel Baranger
Université Paris-Saclay, CentraleSupélec
ENS Paris-Saclay, CNRS,
LMPS - Laboratoire de Mécanique Paris-Saclay,
4, Avenue des Sciences, 91190, Gif-sur-Yvette, France
emmanuel.baranger@ens-paris-saclay.fr

Ludovic Chamoin
Université Paris-Saclay, CentraleSupélec
ENS Paris-Saclay, CNRS,
LMPS - Laboratoire de Mécanique Paris-Saclay,
4, Avenue des Sciences, 91190, Gif-sur-Yvette, France

IUF - Institut Universitaire de France
1 rue Descartes, 75231, Paris Cedex 5 France
ludovic.chamoin@ens-paris-saclay.fr

Abstract

his article proposes a new approach to train physics-augmented neural networks with observable data to represent mechanical constitutive laws. To train the neural network and learn thermodynamics potentials, the proposed method does not rely on strain-stress or strain-free energy pairs but needs only partial strain or displacement measurements inside the structure. The neural network is trained thanks to an unsupervised procedure in which the modified Constitutive Relation Error (mCRE) is minimized. The mCRE functional provides a bias-aware data assimilation framework with a rich physical sense as the Constitutive Relation Error (CRE) part can be interpreted as a modeling error continuously defined over the structure, and can be used as a prediction quality in the inference phase. This article also extends previous works on the mCRE by introducing a new minimization procedure in the case of nonlinear state laws. As typical structural health monitoring applications may require that the neural networks should be trained online, an important focus is thus made on automatic and adaptive tuning of sensitive hyperparameters (learning rate, weighting between losses, number of epochs and initialization). It is shown that when the training database is rich enough with respect to the loading cases, the proposed method achieves remarkable performance regarding the quality of the learned model, noise robustness, and low sensitivity to user-defined hyperparameters. The method is evaluated on two test cases: a non-quadratic potential in the small strain regime with synthetic optic fiber measurements, and a Mooney-Rivlin model in the hyperelastic case with synthetic digital image correlation observations.

***K*eywords** Constitutive modeling, Physics-augmented Neural Network, Unsupervised learning, Constitutive relation error, Evolution laws, Data assimilation

1 Introduction

Structural damage is a critical engineering issue. It may affect a large scope of structures, including wind turbines, aircraft or space structures. Consequently, Structural Health Monitoring (SHM) with early damage detection and diagnosis of its evolution is fundamental to control the integrity of sensor-equipped structures throughout their life cycle, thus increasing durability and safety. To exploit the best between numerics and measurements, the trend is to design new processes in which physical systems and their numerical simulator (twin) dynamically and seamlessly exchange information in real-time through a feedback loop. This paradigm, called Dynamic Data Driven Application Systems (DDDAS) [1, 2] aims

- to continuously predict the evolution of the physical phenomena of interest and drive the system accordingly (inference phase);
- to dynamically update the computational model using in situ measurements which are assimilated on-the-fly (training phase).

Nowadays, accurate experimental information can be obtained on the damage state with observations from optical fibers or imaging, which give high spatial resolution [3, 4]. Physics-based damage modeling and simulation have also been the object of significant advances for more than forty years. They represent a rich history of engineering sciences, and they are nowadays used as virtual twins for the numerical prediction of damage initiation and growth, until failure, inside structures (replacing costly experimental campaigns) [5]. To accurately predict structural health, constitutive modeling is essential. Traditionally, experimental information was used to update the parameters of a constitutive model, which form was fixed. Parameter identification of constitutive model has been studied for the past decades and aims to find the model parameters $\mathbf{p}$ of a constitutive relation that match the best with the observations [6]. A wide variety of methods has been designed so far for parameter identification, such as the Finite Element Model Updating (FEMU) [7], the Equilibrium Gap Method (EGM) [8], the Virtual Fields Method (VFM) [9], the Reciprocity Gap Method (RGM)[10], or the modified constitutive relation error (mCRE)[11], which is the method of interest here.

The mCRE provides an interesting framework for parameter identification and updating as the associated functional incorporates a strong physical sense that can be interpreted as a modeling error. The strategy enforces reliable information (*e.g.* equilibrium) while releasing uncertain information (*e.g.* constitutive law or measurement values). It is thus a hybrid approach that leans on both physics-based and data-based information. It is a flexible compromise between purely data-based approaches[12] which require high data quantity, have high variance and low bias, and model-based approaches which require low data quantity and have low variance and high bias. Numerous researches dealing with mCRE have been conducted, involving forced vibrations dynamics [13, 14, 15], transient dynamics [16, 17] and nonlinear material behavior [18, 19]. Additionally, the mCRE framework has proved its robustness to corrupted [20] and noisy [21] measurements. The classical mCRE approach, as other inversion methods, postulates a constitutive model (even though it is released to inform on possible bias), which has the advantage of being physically interpretable. Yet, the postulated model form does not always allow to describe the observed complexity correctly. Searching for the best constitutive relation in the structure, the present work proposes a mCRE framework in which the form of the model is not assumed, thus naturally correcting model bias using neural networks known as universal approximators [22].

In [23], the representation of the constitutive relation through a neural network was among the pioneering contributions. At that time, the learning of strain-stress mapping was purely data-driven, which means that no physical knowledge was neither enforced in the network architecture nor informed in the loss function. Recently, physics-informed machine learning has shown interesting results in several scientific fields [24]. Hybrid modeling is a good compromise between these two approaches. Coupling techniques between deep learning and physical models help alleviate frequent concerns in the use of neural networks, such as lack of physical consistency, lack of generalization, and difficulty to train (quantity of data, hyperparameters tuning) [25]. A first possibility, introduced in [26] as a "physics-informed" approach, is to include physical knowledge in the loss function used to train the neural networks. This idea is to write the loss function as the sum of a measure discrepancy term and a penalization of the non-satisfaction of physics. A second idea, often mentioned as "physics-augmented" is to enforce physical knowledge in the architecture. Several frameworks have been proposed to take into account physical properties such as thermodynamics constraints in the architecture [27, 28, 29, 30, 31]. All the latter are based on the convexity of the free energy imposed by an input-convex neural network (ICNN) introduced in [32]. A third one, often referred to as transfer learning, consists of informing prior knowledge in the network initialization. This idea helps increase the quantity of data used for training and reduce the sensitivity of a random initialization on the training result. The proposed mCRE strategy naturally integrates all these recent trends, by the definition of the loss function, the chosen architecture of the neural network, and the physics-guided initialization.

Dealing with constitutive modeling, coupling physical knowledge with neural networks is now an emerging field. The recent contributions can be separated according to the data used for training. A first community aims to train neural networks in a supervised learning procedure with a strain-stress database (or strain-free energy) generated from a known constitutive model [27, 28, 30, 33, 34, 35, 36]. As the forward pass of a neural network can be easily parallelized, the use of a neural network after training can achieve high gain in computational time when the initial model is costly to compute, for example, in the case of microscale modeling [37, 38]. Measuring strain-stress couples with complex loading is a challenge today, so the supervised training procedure cannot be used to train neural networks. In [29] and [39], methods tackled the issue of unsupervised training of neural networks for constitutive modeling. Full-field displacement observations were used to compute strain employed as an input of a neural network trained to output stress with a loss function that penalizes the non-satisfaction of the equilibrium (boundary conditions are known). The advantage of the present method over the latter is the use of a modeling error directly focused on what needs to be learned: the constitutive relation. A comparison of the current method and [**?**] is available in [40]. Furthermore, this constitutive relation error localizes the model bias in the structure. The present work requires partial observations, such as the case with optic fiber measurement.

The following uses the mCRE framework in the case of a free model form: the constitutive relation is described with physics-augmented neural networks that guarantee physical properties enforced in the architecture. The contribution of this paper is not the use of a new physics-augmented architecture, but the training of an existing architecture [27, 28, 29, 30, 41] in an unsupervised fashion. This adaptation of the mCRE framework requires a new minimization procedure proposed here. Furthermore, the DDDAS context imposes that the neural network can be trained online, so this article also proposes an automatic tuning of sensitive user-defined parameters such as initialization, learning rate and ponderation between losses. The method focuses on isothermal constitutive state laws (evolution laws will be investigated later) and is tested on synthetic data with 2D geometry. The boundary conditions and the measurement noise level are assumed to be known, even though the mCRE framework is suited to deal with uncertain boundary conditions (see for example [42]). The validation of the proposed method is based on the quality of the learned constitutive relation, noise robustness, and sensitivity to user-defined hyperparameters.

The reminder of the paper is organized as follows; Section 2 recalls basics on the mCRE and introduces the new minimization procedure suited for nonlinear state laws. Section 3 details the specificities of using neural networks to describe the constitutive relation (how physics is enforced in the architecture and hyperparameters tuning strategy). Section 4 evaluates the method and discusses the results with two test cases: a non-quadratic potential in the small strain regime (with optic fiber observations) and a Mooney-Rivlin potential in hyperelasticity (with digital image correlation). Eventually, conclusions and prospects are drawn in Section 5.

2 The modified constitutive relation error (mCRE) concept and minimization for nonlinear constitutive relations

This section defines the problem (Section 2.1), recalls the basics on the CRE (Section 2.2) and mCRE (Section 2.3) concepts, and introduces the proposed minimization of mCRE suited for nonlinear state laws (Section 2.4).

2.1 Problem definition

To define the problem notations, let us consider a body in initial configuration $\Omega \subset \mathbb{R}^d (d = 1, 2, 3)$ (Figure 1) with boundary $\partial\Omega$ and isothermal environment. Dirichlet boundary conditions are imposed on $\partial\Omega_1 \subset \partial\Omega$ by means of a displacement field $\mathbf{u}_d$. Neumann boundary conditions are prescribed on $\partial\Omega_2 \subset \partial\Omega$ by means of a force field $\mathbf{f}_s^d$. A body force field $\mathbf{f}_v^d$ may also be prescribed in Ω. Additionally, strain measurements $\mathbf{E}_{obs}$ (in the case of observations from optic fibers) or displacement measurements $\mathbf{u}_{obs}$ (in the case of digital image correlation) are available; Section 4 shows one case per measurement type.

The solution to the mechanical problem is the couple $(\mathbf{u}, \mathbf{S})$ of displacement and second Piola-Kirchoff stress fields that satisfies the three following groups of equations:

- **kinematic admissibility** defines the space $\boldsymbol{\mathcal{U}}_{ad}$ of displacement fields satisfying the Dirichlet boundary conditions:

$$\mathbf{u}_{|\partial\Omega_1} = \mathbf{u}_d$$

- **static admissibility** defines the space $\boldsymbol{\mathcal{S}}_{ad}$ of stress fields satisfying the equilibrium :

$$\int_\Omega \mathbf{S}(\mathbf{u}) : \left(\mathbf{F}(\mathbf{u})^T.\nabla\mathbf{v}\right) = \int_\Omega \mathbf{f}_d^v.\mathbf{v} + \int_{\partial\Omega_2} \mathbf{f}_d^s.\mathbf{v} \;\; \forall \mathbf{v} \in \boldsymbol{\mathcal{U}}_0$$

  with $\mathbf{F}$ the deformation gradient and $\mathcal{U}_0$ the space of kinematic admissibility with homogeneous Dirichlet conditions.

- **constitutive relation** :

$$\mathbf{S} = \frac{\partial\psi(\mathbf{E};\ \mathbf{p})}{\partial\mathbf{E}}$$

  with the strain tensor $\mathbf{E}(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^T + \nabla\mathbf{u}^T\nabla\mathbf{u})$ and $\mathbf{p}$ the parameters of the constitutive law. Such a notation for the constitutive law, involving a convex potential ψ depending on state variables alone, is provided by the first principle of thermodynamics.

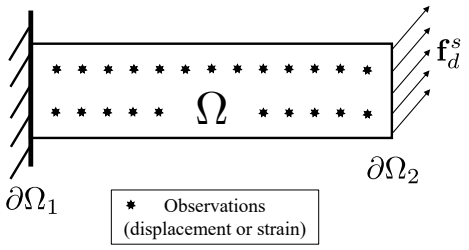

Figure 1: Configuration studied.

The minimization of the mCRE aims to identify the constitutive relation parameters $\mathbf{p}$ that fit the best to experimental data. Here Dirichlet and Neumann's conditions are known, as well as the observations $\mathbf{E}_{obs}$ or $\mathbf{u}_{obs}$ that are affected by measurement noise.

2.2 CRE concept

The CRE concept was introduced in the 70s in the context of finite element verification [43] (*i.e.* a posteriori error estimation and mesh adaptation). A detailed explanation can be found in [44] and the general idea is to split the equations of Section 2.1 following the reliability of the information. The constitutive relation is considered as the unreliable equation so that an admissible pair $(\hat{\mathbf{u}}, \hat{\mathbf{S}}) \in (\boldsymbol{\mathcal{U}}_{ad} \times \boldsymbol{\mathcal{S}}_{ad})$ is constructed (see [44]) and the constitutive relation error is defined as follows:

$$\mathcal{E}^2_{CRE}(\hat{\mathbf{u}}, \hat{\mathbf{S}}) = \int_\Omega \left(\psi(\hat{\mathbf{u}}) + \psi^*(\hat{\mathbf{S}}) - \hat{\mathbf{S}} : \mathbf{E}(\hat{\mathbf{u}})\right) \tag{1}$$

where $\psi^*(\hat{\mathbf{S}})$ stands for the Legendre-Fenchel dual potential defined by $\psi^*(\hat{\mathbf{S}}) = \sup_{\mathbf{E}} [\hat{\mathbf{S}} : \mathbf{E} - \psi(\mathbf{E})]$.

