Artificial intelligence powered data-driven method for Shape Memory Alloy behavior modeling

Wei Yan, Tarak Ben Zineb, Hamid Zahrouni, Heng Hu

CSMA 2024
16ème Colloque National en Calcul des Structures
13-17 Mai 2024, Presqu'île de Giens (Var)

Artificial intelligence powered data-driven method for Shape Memory Alloy behavior modeling

W. Yan[1,2], T. Ben Zineb[2], H. Zahrouni[1],H. Hu[3]

[1] *Université de Lorraine, CNRS, Arts et Métiers ParisTech, LEM3, F-57000 Metz, France*
[2] *Université de Lorraine, CNRS, Arts et Métiers ParisTech, LEM3, F-54000 Nancy, France*
[3] *School of Mathematics and Statistics, Ningxia University, Yinchuan, 750021, Ningxia, PR China*

Résumé — In this paper, we propose an artificial intelligence (AI) powered data-driven computing method (DDCM) for shape memory alloy (SMA) behavior modeling. DDCM can directly use material behavior data to simulate thermomechanical behavior, bypassing complex constitutive modeling. The recurrent neural network (RNN) model are utilized to expand existing data collected from numerical simulations with simple loading and unloading conditions. The combination of DDCM and AI provides an alternative way for material development with less prior knowledge and good problem migration.
Mots clés — Shape memory alloy, data-driven computing, artificial intelligence.

1 Introduction

SMAs are widely used smart materials in recent years because of their unique material behavior. However, this behavior involves strong nonlinearities and the coupling of multiple physical fields, results in its complicated constitutive modeling [1]. The existed models always include a large number of internal variables, lead that the computational cost in simulation become unaffordable. This work attempts to bypass complex constitutive modeling through model-free data-driven computing, and introduce artificial intelligence technique to improve simulation efficiency [2, 3].

Data preparation of the SMA was carried out by several simple loading and unloading tests. Then, an artificial neural network are utilized to expand existing data. Finally, the model-free data-driven computing of shape memory alloys were driven directly using data rather than constitutive models. Fig. 1 gives an overview of the DDCM computational framework for SMA behavior modeling.

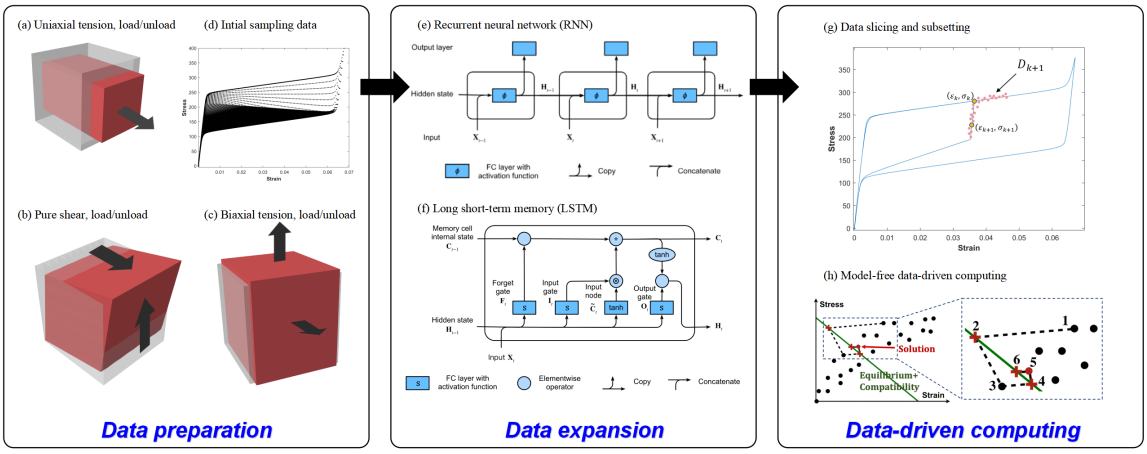

FIGURE 1 – A flowchart of the AI powered DDCM for SMAs. (a) Uniaxial tension test ; (b) Pure shear test ; (c) Biaxial tension test ; (d)Initial sampling data from (a) to (c) ; (e) Architecture of recurrent neural network (RNN) [4] ; (f) Computing the hidden state in an LSTM model [4] ; (h) Data slicing and subsetting for local material data ; (g) Fixed-point computing in DDCM.

