SMAI-JCM
SMAI Journal of Computational Mathematics

Learning Memory And Material Dependent Constitutive Laws

Kaushik Bhattacharya, Lianghao Cao, George Stepaniants, Andrew M. Stuart & Margaret Trautner

Volume 12 (2026), p. 219-267.
https://doi.org/10.5802/smai-jcm.148

The SMAI Journal of Computational Mathematics is a member of the Centre Mersenne for Open Scientific Publishing
http://www.centre-mersenne.org/
Submissions at https://smai-jcm.centre-mersenne.org/ojs/submission
e-ISSN: 2426-8399

SMAI Journal of Computational Mathematics
Vol. 12, 219-267 (2026)

Learning Memory And Material Dependent Constitutive Laws

Kaushik Bhattacharya [1]
Lianghao Cao [2]
George Stepaniants [3]
Andrew M. Stuart [4]
Margaret Trautner [5]

[1] Mechanical and Civil Engineering, California Institute of Technology, Pasadena, CA, USA
E-mail address: bhatta@caltech.edu
[2] Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA
E-mail address: lianghao@caltech.edu
[3] Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA
E-mail address: gstepan@caltech.edu
[4] Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA
E-mail address: astuart@caltech.edu
[5] Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA
E-mail address: trautner@caltech.edu.

Abstract. We propose and study a neural operator framework for learning memory- and material microstructure-dependent constitutive laws for heterogeneous materials. We work in the two-scale setting, where homogenization theory provides a systematic approach to deriving macroscale constitutive laws, obviating the need to resolve the complex microstructure repeatedly. However, the unit-cell problems that define these constitutive models are typically not amenable to explicit evaluation. It is therefore of interest to learn constitutive models from data generated by the unit cell problem. Our proposed framework models homogenized constitutive laws with both memory- and microstructure-dependence using Markovian recurrent and Fourier neural operators. The homogenization problem for Kelvin–Voigt viscoelastic materials is studied to provide firm theoretical foundations for our model. The theoretical properties of the cell problem in this Kelvin–Voigt setting motivate the proposed learning framework, and are also used to prove a universal approximation theorem for the learned macroscale constitutive model. Numerical experiments show that the proposed learning framework accurately learns memory- and microstructure-dependent viscoelastic and elasto-viscoplastic constitutive models, beyond the setting of the theory. Furthermore, we show that the learned constitutive models can be successfully deployed in macroscale simulations of material deformation for different microstructures without retraining.

2020 Mathematics Subject Classification. 35B27, 65M60, 68T07, 74D05, 74D10, 74Q10, 74Q15.

Keywords. Constitutive modeling, homogenization, memory, microstructure, Kelvin–Voigt, viscoelasticity, elasto-viscoplasticity, neural operator.

1. Introduction

Physical materials have multiple scales [64], including the atomic scale governed by quantum interactions between atoms, the microscale encompassing fine microstructures such as grains or defects, and the macroscale, which describes a material in bulk and analyzes its observable physical properties. Modeling this cascade of information across scales is a problem of immense interest in the materials science community, as it has the potential to connect structure at the atomic and microscales to

This work is supported by the ONR MURI on Data-Driven Closure Relations N00014-23-1-2654. GS is supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship (MSPRF) under award number 2402074. KB and AMS are supported by the ONR SciAI Center under grant N00014-23-1-2729. AMS is also supported by a Department of Defense Vannevar Bush Faculty Fellowship.
https://doi.org/10.5802/smai-jcm.148

observable material properties [55]. This, in turn, allows the prediction of macroscale phenomena by appropriately summarizing the smaller scales [66]. Multiscale materials can be simulated by constructing a hierarchy of physical models that capture the relevant phenomena at each scale and allowing these scales to interact pairwise. This simulation method is very accurate when there is clear scale separation but is prohibitively expensive, as it requires simulating the dynamics of an atomic- or microscale problem within each representative volume element of the coarser scale [66].

The theory of homogenization [1, 5, 54] defines macroscale constitutive laws that are found by averaging over smaller scales and, thus, provides an alternative, cheaper route for simulating the behavior of multiscale materials. In this work, we consider heterogeneous materials in a two-scale setting with a microscale $\varepsilon \ll 1$, often determined by the typical grain or defect size. The relationship between average stress and average strain, over a unit cell of scale $\mathcal{O}(\varepsilon)$, defines a constitutive law that governs the material at the macroscale of $\mathcal{O}(1)$. Once this macroscale constitutive law is found, we can avoid solving for the microscale dynamics entirely.

In many settings, homogenization does not yield usable closed-form expressions for the relationship between cell-averaged strain and stress. In this context, it is of interest to *learn* the relationship from data. For viscoelastic and inelastic materials, the stress–strain relationship often involves memory[1] by taking strain history, a temporally varying function, as input to determine stress at a given time. Furthermore, it is of interest to learn how the constitutive relationship depends on the properties of the microstructure represented by a spatially varying function. Learning this dependence obviates the need to solve a different homogenization problem for each different material and has the potential to further accelerate computations. This paper aims to develop, analyze, and numerically study neural operator architectures, which take as input temporally and spatially varying functions, suitable for learning memory- and microstructure-dependent macroscale constitutive laws.

In Section 1.1, we summarize our contributions and provide an overview of the remainder of the paper. Section 1.2 contains a literature review, detailing the context for our contributions. In Section 1.3 we define the notation used throughout the paper.

1.1. **Contributions and Paper Overview**

This work makes the following novel contributions to data-driven constitutive modeling in the homogenization setting:

(C1) We propose a recurrent neural operator constitutive model that predicts the stress-strain relationship across a wide range of microstructures without retraining.

(C2) In the context of one-dimensional Kelvin–Voigt linear viscoelasticity, we prove Lipschitz continuity of the map from properties of the microstructure to the homogenized stress. Using this, we derive a universal approximation theorem proving the existence of a recurrent neural operator that uniformly approximates the stress dynamics of a large class of one-dimensional microstructures up to a desired error (Theorem 4.7).

(C3) We provide insight into advantageous choices of measures from which to draw training data, namely the strain trajectories and microstructures, when learning a model.

(C4) We showcase the empirical success of our proposed constitutive model in both one-dimensional linear viscoelastic (where the theory applies directly) and one-dimensional nonlinear elasto-viscoplastic materials (where it does not). We (i) accurately predict stress responses when incorporating both the strain trajectory and properties of the microstructure as inputs; and

[1] Often referred to as history- or path-dependence in the mechanics literature.

(ii) deploy the learned model in macroscale simulations for different microstructures. The implementation is available through the link below.

`https://github.com/lcao11/learning_homogenization`

In Section 2, we describe our hypotheses about the exact homogenized constitutive law that we wish to approximate, and we describe the new class of neural operators, FNM–RNO, that we introduce, study, analyze, and test in the remainder of the paper; this is contribution (C1). In Section 3 we introduce the two-scale Kelvin–Voigt model of one-dimensional viscoelastic materials, concentrating on describing the cell problem. This is a constitutive model for which we are able to prove a universal approximation theorem for FNM–RNO, motivating its form. Section 4 starts by studying the Lipschitz properties of the cell problem with respect to the microstructure, the first part of the contribution (C2). We then use this continuity to prove our universal approximation result (Theorem 4.7), showing that the cell problem partial differential equation (PDE) solution operator can be efficiently approximated by an FNM–RNO, the second part of the contribution (C2). Section 5 addresses contributions (C3) and (C4) by describing numerical results implementing this FNM–RNO model to determine the model for the dynamics of viscoelastic and elasto-viscoplastic homogenized materials, and then simulate from this model. The supplementary material contains detailed technical discussions on the equivalence of different cell problems, Lipschitz properties of the cell problem PDE, universal approximation result, and training formulations.

1.2. **Literature Review and Background**

Our work touches upon several classical and modern topics in the constitutive modeling of materials, including homogenization methods, memory-dependent modeling, machine learning of constitutive laws, and model dependence on microstructure. We discuss prior research in each of these areas below.

Theoretical and Numerical Homogenization. In its simplest formulation, homogenization theory [54] studies elliptic or parabolic PDEs whose coefficients vary periodically on a small lengthscale $\varepsilon \ll 1$. Such PDEs are *multiscale* since their solutions have coarse-grained features as well as fine-grained features of scale ε. Homogenization uses a power series expansion to determine the limit of the PDE solution as the lengthscale ε tends to zero. This results in a new averaged or *homogenized* PDE of the same form with an *effective* coefficient function that is now independent of the microscale ε. This effective coefficient is determined by a boundary value problem (BVP) called the *cell problem*, which is solved at the microscale level. Homogenization theory can be extended to PDEs with random or even nonperiodic coefficients [13, 36], and care must be taken to establish convergence of the true solution to the homogenized limit [3, 13, 54]. The texts of Milton [47] and Zohdi & Wriggers [66] give a comprehensive review of effective material properties that result from homogenization.

The effective modulus of a homogenized material is determined by solving a cell boundary value problem at the level of the microscale ε. Predicting the macroscale behavior of a material by numerically solving this cell problem BVP is called *numerical homogenization* and is a core focus of *computational micromechanics* [66]. Cell problems are typically solved with periodic, Dirichlet, or Neumann boundary conditions using spectral [48, 49] or finite element methods [26, 60]. The cell problem arising from homogenization enables us to compute the effective properties of a multiscale material, but this computation must be repeated for each new microstructure, making it an expensive procedure. We discuss below how machine learning methods allow us to perform homogenization over a range of microstructures simultaneously.