In the case of a quadratic potential, when $\psi(\mathbf{E}) = \frac{1}{2}\mathbf{E} : \mathbf{K} : \mathbf{E}$, the dual potential reads $\psi^*(\mathbf{S}) = \frac{1}{2}\mathbf{S} : \mathbf{K}^{-1} : \mathbf{S}$, the constitutive relation is $\mathbf{S} = \mathbf{K}\mathbf{E}(\mathbf{u})$ and the CRE is written :

$$\mathcal{E}^2_{CRE}(\hat{\mathbf{u}}, \hat{\mathbf{S}}) = \int_\Omega (\hat{\mathbf{S}} - \mathbf{K}\mathbf{E}(\hat{\mathbf{u}})) : \mathbf{K}^{-1} : (\hat{\mathbf{S}} - \mathbf{K}\mathbf{E}(\hat{\mathbf{u}})) = ||\hat{\mathbf{S}} - \mathbf{K}\mathbf{E}(\hat{\mathbf{u}})||^2_{\mathbf{K}^{-1}} \tag{2}$$

where $\mathbf{K}$ is the symmetric positive definite Hooke tensor and $||\bullet||_{\mathbf{K}^{-1}}$ is the energy norm on stress fields. Here it is clear that the error is on the constitutive relation.

Figure 2 shows two interpretations of the CRE concept. On the left, in the space of strain-stress couples, we can see that for an admissible pair $(\hat{\mathbf{u}}, \hat{\mathbf{S}})$, the CRE is a distance to the space of stress-strain couples that satisfy the constitutive relation. This is the meaning of (2) in the case of a quadratic potential. On the right, for a given point $(\hat{\mathbf{E}}, \hat{\mathbf{S}})$, the area in blue is $\psi^*(\hat{\mathbf{S}})$, the area in orange is $\psi(\hat{\mathbf{u}})$, the area delimited by the red rectangle is $\hat{\mathbf{S}} : \hat{\mathbf{E}}$. Thus the hatched area is $\psi(\hat{\mathbf{u}}) + \psi^*(\hat{\mathbf{S}}) - \hat{\mathbf{S}} : \mathbf{E}(\hat{\mathbf{u}})$.

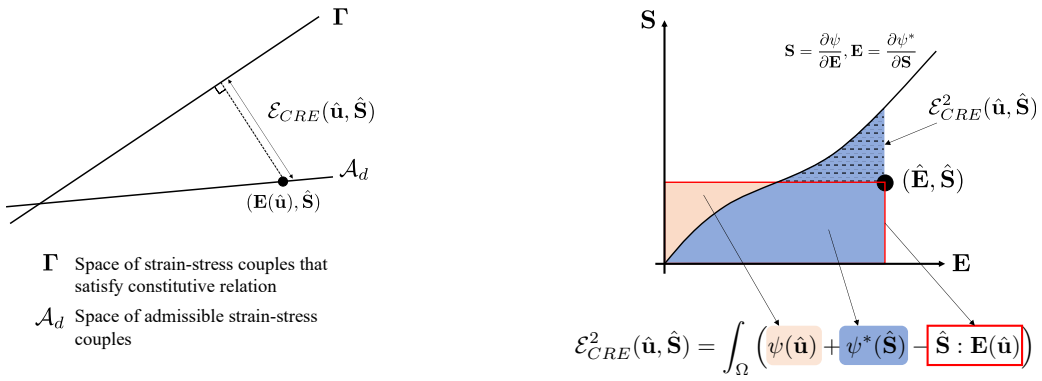

Figure 2: Left: interpretation of the CRE in the space of strain-stress couples. Right: interpretation of the CRE with the stress-strain curve.

2.3 Extension of the CRE concept to the context of parameters updating: the modified Constitutive Relation Error

In the 90's the CRE concept was used in the context of inverse problems and parameter updating. In the first approach, observations were enforced in the definition of the admissibility space and optimal parameters were found by minimizing the CRE cost-function [45], but this approach was not well suited to the case of important measurement noise. In [11], a more flexible approach (the one used in this paper) was proposed to enforce only reliable information on the admissible space $\mathcal{A}_d$, thus consistently extending the general framework. The optimal model parameters $\mathbf{p}_{opt}$ are defined as :

$$\mathbf{p}_{opt} = \underset{\mathbf{p}}{\operatorname{argmin}} \left[\min_{(\hat{\mathbf{u}}, \hat{\mathbf{S}}) \in \mathcal{A}_d} \mathcal{E}^2_{mCRE}(\hat{\mathbf{u}}, \hat{\mathbf{S}};\ \mathbf{p}) \right] \tag{3}$$

with

$$\mathcal{E}^2_{mCRE}(\hat{\mathbf{u}}, \hat{\mathbf{S}};\ \mathbf{p}) = \int_\Omega \left(\psi(\hat{\mathbf{u}}; \mathbf{p}) + \psi^*(\hat{\mathbf{S}}; \mathbf{p}) - \hat{\mathbf{S}} : \mathbf{E}(\hat{\mathbf{u}}) \right) + \alpha ||\Pi \mathbf{u} - \mathbf{u}_{obs}||^2 \tag{4}$$

where $\mathcal{A}_d = (\boldsymbol{\mathcal{U}}_{ad} \times \boldsymbol{\mathcal{S}}_{ad})$ (even though it is possible to consider loading as uncertain information, see [4]), α is a scaling factor, and Π is an operator that extracts observation outputs from the displacement field $\mathbf{u}$, so that the model can be compared with the observations. For example, in the discretized version of the mCRE with displacement observations, Π is a diagonal matrix in which the diagonal coefficient is 1 when the degree of freedom is observed and 0 otherwise.

The minimization of (3) depends on the type of model considered. Previous works have been done on linear dynamics models [11, 13, 17, 46], for defect identification [47], with uncertain measurements or loading [15, 48], coupled with data-assimilation tools [49] or with nonlinear evolution laws [18]. The present work extends the work of [18] by focusing on nonlinearities in state laws that require a different minimization procedure. The following section introduces this new minimization algorithm.

2.4 Minimization of mCRE for nonlinear state laws

In the present context, the minimization of the mCRE aims to find the parameters $\mathbf{p}_{opt}$ such that :

$$\mathbf{p}_{opt} = \underset{\mathbf{p} \in \mathcal{P}}{\operatorname{argmin}} \left[\min_{(\mathbf{u}, \mathbf{S}) \in \mathcal{A}_d} \left[\int_\Omega \left(\psi(\hat{\mathbf{u}}; \mathbf{p}) + \psi^*(\hat{\mathbf{S}}; \mathbf{p}) - \hat{\mathbf{S}} : \mathbf{E}(\hat{\mathbf{u}}) \right) + \alpha ||\Pi \mathbf{E}(\hat{\mathbf{u}}) - \mathbf{E}_{obs}||^2 \right] \right] \tag{5}$$

where the potential ψ and the space of parameters $\mathcal{P}$ are such that the constitutive model guarantees the physical conditions listed in Section 3.1. This section describes the minimization in a unified framework suited for :

- an explicitly given parametrized constitutive relation as it is traditionally the case with the mCRE framework [4];
- a constitutive relation described by a neural network (in this case, parameters $\mathbf{p}$ should be understood as weights and biases of the network). The description of a constitutive relation by a physics-augmented neural

network enables to search for the potential ψ in a space of potentials that satisfy the requirements of a constitutive law (see Section 3.1).

This minimization consists in an iterative process where iteration n is composed of :

- a first step performed with fixed parameters $\mathbf{p}^{(n)}$, in which a new admissible field couple $(\hat{\mathbf{u}}, \hat{\mathbf{S}})^{(n+1)}$ is found to minimize the mCRE.
- a second step performed with fixed admissible field couple $(\hat{\mathbf{u}}, \hat{\mathbf{S}})^{(n+1)}$, in which parameters are updated with a gradient descent step.

Both steps are detailed in the following section for iteration n. A novelty of this approach lies in the way the Step 1 is performed: in the present context this Step 1 is nonlinear and the minimization is performed with a Newton scheme.

2.4.1 Step 1

In this step parameters $\mathbf{p}^{(n)}$ are fixed and the goal is to find :

$$(\hat{\mathbf{u}}, \hat{\mathbf{S}})^{(n+1)} = \underset{(\mathbf{u},\mathbf{S})\in\mathcal{A}_d}{\operatorname{argmin}} \left[\mathcal{E}^2_{mCRE}(\hat{\mathbf{u}}, \hat{\mathbf{S}};\ \mathbf{p}^{(n)})\right] \tag{6}$$

with

$$\mathcal{E}^2_{mCRE}(\hat{\mathbf{u}}, \hat{\mathbf{S}};\ \mathbf{p}^{(n)}) = \int_\Omega \Big(\psi(\hat{\mathbf{u}}; \mathbf{p}^{(n)}) + \psi^*(\hat{\mathbf{S}}; \mathbf{p}^{(n)}) - \hat{\mathbf{S}} : \mathbf{E}(\hat{\mathbf{u}})\Big) + \alpha||\Pi\mathbf{E}(\hat{\mathbf{u}}) - \mathbf{E}_{obs}||^2 \tag{7}$$

To express the Legendre-Fenchel transform $\psi^*(\hat{\mathbf{S}})$ (defined in Section 2.2) in a more convenient way, a displacement field $\hat{\mathbf{v}}$ is introduced. With the convexity of ψ with respect to $\mathbf{E}$ the $\sup$ is reached for $\hat{\mathbf{v}}$ such that :

$$\hat{\mathbf{S}} = \frac{\partial \psi}{\partial \mathbf{E}}\Big|_{\mathbf{E}(\hat{\mathbf{v}})} \tag{8}$$

so that,

$$\psi^*(\hat{\mathbf{S}}) = \frac{\partial \psi}{\partial \mathbf{E}}\Big|_{\mathbf{E}(\hat{\mathbf{v}})} : \mathbf{E}(\hat{\mathbf{v}}) - \psi(\mathbf{E}(\hat{\mathbf{v}})) \tag{9}$$

By replacing in (7) the mCRE writes :

$$\mathcal{E}^2_{mCRE}(\hat{\mathbf{u}}, \hat{\mathbf{v}}; \mathbf{p}) = \int_\Omega \Big(\psi(\hat{\mathbf{u}}; \mathbf{p}^{(n)}) - \psi(\hat{\mathbf{v}}; \mathbf{p}^{(n)}) + \frac{\partial \psi}{\partial \mathbf{E}}\Big|_{\mathbf{E}(\hat{\mathbf{v}})} : (\mathbf{E}(\hat{\mathbf{v}}) - \mathbf{E}(\hat{\mathbf{u}}))\Big) + \alpha||\Pi\mathbf{E}(\hat{\mathbf{u}}) - \mathbf{E}_{obs}||^2 \tag{10}$$

To search the couple $(\hat{\mathbf{u}}, \hat{\mathbf{S}})^{(n+1)}$ in the admissible space $\mathcal{A}_d$:

- $\hat{\mathbf{u}}$ is discretized with the following decomposition :

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_a \\ \mathbf{U}_d \end{bmatrix} \tag{11}$$

  where $\mathbf{U}_a$ is the discretized displacement field at active degrees of freedom and $\mathbf{U}_d$ the discretized displacement field for the Dirichlet boundary conditions.
- the mCRE minimization is performed under the constraint :

$$h(\mathbf{v}, \boldsymbol{\lambda}) = \int_\Omega \mathbf{S}(\mathbf{v}) : \Big(\mathbf{F}(\mathbf{v})^T.\nabla\boldsymbol{\lambda}\Big) - \int_\Omega \mathbf{f}^v_d.\boldsymbol{\lambda} - \int_{\partial\Omega_2} \mathbf{f}^s_d.\boldsymbol{\lambda} = \mathbf{0}\ \ \forall\boldsymbol{\lambda} \in \boldsymbol{\mathcal{U}}_0 \tag{12}$$

  so that $\hat{\mathbf{S}}$ satisfies the static admissibility.

This constrained minimization is performed by means of a Lagrangian functional:

$$\mathcal{L} = \mathcal{E}^2_{mCRE}(\hat{\mathbf{u}}, \hat{\mathbf{v}}) + h(\mathbf{v}, \boldsymbol{\lambda}) \tag{13}$$

with the following Newton scheme [50]:

$$\begin{bmatrix} \mathbf{u}_{k+1} - \mathbf{u}_k \\ \mathbf{v}_{k+1} - \mathbf{v}_k \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 \mathcal{E}^2_{mCRE}}{\partial \mathbf{u}^2} & \frac{\partial^2 \mathcal{E}^2_{mCRE}}{\partial \mathbf{u} \partial \mathbf{v}} & \mathbf{0} \\ \frac{\partial^2 \mathcal{E}^2_{mCRE}}{\partial \mathbf{v} \partial \mathbf{u}} & \frac{\partial^2 \mathcal{E}^2_{mCRE}}{\partial \mathbf{v}^2} & \frac{\partial h(\mathbf{v}, \boldsymbol{\lambda})}{\partial \mathbf{v}} \\ \mathbf{0} & \frac{\partial h(\mathbf{v}, \boldsymbol{\lambda})}{\partial \mathbf{v}} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial \mathcal{E}^2_{mCRE}}{\partial \mathbf{u}} \\ \frac{\partial \mathcal{E}^2_{mCRE}}{\partial \mathbf{v}} \\ h(\mathbf{v}, \boldsymbol{\lambda}) \end{bmatrix} \tag{14}$$

The right-hand side of (14) is evaluated with the quantities of iteration k.