2 Methods

2.1 SMA thermomechanical model

The 3D SMA constitutive model developed by Chemisky et al. [5], is implemented in Abaqus as a C++ umat. The model describes microscopic mechanisms of the phase transformation, the martensite reorientation and the twin accommodation, with an additive decomposition of strain :

$$\begin{aligned} &\varepsilon = \varepsilon^e + \varepsilon^{th} + \varepsilon^{tr} + \varepsilon^{twin} \\ &\varepsilon^e = \mathbb{S} : \boldsymbol{\sigma}, \quad \varepsilon^{th} = \boldsymbol{\alpha}(T - T_{ref}), \quad \varepsilon^{tr} = f\bar{\varepsilon}^{tr}, \quad \varepsilon^{twin} = f^{FA}\bar{\varepsilon}^{twin}. \end{aligned} \tag{1}$$

where ε^e, ε^{th}, ε^{tr} and ε^{twin} denote strains related to elasticity, thermal expansion, transformation and self-accommodated twinning of martensite variants, $\mathbb{S}$ is the compliance tensor, $\boldsymbol{\alpha}$ is the thermal expansion tensor, T_{ref} is the reference temperature, f and f^{FA} denote the martensitic volume fraction $\frac{V_M}{V}$ and the formed self-accommodated martensitic volume $\frac{V^{FA}}{V}$, respectively. The strains $\bar{\varepsilon}^{tr}$ and $\bar{\varepsilon}^{twin}$ represent the mean strain over the martensite volume and formed self-accommodated martensitic volume, respectively.

Considering the equilibrium temperature $T_0 = \frac{\delta U}{\delta S}$ and a linear variation of entropy $B = -\delta S$, the thermodynamical potential (variation of Gibs free enregy) could be rewritten, neglecting the term related to specific heat :

$$\begin{aligned} \Delta G = &-\Delta T S^A + B(T - T_0)f - \frac{1}{2}\boldsymbol{\sigma} : \mathbb{S} : \boldsymbol{\sigma} - \boldsymbol{\sigma} : \boldsymbol{\alpha}\Delta T - \boldsymbol{\sigma} : \varepsilon^{tr} f \\ &- \boldsymbol{\sigma} : \varepsilon^{twin} f^{FA} + \frac{1}{2} f H_\varepsilon \bar{\varepsilon}^{tr} : \bar{\varepsilon}^{tr} + \frac{1}{2} H_f f^2 + \frac{1}{2} f^{FA} H_{twin} \bar{\varepsilon}^{twin} : \bar{\varepsilon}^{twin} \end{aligned} \tag{2}$$

The time derivative of Gibbs free energy gives the definition of the general thermodynamic force A_k conjugated to the in internal variables V_k ($A_k = \frac{\partial \Delta G}{\partial V_k}$). A natural choice is to consider four internal variables, f, f^{FA}, $\bar{\varepsilon}^{tr}$ and $\bar{\varepsilon}^{twin}$. As the self-accommodated martensitic behavior depends on the activated martensite variants, the rate of f^{FA} is dependent on that of f [5]. Thus, there is no driving force associated to f^{FA}, and the influence of the terms related to f^{FA} are included in the transformation driving force A_f. To obtain the evolution of internal variables, the dissipation is defined as :

$$\phi = -A_f \dot{f} + \mathbf{A}_{\varepsilon^{tr}} : \dot{\bar{\varepsilon}}^{tr} - \mathbf{A}_{\bar{\varepsilon}^{twin}} : \dot{\bar{\varepsilon}}^{twin} - \vec{q} \cdot \frac{\overrightarrow{\text{grad}\, T}}{T} \geqslant 0 \tag{3}$$