Memory and Internal Variables. Viscoelastic materials exhibit both viscous and elastic behavior; hence, their strain and stress dynamics explicitly depend on time. In particular, the application of a sudden strain deformation or stress load at one location is remembered throughout the material at all future times, and this memory is quantified by exponentially decaying *memory kernels* called the creep compliance and relaxation modulus functions [19, 62]. This observation that viscoelastic materials have fading memory was formally shown to hold under very general mathematical assumptions in a series of papers by Coleman & Noll [14, 15].

Fading memory also arises in multiscale viscoelastic materials whose microstructure varies periodically on a cell of size ε. Even though the original multiscale material locally exhibits no memory effects in strain or stress (e.g., Markovian behavior), homogenizing by averaging at the ε scale and taking $\varepsilon \to 0$ introduces local creep compliance and relaxation modulus functions that dictate long-term memory in the strain and stress dynamics at every point of the homogenized material. This remarkable result was first proven by Sanchez–Palencia on the Kelvin–Voigt model using semigroup theory [57, Chapter 6]. Further extensions to thermo-viscoelasticity were proven in the seminal paper of Francfort and Suquet [21]. Tartar [61] showed that the memory kernel (relaxation modulus) relating strain to stress after homogenization is given by a possibly infinite sum of exponentials. Suquet and coauthors [11, 38, 39] discuss more about the structure of these memory kernels and their approximation by finite sums of exponentials. In one-dimensional piecewise-constant materials, the creep compliance and relaxation modulus memory kernels are exactly given by a finite sum of exponentials, and this has been rediscovered in various classical and modern texts [7, 25]. Approximating these kernels by a finite sum of exponentials is known as a *Prony series*, and this technique has been thoroughly explored both in theory [39, 58, 62] and experiments [32, 37, 51, 59].

Viscoelastic materials whose memory kernels are given by finite sums of exponentials can be transformed into differential equations with an internal state vector whose dimension is equal to the number of exponential terms. These internal state variables incorporate the material history in a *Markovian* manner, leading to more efficient simulations of material stress-strain dynamics [7, 42]. Internal variables also arise in models of plastic [56] and viscoplastic [42] materials, and reviews of this subject can be found in [8, 28]. Hence, memory in materials is fundamentally linked to internal-variable and differential-equation representations. A review of the equivalence between such model representations can be found in [18].

Machine Learning of Constitutive Models. Two central applications of data-driven methods in materials science are the discovery of unknown constitutive laws and the acceleration of composite multiscale material simulations [43]. Data-driven learning of constitutive laws is an actively developing field that has incorporated a variety of tools, including model-constrained parameter inference [2], physics-informed machine learning [27], probabilistic machine learning [23], deep learning [44], and operator learning [6]. We refer readers to a recent comprehensive review paper [22] on this topic.

For inelastic materials, a constitutive model must use the strain history to predict stress evolution; the relationship between strain and stress is no longer instantaneous. Learning such a mapping between strain and stress time series has been approached with several data-driven architectures. Liu et al. [41] featurize strain and stress time series through principal component analysis and learn a mapping between these feature spaces. This approach has the benefit of being invariant to the level of time discretization of the data, but suffers from a lack of causality in its learned strain-to-stress map. Causality can be enforced using recurrent neural networks (RNNs), and the LSTM [24] and GRU [50] recurrent architectures have been very effective at learning strain-to-stress maps with history dependence.

A natural approach to enforce both causality and independence to time discretization is to model the constitutive law as a differential equation, which is forced by the strain trajectory and whose output is

the stress. Compared to the non-physical architecture of RNN models, this approach is guided by the internal variable theories of memory-dependent materials discussed above and has motivated a large application of neural ODEs [12, 30], also referred to as recurrent neural operators [42], to constitutive modeling of materials [7, 30, 31, 42, 65].

Microstructure-Dependent Architectures. Since data-driven constitutive models must be retrained for each microstructure, there is a need to develop *microstructure-dependent* architectures that can predict constitutive laws of new materials without retraining. One approach is to allow data-driven models to depend on summary statistics of the material, such as the volume fraction, elastic moduli of different phases, and mean fiber and grain sizes and distances. This idea has been used in several important architectures such as the Deep Material Network [44] and material-dependent recurrent networks [50]. Bishara et al. [9] provide a good review of such methods. These approaches assume that a microstructure can be adequately described by predetermined statistics, which are mostly applicable to n-phase media, and thus do not generalize to more complex spatially varying microstructures.

The fact that a microstructure must generally be interpreted as a full *functional* input into a data-driven constitutive model has been considered in a few recent papers. In [6], the Fourier neural operator (FNO) architecture was trained on elastic multiscale materials to learn a map from their microstructure as a spatially varying function on the cell domain to the solution of the cell problem. Crucially, the regularity or Lipschitz continuity of the cell problem resulting from homogenization was used to prove that this FNO architecture is a universal approximator: it can predict the homogenized elastic modulus across a range of microstructures with uniformly bounded error. In this paper, we show that cell problems of *viscoelastic* materials also satisfy Lipschitz regularity conditions, enabling us to build differential equation FNO architectures with universal approximation guarantees across a range of microstructures.

Jones et al. [30] was the first work to propose a neural ODE architecture that was microstructure-dependent and could thus simulate stress-strain dynamics with internal state variables across a wide array of microstructures. Their approach was to featurize the microstructure function using a graph convolutional neural network and to augment the initial conditions of the internal state variables with this feature vector. This way of encoding the microstructure in the initial conditions of the internal state variables is motivated by the improved numerical performance of augmented neural ODEs [17]. In contrast, the theory of viscoelastic [7] and viscoplastic [42] materials instead shows that material dependence must be encoded in the *functional form* of the differential equation driving the internal state variables rather than in their initial conditions. This is the approach we take here, which allows us to develop neural operator architectures that accurately predict strain-stress dynamics across a variety of microstructures and, crucially, to derive approximation guarantees proving the existence of such architectures that can attain uniform stress prediction error over a large microstructure class.

1.3. **Notation**

Sets. We denote the set of all integers by $\mathbb{Z}$. The set of d-dimensional integer vectors is given by $\mathbb{Z}^d$. We denote by $\mathbb{N}$ the natural numbers including zero, and by $\mathbb{N}_{>0}$ the natural numbers excluding zero. For $M \in \mathbb{N}_{>0}$ we denote $[M]$ as the set of indices $\{1, \dots, M\}$.

Euclidean Spaces. Let m be an arbitrary positive integer. Define $\mathbb{R}$ as the real line and $\mathbb{R}_+$ as the positive real line, including zero. The set of complex numbers is denoted by $\mathbb{C}$. Let $\mathbb{R}^m$ denote the m-dimensional Euclidean space and $\mathbb{R}^{m\times m}$ be the space of $m \times m$ real matrices. We also use $\mathbb{R}_+$ and $\mathbb{R}^m_+$ for the space of nonnegative real numbers and m-dimensional vectors with nonnegative entries, respectively. We define $\mathbb{D}^m$ to be the space of $m \times m$ real diagonal matrices and $\mathbb{D}^m_+$ to be the space of real diagonal matrices with nonnegative diagonal entries. We write the Euclidean inner product and

norm on $\mathbb{R}^d$ as $\langle \cdot, \cdot \rangle_2$ and $\| \cdot \|_2$ respectively. For matrices in $\mathbb{R}^{d\times d}$ we write the Frobenius norm as $\| \cdot \|_F$ and the infinity norm on matrices and vectors as $\| \cdot \|_\infty$.

Function Spaces. We define the domain $\Omega \subset \mathbb{R}^d$ to be a bounded open set and $\mathcal{T} \subset \mathbb{R}_+$ to be a time interval which can be finite or infinite and open or closed. Define the Hilbert space $L^2(\Omega; \mathbb{R}^m)$ whose inner product and norm we denote by $\langle \cdot, \cdot \rangle_{L^2}$ and $\| \cdot \|_{L^2}$ respectively. We also define the space $L^1(\Omega; \mathbb{R}^m)$ equipped with the norm $\| \cdot \|_{L^1}$. Define the space $L^\infty(\mathcal{T}; \mathbb{R}^m)$ and denote its norm by $\| \cdot \|_{L^\infty}$. The space $W^{k,p}(\mathcal{T}) = W^{k,p}(\mathcal{T}; \mathbb{R})$, or multivariate generalizations taking values in $\mathbb{R}^d$ or $\mathbb{R}^{d\times d}$, denotes the Sobolev space of functions defined on the time interval $\mathcal{T}$ with weak derivatives up to order k which are all in L^p, $1 \le p \le \infty$. We also introduce the function space $\mathcal{Z} = L^\infty(\mathcal{T}; L^2(\Omega; \mathbb{R}^m))$ with the norm $\|u\|_{\mathcal{Z}} = \operatorname{ess\,sup}_{t\in\mathcal{T}}(\|u(\cdot, t)\|_{L^2})$. We denote by $H^1_0(\Omega; \mathbb{R}^m)$ the closure of the space of smooth functions compactly supported in Ω in the Sobolev space $W^{1,2}(\Omega)$.

When working with function spaces such as $H^1_0(\Omega; \mathbb{R}^m)$ or $L^1(\Omega; \mathbb{R}^m)$, we often write $H^1_0(\Omega)$ or $L^1(\Omega)$ when it is clear from the context that the functions take values in $\mathbb{R}^m$. We denote by $\mathbb{1}$ the constant unit function on Ω. In much of our theory, we work with the specific choice $\Omega = [0, 1]$. Additionally, we extend all the preceding definitions of functions defined on Ω to functions defined on the d-dimensional torus, denoted $\mathbb{T}^d$.