The first-order derivatives are:

$$\begin{cases} \frac{\partial \mathcal{E}^2_{mCRE}}{\partial \mathbf{u}} &= \int_\Omega \Big(\frac{\partial \mathbf{E}}{\partial \mathbf{u}} (\frac{\partial \psi}{\partial \mathbf{E}} \big|_{\mathbf{E}(\mathbf{u})} - \frac{\partial \psi}{\partial \mathbf{E}} \big|_{\mathbf{E}(\mathbf{v})}) \Big) + \alpha \frac{\partial \mathbf{E}}{\partial \mathbf{u}} (\Pi \mathbf{E}(\mathbf{u}) - \mathbf{E}_{obs})) \\ \frac{\partial \mathcal{E}^2_{mCRE}}{\partial \mathbf{v}} &= \int_\Omega \frac{\partial \mathbf{E}}{\partial \mathbf{v}} \frac{\partial^2 \psi}{\partial \mathbf{E}^2} \big|_{\mathbf{E}(\mathbf{v})} (\mathbf{E}(\mathbf{v}) - \mathbf{E}(\mathbf{u})) \\ h(\mathbf{v}, \boldsymbol{\lambda}) &= - \int_\Omega \frac{\partial \psi}{\partial \mathbf{E}} \big|_{\mathbf{E}(\mathbf{v})} : \mathbf{F}(\mathbf{v}) \nabla \boldsymbol{\lambda} - \mathbf{f}^v_d \boldsymbol{\lambda} - \int_{\partial_2 \Omega} \mathbf{f}^s_d \boldsymbol{\lambda} \end{cases} \tag{15}$$

Concerning second-order derivatives, terms that vanish at convergence of the Newton scheme are neglected so that:

$$\begin{cases} \frac{\partial^2 \mathcal{E}^2_{mCRE}}{\partial \mathbf{u}^2} = \int_\Omega \Big(\frac{\partial \mathbf{E}}{\partial \mathbf{u}} \Big)^2 \frac{\partial^2 \psi}{\partial \mathbf{E}^2} \big|_{\mathbf{E}(\mathbf{u})} + \alpha \Big(\frac{\partial \mathbf{E}}{\partial \mathbf{u}} \Big)^2 \Pi . \Pi^T \\ \frac{\partial^2 \mathcal{E}^2_{mCRE}}{\partial \mathbf{v}^2} = \int_\Omega \Big(\frac{\partial \mathbf{E}}{\partial \mathbf{v}} \Big)^2 \frac{\partial^2 \psi}{\partial \mathbf{E}^2} \big|_{\mathbf{E}(\mathbf{v})} \\ \frac{\partial^2 \mathcal{E}^2_{mCRE}}{\partial \mathbf{u} \partial \mathbf{v}} = \int_\Omega - \frac{\partial \mathbf{E}}{\partial \mathbf{v}} \frac{\partial \mathbf{E}}{\partial \mathbf{u}} \frac{\partial^2 \psi}{\partial \mathbf{E}^2} \big|_{\mathbf{E}(\mathbf{v})} \\ \frac{\partial h(\mathbf{v}, \boldsymbol{\lambda})}{\partial \mathbf{v}} = \int_\Omega \frac{\partial \mathbf{E}}{\partial \mathbf{v}} \frac{\partial^2 \psi}{\partial \mathbf{E}^2} \big|_{\mathbf{E}(\mathbf{v})} \mathbf{F}(\mathbf{v}) \nabla \boldsymbol{\lambda} + \frac{\partial \psi}{\partial \mathbf{E}} \big|_{\mathbf{E}(\mathbf{v})} \frac{\partial \mathbf{F}(\mathbf{v})}{\partial \mathbf{v}} \nabla \boldsymbol{\lambda} \end{cases} \tag{16}$$

with $\mathbf{F}(\mathbf{v}) = (1 + \nabla \mathbf{v})$.

All the integrals are performed using finite element discretization (A shows the discretization in the case of linear elasticity). The initialization of the Newton scheme is done with the fields $(\mathbf{u}, \mathbf{v})$ obtained at the end of step 1 with the previous parameters. Step 1 terminates when the relative update on the displacement field is below a tolerance (usually set to 10^{-4}).

2.4.2 Step 2

This step consists of the updating of the parameters $\mathbf{p}$ with a gradient descent step :

$$\mathbf{p}^{(n+1)} = \mathbf{p}^{(n)} - l_r \frac{d \mathcal{E}^2_{mCRE}(\hat{\mathbf{u}}^{(n+1)}, \hat{\mathbf{v}}^{(n+1)}; \mathbf{p}^{(n)})}{d \mathbf{p}} \tag{17}$$

with $(\hat{\mathbf{u}}^{(n+1)}, \hat{\mathbf{v}}^{(n+1)})$ the fields obtained at the end of step 1 of the $n+1$ iteration of mCRE minimization. For the sake of simplicity, $(\hat{\mathbf{u}}^{(n+1)}, \hat{\mathbf{v}}^{(n+1)})$ are denoted $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ in the following. The gradient of mCRE with respect to the parameters is computed with the adjoint-state method :

$$\frac{d \mathcal{E}^2_{mCRE}(\hat{\mathbf{u}}, \hat{\mathbf{v}}; \mathbf{p})}{d \mathbf{p}} = \frac{d \mathcal{L}(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \boldsymbol{\lambda}; \mathbf{p})}{d \mathbf{p}} = \frac{d \hat{\mathbf{u}}}{d \mathbf{p}} \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{u}}} + \frac{d \hat{\mathbf{v}}}{d \mathbf{p}} \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{v}}} + \frac{d \boldsymbol{\lambda}}{d \mathbf{p}} \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} + \frac{\partial \mathcal{L}}{\partial \mathbf{p}} = \frac{\partial \mathcal{L}}{\partial \mathbf{p}} \tag{18}$$

$$\frac{d \mathcal{E}^2_{mCRE}(\hat{\mathbf{u}}, \hat{\mathbf{v}}; \mathbf{p})}{d \mathbf{p}} = \int_\Omega \frac{\partial \psi}{\partial \mathbf{p}} \big|_{\mathbf{E}(\mathbf{u})} - \frac{\partial \psi}{\partial \mathbf{p}} \big|_{\mathbf{E}(\mathbf{v})} \tag{19}$$

with $\mathcal{L}$ the Lagrangian associated to the constrained minimization problem of step 1 defined in Section 2.4.1 (minimization defined in (6) with the constraint defined in (12)):

It is worth noticing in (19) that the parameters are updated following the computation of the gradient in the whole structure and not only where there is a measurement.

To summarize this section and introduce the next one, Figure 3 illustrates the general methodology for further training the physics-augmented neural network with the mCRE.

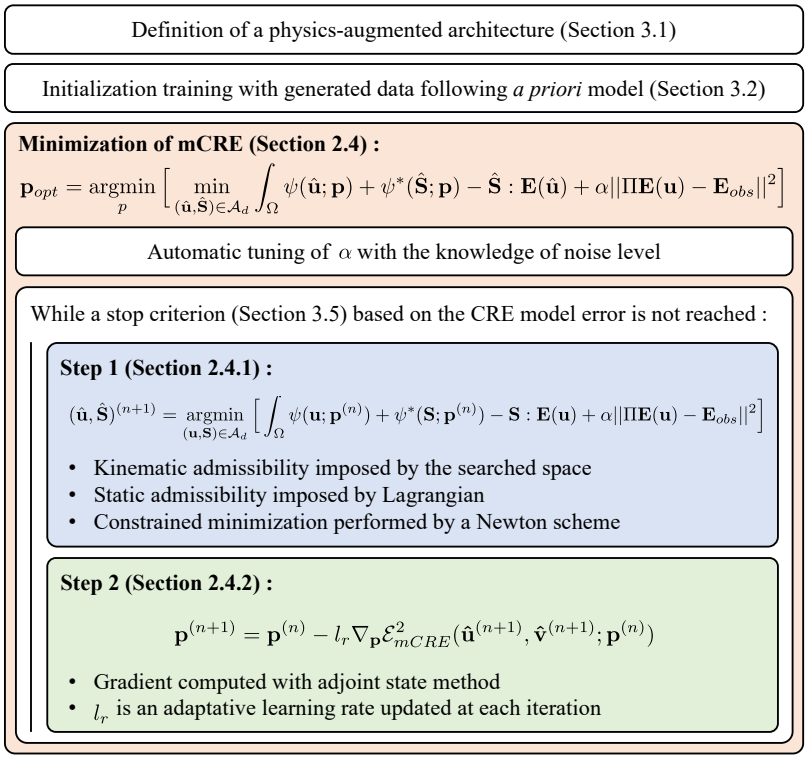

Figure 3: Description of the method developed.

3 Physics-augmented Neural Networks for constitutive modeling

The previous section has introduced the minimization of the mCRE in the case of nonlinear state laws, either in the case of an explicitly given model or for a constitutive relation described by a neural network. Here, the use of neural networks enables to release the form of the constitutive relation: the potential ψ is searched among all potentials that satisfy the requirements of a constitutive law defined in Section 3.1. Here, the parameters $\mathbf{p}$ are the weights and biases of the neural network.

With neural networks, a particular attention should be paid to the consistency of the prediction with respect to physical principle such as thermodynamics. Section 3.1 shows how physics is enforced in the architecture. In the case of a neural network, the minimization task is also much harder because of the large number of parameters. Indeed, neural networks are known to be hard to train [51], partly because of the sensitivity to user-defined hyperparameters. Moreover, the DDDAS framework prevents the user from changing hyperparameters and performing several times the training. In the following, this issue is tackled by:

- a physics-guided initialization described in Section 3.2;
- a Morozov-based automatic tuning of the weighting between losses introduced in Section 3.3;
- an empirical adaptive learning rate rule described in Section 3.4;
- a CRE based stop criterion used to avoid the predefined choice of the number of epochs (Section 3.5).

3.1 Definition of the physics-augmented architecture

The architecture used is shown in Figure 4. This architecture tends to be widely used for constitutive modeling [27, 28, 29, 30, 41].

This neural network guarantees the convexity of the free-energy ψ with respect to $\mathbf{E}$ thanks to a special architecture: input-convex neural networks (ICNN) [32]. The weights between intermediate layers are positive (the weights that map the input layer to intermediate layers can be negative), and the activation functions are convex and non-decreasing. A given layer is connected to the previous one through positive weights and connected to the input through weights in $\mathbb{R}$. The activation function used is $\mathcal{A}$:

$$\mathcal{A}(X_i) = \begin{cases} 0 \text{ if } X_i \leq 0 \\ X_i^2 \text{ if } X_i > 0 \end{cases} \tag{20}$$

where X_i denotes the i-th component of the vector $\mathbf{X}$.

This activation function is inspired by the ReLU function. Yet, the second derivative of the activation function should not be zero on the whole definition domain, otherwise, the tangent operator required in the optimization process would be zero. This motivates the modification of the ReLU function to a "ReLU square". The main focus of this work is not on the physical constraints in the architecture of the neural network but on the optimization procedure to tune it with observable data. A detailed guide for enforcing physical requirements for a constitutive model can be found in [41] for example.

In this article, the first test case takes components of the Green-Lagrange tensor as network input, while the second one uses the following invariant in the input of the network :

- $J = \det(\mathbf{2E} + \mathbf{I})$
- $\tilde{I}_1 - 3$, with $\tilde{I}_1 = J^{-2/3}\,\mathrm{tr}(\mathbf{2E} + \mathbf{I})$
- $\tilde{I}_2 - 3$, with $\tilde{I}_2 = \frac{1}{2}J^{-4/3}[\mathrm{tr}(2\mathbf{E} + \mathbf{I})^2 - \mathrm{tr}((\mathbf{2E} + \mathbf{I})^2)]$

The stress is obtained by automatic differentiation of the free energy:

$$\mathbf{S} = \frac{\partial \psi}{\partial \mathbf{E}} \tag{21}$$

$\psi(\mathbf{E})$ and $\mathbf{S} = \frac{\partial \psi}{\partial \mathbf{E}}$ should vanish at zero deformation ($\mathbf{E} = \mathbf{0}$) so the potential is written :

$$\psi(\mathbf{E}) = \psi_{NN}(\mathbf{E}) - \psi_{NN}(\mathbf{E} = \mathbf{0}) - \frac{\partial \psi_{NN}}{\partial \mathbf{E}}\Big|_{\mathbf{E}=\mathbf{0}} : \mathbf{E} \tag{22}$$

where ψ_{NN} is the output of the neural network.

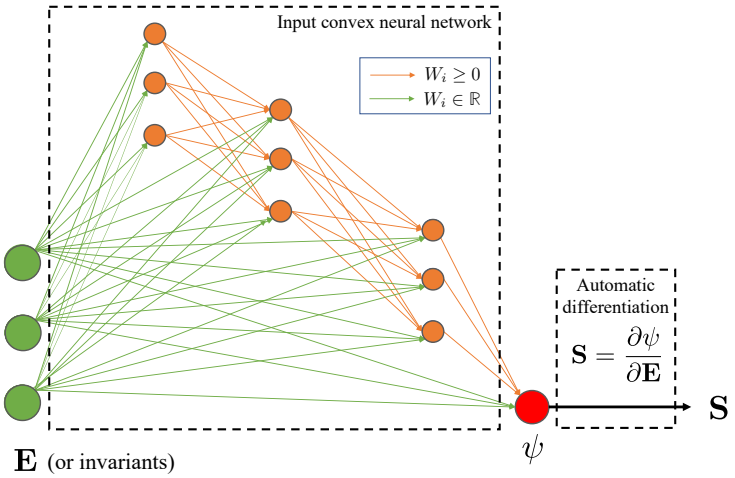

Figure 4: Input convex thermodynamics-consistent architecture used.