which is decomposed into two parts that the mechanical dissipation consisting of the first three items and the remaining thermal dissipation. The mechanical dissipation associated with twin accommodation is so low that it could be neglected, compared with the dissipation coming from transformation and orientation mechanisms. The yield force of transformation is a function of B, T_0 and additional parameters F^f_{max}, B_f and B_r, and a linear function of the equivalent transformation strain is introduced to take into account the stabilization of martensite. The criteria of forward and reverse transformation could be obtained :

$$-A_f = F_f^{max} + (B_f - B) \cdot (T - T_0) - H_s |\bar{\varepsilon}^{tr}|_\varepsilon \tag{4}$$

$$-A_f = -F_f^{max} + (B_r - B) \cdot (T - T_0) - H_s |\bar{\varepsilon}^{tr}|_\varepsilon \tag{5}$$

where H_s is the dependence parameter, and $|\cdot|_\varepsilon$ corresponds to the Mises equivalent strain. An isotropic criterion is chosen to describe the orientation surface :

$$|\mathbf{A}_{\varepsilon^{tr}}|_\sigma = f \cdot F_\varepsilon^{crit} \tag{6}$$

where $|\cdot|_\sigma$ corresponds to the Mises equivalent stress. Considering the physical limitations, the modified driving forces expressed with Lagrange multipliers are defined as :

$$\begin{aligned} &F_f = -A_f - \lambda_0 - \lambda_1 = \boldsymbol{\sigma} : \bar{\varepsilon}^{tr} - B(T - T_0) - \frac{1}{2} H_\varepsilon \bar{\varepsilon}^{tr} : \bar{\varepsilon}^{tr} - H_f f \\ &+ \zeta^{FA}(\boldsymbol{\sigma} : \bar{\varepsilon}^{twin} - \frac{1}{2} H_{twin} \bar{\varepsilon}^{twin} : \bar{\varepsilon}^{twin}) - \lambda_0 - \lambda_1 \\ &\mathbf{F}_\varepsilon = -\mathbf{A}^*_{\varepsilon^{tr}} - \lambda_2 \bar{\varepsilon}^{tr} = \boldsymbol{\sigma}^D - H_\varepsilon \bar{\varepsilon}^{tr} - \lambda_2 \bar{\varepsilon}^{tr} \\ &\mathbf{F}_{\varepsilon^{twin}} = -\mathbf{A}^*_{\varepsilon^{twin}} = -f^{FA}\boldsymbol{\sigma}^D + f^{FA} H_{twin} \bar{\varepsilon}^{twin} \end{aligned} \tag{7}$$

in which $\mathbf{A}_{\varepsilon^{tr}} = f \cdot \mathbf{A}^*_{\varepsilon^{tr}}$ and $\mathbf{A}_{\varepsilon^{twin}} = f \cdot \mathbf{A}^*_{\varepsilon^{twin}}$. Here, the evolution for phase transformation controlled by the predefined yield and the driving forces has been formulated. These equations are considered when all the thermodynamic transformation forces are activated. It is necessary to take into account special cases when transformation and/or orientation are not activated and the effects of internal loops. More details could be referred in [5]. These equations, depending on the activation condition for the different mechanisms, lead to a system of equations to be solved for stress, and all the considered internal variables. For a given increment of strain and temperature, the set of equations is solved using the return mapping algorithm (RMA) [6].