We denote the *total variation* of a function $u \in L^1(\mathbb{T})$ by

$$|u|_{\mathsf{BV}} = \sup\left\{ \sum_{i=0}^{N-1} |u(x_{i+1}) - u(x_i)| \;\middle|\; 0 = x_0 < x_1 < \ldots < x_N = 1,\ N \ge 1 \right\} \tag{1.1}$$

and the set of functions of bounded variation on $\mathbb{T}$ as

$$\mathsf{BV} = \{u \in L^1(\mathbb{T}) :\ |u|_{\mathsf{BV}} < \infty\}. \tag{1.2}$$

Maps. Let A (resp. B) be a map from input domain $\mathcal{X}_A$ (resp. $\mathcal{X}_B$) into a space of functions defined over $\mathbb{T}^d$ and taking values in $\mathbb{R}^{d_A}$ (resp. $\mathbb{R}^{d_B}$). The notation (A, B) denotes the map from $\mathcal{X}_A \times \mathcal{X}_B$ into the function space on domain $\mathbb{T}^d$ taking values in $\mathbb{R}^{d_A+d_B}$ such that, for $(x_a, x_b) \in \mathcal{X}_A \times \mathcal{X}_B$, $((A, B)(x_a, x_b))_j$ equals $A(x_a)_j$ for $j \in [d_A]$ and equals $B(x_b)_{j+d_A}$ for $j \in [d_B]$.

Trajectories. For any time-dependent function g we denote by $\{g(t)\}_{t\in\mathcal{T}}$ the set that includes pointwise evaluation of $g(t)$ and its time-derivative for all $t \in \mathcal{T}$. When it is clear in the appropriate context, we write g as shorthand for $\{g(t)\}_{t\in\mathcal{T}}$. We use $\dot{g}$ to indicate a time derivative of the trajectory g. In particular, $\dot{\bar{\epsilon}}$ is a time derivative of $\bar{\epsilon}$. Note, however, that in the context of elasto-viscoplasticity, we use the commonly adopted convention that $\dot{\epsilon}_{p0}$ denotes the rate constant; in particular, it is not the derivative of a time-dependent function.

Subsets of Banach Spaces. Denote by $\mathcal{M}^B_{f_{\min}, f_{\max}}$ the set of functions $f \in \mathsf{BV}(\Omega)$ satisfying

$$|f|_{\mathsf{BV}} \le B, \quad \operatorname*{ess\,sup}_{y\in\Omega} f(y) \le f_{\max}, \quad \operatorname*{ess\,inf}_{y\in\Omega} f(y) \ge f_{\min} \tag{1.3}$$

for some $0 < f_{\min} \le f_{\max} < \infty$ and $B > 0$. And we denote by $\mathcal{C}_{g_{\max}, \dot{g}_{\max}}$ the set of functions $g \in W^{1,\infty}(\mathcal{T})$ satisfying

$$\operatorname*{ess\,sup}_{t\in\mathcal{T}} |g(t)| \le g_{\max}, \quad \operatorname*{ess\,sup}_{t\in\mathcal{T}} |\dot{g}(t)| \le \dot{g}_{\max} \tag{1.4}$$

for some constants $0 < g_{\max}, \dot{g}_{\max} < \infty$.

2. Proposed Approximate Constitutive Law

In Section 2.1, we introduce the concept behind the derivation of homogenized constitutive laws in a general setting. In Section 2.2, we propose a form of neural operator architecture to learn the homogenized constitutive law. The general setting encompasses the particular case of Kelvin–Voigt linear viscoelasticity, which we use to motivate and to analyze the proposed architecture for the constitutive model, but it is not restricted to this case.

2.1. Homogenized Constitutive Law

Let $\mathcal{D} \subset \mathbb{R}^d$ denote a bounded open set and consider the following multiscale material model:

$$\rho \partial_t^2 u_\varepsilon(x,t) = \nabla_x \cdot \sigma_\varepsilon(x,t) + b(x,t) \qquad x \in \mathcal{D}, t \in \mathcal{T} \tag{2.1a}$$

$$\sigma_\varepsilon(x,t) = \Psi(\{\nabla_x u_\varepsilon(x,s)\}_{s\in\mathcal{T}}; M, x, \varepsilon)(t) \qquad x \in \mathcal{D}, t \in \mathcal{T} \tag{2.1b}$$

$$u_\varepsilon(x,0) = \partial_t u_\varepsilon(x,0) = 0 \qquad x \in \mathcal{D} \tag{2.1c}$$

$$u_\varepsilon(x,t) = 0 \qquad x \in \partial\mathcal{D}, t \in \mathcal{T}. \tag{2.1d}$$

Here u_ε denotes displacement and σ_ε stress, and ε is a small parameter defining the spatial microscale; the microstructural properties encapsulated in the spatial fields M vary on this scale. Function Ψ is a multiscale constitutive model taking as input the history of the strain $\nabla_x u_\varepsilon$[2]. Recall, that for any time-dependent function g we use the notation $\{g(s)\}_{0\le s\le t}$ to include pointwise evaluation of $g(s)$ and its time-derivative for all $0 \le s \le t$.

The objective of homogenization [5, 10, 54] is to remove the small parameter ε and obtain homogenized constitutive law Ψ_0 and homogenized displacement u and stress σ, related by an equation of the form

$$\rho \partial_t^2 u(x,t) = \nabla_x \cdot \sigma(x,t) + b(x,t) \qquad x \in \mathcal{D}, t \in \mathcal{T} \tag{2.2a}$$

$$\sigma(x,t) = \Psi_0(\{\nabla_x u(x,s)\}_{s\in\mathcal{T}}; M)(t) \qquad x \in \mathcal{D}, t \in \mathcal{T} \tag{2.2b}$$

$$u(x,0) = \partial_t u(x,0) = 0 \qquad x \in \mathcal{D} \tag{2.2c}$$

$$u(x,t) = 0 \qquad x \in \partial\mathcal{D}, t \in \mathcal{T}. \tag{2.2d}$$

When designed properly, this homogenized model delivers $(u, \sigma) \approx (u_\varepsilon, \sigma_\varepsilon)$. However, as it does not involve small parameter ε, it offers considerable computational cost savings over (2.1).

Remark 2.1. We note that Ψ, Ψ_0 are *causal* functions of the strain trajectory meaning that $\Psi(\,\cdot\,)(t)$ (respectively $\Psi_0(\,\cdot\,)(t)$) depends only on the strain history $\{\nabla_x u_\varepsilon(x,s)\}_{0\le s\le t}$, (respectively $\{\nabla_x u(x,s)\}_{0\le s\le t}$) up to time t.

In many situations, an exact expression for Ψ_0 is not available. We consider a setting in which the microscale is periodic. Hence, the microstructure takes $M : \mathbb{T}^d \to \mathbb{R}^{d_M}$ where d_M is the dimension of the output microstructural properties, viewed as living in a space isomorphic to a Euclidean space, and where the explicit x-dependence in Ψ is through $M(x/\varepsilon)$ only. This results in Ψ_0 not depending explicitly on x. In the settings we consider, if Ψ_0 depends on both x and $\frac{x}{\varepsilon}$, the macroscale variable x enters as a parameter, and the homogenization argument still holds [54, Section 12.7]. Indeed, this more general extension is a potential use case for our work; the macroscale variable x determines a different microstructure $M(x)$. Since our framework allows for changes in microstructure, the same model could be used to resolve microscale dynamics in the presence of macroscale material variation.

We let $\overline{\sigma}(t)$ (resp. $\overline{\epsilon}(t)$) denote time-dependent stress (resp. strain) found by averaging σ_ε and $\nabla_x u_\varepsilon$ over the unit cell with side of length ε. Then we can write $\Psi_0 : (\overline{\epsilon}, M) \mapsto \overline{\sigma}$; in particular, Ψ_0 takes

[2]Because we primarily work in one spatial dimension in this paper, we have, for simplicity of exposition, not expressed the constitutive law in terms of the symmetrized gradient of the displacement.

as input both a time-dependent function (strain) and spatially varying functions M capturing the microstructure. Our goal in this paper is to define, analyze, and numerically study a methodology to determine Ψ_0 from numerically generated data. This data will be found by studying PDE (2.1) on a unit cell (one period) and creating pairs of stress-strain histories, both averaged over the unit cell, for a variety of material properties M. From this, we wish to learn an approximation of Ψ_0.

2.2. **Neural Operator Constitutive Law**

Here we define the neural network architecture $\Psi^{\text{FNM–RNO}}$, designed to approximate the homogenized constitutive law Ψ_0 in (2.2b). Since Ψ_0 does not depend explicitly on x, the same is true of our model $\Psi^{\text{FNM–RNO}}$. There are natural generalizations to allow for x dependence in $\Psi^{\text{FNM–RNO}}$, to account for situations where Ψ_0 depends explicitly on x, but we do not consider these here.