3.2 Reducing the sensitivity to initialization: physics-guided initialization with *a priori* knowledge

The training of neural networks is known to be sensitive to initialization [51], as well as the minimization of the mCRE in the case of parameter updating [4]. Instead of random initialization, a first training is performed to initialize the network with *a priori* model with the expected material parameters. Indeed, with a random initialization, the mCRE minimization fails to converge in most of the experiments performed. A database $\{\tilde{\mathbf{E}}, \tilde{\psi}\}$ is generated from a known constitutive model assumed to be close to the target one. The initialization training is performed in a classical supervised way with the Adam optimizer [52] in order to minimize the mean square error loss. The positivity constraint on the weights is already enforced through gradient clipping. The choice of the *a priori* model can depend on the structure and type of sollicitations considered. In Section 4, two different initializations are used (quadratic potential under small strain regime and Neo-Hookean model in the case of hyperelasticity).

The generated database is created to be rich enough to describe the studied range. The weights and biases of the neural networks obtained at the end of the training are used as the initialization of the mCRE minimization algorithm described in the following section.

3.3 Reducing the sensitivity to weighting between losses: automatic tuning of α

A frequent concern in multiple losses optimization is the tuning of the weighting between losses [53, 54]. This concern is also known in the case of mCRE minimization, with the parameter α [55]. Here, the physical sense of the mCRE can provide a way to tune α: α is automatically tuned before the training and constantly adapted during the training. The general philosophy of the tuning is that the model should not be updated below the noise level, considering the noise level of the measurement tools is known a priori. Hence, the field $\hat{\mathbf{u}}$ constructed at the end of the Newton scheme (14) in step 1 should get as close as possible to the measure, but should not fit the noise. To quantify this criterion, α is re-written :

$$\alpha = \alpha' \frac{1}{n_{obs}\sigma^2}$$

where n_{obs} is the number of observations and σ the standard deviation of the measurement noise. Replacing in (6) :

$$(\hat{\mathbf{u}}, \hat{\mathbf{S}})^{(n+1)} = \underset{(\mathbf{u},\mathbf{S})\in\mathcal{A}_d}{\operatorname{argmin}} \left[\int_\Omega \psi(\mathbf{u}; \mathbf{p}^{(n)}) + \psi^*(\mathbf{S}; \mathbf{p}^{(n)}) - \mathbf{S} : \mathbf{E}(\mathbf{u}) + \alpha' \frac{1}{n_{obs}\sigma^2} ||\Pi\mathbf{E}(\mathbf{u}) - \mathbf{E}_{obs}||^2 \right] \tag{23}$$

In order for $\hat{\mathbf{u}}$ not to fit the observations below the noise level, the order of magnitude of $\frac{1}{n_{obs}\sigma^2}||\Pi\mathbf{E}(\hat{\mathbf{u}}) - \mathbf{E}_{obs}||^2$ should be 1, as stated in the Morozov criterion [56]. Before the training of the neural network, several steps 1 are performed to find the proper value of α'. The greater α' the lower the data discrepancy error after step 1. This tuning is performed by dichotomy and details can be found in B. During the minimization of the mCRE, as soon as the normalized data-driven discrepancy is out of the admissible bound after step 1, α' is returned following this procedure and a new step 1 is performed with the new α'.

This simple technique is powerful to prevent the network from overfitting as the predicted field $\mathbf{u}$ cannot fit the observations below the noise level.

3.4 Reducing the sensitivity to learning rate: automatic tuning of learning rate

Another well-known sensitive parameter in neural networks training is the choice of the learning rate l_r [57]. A too small value of l_r can cause the training to be extremely slow, whereas a too large value can cause instability during the training with the risk of not reaching convergence. The value of the learning rate should be analyzed with regard to the progress made during one training step. Here, the two steps minimization procedure provides an interesting indicator on the update speed of the training. Indeed, the number of iterations performed in step 1 depends on the modification of the space of solutions that satisfy the constitutive relation for given parameters ($\mathbf{\Gamma_p}$) caused by the parameters update in step 2 (because step 1 is initialized with $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ found at the end of the previous step 1), as it is shown in Figure 5.

According to multiple experiments, a good compromise is located around 4 iterations in step 1. It is thus possible to automatically adapt the learning rate with the following empirical-based rule:

$$lr \leftarrow lr \times \texttt{update\_coefficient(number\_iterations\_step\_1)} \tag{24}$$

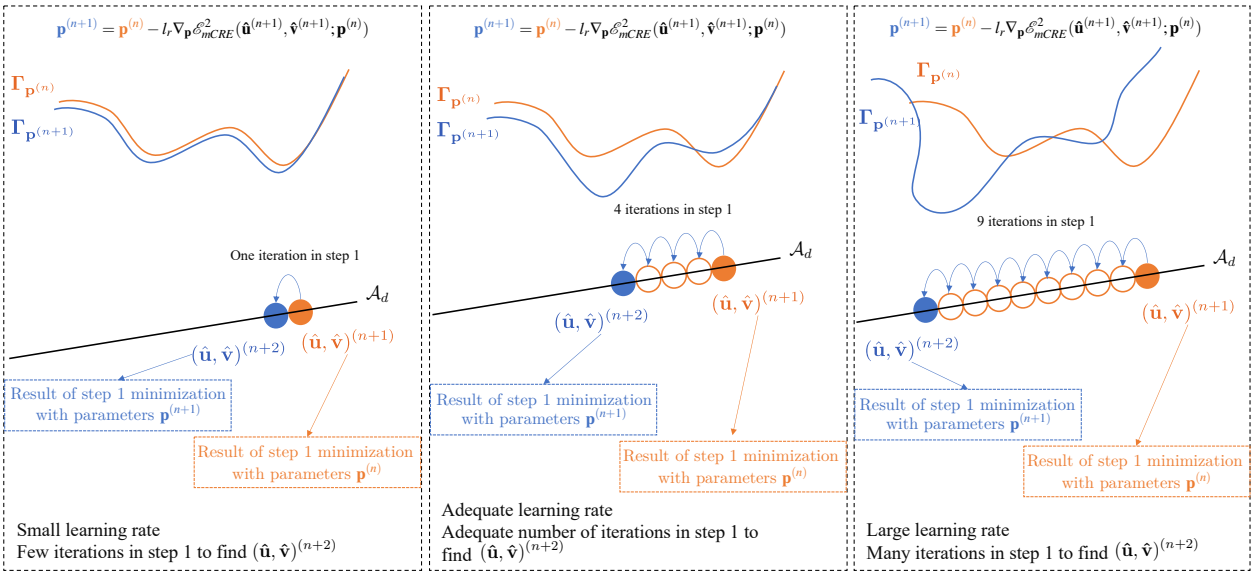

Figure 5: Influence of the learning rate on the number of iterations in step 1.

An example of the values of `update_coefficient(number_iterations_step_1)` is presented in Figure 6. These values have been obtained empirically by testing a significant amount of values on different test cases compared to the ones presented in this paper. An ongoing work is currently performed to design an optimization process that does not rely on empirically chosen coefficients. The main advantage of this rule is that the convergence of the method is no more sensitive to the user learning rate choice, as shown in Section 4.3, which enables to train the network online. Note that this rule can be adapted depending on the test case: a way to adapt this rule is to set the target number of iterations to the number of iterations needed to solve the FEM problem for the initialized network.

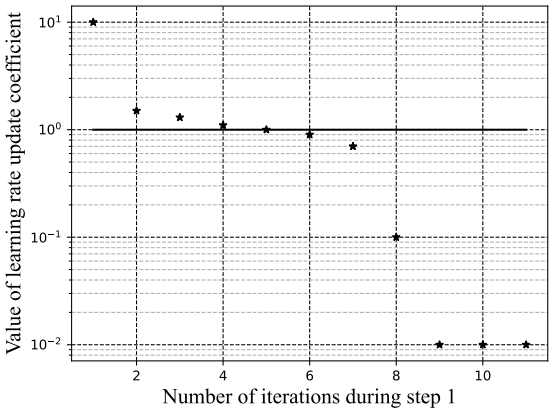

Figure 6: Example of the value of the coefficient used in the empirical learning rate update rule as function of the number of iterations during step 1.

3.5 Reducing the sensitivity to the number of epochs: definition of a physics-based stop criterion

Neural networks training requires a stop criterion. A naive technique, but yet efficient in many approaches, is to predefine a number of epochs before the training and to retune this parameter *a posteriori* as long as the predictions are not satisfactory [58]. This technique is not suited here as the training should be performed online for DDDAS. Another widely used rule is to define a criterion to present overfitting. The idea is to stop the training when the loss computed on the validation dataset is stagnating while the loss on the training dataset is still decreasing. This criterion can not work here because of the unsupervised nature of the training.

The mCRE minimization is a really interesting framework as it provides a strong physical sense that can be used to define a stop criterion. In the case of neural networks that are theoretically universal approximators, the method can be designed as if there were no model bias. Yet, to the authors' knowledge, this property has not yet been shown in the case of physics-augmented constitutive modeling. The assumption here is that a physics-augmented neural network (constructed to enforce in the architecture all the requirements of a constitutive model) can approximate any physics-consistent constitutive model (which matches the same requirements). Assuming there is no model bias, the network can represent any constitutive model that satisfies the imposed constraints. As this is further shown in Section 4.1.3, the CRE term can be interpreted as a modeling error. As such, it is perfectly suited to correct model bias. The CRE is homogeneous to an energy (see Figure 2) and can thus be compared to the energy in the structure, the normalized CRE writes :

$$\mathcal{E}^2_{\text{normalized\_}CRE} = \frac{\mathcal{E}^2_{CRE}}{\int_\Omega \psi(\mathbf{E}(\mathbf{v}))} \tag{25}$$

A stop criterion (for the overall convergence of the method) can then be established based on the relative energy error in the system:

$$\mathcal{E}^2_{\text{normalized\_}CRE} < \text{tol} \tag{26}$$

where tol is a user-defined target value of the stop criterion.

This criterion is also used to estimate how far from convergence is the minimization process: near convergence the number of iterations in step 1 is naturally lower. This can be a problem in the previous learning rate update rule, because the learning rate can be set to high values which can lead to a leap far from the minimum. To alleviate this concern, near convergence the learning rate update rule is modified so that the learning rate is only updated if it tends to decrease: the update coefficients greater than 1 are set to 1 near convergence.

3.6 Extension to the case of multiple loadings in the database

So far, the method has been presented in the case of only one loading case. In practice, the training database is constituted of multiples loading cases so the method should be modified :

- Step 1 and step 2 are consecutively performed for each loading case.
- For the stop criterion, the method stops when the criterion is reached on all the loading cases. When a loading case reaches the criterion, the weight and bias of the neural network are no more updated with this loading case. In each epoch, a verfication is made to check if the criterion is still reached for this loading case: if the criterion is not reached, the weight and bias are then updated.
- Concerning the learning rate, the same value is applied for all the loading cases. The update rule is applied after each epoch (when all the loading cases have been treated) and takes into account the mean of iterations performed in all steps 1.

4 Results

This section presents the results of the method evaluated on two test cases with synthetic data: a non-quadratic potential in the small strain regime is studied with optic fiber measurements in Section 4.1, and a Mooney-Rivlin model is searched in the hyperelastic case with digital image correlation observations in Section 4.2. Then relevance of adaptive learning rate in Section 4.3 and adaptive α in Section 4.4 are investigated, as well as noise robustness in Section 4.5.

4.1 First test case: Non-quadratic potential in the small strain regime

This part illustrates the performance of the method on a first test case, with a non-quadratic potential to be learned under the small strain regime. Section 4.1.1 and 4.1.2 respectively present the training and validation databases, Section 4.1.3 present the constitutive relation error before the training of the neural network with the mCRE procedure (*i.e.* after the initialization training), and Section 4.1.4 proposes an evaluation of the method on this test case.

4.1.1 Training database

The training database is constituted of synthetic data generated with a finite element simulation :

- constitutive relation: the chosen reference non-quadratic potential distinguishes the behavior in tension from the one in compression along the longitudinal axis (axis 1):

$$\psi(\mathfrak{e}) = \frac{1}{2}\tilde{E} < \mathfrak{e}_{11} >_+^2 + \frac{1}{2}E < \mathfrak{e}_{11} >_-^2 + \frac{1}{2}E\,\mathfrak{e}_{22}^2 + G\,\mathfrak{e}_{21}^2 \tag{27}$$

  where $< \bullet >_+$ and $< \bullet >_-$ stand respectively for the positive and negative part, $\tilde{E} = 12GPa$ and $E = 20GPa$ are the tension and compression Young moduli, $G = 8GPa$ is the Coulomb modulus and $\mathfrak{e}_{ij}$ are components of the linearized strain tensor $\mathfrak{e} = \frac{1}{2}(\nabla\mathbf{u} + (\nabla\mathbf{u})^T)$.
- geometry: the geometry considered is the 2D (plane strain) cantilever beam presented in Figure 7.
- loading: the database includes 10 different loading cases $\mathbf{f}_d^s$ with a combination of pure bending, tension and compression.

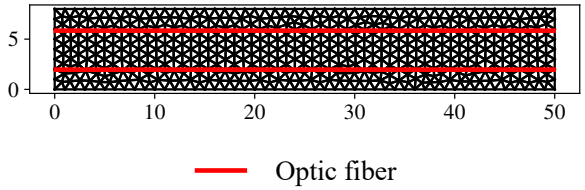

Figure 7: Geometry, mesh and position of optic fibers.

After the finite element simulation, the component of the strain in the direction of the optic fiber is saved. White Gaussian noise is then added to the synthetic measurements, with a level discussed in Section 4.5. It is worth noticing that only the noisy observations and boundary conditions are used in the mCRE minimization procedure: the analytical expression of the target potential is not used in the training but only for validation purposes. Thus this training can be qualified as unsupervised training.