2.2 Model-free data-driven computing

The idea of model-free data-driven computing is to directly formulate boundary value problem in terms of the material data without requiring explicit material modeling [7]. A distance functional of material data $\left(\boldsymbol{\varepsilon}^i, \boldsymbol{\sigma}^i\right)$ and mechanical state $(\boldsymbol{\varepsilon}, \boldsymbol{\sigma})$ are defined to describe the mutual matching of constitutive relation and system equilibrium. The local functional Ψ^e in the material point e is formulated as :

$$\Psi^e(\boldsymbol{\varepsilon}_e, \boldsymbol{\sigma}_e) = \min_{(\boldsymbol{\epsilon}^i_e, \boldsymbol{\sigma}^i_e) \in \mathcal{D}} \frac{1}{2} \left[\left(\boldsymbol{\varepsilon}_e - \boldsymbol{\varepsilon}^i_e\right) : \mathbb{C} : \left(\boldsymbol{\varepsilon}_e - \boldsymbol{\varepsilon}^i_e\right) + \left(\boldsymbol{\sigma}_e - \boldsymbol{\sigma}^i_e\right) : \mathbb{C}^{-1} : \left(\boldsymbol{\sigma}_e - \boldsymbol{\sigma}^i_e\right) \right] \tag{8}$$

which is minimized by searching over all data points in the database $\mathcal{D}$. The pseudo-material tensor $\mathbb{C}$ is a given numerical tensor and serves as a weight to ensure that the related strain and stress terms have the same dimension [8]. The core algorithm of data-driven method is to minimize the global distance function under compatibility and equilibrium constraints :

$$\text{Minimize :} \quad \sum_{e=1}^{m} w_e \Psi^e(\boldsymbol{\varepsilon}_e, \boldsymbol{\sigma}_e) \tag{9a}$$

$$\text{subject to :} \quad \boldsymbol{\varepsilon}_e = \mathbf{B}_e \cdot \mathbf{u} \quad \text{and} \quad \sum_{e=1}^{m} w_e \mathbf{B}_e : \boldsymbol{\sigma}_e = \mathbf{f} \tag{9b}$$

in which w_e is the integration weight, $\mathbf{B}_e$ is the matrix of spatial derivatives of the element shape functions, $\mathbf{u}$ and $\mathbf{f}$ are the displacement and the body force vectors, respectively.

In the data-driven problem, the compatibility constraints can be satisfied by expressing the strains in terms of displacements while the equilibrium constraints can be enforced by means of Lagrange multipliers. A system of Euler-Lagrange equations could be obtained by taking the variations of unknowns in the data-driven problem. This problem could be reformulated as a standard form that was easily solved with finite element method. Then, a fixed-point iterations algorithm [9] was used to compute the material datasets, mechanical stresses, and the states mapping as shown in Fig. 1 (h).

A natural extension of the data-driven paradigm just described concerns inelastic materials whose response is irreversible and history dependent [10]. The instantaneous response of inelastic materials is characterized by its dependence on the past history of deformation. By virtue of this history dependence, the set of stress–strain pairs attainable at a material point depends itself on time. As shown in Fig. 1 (g), the instantaneous local material data set could be defined as :

$$\mathcal{D}_{e,k+1} = \{(\boldsymbol{\varepsilon}_{e,k+1}, \boldsymbol{\sigma}_{e,k+1}) : \text{past local history}\} \tag{10}$$

The challenge now is to formulate practical methods for characterizing the history dependence of local material datasets, which can be achieved through the LSTM network mentioned in the next section.

2.3 Long Short-Term Memory (LSTM) network

The SMA thermomechanical behavior is a typical load-path dependent problem, the RNN architecture is introduced to handle this load-path dependency or history dependency [11]. RNN captures the dynamics of sequences via recurrent connections, which can be thought of as cycles in the network of nodes. As the unfolded view in Fig. 1 (e), it is unrolled across time steps with the same underlying parameters applied at each step. However, recording the information of too many time steps will cause the gradient vanish or explode in training. Therefore it is better to turn to Long-Short Term Memory

(LSTM) network that is capable to filter out unimportant information of long time series data. LSTM network resemble standard recurrent neural networks but here each ordinary recurrent node is replaced by a memory cell. Each memory cell contains an internal state that a node with a self-connected recurrent edge of fixed weight 1, ensuring that the gradient can pass across many time steps without vanishing or exploding.