Definition 2.2 (FNM–RNO Architecture)**.** Define the mapping

$$\begin{aligned} \Psi^{\text{FNM–RNO}} : C^1(\mathcal{T};\mathbb{R}^{d\times d}) \times L^2(\mathbb{T}^d;\mathbb{R}^{d_M}) &\longrightarrow C(\mathcal{T};\mathbb{R}^{d\times d}), \\ (\{\overline{\epsilon}(t)\}_{t\in\mathcal{T}}, M) &\longmapsto \{\overline{\sigma}(t)\}_{t\in\mathcal{T}}, \end{aligned} \tag{2.3}$$

through the equations

$$\overline{\sigma}(t) = F_{\text{FNM}}(\overline{\epsilon}(t), \dot{\overline{\epsilon}}(t), \xi(t); M), \quad t \in \mathcal{T}, \tag{2.4a}$$

$$\dot{\xi}(t) = G_{\text{FNM}}(\overline{\epsilon}(t), \xi(t); M), \quad t \in \mathcal{T}, \tag{2.4b}$$

$$\xi(0) = 0 \tag{2.4c}$$

where

$$\begin{aligned} F_{\text{FNM}} &: \mathbb{R}^{d\times d} \times \mathbb{R}^{d\times d} \times \mathbb{R}^L \times L^2(\mathbb{T}^d;\mathbb{R}^{d_M}) \longrightarrow \mathbb{R}^{d\times d} \\ G_{\text{FNM}} &: \mathbb{R}^{d\times d} \times \mathbb{R}^L \times L^2(\mathbb{T}^d;\mathbb{R}^{d_M}) \longrightarrow \mathbb{R}^L. \end{aligned} \tag{2.5}$$

Here $\xi \in \mathbb{R}^L$ denotes the internal state variable with dimension $L \in \mathbb{N}_{>0}$ and $M \in L^2(\mathbb{T}^d;\mathbb{R}^{d_M})$ is a vector-valued function that specifies the microstructure on the unit cell.

As defined, $\Psi^{\text{FNM–RNO}}$ is a causal function of the strain trajectory (see Remark 2.1) since it is given by the simulation of a differential equation forced by the strain trajectory. The functions $F_{\text{FNM}}, G_{\text{FNM}}$ are chosen to be Fourier Neural Mappings (FNMs), introduced in [29]; these are neural networks that act on functions as well as vector inputs. They are generalizations of Fourier Neural Operators (FNOs) [40], neural networks that map functions to functions by composing a set of nonlinear operators $\{\mathscr{L}_t\}_{t=1}^T$, each built from a linear operator acting as multiplier in Fourier space composed with a pointwise nonlinear operation acting on the original grid space of the input function.

Our architecture, shown in Figure 2.1, builds on the original FNM architecture defined in [29] which accepts a function-valued input and returns a vector-valued output (see three rightmost boxes in Figure 2.1). As noted in the definition of $F_{\text{FNM}}, G_{\text{FNM}}$, our neural operator needs to accept both the microstructure function as well as the strain, strain rate, and internal variable vector quantities. Our architecture enables this by taking the vector inputs, multiplying them by Fourier modes to lift them to functions on the same domain $\mathbb{T}^d$ as the microstructure, concatenating all functions (including the original microstructure) along a new channel dimension, and then inputting this multi-channel function into the classic FNM architecture (see two leftmost boxes in Figure 2.1). Our generalized model has the potential to learn both history dependence, through the recurrent structure in time via the vector inputs, as well as material dependence through the FNMs. We prove in Lemma D.5 that such FNMs have a bounded Lipschitz constant with respect to their vector inputs, and therefore imply by regularity of ODEs that $\overline{\sigma} \in C(\mathcal{T};\mathbb{R}^{d\times d})$.

Remark 2.3. Our FNM architecture accepts a function input, a vector input, and returns a vector output. When dealing with multiple functions or vector inputs, we assume they are concatenated to form a single input to the FNM. When dealing with matrix inputs or outputs, we assume they are flattened to vector inputs or outputs. Hence, $\bar{\epsilon}, \dot{\bar{\epsilon}}, \bar{\sigma} \in \mathbb{R}^{d\times d}$ are flattened to become vectors in $\mathbb{R}^{d^2}$.

For multiple function inputs, their evaluations at spatial points are concatenated (pointwise flattening). Hence, the input microstructure M is defined generally as a vector-valued function with a d_M-dimensional output. This allows us to use the FNM–RNO architecture to model a wide class of inelastic materials. For example, microstructural properties of viscoelastic materials are defined by the spatial elasticity and viscosity tensor fields $E, \nu : \mathbb{T}^d \to \mathbb{R}^{d\times d\times d\times d}$ which can be flattened and concatenated to define $M = (E, \nu) : \mathbb{T}^d \to \mathbb{R}^{d_M}$ where in this case $d_M = 2d^4$. Of course, under additional symmetries, such as isotropy conditions on the elasticity and viscosity tensors, they can be summarized as a microstructure function M with a much lower dimensionality d_M. In the setting of elasto-viscoplastic materials (see Section 5.4), their microstructure is defined by four functions: the elasticity tensor (Young's modulus) $E : \mathbb{T}^d \to \mathbb{R}^{d\times d\times d\times d}$, strain rate constant $\dot{\epsilon}_{p0} : \mathbb{T}^d \to \mathbb{R}^{d\times d}$, yield stress $\sigma_Y : \mathbb{T}^d \to \mathbb{R}$, and rate exponent $n : \mathbb{T}^d \to \mathbb{R}$. Hence, the microstructure can concatenate these four functions as $M = (E, \dot{\epsilon}_{p0}, \sigma_Y, n) : \mathbb{T}^d \to \mathbb{R}^{d_M}$ where in this case $d_M = d^4 + d^2 + 2$.

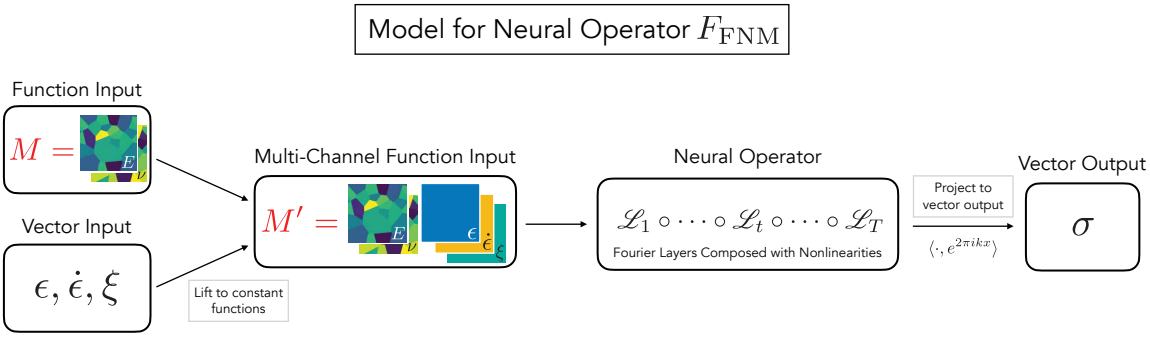


Figure 2.1. Our Fourier Neural Mapping (FNM) architecture accepts the microstructure $M : \mathbb{T}^d \to \mathbb{R}^{d_M}$ as well as the vector inputs consisting of the flattened strain and strain rate tensors and the necessary internal variables. Vector inputs are lifted to a function space and concatenated to M to form a multi-channel function, which is processed through nonlinear Fourier layers, and the output function is projected linearly to an output vector; see Definition C.1 for all details of these transformations. The figure above displays the architecture for F_{FNM} in the case of the Kelvin–Voigt model, but the definition of G_{FNM} as well as extension of this model to other material rheologies (elasto-viscoplastic) requires a minor change to the input and output function/vector variables.

To summarize, the neural operators $F_{\text{FNM}}, G_{\text{FNM}}$ are Fourier Neural Mappings as defined above. To fit these operators to input-output data, they contain learnable parameters comprising all the weights and biases in the layers of our neural network architecture, which are optimized during training. The first group of learnable parameters consists of the linear weights in the lifting step, which combine vector inputs before lifting them to function space via Fourier-mode multiplication. The second group of parameters consists of the weights, biases, and Fourier multipliers of the Fourier layers $\mathscr{L}_t$. Lastly, the third group of parameters consists of linear weights that recombine the projections of the output function onto Fourier modes to produce an output vector. The interested reader can find a detailed definition of the FNM architecture and its learnable parameters in Appendix C.

3. **Kelvin–Voigt Viscoelasticity**

We now introduce the classical Kelvin–Voigt (KV) model for a multiscale viscoelastic material. The structure of the FNM–RNO architecture $\Psi^{\text{FNM–RNO}}$ introduced in the previous section is motivated by the homogenized form Ψ_0 of the Kelvin–Voigt model described below. In fact, we will later prove that this neural architecture approximates the homogenized constitutive law of Kelvin–Voigt viscoelasticity to arbitrary accuracy.

We begin in Section 3.1 by introducing the multiscale Kelvin–Voigt model in one dimension and describe the structure of its average strain-to-stress map Ψ_0 resulting from homogenization. We then show in Section 3.2 that for piecewise-constant microstructures, the map Ψ_0 has an explicit analytical form, with memory captured through a differential equation forced by the strain, whose parameters depend continuously on the microstructure pieces.

3.1. **Homogenization and Cell Problem**

Let $E, \nu : \mathbb{T} \to \mathbb{R}$ and $E_\varepsilon(x) = E(\frac{x}{\varepsilon})$ and $\nu_\varepsilon(x) = \nu(\frac{x}{\varepsilon})$ where $\varepsilon \ll 1$ denotes a small spatial length-scale. One-dimensional, multiscale Kelvin–Voigt viscoelasticity is governed by the following partial differential equation on a spatial domain $\mathcal{D} = [0, D]$ and time interval $\mathcal{T} = [0, T]$:

$$\begin{aligned}
\rho \partial_t^2 u_\varepsilon(x,t) &= \partial_x \sigma_\varepsilon(x,t) + b(x,t) && x \in \mathcal{D}, t \in \mathcal{T} && \text{(3.1a)}\\
\sigma_\varepsilon(x,t) &= E_\varepsilon(x)\partial_x u_\varepsilon(x,t) + \nu_\varepsilon(x)\partial_{xt}^2 u_\varepsilon(x,t) && x \in \mathcal{D}, t \in \mathcal{T} && \text{(3.1b)}\\
u_\varepsilon(x,0) &= \partial_t u_\varepsilon(x,0) = 0 && x \in \mathcal{D} && \text{(3.1c)}\\
u_\varepsilon(0,t) &= u_\varepsilon(D,t) = 0 && t \in \mathcal{T}. && \text{(3.1d)}
\end{aligned}$$

Thus, the material properties $E_\varepsilon, \nu_\varepsilon$ depend only on the microscale variable $y = \frac{x}{\varepsilon}$ and have no dependence on the macroscale variable x independent of y. We note that the external forcing b is assumed independent of ε. These assumptions can be relaxed, but doing so increases computational complexity when learning homogenized models.