4.1.2 Validation database

The validation database is constituted of 6 deformations paths: uniaxial tension (UT), uniaxial compression (UC), biaxial tension (BT), biaxial compression (BC), simple shear (SS), and pure shear (PS). Equations (28) show the deformation gradient for these 6 cases, with $\gamma \in [0, 1e^{-3}]$ a loading parameter. 100 points per loading are used in the validation database. This validation process is the same as the one used in [29]. These deformation paths are only used for validation and are not used during training. For validation, a comparison is made between the learned and the ground truth potential along these deformation paths.

$$\mathbf{F}^{UT}(\gamma) = \begin{bmatrix} 1+\gamma & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{F}^{UC}(\gamma) = \begin{bmatrix} \frac{1}{1+\gamma} & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{F}^{BT}(\gamma) = \begin{bmatrix} 1+\gamma & 0 \\ 0 & 1+\gamma \end{bmatrix},$$

$$\mathbf{F}^{BC}(\gamma) = \begin{bmatrix} \frac{1}{1+\gamma} & 0 \\ 0 & \frac{1}{1+\gamma} \end{bmatrix}, \mathbf{F}^{SS}(\gamma) = \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix}, \mathbf{F}^{PS}(\gamma) = \begin{bmatrix} 1+\gamma & 0 \\ 0 & \frac{1}{1+\gamma} \end{bmatrix} \tag{28}$$

4.1.3 Error after initialization training

Before minimizing the mCRE, the network is initialized to represent linear elasticity thanks to a classical supervised training described in Section 3.2 The potential used to generate data for the initialization training is the following :

$$\psi(\mathfrak{e}) = \frac{1}{2}E\,\mathfrak{e}_{11}^2 + \frac{1}{2}E\,\mathfrak{e}_{22}^2 + G\,\mathfrak{e}_{21}^2 \tag{29}$$

with $E = 24GPa$ the Young modulus, $G = 11GPa$ the Coulomb modulus and $\mathfrak{e}_{ij}$ the components of the linearized strain tensor.

Table 1 summarizes the parameters used for this initialization training. At the end of this training, the network is well suited to represent linear elasticity as shown in Figure 8 (top). However, along axis 1, this potential does not represent

Number of layers	4
Number of neurons per hidden layer	50
Optimizer	Adam
Number of epochs	1e5
Learning rate (cyclic l_r scheduler [59])	$l_r \in [1e^{-5}, 1e^{-4}]$, cycle of 100 epochs

Table 1: Parameters used for the initialization training.

well the non-quadratic potential: this is the goal of the mCRE minimization to finetune this neural network to accurately represent the non-quadratic potential. Figure 9 shows, for a bending loading, the fields $\mathbf{S}_{11}$ and $\mathbb{e}_{11}$ associated to displacement fields $\mathbf{u}$ and $\mathbf{v}$. To understand the role of fields $\mathbf{u}$ and $\mathbf{v}$ the reader can refer to the linear elastic case in A. Figure 9 demonstrates that the discrepancy between fields associated with $\mathbf{u}$ and $\mathbf{v}$ are localized where the $\mathbb{e}_{11}$ is positive, *i.e.* where the material is in tension along axis 1. This discrepancy is explained by the difference between the non-quadratic potential and the neural network in tension that can be observed in Figure 8. At the bottom of Figure 9 the normalized CRE map shows the localization of the error. This CRE map is a powerful tool because :

- the analytical ground truth potential is not needed to compute the error,
- an error level is available everywhere in the structure, even where there is no observation,
- it can deal with noisy and partial observations,
- this error is localized in the zone of model bias (here in the zone in tension along axis 1),

For all these reasons, the CRE is a strong indicator of model bias, and can even be used in the inference phase. In the following, results after the minimization of mCRE are presented and discussed.

4.1.4 Evaluation of the method

Figure 8 (bottom) shows the potential predicted by the neural network before and after the mCRE minimization: the model bias is corrected as the tension behavior along axis 1 is now correctly predicted. This correction of model bias is possible as the loading database contains solicitations that make appear tension along axis 1. Figure 10 shows the convergence of the mean (over the loading cases) of the mCRE and normalized CRE as such as the evolution of relative error on the validation dataset. The similarity in the evolution between mCRE and normalized CRE is explained by the fact that α is adapted during training to constraint the normalized data discrepancy term of the mCRE to be close to 1 (see Section 3.3).

About the evolution of these criteria, the plateau at the beginning of training is caused by the small value of the initial learning rate. With the adaptive learning rate rule described in Section 3.4, the learning rate gets high enough to achieve decreasing of mCRE after several epochs (here around 12 epochs).

4.2 Second test case: Mooney-Rivlin hyperelastic model

4.2.1 Training database

The training database is constituted of synthetic data generated with a finite element simulation :

- constitutive relation: Mooney-Rivlin potential [60, 61]:

$$\psi(\mathbf{C}) = A(\tilde{\mathbf{I}}_1 - 3) + B(\tilde{\mathbf{I}}_2 - 3) + C(\tilde{\mathbf{I}}_1 - 3)(\tilde{\mathbf{I}}_2 - 3) \tag{30}$$

  with

$$\begin{aligned} J &= \det(\mathbf{C}) \\ \tilde{I}_1 &= J^{-2/3}\,\mathrm{tr}(\mathbf{C}) \\ \tilde{I}_2 &= \frac{1}{2}J^{-4/3}[\mathrm{tr}(\mathbf{C})^2 - \mathrm{tr}(\mathbf{C^2})] \\ \mathbf{C} &= 2\mathbf{E} + \mathbf{I} \end{aligned}$$

  and $A = 1.9$, $B = 0.4$ and $C = 1$ some material parameters.

- geometry: the geometry considered is a 2D (plane strain) square.

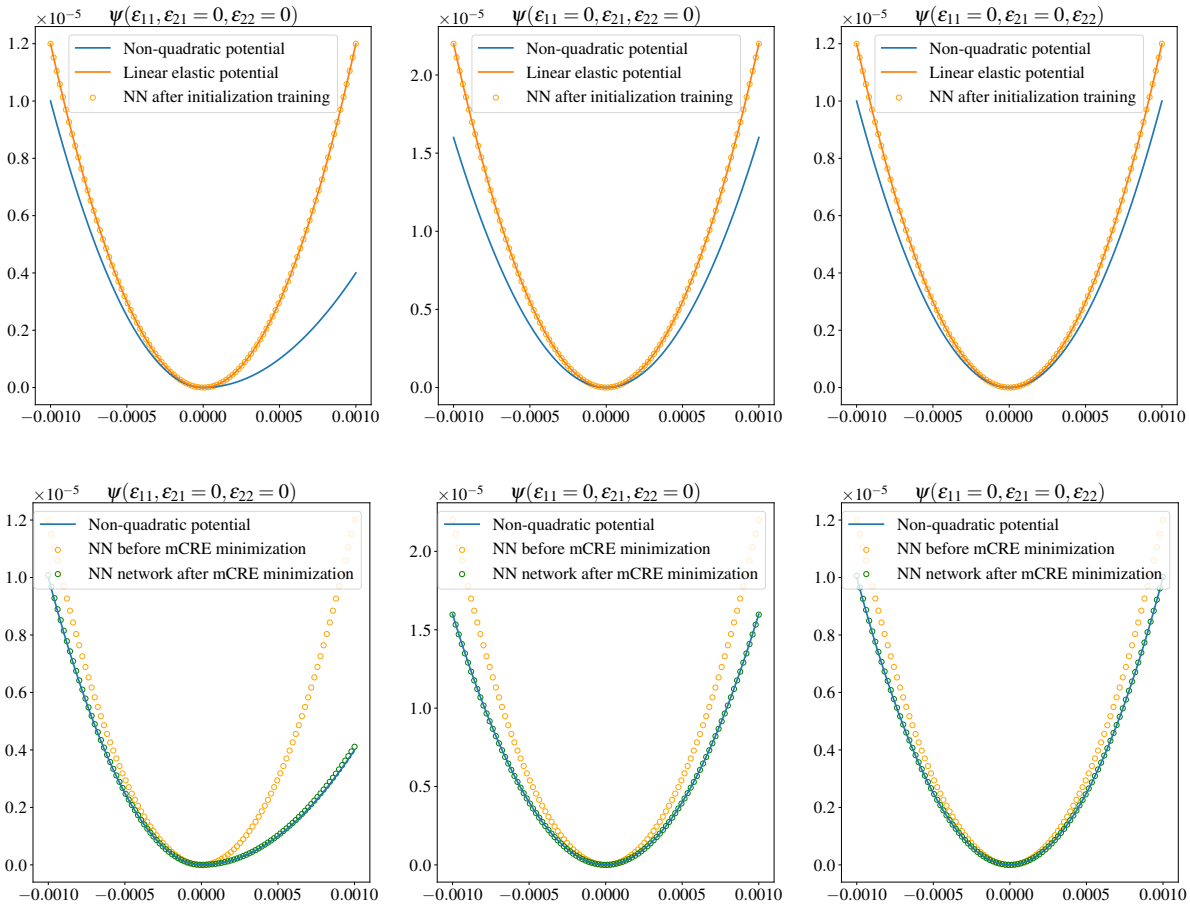

Figure 8: Top: Potentials evolution with respect to the components of the linearized strain tensor after the initialization training to represent linear elasticity. This is the prediction of the neural network before the mCRE minimization. Bottom: Potentials evolution with respect to the components of the linearized strain tensor before and after mCRE minimization.

- loading: the database includes 30 different loading cases $\mathbf{f}_d^s$ with uniaxial tension and uniaxial compression. The training database is deliberately chosen with few loading cases to quantify the generalization to out-of-training database loading cases.

After the finite element simulation, the full displacement is stored and white Gaussian noise is then added to the synthetic measurements with a noise level of $0,1\%$. Here again, only the noisy observations and boundary conditions are used in the mCRE minimization procedure: the analytical expression of the target potential is not used in the training but only for validation purposes. Even though the method presented in Section 3 is detailed in the case of strain observations, it extends easily to the case of displacement observations (as it is presented in Section 2).

4.2.2 Validation database

The validation database is constructed according to the same procedure as the one described in section 4.1.2 with the deformation paths (28). The difference is that here, the target model describes hyperelastic behavior so the range of γ should be different: $\gamma \in [0, 1]$.

4.2.3 Evaluation of the method, case of incomplete training database

For this test case in which the network has to learn a Mooney-Rivlin constitutive model, the method is evaluated when the training database does not include enough deformation paths to describe all the part of the constitutive model. The interest of this incomplete training database is to see the generalization outside the training database. Before the

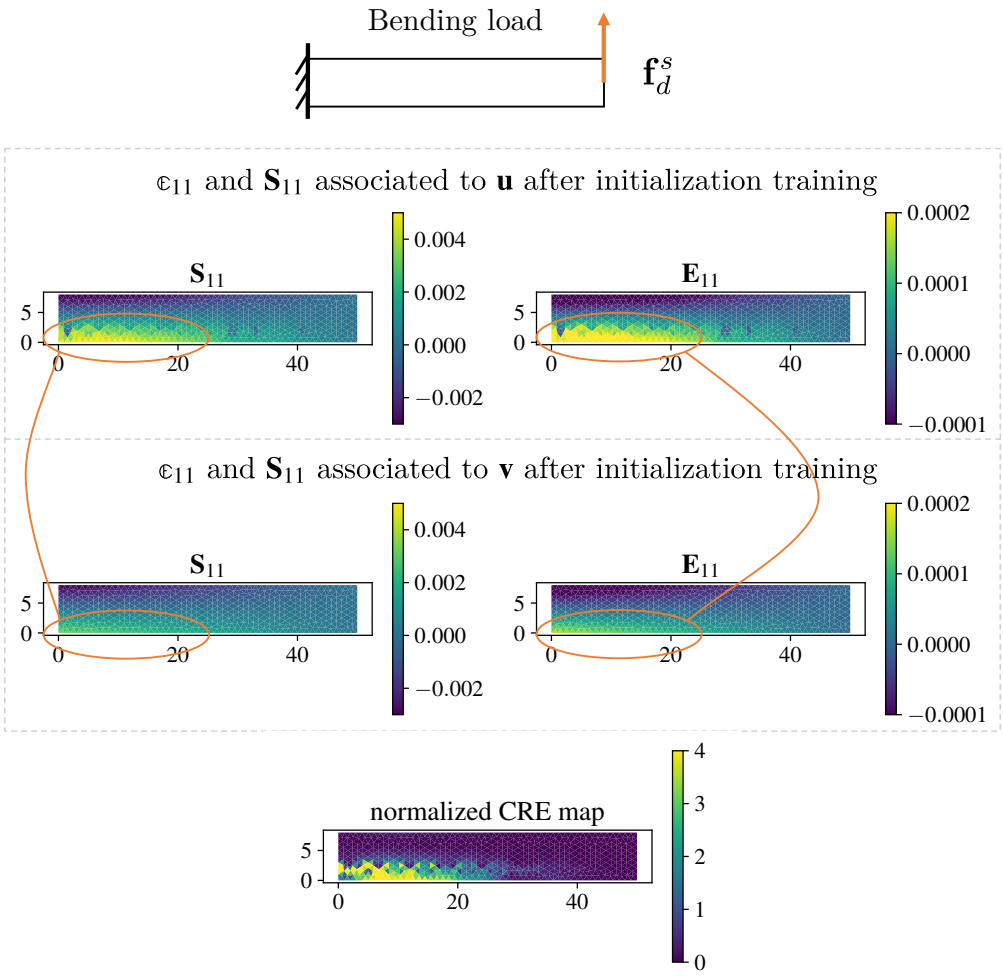

Figure 9: For a bending loading (top), fields $\mathbf{S}_{11}$ and $\mathfrak{e}_{11}$ associated with displacement fields $\mathbf{u}$ and $\mathbf{v}$ (middle) and the normalized CRE map (bottom). These observations depend on the initialization choice.

minimization of the mCRE, the network is initialized thanks to an initialization training described in Section 3.2 with the Neo-Hookean model :

$$\psi(\mathbf{E}) = A(\tilde{I}_1 - 3) + B(J-1)^2 \tag{31}$$

with $J = \det(\mathbf{C})$, $\tilde{I}_1 = J^{-2/3}\,\mathrm{tr}(\mathbf{C})$, $\mathbf{C} = 2\mathbf{E} + \mathbf{I}$, $A = 1.9$ and $B = 2.4$ are material parameters.