As shown in Fig. 1 (f), a basic LSTM cell consists of four main gates : an input gate (for updating and adding new information), a forget gate (for deciding which information is kept or forgotten from the previous accumulated information), a memory gate (for giving the network a Long term memory of past events) and an output gate (for predicting the next hidden state). The compact forms of LSTM gates [4] are given by the following expressions :

$$\begin{aligned}
\mathbf{F}_t &= s(\mathbf{X}_t\mathbf{W}_{xf} + \mathbf{H}_{t-1}\mathbf{W}_{hf} + \mathbf{b}_f)\\
\mathbf{I}_t &= s(\mathbf{X}_t\mathbf{W}_{xi} + \mathbf{H}_{t-1}\mathbf{W}_{hi} + \mathbf{b}_i)\\
\mathbf{O}_t &= s(\mathbf{X}_t\mathbf{W}_{xo} + \mathbf{H}_{t-1}\mathbf{W}_{ho} + \mathbf{b}_o)\\
\tilde{\mathbf{C}}_t &= \tanh(\mathbf{X}_t\mathbf{W}_{xc} + \mathbf{H}_{t-1}\mathbf{W}_{hc} + \mathbf{b}_c)\\
\mathbf{C}_t &= \mathbf{F}_t \odot \mathbf{C}_{t-1} + \mathbf{I}_t \odot \tilde{\mathbf{C}}_t\\
\mathbf{H}_t &= \mathbf{O}_t \odot \tanh(\mathbf{C}_t)
\end{aligned} \tag{11}$$

where $\mathbf{W}_{xf}, \mathbf{W}_{xi}, \mathbf{W}_{xo}, \mathbf{W}_{xc} \in \mathbb{R}^{d\times h}$ and $\mathbf{W}_{hf}, \mathbf{W}_{hi}, \mathbf{W}_{ho}, \mathbf{W}_{hc} \in \mathbb{R}^{h\times h}$ are weights and $\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_o, \mathbf{b}_c \in \mathbb{R}^{1\times h}$ are bias. The operator $\odot$ denotes Hadamard (elementwise) product operator, $s(\cdot)$ is the sigmoid function and $tanh(\cdot)$ is the hyperbolic tangent function. A stacked LSTM architecture has been built to enhance its capabilities to capture more complex behaviors, which is deep not only in the time direction but also in the input-to-output direction.

3 Results

3.1 Data preparation

A 3D cube composed of SMA is used to collect stress-strain data sequence. As an example shown in Fig. 2, several loading-unloading cycles of tension are performed by this model with an hexahedral linear element with reduced integration (C3D8R). The hourglass stiffness parameter is taken as 134.6 to prevent zero-energy modes. Following the constitutive model introduced in the previous section, the material parameters of a NiTi alloy are given in Table. 1.

TABLE 1 – Material parameters of SMA

$E(MPa)$	70000	$B_f(MPa\,^\circ C^{-1})$	5	$H_f(MPa)$	4
ν	0.3	$B_r(MPa\,^\circ C^{-1})$	6	$H_{tr}(MPa)$	1000
ε^{tr}_{trac}	0.06	$M_s(^\circ C)$	20	$H_{tw}(MPa)$	40000
$\varepsilon^{tr\ FA}_{trac}$	0.05	$A_f(^\circ C)$	50	$H_s(MPa)$	50
ε^{tr}_{comp}	0.04	$F_\varepsilon(MPa)$	100		

A displacement load is applied on the top face, which gradually increases until it reaches a given maximum value and is then released. The given maximum takes 100 values evenly in the range of 0.0007 to 0.07, resulting in a stress-strain data sequence of 100 loading paths. Each path length is 400 because the load is applied or released equally through 200 steps. These 100 samples, the time series of length 400, are used to train the LSTM network and are expanded in the next section.