Equations (3.1) are a specific instance of the general setting of (2.1). In this specific setting, the homogenization procedure is outlined in [7, Section 2.2], a one-dimensionalization of the general case of homogenization for Kelvin–Voigt viscoelasticity developed in [21]. The homogenized operator

$$\begin{aligned}
\Psi_0 : C^1(\mathcal{T};\mathbb{R}) \times L^2(\mathbb{T};\mathbb{R}^2) &\longrightarrow C(\mathcal{T};\mathbb{R}),\\
(\{\overline{\epsilon}(t)\}_{t\in\mathcal{T}}, E, \nu) &\longmapsto \{\overline{\sigma}(t)\}_{t\in\mathcal{T}},
\end{aligned} \tag{3.2}$$

mapping strain to stress is given by the solution of the *cell problem*

$$\begin{aligned}
\overline{\sigma}(t) &= \int_\Omega \sigma(y,t)\,\mathrm{d}y, && t \in \mathcal{T}, && \text{(3.3a)}\\
-\partial_y \sigma(y,t) &= 0, && (y,t) \in \Omega \times \mathcal{T}, && \text{(3.3b)}\\
\sigma(y,t) &= E(y)\partial_y u(y,t) + \nu(y)\partial_{yt} u(y,t), && (y,t) \in \Omega \times \mathcal{T}, && \text{(3.3c)}\\
u(0,t) &= 0,\ u(1,t) = \overline{\epsilon}(t), && t \in \mathcal{T}, && \text{(3.3d)}\\
u(y,0) &= 0, && y \in \Omega, && \text{(3.3e)}
\end{aligned}$$

where $\Omega = [0,1]$ and $\mathcal{T} = [0,T]$ and the boundary condition $\overline{\epsilon}(t)$ satisfies $\overline{\epsilon}(0) = 0$. This version of the equations is derived in [7, Lemma 3.12]. The boundary condition $\overline{\epsilon}$ is suggestively written since the spatially averaged strain $\overline{\epsilon}(t) = \int_\Omega \partial_y u(y,t)\,\mathrm{d}y$ is exactly the boundary condition. In one dimension, $\sigma(y,t) = \sigma(t)$ is not spatially dependent due to the balance of forces in (3.3b). Thus, the spatially averaged stress is $\overline{\sigma} = \sigma$. We remark that the cell problem is not well-defined for all spatially varying microstructural properties $(E, \nu) \in L^2(\mathbb{T};\mathbb{R}^2)$ and hence in our theoretical analysis of this PDE, we

constrain these properties to be positive, bounded from above and below, and of bounded variation (see Assumptions 3.2 below).

A useful procedure to analyze this system is to decompose the solution of our cell problem into a heterogeneous periodic component and a homogeneous nonperiodic component as

$$u(y,t) = p(y,t) + \overline{\epsilon}(t)y, \tag{3.4}$$

where $p(y,t)$ satisfies the *detrended* cell problem

$$\partial_y\Big(\nu\partial_{yt}p + E\partial_y p\Big) = -\dot{\overline{\epsilon}}\partial_y\nu - \overline{\epsilon}(t)\partial_y E(y), \qquad (y,t)\in\Omega\times\mathcal{T}, \tag{3.5a}$$

$$p(0,t) = p\ (1,t) = 0, \qquad t\in\mathcal{T}, \tag{3.5b}$$

$$p(y,0) = 0, \qquad y\in\Omega. \tag{3.5c}$$

While we refer to $p(y,t)$ as the periodic component of the solution, it is more precisely the solution to the Dirichlet detrended cell problem above with homogeneous boundary conditions. Our analysis in Section 4.1 and Appendix B proves Lipschitz regularity of the cell problem in (3.3), and these arguments rely on integration by parts formulae, which are easier to express in terms of $p(y,t)$.

We define a norm under which we can study the magnitudes of the solutions u and p to the original and detrended cell problems above. Following the notation in [7], we define the ξ-dependent quadratic form

$$q_\xi(u,v) := \int_\Omega \xi(y)\partial_y u(y)\partial_y v(y)\,\mathrm{d}y \tag{3.6}$$

for arbitrary $\xi\in L^\infty(\Omega;(0,\infty))$. Define the bounds

$$\xi_{\max} := \operatorname*{ess\,sup}_{x\in\Omega}\xi(x) < \infty, \quad \xi_{\min} := \operatorname*{ess\,inf}_{x\in\Omega}\xi(x) > 0. \tag{3.7}$$

Note that $q_\xi(\,\cdot\,,\cdot\,)$ defines an inner product with resulting norm

$$\|u\|^2_{H^1_0,\xi} := q_\xi(u,u). \tag{3.8}$$

In the case that $\xi = \mathbb{1}(\,\cdot\,)$ is the constant unit function, we write

$$\|u\|^2_{H^1_0} := q_{\mathbb{1}}(u,u). \tag{3.9}$$

The norms $\|\cdot\|_{H^1_0,\xi}$ are equivalent for all $\xi\in L^\infty(\Omega;(0,\infty))$ as shown in the following:

Lemma 3.1 ([7, Lemma 1.1])**.** *For any $\xi,\zeta\in L^\infty(\Omega;(0,\infty))$ satisfying properties* (3.7)*, the norms $\|\cdot\|_{H^1_0,\xi}$ and $\|\cdot\|_{H^1_0,\zeta}$ are equivalent in the sense that*

$$\frac{\zeta_{\min}}{\xi_{\max}}\|u\|^2_{H^1_0,\xi} \le \|u\|^2_{H^1_0,\zeta} \le \frac{\zeta_{\max}}{\xi_{\min}}\|u\|^2_{H^1_0,\xi}. \tag{3.10}$$

Hence, we can use any inner product $q_\xi(u,v)$ for ξ satisfying (3.7) since they are all equivalent. Given that E,ν are the coefficient functions of our cell problem (3.3), we frequently use the weighted inner products q_E, q_ν. In particular, we can write the *weak form* of our cell problem in (3.3), seeking solution $u\in C^1(\mathcal{T};H^1_0(\Omega;\mathbb{R}))$ satisfying

$$q_\nu(\partial_t u,\varphi) + q_E(u,\varphi) = 0, \quad \forall\,\varphi\in H^1_0(\Omega;\mathbb{R}), t\in\mathcal{T}, \tag{3.11a}$$

$$u = 0, \quad t = 0. \tag{3.11b}$$

Finally, we note that the solution to the cell problem u can be interpreted as a function of time that maps into $L^2(\Omega;\mathbb{R})$ so it lives in $\mathcal{Z} = L^\infty(\mathcal{T};L^2(\Omega;\mathbb{R}))$ equipped with the norm $\|u\|_{\mathcal{Z}} = \operatorname{ess\,sup}_{t\in\mathcal{T}}(\|u(\,\cdot\,,t)\|_{L^2})$. Recall the notation for the sets $\mathcal{M}^B_{\cdot,\cdot}$ and $\mathcal{C}_{\cdot,\cdot}$ from Section 1.3.

3.2. **Microstructure Dependence**

The goal of this and the next section is to study the properties of, and approximate, the homogenized map $(\bar{\epsilon}, E, \nu) \mapsto \overline{\sigma}$, where $\bar{\epsilon}$ and $\overline{\sigma}$ are shorthand for $\{\bar{\epsilon}(t)\}_{t\in\mathcal{T}}$ and $\{\overline{\sigma}(t)\}_{t\in\mathcal{T}}$ respectively, defined by (3.3). This map allows us to study how the average stress $\overline{\sigma}$ depends on the spatially varying microstructural properties E and ν and the average strain boundary condition $\bar{\epsilon}$. We use the assumptions:

Assumptions 3.2. *We make the following assumptions on E, ν, and $\bar{\epsilon}$ throughout:*

(1) Assume that for the constants $0 < E_{\min} \le E_{\max} < \infty$, $0 < \nu_{\min} \le \nu_{\max} < \infty$ and $B > 0$ we have that $E \in \mathcal{M}^B_{E_{\min},E_{\max}}$ and $\nu \in \mathcal{M}^B_{\nu_{\min},\nu_{\max}}$, where $\mathcal{M}^B_{\cdot,\cdot}$ is defined in (1.3).

(2) Assume that for the constants $0 < \bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max} < \infty$ we have that $\bar{\epsilon} \in \mathcal{C}_{\bar{\epsilon}_{\max},\dot{\bar{\epsilon}}_{\max}}$, where $\mathcal{C}_{\cdot,\cdot}$ is defined in (1.4).

Remark 3.3. The lower and upper bounds on the material microstructure, strain, and strain rate are used throughout the paper to enforce well-posedness of the respective cell problem equations as well as to bound the magnitude of the stress response. In contrast, the requirement that materials be of bounded variation is used only in the derivation of our universal approximation theory, Section 4, to enforce that the set of material microstructures can be compactly embedded in L^2.