Figure 11 shows the potential before and after the mCRE minimization along two deformation paths. On the left, it is uniaxial compression along one axis (whereas it was the other axis in the training database) and the learned potential is close to the target. This generalization from one axis to another is explained by the fact that the isotropic invariants were put inside the network. On the right, the response to pure shear is not correctly described as this does not appear in the training database. Yet, this lack of precision is compensated by the mCRE framework that can be used to assimilate data in the online phase by the mean of the predicted fields associated with $\mathbf{u}$ (trade-off between the prediction from the trained model and assimilated data). Moreover, even in the prediction phase when data are assimilated the CRE modeling error is available to give a confidence bound on the prediction.

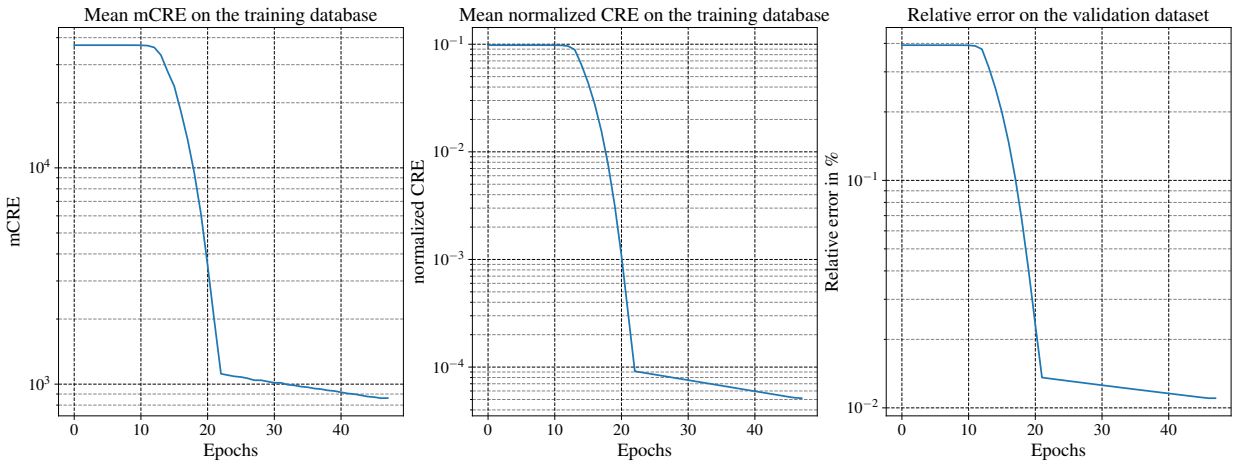

Figure 10: left: evolution of the mean (over the loading cases) mCRE during the training; center: evolution of the mean (over the loading cases) normalized CRE during the training, right: evolution of the relative error on the validation database during training.

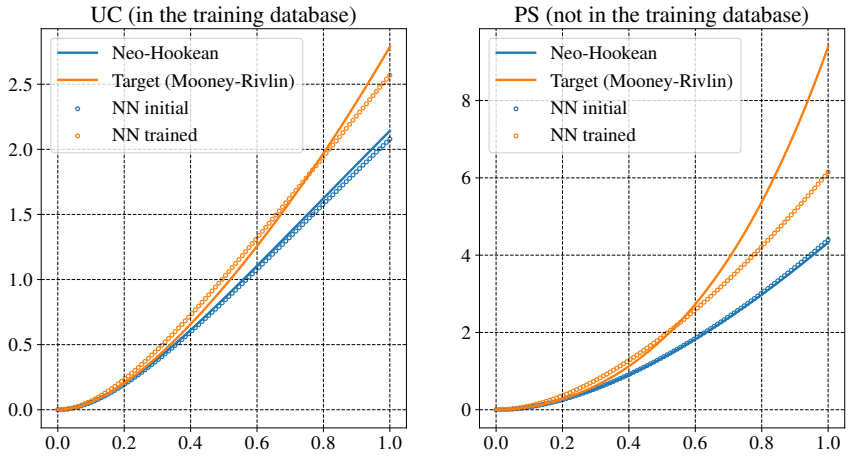

Figure 11: Potential before and after the mCRE minimization along two deformation paths: on the left is uniaxial compression and on the right is pure shear.

4.3 Relevance of adaptive learning rate

The learning rate parameter is a crucial parameter for the convergence of deep learning training: Figure 12 (left) shows the evolution of the relative error on the validation dataset during the training for several different values of l_r. This figure highlights the sensitivity of this parameter on good convergence. In most deep learning use cases, this parameter is tuned by the user until the results are satisfactory. Here, the DDDAS context [2] imposes that the training has to be performed online, thus imposing an automatic tuning of this parameter. Figure 12 (middle) shows the evolution of the relative error on the validation dataset for different initial learning rates. This shows that the adaptive learning rate rule drastically reduces the sensitivity of the choice of the learning rate as very different initial learning rates achieve close convergence whereas it was not the case without the adaptive learning rate. Figure 12 (right) demonstrates that the learning rates all stabilize to a close value, even with an important difference in the initial learning rate choice. This is the case only when the initial learning rate is small: if the initial learning rate is too large, the first epoch will update the network far from the initialization with *a priori* knowledge, so it is better to set a low initial value. Indeed, even if there is a high difference between the low initial learning rate and the proper one, the learning rate will adapt only in a few iterations.

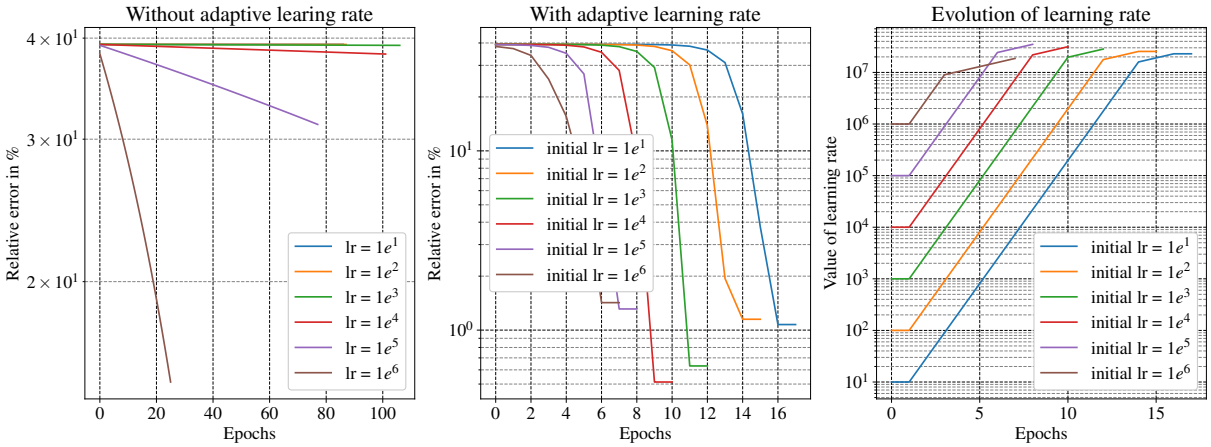

Figure 12: Left and middle: evolution of the relative error on the validation dataset (left: sensitivity of the fixed learning rate without adaptive learning rate, middle: sensitivity of the initial learning rate when the adaptive rule is applied). Right: evolution of the learning rate when the adaptive rule is applied.

4.4 Relevance of adaptive weighting between losses

This part presents the interest of the adaptive tuning of α. First, Figure 13 helps to understand the role played by α. It shows the strain and stress fields associated with displacement fields $\mathbf{u}$ at the end of step 1 for different values of α. Equation (35) shows the role of α in the linear elastic case: $\mathbf{U}$ is a compromise between the observed data and the model for the current parameters, and α sets up the balance between these two terms. A similar interpretation can be done in the case of non-quadratic potentials as shown in Figure 13:

- For a small value of α (here when $\alpha = 1$), $\mathbf{E}_{11}$ and $\mathbf{S}_{11}$ are really closes to the one obtained with $\mathbf{v}$ (no compromise with data). The associated CRE map shows there is no error, which means the value of α is too low.
- On the opposite, $\alpha = 10\ 000$ shows an important difference between $\mathbf{u}$ and $\mathbf{v}$ but the solution associated with $\mathbf{u}$ seems noisy.
- In this case a good choice of α is between 100 and 10 000: the automatic tuning gives a value of $\alpha = 2\ 000$.

Figure 14 (left) illustrates how the convergence is affected by different values of α. This shows that convergence is sensitive to the chosen values of α, which motivates the need for adaptive tuning. The right-hand side of Figure 14 shows the convergence for the same initial values of α. As α is always chosen such that the $\mathbf{u}$ field fits the observations to the noise level, the convergence is the same for these 3 initial values of α.

Nevertheless, this adaptive tuning comes with a high computational cost: when $\mathbf{u}$ does not fit the observations to noise level, multiple steps 1 have to be performed in the dichotomy to find the proper value of α. This constitutes an important drawback of the method and should be further investigated.

4.5 Robustness to noise

The mCRE for parameter identification has already been shown to be very effective in the presence of noise measurement [20, 62]. This article confirms the previous observations in the case of neural network training. Figure 15 (left) shows the evolution relative error on the validation dataset for different noise levels: even with high noise levels, the relative error on the validation dataset is acceptable. Figure 15 (right) shows the potential learned with high noise. For the 40% noise case, the potential is not representative of reality but is still close and better than before the training. This is achievable thanks to the adaptive α tuning which requires the knowledge of the noise level.

5 Conclusion

This research paper introduced a novel approach to train physics-augmented neural networks for constitutive law representation using partial strain or displacement measurements and boundary conditions. All reliable information

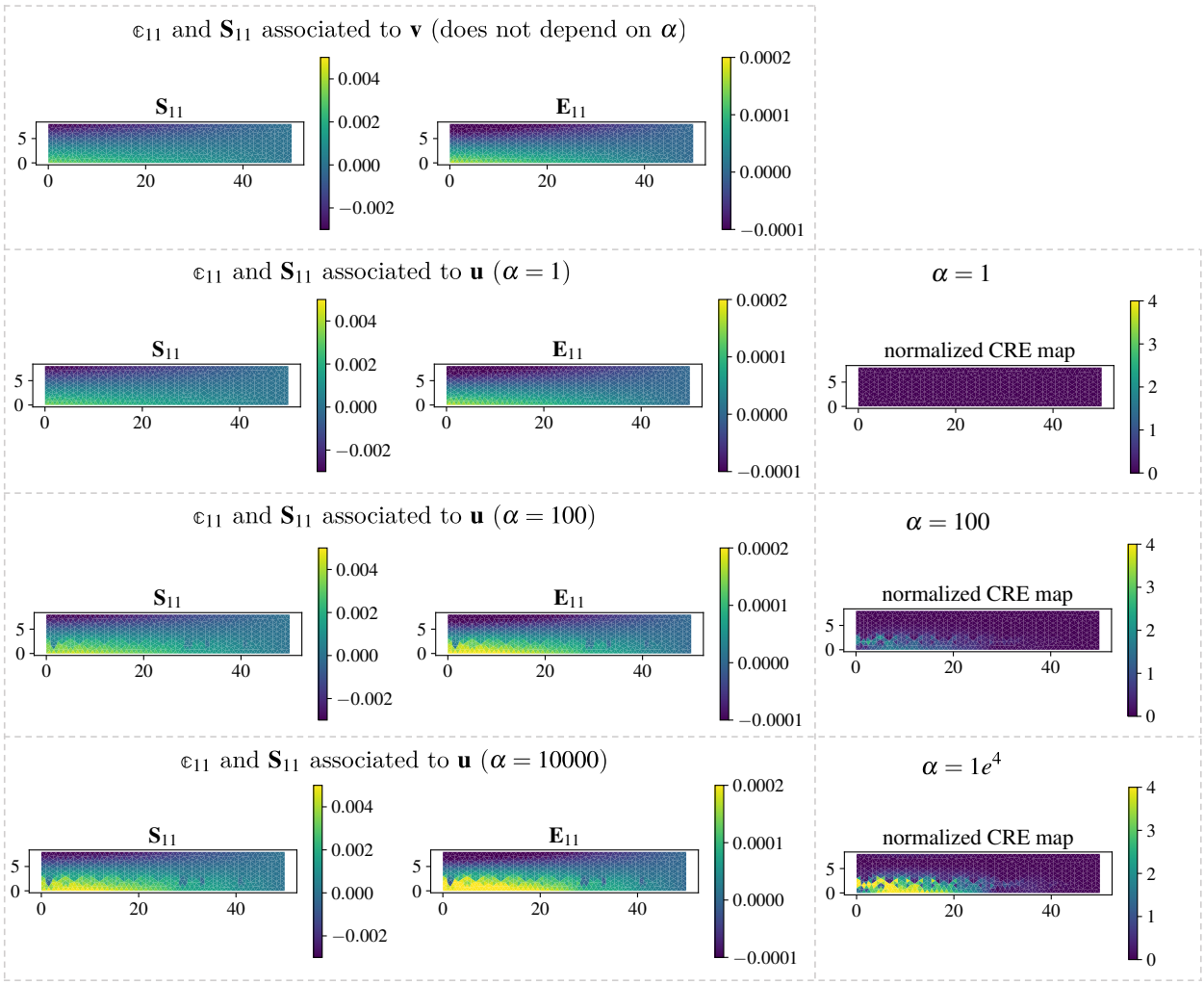

Figure 13: Influence of α on the strain and stress fields associated to displacement field $\mathbf{u}$ at the end of step 1 (see section 2.4.1) and on the associated CRE map.

such as thermodynamics or equilibrium properties was enforced either in the network or in the minimization procedure. The unsupervised training procedure was based on the modified Constitutive Relation Error (mCRE), a widely studied functional in the parameters identification field. It provided a strong physical sense as the CRE can be interpreted as a modeling error everywhere in the structure.