3.2 Data expansion

The generated data are split into training (80% of total samples), validation (10%) and test data sets (10%). The LSTM network implementation is performed in matlab with the Deep Learning Toolbox. Fig. 3 presents a comparison between the predicted stress components by the stacked LSTM network and the test data. Although the performance of this LSTM network is not completely satisfactory, it can be further improved when more training data is introduced. The result proves the potential of RNN as

a surrogate model capable of characterizing the history dependent mechanical behaviors and expanding existing data.

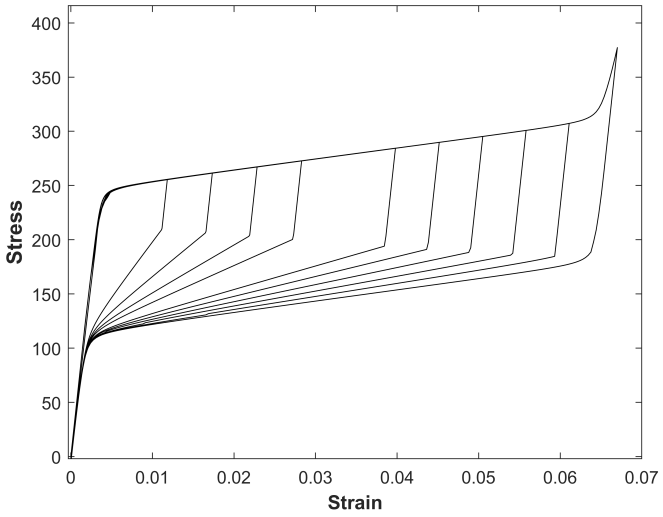

FIGURE 2 – An example of data sampling from uniaxial tension loading paths.

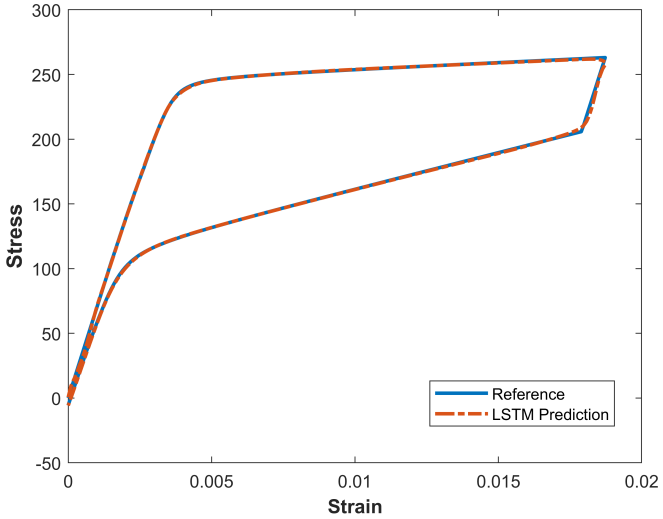

FIGURE 3 – Comparison between the predicted stress components by the stacked LSTM network and the test data.

3.3 Data-driven computing

In the data-driven computing for history-dependent materials, the instantaneous local data set could be obtained through the LSTM network trained previously. Fig. 4 shows the comparison of data-driven computing with the traditional plastic FEM computing. The nonlinear behaviors of SMA, the evolution of the non-complete and complete transformation, have been described accurately by DDCM. The results show that data-driven computing can achieve the same level of accuracy as traditional methods. In the following work, the efficiency of data-driven computing will be further improved by introducing big data techniques [12].