By definition of the sets $\mathcal{M}^B_{\cdot,\cdot}$ and $\mathcal{C}_{\cdot,\cdot}$ from Section 1.3 we see that the bounds implied by the preceding assumptions hold using the ess sup, ess inf over the cell problem domain $y \in \Omega$ or the time domain $t \in \mathcal{T}$. In the remainder of the paper, we will write sup and inf for notational brevity, but essential suprema and infima are implied.

As proved in the seminal paper of Francfort and Suquet [21], the map Ψ_0 in (3.3) takes the integral equation form

$$\overline{\sigma}(t) = \Psi_0(\{\bar{\epsilon}(t)\}_{t\in\mathcal{T}}; E, \nu) := E'\bar{\epsilon}(t) + \nu'\dot{\bar{\epsilon}}(t) - \int_0^t K(t-s)\bar{\epsilon}(s)\,\mathrm{d}s \tag{3.12}$$

and the argument employed in [21] shows *existence* of this law in any dimension, albeit without explicit analytical formulas.

For one-dimensional materials, the homogenization analysis simplifies and allows us to obtain an explicit analytic characterization of E', ν', and K. For the 1D cell problem (3.3), the parameters E', ν' (which we refer to as the *Markovian parameters* in what follows, as they do not involve memory) are derived in [7, Appendix B.1] and shown to take the following form:

$$E' = \frac{\int_0^1 \frac{E(y)}{\nu(y)^2}\,\mathrm{d}y}{\left(\int_0^1 \frac{1}{\nu(y)}\,\mathrm{d}y\right)^2}, \quad \nu' = \frac{1}{\int_0^1 \frac{1}{\nu(y)}\,\mathrm{d}y}. \tag{3.13}$$

The memory kernel is given in the Laplace domain as

$$\mathcal{L}[K](s) = E' + \nu' s - \left(\int_0^1 \frac{\mathrm{d}y}{E(y) + \nu(y)s}\right)^{-1}, \tag{3.14}$$

where $\mathcal{L}[K] : \mathbb{R} \to \mathbb{R}$ is the Laplace transform of $K : \mathbb{R} \to \mathbb{R}$. Using novel analytical results for reciprocal Cauchy transforms of measures [16], the inverse Laplace transform of this formula can be taken, to derive a closed-form expression for $K(t) = \int_{\mathbb{R}} e^{-\alpha t}\,\mathrm{d}\mu(\alpha)$ given by a continuum integral over exponential decays. These exponents are weighted by a measure μ that depends explicitly on the E, ν microstructure parameters.

When the microstructure properties are piecewise-constant functions, the memory kernel $K(t)$ becomes a finite sum of exponentials, as shown by the following result:

Proposition 3.4 (Theorem 3.6 in [7]). *Assume (E, ν) are piecewise-constant with L pieces of positive lengths $\{d_\ell\}_{\ell\in[L]}$ where*

$$E(y) = E_i,\ \nu(y) = \nu_i, \quad y \in \left[\sum_{l=1}^{i-1} d_l, \sum_{l=1}^{i} d_l\right) \tag{3.15}$$

with $i \in [L]$, and where the piece lengths add up to $\sum_{i=1}^{L} d_i = 1$. Then the map from $\{\overline{\epsilon}(t)\}_{t\in\mathcal{T}}$ to $\{\overline{\sigma}(t)\}_{t\in\mathcal{T}}$ is given by the integro-differential Volterra equation

$$\overline{\sigma}(t) = E'_{\mathrm{pc}}\overline{\epsilon}(t) + \nu'_{\mathrm{pc}}\dot{\overline{\epsilon}}(t) - \int_0^t K_{\mathrm{pc}}(t-s)\overline{\epsilon}(s)\,\mathrm{d}s \tag{3.16}$$

where the memory kernel is given by

$$K_{\mathrm{pc}}(t) = \sum_{l=1}^{L-1} \beta_l e^{-\alpha_l t}. \tag{3.17}$$

The Markovian parameters (3.13) *are given by*

$$E'_{\mathrm{pc}} = L\frac{\sum_{i=1}^{L} d_i \frac{E_i}{\nu_i^2}}{\left(\sum_{i=1}^{L} \frac{d_i}{\nu_i}\right)^2}, \quad \nu'_{\mathrm{pc}} = \frac{1}{\sum_{i=1}^{L} \frac{d_i}{\nu_i}}. \tag{3.18}$$

Defining the two polynomials

$$P(s) = \prod_{i=1}^{L}(E_i - \nu_i s), \quad Q(s) = \sum_{i=1}^{L} d_i \prod_{j\neq i}(E_j - \nu_j s). \tag{3.19}$$

the exponential decays $\{\alpha_l\}_{l=1}^{L-1}$ of the memory kernel are defined as the roots of $Q(s)$ and the exponent coefficients $\{\beta_l\}_{l=1}^{L-1}$ are given by the residues around the poles of the rational function $P(s)/Q(s)$ which are all positive valued. Finally, the Volterra equation (3.16) *relating strain-to-stress is equivalent to the differential equation model*

$$\begin{aligned} \overline{\sigma}(t) &= E'_{\mathrm{pc}}\overline{\epsilon}(t) + \nu'_{\mathrm{pc}}\dot{\overline{\epsilon}}(t) - \langle \mathbb{1}_{L-1}, \xi(t)\rangle, && t \in \mathcal{T} \\ \dot{\xi}(t) &= -A\xi(t) + b\overline{\epsilon}(t), && t \in \mathcal{T} \\ \xi(0) &= 0 \end{aligned} \tag{3.20}$$

where the matrix A is diagonal with positive entries $\{\alpha_l\}_{l=1}^{L-1}$ and the coefficient vector $b = \{\beta_l\}_{l=1}^{L-1} \in \mathbb{R}_+^{L-1}$.

Next, we build on the preceding proposition to derive the form of the coefficients β_l explicitly, and then to conclude that the parameters of the differential equation (3.20) depend continuously on the piecewise-constant material parameterization.

Lemma 3.5. *The vector of coefficients $b = \{\beta_l\}_{l=1}^{L-1} \in \mathbb{R}_+^{L-1}$ from* (3.20) *are defined by the following closed form for the inverse of the components:*

$$\beta_l^{-1} = \sum_{i=1}^{L} \frac{d_i}{\nu_i} \cdot \frac{1}{\left(\frac{E_i}{\nu_i} - \alpha_l\right)^2}. \tag{3.21}$$

From this we conclude $(E'_{\mathrm{pc}}, \nu'_{\mathrm{pc}}, A, b)$ are continuous functions of the microstructural parameters $\{(d_i, E_i, \nu_i)\}_{i=1}^{L}$ provided these parameters are all strictly positive.

Proof. First, we establish expression (3.21). Taking the polynomials $P(s), Q(s)$ defined in Proposition 3.4, and recalling that the β_l are the residues around the poles of $P(s)/Q(s)$, we see that

$$\beta_l = \lim_{s\to\alpha_l} \frac{P(s)}{Q(s)}(s-\alpha_l) = -\lim_{s\to\alpha_l} \frac{\prod_{i=1}^L (E_i-\nu_i s)(\alpha_l - s)}{\sum_{i=1}^L d_i \prod_{j\neq i}(E_j-\nu_j s)}.$$

Now applying l'Hôpital's rule, we get that

$$\beta_l = -\lim_{s\to\alpha_l} \frac{\alpha_l - s}{\sum_{i=1}^L \frac{d_i}{E_i-\nu_i s}} = \frac{1}{\sum_{i=1}^L \frac{d_i\nu_i}{(E_i-\nu_i\alpha_l)^2}} = \frac{1}{\sum_{i=1}^L \frac{d_i}{\nu_i}\cdot\frac{1}{\left(\frac{E_i}{\nu_i}-\alpha_l\right)^2}}.$$

Note that the roots $\{\alpha_l\}_{l=1}^{L-1}$ are clearly continuous functions of the microstructural parameters $\{(d_i, E_i, \nu_i)\}_{i=1}^L$, as long as these parameters are strictly positive, because they depend continuously on the coefficients of the polynomial Q. Likewise, the coefficients $\{\beta_l\}_{l=1}^{L-1}$ and the Markovian parameters $E'_{\text{pc}}, \nu'_{\text{pc}}$ are continuous functions of the material parameters when these parameters are strictly positive. The continuity of the coefficients β_l is easy to see except at the possible poles of its denominator, where $\alpha_l = \frac{E_l}{\nu_l}$; but noting that β_l must tend to zero at such poles, as a function of the material parameters, establishes continuity there. ∎

Examination of Proposition 3.4 shows that the solution map $(\overline{\epsilon}; E, \nu) \mapsto \overline{\sigma}$ is invariant under permutation of the pieces in a piecewise-constant microstructure. This allows us to sort the pieces for mathematical convenience, as in the following lemma.

Lemma 3.6. *Sort the ratios $\{\frac{E_l}{\nu_l}\}_{l=1}^L$ in increasing order. Then the roots $\{\alpha_l\}_{l=1}^{L-1}$ may also be sorted in increasing order, and satisfy the bounds*

$$\frac{E_l}{\nu_l} \leq \alpha_l \leq \frac{E_{l+1}}{\nu_{l+1}}, \quad \ell \in [L-1]; \tag{3.22}$$

equality is achieved if and only if $\frac{E_l}{\nu_l} = \frac{E_{l+1}}{\nu_{l+1}}$. In this ordering it also follows that

$$\beta_l \leq \frac{1}{\sum_{i=1}^L \frac{d_i}{\nu_i}}\left(\frac{E_L}{\nu_L} - \frac{E_1}{\nu_1}\right)^2 \tag{3.23}$$

where $\{\frac{E_l}{\nu_l}\}_{l=1}^L$ are ordered increasingly.