The automatic hyperparameter tuning is another significant feature of the proposed method, which is particularly relevant for online training in the Dynamic Data Driven Application Systems (DDDAS) paradigm. The evaluation of the method has shown to be effective in producing accurate models in the case of nonlinear state laws, excellent noise robustness and low sensitivity to user-defined hyperparameters. Overall, this proposed method offers a potentially valuable tool for predicting the behavior of materials and structures under external loadings, thus opening up new avenues for research in the field of computational mechanics.

Yet, compared to classical supervised training, this method has a relatively high computational cost. This is caused by the need to construct admissible displacement and stress fields before each gradient descent step. The method was introduced in 2D, and even though it can be easily generalized to 3D, the 3D computational cost seems prohibitive (the total training time is presently around 2 hours in 2D). A more efficient automatic tuning of weighting between losses, or extending previous works on the coupling between reduced order modeling techniques and mCRE [63], could be a research direction to alleviate this concern. To efficiently update the model and predict the quantities of interest, the present work could be coupled with previous work on mCRE to perform real-time online SHM with Kalman filtering [49, 64] and integrate ideas from [65] in which some layers are frozen during the online training.

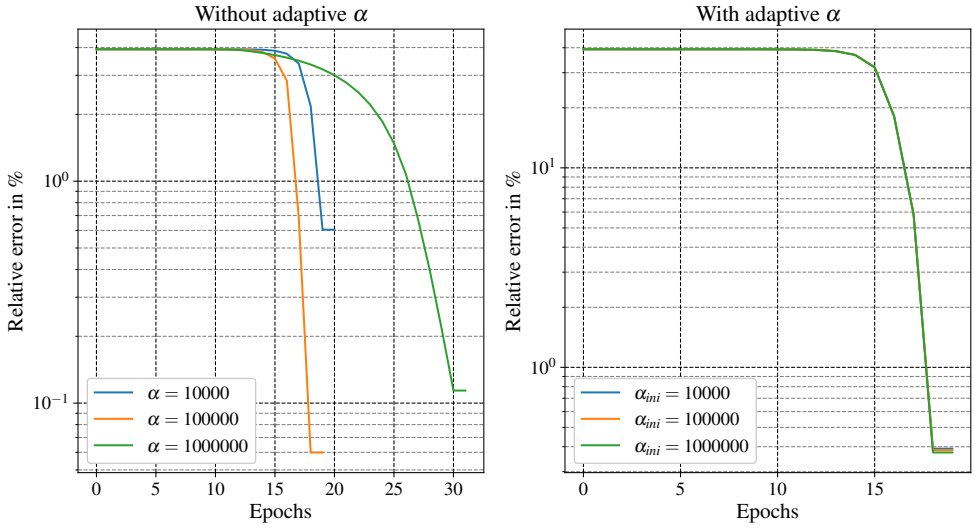

Figure 14: Evolution of the relative error on the validation dataset. Left: sensitivity of the parameters α without adaptive rule. Right: sensitivity of the initial α rate when the adaptive rule is applied.

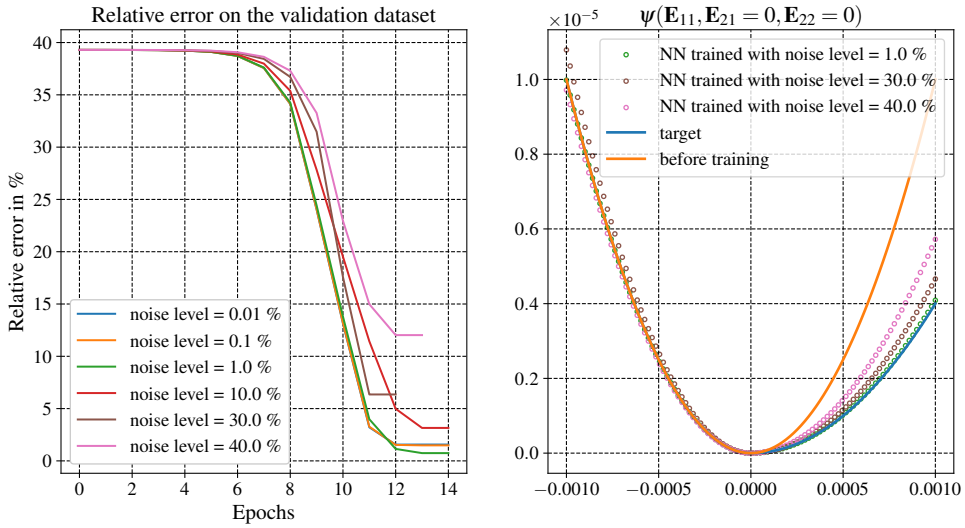

Figure 15: Influence of the noise level on the quality of the training. Left: evolution of the relative error on the validation set, right: $\psi(\mathbf{E}_{11}, \mathbf{E}_{21} = 0, \mathbf{E}_{22} = 0)$ for different noise levels.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101002857).

References

[1] E. P. Blasch, F. Darema, D. Bernstein, Introduction to the Dynamic Data Driven Applications Systems (DDDAS) Paradigm, Springer International Publishing, Cham, 2022, pp. 1–32.

[2] L. Chamoin, DREAM-ON: Merging advanced sensing techniques and simulation tools for future structural health monitoring technologies, in: The Project Repository Journal, Vol. 10, 2021, pp. 124–127. URL https://hal.archives-ouvertes.fr/hal-03304265

[3] L. Chamoin, S. Farahbakhsh, M. Poncelet, An educational review on distributed optic fiber sensing based on Rayleigh backscattering for damage tracking and structural health monitoring, Measurement Science and Technology, 2022, 24–25. URL https://hal.archives-ouvertes.fr/hal-03701630

[4] H. N. Nguyen, New numerical strategies for robust, consistent, and computationally efficient model identification from full-field measurements, PhD Thesis, Université Paris-Saclay, May 2021. URL https://tel.archives-ouvertes.fr/tel-03260879

[5] P. Liu, Damage Modeling of Composite Structures, Elsevier, 2021, pp. 377–384. URL https://www.sciencedirect.com/science/article/pii/B9780128209639099946.

[6] M. Bonnet, A. Constantinescu, Inverse problems in elasticity, Inverse Problems 21 (2), 2005. URL https://dx.doi.org/10.1088/0266-5611/21/2/R01.

[7] M. Grédiac, The use of full-field measurement methods in composite material characterization: interest and limitations, Composites Part A: Applied Science and Manufacturing, 2004. doi:0.1016/j.compositesa.2004.01.019.

[8] D. Claire, F. Hild, S. Roux, A finite element formulation to identify damage fields: the equilibrium gap method, International Journal for Numerical Methods in Engineering, 61:189–208, 2004. URL https://doi.org/10.1002/nme.1057.

[9] M. Grédiac, E. Toussaint, F. Pierron, Special virtual fields for the direct determination of material parameters with the virtual fields method. International Journal of Solids and Structures, 39(10):2691–2705, 2002. URL https://doi.org/10.1016/S0020-7683(02)00127-0.

[10] S. Andrieux, A. Abda, H. Bui, Reciprocity principle and crack identification. Inverse Problems, 15(1):59–65, 1999. URL https://dx.doi.org/10.1088/0266-5611/15/1/010.

[11] N. Maia, M. Reynier, P. Ladevèze, Error localization for updating finite element models using frequency-response-functions, 1994. URL https://api.semanticscholar.org/CorpusID:18857921.

[12] T. Kirchdoerfer, M. Ortiz, Data-driven computational mechanics, Computer Methods in Applied Mechanics and Engineering, 304:81–101, 2016. URL https://doi.org/10.1016/j.cma.2016.02.001.

[13] P. Ladevèze, A. Chouaki, Application of a posteriori error estimation for structural model updating, Inverse Problems 15, 49, 1999. URL https://doi.org/10.1088/0266-5611/15/1/009.

[14] B. Banerjee, T. F. Walsh, W. Aquino, M. Bonnet, Large scale parameter estimation problems in frequency-domain elastodynamics using an error in constitutive equation functional, Computer Methods in Applied Mechanics and Engineering 253, 60–72, 2013. URL https://www.sciencedirect.com/science/article/pii/S0045782512002770.

[15] A. Deraemaeker, P. Ladevèze, T. Romeuf, Model validation in the presence of uncertain experimental data, Engineering Computations 21, 808–833, 2004. URL ttps://doi.org/10.1108/02644400410554335.

[16] O. Allix, P. Feissel, P. Thévenet, A delay damage mesomodel of laminates under dynamic loading: Basic aspects and identification issues, Computers & Structures 81, 1177–1191, 2003. URL https://doi.org/10.1016/S0045-7949(03)00035-X

[17] M. Bonnet, W. Aquino, Three-dimensional transient elastodynamic inversion using an error in constitutive relation functional, Inverse Problems 31 (3), 035010, 2015. URL https://dx.doi.org/10.1088/0266-5611/31/3/035010

[18] B. Marchand, L. Chamoin, C. Rey, Parameter identification and model updating in the context of nonlinear mechanical behaviors using a unified formulation of the modified constitutive relation error concept, Computer Methods in Applied Mechanics and Engineering 345, 1094–1113, 2019. URL https://www.sciencedirect.com/science/article/pii/S0045782518304511

[19] H. Nguyen, L. Chamoin, C. Minh, mcre-based parameter identification from full-field measurements: Consistent framework, integrated version, and extension to nonlinear material behaviors, Computer Methods in Applied Mechanics and Engineering 400, 115461, 2022. doi:10.1016/j.cma.2022.115461.

[20] O. Allix, P. Feissel, H. M. Nguyen, Identification strategy in the presence of corrupted measurements, Engineering Computations 22 (5/6), 487–504, 2005. URL https://doi.org/10.1108/02644400510602989

[21] S. Huang, P. Feissel, P. Villon, Modified constitutive relation error: An identification framework dealing with the reliability of information, Computer Methods in Applied Mechanics and Engineering 311, 1–17, 2016. URL https://doi.org/https://doi.org/10.1016/j.cma.2016.06.030.

[22] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (5), 359–366, 1989. URL https://doi.org/https://doi.org/10.1016/0893-6080(89)90020-8.

[23] J. Ghaboussi, X. Wu, G. K. PhD, Neural network material modelling, Statyba 5 (4) 250–257, 1999. URL https://doi.org/10.1080/13921525.1999.10531472

[24] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nature Reviews Physics 3 (6), 2021. URL https://doi.org/10.1038/s42254-021-00314-5.

[25] J. Willard, X. Jia, S. Xu, M. Steinbach, V. Kumar, Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. ACM Comput. Surv. 55, 4, Article 66, 2023. URL https://doi.org/10.1145/3514228.

[26] 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, Journal of Computational Physics 378, 686–707, 2019. URL https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045.

[27] D. Klein, M. Fernández, R. Martin, P. Neff, O. Weeger,Polyconvex anisotropic hyperelasticity with neural networks, Journal of the Mechanics and Physics of Solids, Volume 159, 2022. URL https://doi.org/10.1016/j.jmps.2021.104703.

[28] F. As'ad, P. Avery, C. Farhat, A mechanics-informed artificial neural network approach in data-driven constitutive modeling, International Journal for Numerical Methods in Engineering 123, 2022. URL https://doi.org/10.1002/nme.6957.

[29] P. Thakolkaran, A. Joshi, Y. Zheng, M. Flaschel, L. De Lorenzis, S. Kumar, NN-euclid: Deep-learning hyperelasticity without stress data, Journal of the Mechanics and Physics of Solids 169, 2022. URL https://doi.org/10.1016/j.jmps.2022.105076.

[30] J. N. Fuhg, C. M. Hamel, K. Johnson, R. Jones, N. Bouklas, Modular machine learning-based elastoplasticity: generalization in the context of limited data, 2022. URL https://doi.org/10.48550/ARXIV.2210.08343.

[31] S. Huang, Z. He, R. Chem, C. Reina, Variational Onsager Neural Networks (VONNs): A thermodynamics-based variational learning strategy for non-equilibrium PDEs, Journal of the Mechanics and Physics of Solids, Volume 163, 2022, 104856, URL https://doi.org/10.1016/j.jmps.2022.104856.

[32] B. Amos, L. Xu, J. Kolter, Input convex neural networks, Proceedings of the 34th International Conference on Machine Learning, 70:146-155, 2017. URL https://proceedings.mlr.press/v70/amos17b.html.

[33] N. N. Vlassis, W. Sun, Sobolev training of thermodynamic-informed neural networks for interpretable elasto-plasticity models with level set hardening, Computer Methods in Applied Mechanics and Engineering 377, 113695, 2021. URL https://doi.org/https://doi.org/10.1016/j.cma.2021.113695.

[34] F. Masi, I. Stefanou, Multiscale modeling of inelastic materials with thermodynamics-based artificial neural networks (tann), Computer Methods in Applied Mechanics and Engineering 398, 115190, 2022. URL https://doi.org/https://doi.org/10.1016/j.cma.2022.115190.