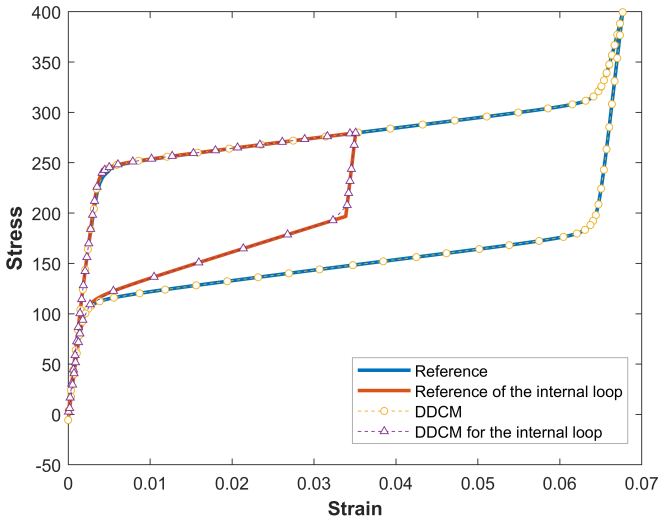

FIGURE 4 – Comparison of data-driven computing with traditional plastic FEM computing

4 Discussion

This paper presented a model-free data-driven method to simulate the behavior of SMA. This method directly relies on material data characterizing the material behavior implicitly, rather than the explicit constitutive model with various internal variables. Thus, it bypass the complicated extraction the constitutive equation from the data and the time-consuming cost of calculating the internal variables, tangent stiffness tensor and further additional information. Data density is extremely related to the accuracy of DDCM. AI techniques perform nicely in expansion from initial sampling data, powered the model-free data-driven computing efficiently. The combination of AI and DDCM avoids the problems in the directly taking the ANN model replace the constitutive model in traditional FEM computing, such as lack of physical explanation, overfitting and extra training cost.

The AI-powered data-driven computational method has demonstrated its potential for mechanical behavior modeling of SMA. This standardized framework can be transferred to model behaviors of piezoelectric materials, hydrogels, an other advanced smart materials by changing the material database.

Références

[1] C. Cisse, W. Zaki, and T. Ben Zineb. A review of constitutive models and modeling techniques for shape memory alloys. *International Journal of Plasticity*, 76 :244–284, 2016.

[2] R. Xu, J. Yang, W. Yan, Q. Huang, G. Giunta, S. Belouettar, H. Zahrouni, T. B. Zineb, and H. Hu. Data-driven multiscale finite element method : From concurrence to separation. *Computer Methods in Applied Mechanics and Engineering*, 363 :112893, 2020.

[3] R. Xu, C. Bouby, H. Zahrouni, T. B. Zineb, H. Hu, and M. Potier-Ferry. 3D modeling of shape memory alloy fiber reinforced composites by multiscale finite element method. *Composite Structures*, 200 :408–419, 2018.

[4] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola. Dive into deep learning. *arXiv preprint arXiv :2106.11342*, 2021.

[5] Y. Chemisky, A. Duval, E. Patoor, and T. Ben Zineb. Constitutive model for shape memory alloys including phase transformation, martensitic reorientation and twins accommodation. *Mechanics of Materials*, 43(7) :361–376, 2011.

[6] J. C. Simo and T. J. R. Hughes. *Computational inelasticity*, volume 7. Springer Science & Business Media, 2006.

[7] T. Kirchdoerfer and M. Ortiz. Data-driven computing in dynamics. *International Journal for Numerical Methods in Engineering*, 113 :1697–1710, 2018.

[8] T. Kirchdoerfer and M. Ortiz. Data-driven computational mechanic. *Computer Methods in Applied Mechanics and Engineering*, 304 :81–101, 2016.

[9] L. Stainier, A. Leygue, and M. Ortiz. Model-free data-driven methods in mechanics : material data identification and solvers. *Computational Mechanics*, 64 :381–393, 2019.

[10] R. Eggersmann, T. Kirchdoerfer, S. Reese, L. Stainier, and M. Ortiz. Model-free data-driven inelasticity. *Computer Methods in Applied Mechanics and Engineering*, 350 :81–99, 2019.

[11] A. Danoun, E. Prulière, and Y. Chemisky. Thermodynamically consistent recurrent neural networks to predict non linear behaviors of dissipative materials subjected to non-proportional loading paths. *Mechanics of Materials*, 173 :104436, 2022.

[12] A. Y. Zomaya and S. Sakr. Handbook of big data technologies. 2017.