Proof. First, to show the interleaving property of the roots, suppose we order the indices $l \in [L]$ in increasing order of $\frac{E_l}{\nu_l}$ and assume that these ratios are unique with no repetitions so they are strictly increasing. Recall the polynomial

$$Q(s) = \sum_{i=1}^L d_i \prod_{j\neq i}(E_j - \nu_j s) \tag{3.24}$$

of which the α_l are roots. Then we have that

$$Q\left(\frac{E_k}{\nu_k}\right) = \prod_{l=1}^L \nu_l \cdot \sum_{l=1}^L \frac{d_l}{\nu_l}\prod_{j\neq l}\left(\frac{E_j}{\nu_j} - \frac{E_k}{\nu_k}\right) = \prod_{l=1}^L \nu_l \cdot \frac{d_k}{\nu_k}\prod_{j\neq k}\left(\frac{E_j}{\nu_j} - \frac{E_k}{\nu_k}\right). \tag{3.25}$$

which implies that

$$\operatorname{sign}\left[Q\left(\frac{E_k}{\nu_k}\right)\right] = (-1)^{k-1} \tag{3.26}$$

Because the polynomial Q has L roots and alternates sign at every $\frac{E_l}{\nu_l}$ and must be nonzero at these points, it follows that the roots must lie strictly in between these points. Hence, we have that

$$\frac{E_l}{\nu_l} < \alpha_l < \frac{E_{l+1}}{\nu_{l+1}} \tag{3.27}$$

where the inequalities above are strict. Now suppose again that we have a list of unique ratios $\{\frac{E_l}{\nu_l}\}_{l=1}^K$ sorted in strictly nondecreasing order, but every element in this list is repeated $\mathcal{N}_l$ times such that $\sum_{l=1}^K \mathcal{N}_l = L$. Then, by factoring out the term $\prod_{l=1}^K (E_l/\nu_l - s)^{\mathcal{N}_l - 1}$ from $Q(s)$ and combining like terms, we arrive at a new polynomial of the same form as $Q(s)$ with all distinct ratios to which we can apply the previous argument above. This proves the interleaving property of the α_l roots. Using the expression for β_l derived in (3.21), we can also immediately establish the upper bound (3.23) assuming again that the ratios $\frac{E_l}{\nu_l}$ are sorted in increasing order. ∎

Lemmas 3.5 and 3.6 show that in the case of piecewise-constant materials, the coefficients of the differential equation constitutive law (3.20) depend continuously on the collection of microstructural parameters $\{(d_i, E_i, \nu_i)\}_{i=1}^L$ and are bounded. We use this fact to show that we can approximate the stress–strain dynamics of continuously varying materials by their piecewise-constant discretizations.

4. **Universal Approximation**

The central result of this section is a universal approximation theorem, for the homogenized stress-strain relation arising in one-dimensional Kelvin–Voigt viscoelasticity, within the class of FNM–RNO mappings. To achieve this, we first establish Lipschitz properties of the cell problem, with respect to its dependence on microstructural parameters; see Section 4.1. We then show that the homogenized constitutive law defined by (3.2) and (3.3) may be approximated by the homogenized constitutive law associated with a piecewise-constant approximation of the microstructure, in Section 4.2. This approximation incurs an error that is uniform across the class of microstructures we consider in Assumptions 3.2, crucially relying on the fact that they are of uniformly bounded variation. This result is then used, in Section 4.3, to establish a universal approximation theorem for our proposed architecture: for any error $\mathfrak{e} > 0$ there exists a choice of parameters in FNM–RNO that leads to an $\mathfrak{e}-$approximation of the map $\{\overline{\epsilon}, E, \nu\} \mapsto \overline{\sigma}$, uniformly across the set of strain and microstructural inputs specified in Assumptions 3.2.

4.1. **Lipschitz Regularity of Cell Problem**

Here we show the Lipschitz regularity of the Kelvin–Voigt cell problem (3.3). Let u_1, u_2 be solutions corresponding to microstructural parameters (E_1, ν_1) and (E_2, ν_2) respectively, both satisfying the conditions in Assumptions 3.2. We can write these cell problems as

$$\partial_y\Big(\nu_i(y)\partial_{yt}u_i(y,t) + E_i(y)\partial_y u_i(y,t)\Big) = \partial_y \sigma_i(y,t) = 0, \qquad (y,t) \in \Omega \times \mathcal{T} \tag{4.1a}$$

$$u_i(0,t) = 0,\ u_i(1,t) = \overline{\epsilon}(t), \qquad t \in \mathcal{T} \tag{4.1b}$$

$$u_i(y,0) = 0, \qquad y \in \Omega. \tag{4.1c}$$

where the strains corresponding to these stresses are given by

$$\sigma_i(y,t) = \nu_i(y)\partial_{yt}u_i(y,t) + E_i(y)\partial_y u_i(y,t), \quad (y,t) \in \Omega \times \mathcal{T} \tag{4.2}$$

Our goal is to bound the difference between the spatially averaged stresses $\overline{\sigma}_1 = \langle \sigma_1, \mathbb{1} \rangle$ and $\overline{\sigma}_2 = \langle \sigma_2, \mathbb{1} \rangle$ of these two cell problems based on the difference of their microstructural parameters. We do this by first bounding the distance between the solutions u_1, u_2 of these two PDEs. Taking the two equations (4.1a) satisfied by u_1 and u_2 we can rewrite them as

$$\partial_y\Big(\nu_1\partial_{yt}u_1 + E_1\partial_y u_1\Big) = 0$$

$$\partial_y\Big(\nu_1\partial_{yt}u_2 + E_1\partial_y u_2\Big) = \partial_y\Big((\nu_1 - \nu_2)\partial_{yt}u_2 + (E_1 - E_2)\partial_y u_2\Big).$$

Defining the difference functions

$$\gamma = u_1 - u_2, \quad \Delta E = E_1 - E_2, \quad \Delta\nu = \nu_1 - \nu_2, \quad g = \Delta\nu\partial_{yt}u_2 + \Delta E\partial_y u_2, \tag{4.3}$$

we can subtract the equations above to get

$$\partial_y\Big(\nu_1\partial_{yt}\gamma + E_1\partial_y\gamma\Big) = -\partial_y g.$$

Choosing any test function $\varphi \in H_0^1(\Omega)$ we can write the weak form of this PDE as

$$q_{\nu_1}(\partial_t\gamma, \varphi) + q_{E_1}(\gamma, \varphi) = -\langle g, \partial_y\varphi\rangle. \tag{4.4}$$

Now, we are ready to state the following Lipschitz bound on the difference between u_1 and u_2. Both cell problem solutions u_1, u_2 can be viewed as functions of time that map into $L^2(\Omega;\mathbb{R})$ so they live in the function space $\mathcal{Z} = L^\infty(\mathcal{T}; L^2(\Omega;\mathbb{R}))$. We will measure their difference under the norm $\|u\|_{\mathcal{Z}} = \operatorname{ess\,sup}_{t\in\mathcal{T}}(\|u(\,\cdot\,,t)\|_{L^2})$.

Lemma 4.1. *Let u_i be the solution to the cell problem* (4.1) *associated with microstructural properties (E_i, ν_i) for $i = 1, 2$ and a time-varying boundary condition $\bar{\epsilon}(t)$ satisfying Assumptions 3.2. Then we have the Lipschitz bound*

$$\|u_1 - u_2\|_{\mathcal{Z}} \le C_1\|\nu_1 - \nu_2\|_{L^2} + C_2\|E_1 - E_2\|_{L^2} \tag{4.5}$$

where the constants $C_1, C_2 > 0$ depend only on $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ and $\bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}$ and are independent of the time interval $\mathcal{T}$.

Proof. In Proposition B.3, we establish that

$$\sup_{t\in\mathcal{T}}\|\gamma\|_{H_0^1,\nu_1} \le \frac{\nu_{\max}}{E_{\min}}\frac{1}{\sqrt{\nu_{\min}}}\|g\|_{\mathcal{Z}}. \tag{4.6}$$

Combining this with Lemma 3.1 gives us

$$\sup_{t\in\mathcal{T}}\|\gamma\|_{H_0^1} \le \frac{\nu_{\max}}{E_{\min}\nu_{\min}}\|g\|_{\mathcal{Z}}.$$

By the Poincaré inequality and the definition of the norm on $\mathcal{Z}$ we have that $\|\gamma\|_{\mathcal{Z}} \le C_p \sup_{t\in\mathcal{T}}\|\gamma\|_{H_0^1}$ for some constant $C_p > 0$ and hence,

$$\|u_1 - u_2\|_{\mathcal{Z}} = \|\gamma\|_{\mathcal{Z}} \le C_p\frac{\nu_{\max}}{E_{\min}\nu_{\min}}\|g\|_{\mathcal{Z}}. \tag{4.7}$$