[35] C. Bonatti, D. Mohr, One for all: Universal material model based on minimal state-space neural networks, Science Advances 7 (26), eabf3658, 2021. URL https://www.science.org/doi/abs/10.1126/sciadv.abf3658

[36] M. Fernández, M. Jamshidian, T. Böhlke, K. Kersting, O. Weeger, Anisotropic hyperelastic constitutive models for finite deformations combining material theory and data-driven approaches with application to cubic lattice metamaterials, Computational Mechanics 67 (2), 653–677, 2021.

[37] C. Bonatti, B. Berisha, D. Mohr, From CP-FFT to CP-RNN: Recurrent neural network surrogate model of crystal plasticity, International Journal of Plasticity, Volume 158, 2022. URL https://doi.org/10.1016/j.ijplas.2022.103430.

[38] O. Ibragimova, A. Brahme, W. Muhammad, J. Lévesque, K. Inal, A new ANN based crystal plasticity model for FCC materials and its application to non-monotonic strain paths. International Journal of Plasticity, 2021. URL https://doi.org/10.1016/j.ijplas.2021.103059.

[39] D. Z. Huang, K. Xu, C. Farhat, E. Darve, Learning constitutive relations from indirect observations using deep neural networks, Journal of Computational Physics 416, 109491, 2020. URL https://doi.org/https://doi.org/10.1016/j.jcp.2020.109491.

[40] E. Zembra, A. Benady, E. Baranger, L. Chamoin. Use of physics-augmented neural networks for unsupervised learning of material constitutive relations - Comparison of the NN-Euclid and NN-mCRE methods. ENS Paris-Saclay; Centrale Supélec. 2023. URL https://sciencespo.hal.science/LMT/hal-04255767v1.

[41] L. Linden, D. Klein, K. A. Kalina, J. Brummund, O. Weeger, M. Kästner, Neural networks meet hyperelasticity: A guide to enforcing physics, 2023. URL https://doi.org/10.48550/arXiv.2302.02403.

[42] M. I. Diaz, W. Aquino, M. Bonnet, A modified error in constitutive equation approach for frequency-domain viscoelasticity imaging using interior data, Computer Methods in Applied Mechanics and Engineering, 296, 129–149, 2015. URL https://doi.org/10.1016/j.cma.2015.07.025.

[43] P. Ladevèze, D. Leguillon, Error estimate procedure in the finite element method and applications, Siam Journal on Numerical Analysis 20, 485–509, 1983. URL https://doi.org/10.1137/0720033.

[44] P. Ladevèze, J.-P. Pelle, Mastering calculation in linear and nonlinear mechanics, 2004. URL https://doi.org/10.1007/b138705.

[45] B. D. Kohn, Robert V Lowe, A variational method for parameter identification, ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique 22 (1), 119–158, 1988. URL http://eudml.org/doc/193521

[46] A. Chouaki, P. Ladevèze, L. Proslier, An updating of structural dynamic model with damping, Inverse Problems in Engineering: Theory and Practice, 335 – 342, 1996. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-0344991205&partnerID=40&md5=1fb55ca60b254b397f6e1a1ec3db8185

[47] H. Bui, A. Constantinescu, Spatial localization of the error of constitutive law for the identification of defects in elastic bodies, Archives of Mechanics 52 (4-5), 511 – 522, 2020. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-0011733386&partnerID=40&md5=c529de354d46520a9091b5a22df71c14.

[48] P. Ladevèze, G. Puel, A. Deraemaeker, T. Romeuf, Validation of structural dynamics models containing uncertainties, Computer Methods in Applied Mechanics and Engineering 195 (4) 373–393, Adaptive Modeling and Simulation, 2006. URL https://doi.org/https://doi.org/10.1016/j.cma.2004.10.011.

[49] B. Marchand, L. Chamoin, C. Rey, Real-time updating of structural mechanics models using kalman filtering, modified constitutive relation error, and proper generalized decomposition, International Journal for Numerical Methods in Engineering 107 (9) 786–810, 2016. URL https://doi.org/https://doi.org/10.1002/nme.5197.

[50] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge university press, 2004. URL https://doi.org/doi:10.1017/CBO9780511804441.

[51] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, 249–256, 2010. URL https://proceedings.mlr.press/v9/glorot10a.html.

[52] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, CoRR, 2015. URL https://api.semanticscholar.org/CorpusID:6628106

[53] R. Bischof, M. Kraus, Multi-objective loss balancing for physics-informed deep learning, 2021. URL https://doi.org/10.13140/RG.2.2.20057.24169.

[54] S. Wang, Y. Teng, P. Perdikaris, Understanding and mitigating gradient pathologies in physics-informed neural networks, SIAM J. Sci. Comput. 43, A3055–A3081, 2020.

[55] M. Diaz, P.-E. Charbonnel, L. Chamoin, Robust energy-based model updating framework for random processes in dynamics: Application to shaking-table experiments, Computers and Structures 264, 2022. URL https://doi.org/10.1016/j.compstruc.2022.106746.

[56] V. Morozov, The error principle in the solution of operational equations by the regularization method, USSR Computational Mathematics and Mathematical Physics 8 (2), 63–87, 1968. URL https://doi.org/https://doi.org/10.1016/0041-5553(68)90034-7.

[57] R. A. Jacobs, Increased rates of convergence through learning rate adaptation, Neural Networks 1 (4), 295–307, 1988. URL https://doi.org/https://doi.org/10.1016/0893-6080(88)90003-2.

[58] I. J. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, Cambridge, MA, USA, 2016, URL http://www.deeplearningbook.org.

[59] L. N. Smith, Cyclical learning rates for training neural networks, 464-472. 10.1109/WACV.2017.58, 2017. URL http://arxiv.org/abs/1506.01186 arXiv:1506.01186.

[60] M. Mooney, A Theory of Large Elastic Deformation, Journal of Applied Physics 11 (9), 582–592, 1940. URL https://doi.org/10.1063/1.1712836

[61] R. S. Rivlin, Large elastic deformations of isotropic materials. iv. further developments of the general theory, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences 241 (835), 379–397, 1948. URL http://www.jstor.org/stable/91391

[62] P. Feissel, O. Allix, Modified constitutive relation error identification strategy for transient dynamics with corrupted data: The elastic case, Computer Methods in Applied Mechanics and Engineering 196, 1968–1983, 2007. URL https://doi.org/10.1016/j.cma.2006.10.005.

[63] L. Chamoin, P.-E. Allier, B. Marchand, Synergies between the constitutive relation error concept and pgd model reduction for simplified v&v procedures, Advanced Modeling and Simulation in Engineering Sciences 3, 2016. URL https://doi.org/10.1186/s40323-016-0073-9.

[64] M. Diaz, P.-E. Charbonnel, L. Chamoin, A new Kalman filter approach for structural parameter tracking: application to the monitoring of damaging structures tested on shaking-tables. Mechanical Systems and Signal Processing, 2023, 182, pp.109529. URL https://doi.org/10.1016/j.ymssp.2022.109529.

[65] Moya, B., Badias, A., Gonzalez, D., Chinesta, F., Cueto, E. A thermodynamics-informed active learning approach to perception and reasoning about fluids. Computational Mechanics, 1-15. (2023) URL https://doi.org/10.1007/s00466-023-02279-x.

A Minimization of the mCRE with quadratic potential

This appendix aims to demonstrate the minimization of mCRE for a quadratic potential case and displacement observations. In this case, the minimization proposed in Section 2.4 gives the same procedure as previous works on mCRE [4]. In this case, the mCRE writes:

$$\mathcal{E}^2_{mCRE}(\hat{\mathbf{u}},\hat{\mathbf{S}}) = \int_\Omega (\hat{\mathbf{S}} - \mathbf{K}\mathbf{E}(\hat{\mathbf{u}})) : \mathbf{K}^{-1} : (\hat{\mathbf{S}} - \mathbf{K}\mathbf{E}(\hat{\mathbf{u}})) + \alpha||\Pi\hat{\mathbf{u}} - \mathbf{u}_{obs}||^2 = ||\hat{\mathbf{S}} - \mathbf{K}\mathbf{E}(\hat{\mathbf{u}})||^2_{\mathbf{K}^{-1}} + \alpha||\Pi\hat{\mathbf{u}} - \mathbf{u}_{obs}||^2 \tag{32}$$

To perform the minimization of step 1, under the static admissibility constraint, the following discretized Lagrangian is introduced :

$$\mathcal{L}^h(\mathbf{U},\mathbf{V},\mathbf{\Lambda}) = \frac{1}{2}(\mathbf{U}-\mathbf{V})^T\mathbb{K}(\mathbf{p})(\mathbf{U}-\mathbf{V}) + \alpha(\Pi\mathbf{U}-\mathbf{U}_{obs})^T(\Pi\mathbf{U}-\mathbf{U}_{obs}) - \mathbf{\Lambda}^T(\mathbb{K}(\mathbf{p})\mathbf{V}-\mathbf{F}) \tag{33}$$

where the capital letters denote the discretized associated field, $\mathbb{K}$ and $\mathbf{F}$ are the global stiffness matrix and global load vector. $\mathbf{V}$ is the discretized vector of the field $\mathbf{v}$ obtained from duality such that $\mathbf{S} = \mathbf{K}(\mathbf{p})\mathbf{E}(\mathbf{v})$ (see [11]). To impose the kinematic admissibility, with finite element discretization, only the free degrees of freedom (dof) are searched for, and the prescribed dofs are imposed in the search space.

The search for the stationarity of the Lagrangian leads to the following system :

$$\begin{cases} \mathbf{0} = \dfrac{\partial\mathcal{L}^h}{\partial\mathbf{U}} & = \mathbb{K}(\mathbf{p})(\mathbf{U}-\mathbf{V}) + \alpha\Pi^T(\Pi\mathbf{U}-\mathbf{U}_{obs}) \\ \mathbf{0} = \dfrac{\partial\mathcal{L}^h}{\partial\mathbf{V}} & = \mathbb{K}(\mathbf{p})(\mathbf{V}-\mathbf{U}) - \mathbb{K}(\mathbf{p})\mathbf{\Lambda} \\ \mathbf{0} = \dfrac{\partial\mathcal{L}^h}{\partial\mathbf{\Lambda}} & = \mathbb{K}(\mathbf{p})\mathbf{V} - \mathbf{F} \end{cases} \tag{34}$$

From $\frac{\partial\mathcal{L}^h}{\partial\mathbf{V}} = \mathbf{0}$ it comes $\mathbf{\Lambda} = \mathbf{V} - \mathbf{U}$, and (34) eventually yields :

$$\begin{cases} \mathbb{K}(\mathbf{p})\mathbf{V} = \mathbf{F} \\ (\mathbb{K}(\mathbf{p}) + \alpha\Pi^T\Pi)\mathbf{U} = \mathbf{F} + \alpha\Pi^T(\mathbf{U}_{obs} - \Pi\mathbf{U}) \end{cases} \tag{35}$$

Step 2 consists of the gradient descent step described in (17). The gradient of the mCRE with respect to the parameters $\mathbf{p}$ is computed with the adjoint-state method. With $(\hat{\mathbf{U}},\hat{\mathbf{V}})$ obtained at the end of step 1, the gradient writes :

$$\frac{d\mathcal{E}^2_{mCRE}(\hat{\mathbf{U}},\hat{\mathbf{V}};\mathbf{p})}{d\mathbf{p}} = \frac{d}{d\mathbf{p}}\mathcal{L}^h(\hat{\mathbf{U}},\hat{\mathbf{V}},\mathbf{\Lambda};\mathbf{p}) = \frac{1}{2}(\hat{\mathbf{U}}-\hat{\mathbf{V}})^T\frac{\partial\mathbb{K}(\mathbf{p})}{\partial\mathbf{p}}(\hat{\mathbf{U}}+\hat{\mathbf{V}}) \tag{36}$$

B Dichotomy algorithm for the update of weighting between losses

Algorithm 1: Pretraining: dichotomy for automatic tuning of α

Function $f(\alpha)$ **is**
  Find $(\hat{\mathbf{u}}, \hat{\mathbf{S}}) = \min_{(\mathbf{u},\mathbf{S})\in\mathcal{A}_d} \left[\int_\Omega \left(\psi(\mathbf{u};\mathbf{p}) + \psi^*(\mathbf{S};\mathbf{p}) - \mathbf{S} : \mathbf{E}(\mathbf{u}) \right) + \alpha \frac{1}{n_{obs}\sigma^2} ||\Pi\mathbf{E}(\mathbf{u}) - \mathbf{E}_{obs}||^2 \right]$ with the procedure described in Section 2.4.1.
  return $\frac{1}{n_{obs}\sigma^2} ||\Pi\mathbf{E}(\mathbf{u}) - \mathbf{E}_{obs}||^2 - 1$
end
Input: α_{min}, α_{max}
$f_{\alpha_a} \leftarrow f(\alpha_a)$
$f_{\alpha_b} \leftarrow f(\alpha_b)$
$\alpha_m \leftarrow (\alpha_{min} + \alpha_{max})/2$;
while $f(\alpha_m) > 0.1$ **do**
  $f_{\alpha_m} \leftarrow f(\alpha_m)$;
  if $f_{\alpha_m} \cdot f_{\alpha_a} < 0$ **then**
    $\alpha_b \leftarrow \alpha_m$
    $f_{\alpha_b} \leftarrow f_{\alpha_m}$
  end
  else
    $\alpha_a \leftarrow \alpha_m$
    $f_{\alpha_a} \leftarrow f_{\alpha_m}$
  end
  $\alpha_m \leftarrow (\alpha_{min} + \alpha_{max})/2$
end
return α_m;