Thus we focus on bounding the norm of g. By Cauchy–Schwarz we write

$$\begin{aligned}
\|g\|_{\mathcal{Z}} &= \|\Delta\nu\partial_{yt}u_2 + \Delta E\partial_y u_2\|_{\mathcal{Z}} \\
&\le \sup_{t\in\mathcal{T}}\|\partial_{yt}u_2\|_{L^2}\|\Delta\nu\|_{L^2} + \sup_{t\in\mathcal{T}}\|\partial_y u_2\|_{L^2}\|\Delta E\|_{L^2} \\
&\le \sup_{t\in\mathcal{T}}\|\partial_t u_2\|_{H_0^1}\|\Delta\nu\|_{L^2} + \sup_{t\in\mathcal{T}}\|u_2\|_{H_0^1}\|\Delta E\|_{L^2} \\
&\le \frac{1}{\sqrt{\nu_{\min}}}\bigg(\sup_{t\in\mathcal{T}}\|\partial_t u_2\|_{H_0^1,\nu_2}\|\Delta\nu\|_{L^2} + \sup_{t\in\mathcal{T}}\|u_2\|_{H_0^1,\nu_2}\|\Delta E\|_{L^2}\bigg)
\end{aligned}$$

where the last line follows again from Lemma 3.1. In Corollary B.2, we bound the solution of the cell problem to show that $\sup_{t\in\mathcal{T}}\|u_2\|_{H_0^1,\nu_2}$ and $\sup_{t\in\mathcal{T}}\|\partial_t u_2\|_{H_0^1,\nu_2}$ are finite, which implies that

$$\|g\|_{\mathcal{Z}} \le C_1'\|\nu_1 - \nu_2\|_{L^2} + C_2'\|E_1 - E_2\|_{L^2}$$

for constants $C_1', C_2' > 0$ that depend only on $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ and $\bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}$. Finally, combining this with (4.7), gives us

$$\|u_1 - u_2\|_{\mathcal{Z}} \le C_1\|\nu_1 - \nu_2\|_{L^2} + C_2\|E_1 - E_2\|_{L^2} \tag{4.8}$$

where constants $C_1, C_2 > 0$ depend only on $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ and $\bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}$. ∎

We now use the lemma above to show that the stress resulting from the cell problem also satisfies Lipschitz regularity with respect to the microstructural parameters.

Lemma 4.2. *Let σ_i be the stress* (4.2) *resulting from the solution u_i of the cell problem* (4.1) *associated with microstructural properties (E_i, ν_i) for $i = 1, 2$ and a time-varying boundary condition $\bar{\epsilon}(t)$ satisfying Assumptions 3.2. Then we have the Lipschitz bound*

$$\|\sigma_1 - \sigma_2\|_{\mathcal{Z}} \leq C_1 \|\nu_1 - \nu_2\|_{L^2} + C_2 \|E_1 - E_2\|_{L^2} \tag{4.9}$$

where constants $C_1, C_2 > 0$ depend only on $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ and $\bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}$ and are independent of the time interval $\mathcal{T}$. Define the spatial averages of the two stresses as $\overline{\sigma}_1 = \langle \sigma_1, \mathbb{1} \rangle$ and $\overline{\sigma}_2 = \langle \sigma_2, \mathbb{1} \rangle$ where $\mathbb{1}$ is the constant function taking value one in Ω. Then

$$\|\overline{\sigma}_1 - \overline{\sigma}_2\|_{L^\infty} \leq C_1 \|\nu_1 - \nu_2\|_{L^2} + C_2 \|E_1 - E_2\|_{L^2}. \tag{4.10}$$

Proof. We define $\gamma, \Delta E, \Delta \nu$ and g as before and note that

$$\begin{aligned} \|\sigma_1 - \sigma_2\|_{\mathcal{Z}} &= \|\nu_1 \partial_{yt} \gamma + E_1 \partial_y \gamma\|_{\mathcal{Z}} + \|g\|_{\mathcal{Z}} \\ &\leq \nu_{\max} \|\partial_{yt} \gamma\|_{\mathcal{Z}} + E_{\max} \|\partial_y \gamma\|_{\mathcal{Z}} + \|g\|_{\mathcal{Z}} \\ &= \nu_{\max} \sup_{t \in \mathcal{T}} \|\partial_t \gamma\|_{H_0^1} + E_{\max} \sup_{t \in \mathcal{T}} \|\gamma\|_{H_0^1} + \|g\|_{\mathcal{Z}}. \end{aligned}$$

We prove in Proposition B.3 that $\sup_{t \in \mathcal{T}} \|\partial_t \gamma\|_{H_0^1}$ and $\sup_{t \in \mathcal{T}} \|\gamma\|_{H_0^1}$ can both be bounded by constant multiples of $\|g\|_{\mathcal{Z}}$. Hence, we can bound

$$\|\sigma_1 - \sigma_2\|_{\mathcal{Z}} \leq C \|g\|_{\mathcal{Z}} \tag{4.11}$$

for a constant $C > 0$ that depends only on $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ and $\bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}$. Finally, using the bound we derived on $\|g\|_{\mathcal{Z}}$ in the proof of Lemma 4.1, this shows that

$$\|\sigma_1 - \sigma_2\|_{\mathcal{Z}} \leq C_1 \|\nu_1 - \nu_2\|_{L^2} + C_2 \|E_1 - E_2\|_{L^2} \tag{4.12}$$

where $C_1, C_2 > 0$ are constants that depend only on $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ and $\bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}$. The desired result about spatial averages follows by noting that σ_i is in fact constant in Ω. ∎

Finally, these results prove that the strain-to-stress map Ψ_0 of a one-dimensional homogenized Kelvin–Voigt material is Lipschitz with respect to the microstructural parameters.

Corollary 4.3. *For any microstructural properties (E_1, ν_1) and (E_2, ν_2) along with average strain input $\bar{\epsilon}$ satisfying Assumptions 3.2, we have that*

$$\|\Psi_0(\bar{\epsilon}; E_1, \nu_1) - \Psi_0(\bar{\epsilon}; E_2, \nu_2)\|_{L^\infty} \leq C_1 \|\nu_1 - \nu_2\|_{L^2} + C_2 \|E_1 - E_2\|_{L^2}. \tag{4.13}$$

Remark 4.4. The bounding coefficients $C_1, C_2 > 0$ depend only on $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ and $\bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}$ and are independent of the time interval $\mathcal{T}$. These coefficients are not sharp but have the expected asymptotic behavior; namely, that in the limit, they increase when $E_{\max}, \nu_{\max}, \bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}$ tend to infinity and also increase when $E_{\min}, \nu_{\min}$ tend to zero. The scaling with $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ is expected as these parameters control the uniform ellipticity of the cell problem. The exact form of these coefficients can be traced by combining the upper bounds in Corollary B.2 and Proposition B.3 with the steps of Lemmas 4.1 and 4.2 above.

4.2. **Approximation of the PDE by Piecewise-Constant Problems**

In this section, we use the Lipschitz property of the cell problem derived above to show that the homogenized Kelvin–Voigt constitutive model can be well-approximated by a strain-to-stress map that takes the microstructural properties (E, ν), discretizes them into piecewise-constant functions,

and then applies the exact analytical formula for the piecewise-constant strain-to-stress map derived in Section 3.2. This strain-to-stress map is denoted by

$$\begin{aligned} \Psi^{\mathrm{pc}} : C^1(\mathcal{T};\mathbb{R}) \times \mathcal{M}^B_{E_{\min},E_{\max}} \times \mathcal{M}^B_{\nu_{\min},\nu_{\max}} &\longrightarrow C(\mathcal{T};\mathbb{R}), \\ (\{\overline{\epsilon}(t)\}_{t\in\mathcal{T}}, E, \nu) &\longmapsto \{\overline{\sigma}_{\mathrm{pc}}(t)\}_{t\in\mathcal{T}}, \end{aligned} \tag{4.14}$$

and is given by a differential equation model

$$\begin{aligned} \overline{\sigma}_{\mathrm{pc}}(t) &= F_{\mathrm{pc}}(\overline{\epsilon}(t), \dot{\overline{\epsilon}}(t), \xi(t); E, \nu) \\ \dot{\xi}_{\mathrm{pc}}(t) &= G_{\mathrm{pc}}(\overline{\epsilon}(t), \xi(t); E, \nu) \\ \xi_{\mathrm{pc}}(0) &= 0. \end{aligned} \tag{4.15}$$

Note that the resulting definition of Ψ^{pc} is then causal (see Remark 2.1). The average strain and strain rate are $\overline{\epsilon}(t), \dot{\overline{\epsilon}}(t) \in \mathbb{R}$ and the ODE evolves a set of n-dimensional internal variables $\xi_{\mathrm{pc}}(t) \in \mathbb{R}^n$. We define the maps

$$\begin{aligned} F_{\mathrm{pc}} : \mathbb{R}^{n+2} \times \mathcal{M}^B_{E_{\min},E_{\max}} \times \mathcal{M}^B_{\nu_{\min},\nu_{\max}} &\longrightarrow \mathbb{R} \\ G_{\mathrm{pc}} : \mathbb{R}^{n+1} \times \mathcal{M}^B_{E_{\min},E_{\max}} \times \mathcal{M}^B_{\nu_{\min},\nu_{\max}} &\longrightarrow \mathbb{R} \end{aligned} \tag{4.16}$$

which have the following semi-linear form:

$$F_{\mathrm{pc}}(\overline{\epsilon}(t), \dot{\overline{\epsilon}}(t), \xi(t); E, \nu) = E'_{\mathrm{pc}}(E,\nu)\overline{\epsilon}(t) + \nu'_{\mathrm{pc}}(E,\nu)\dot{\overline{\epsilon}}(t) - \langle \mathbb{1}, \xi(t)\rangle \tag{4.17a}$$

$$G_{\mathrm{pc}}(\overline{\epsilon}(t), \xi(t); E, \nu) = -A(E,\nu)\xi(t) + b(E,\nu)\overline{\epsilon}(t). \tag{4.17b}$$

The coefficients of $F_{\mathrm{pc}}, G_{\mathrm{pc}}$ act linearly on the strain $\overline{\epsilon}$, strain rate $\dot{\overline{\epsilon}}$, and hidden state ξ_{pc}. Because the entries of $A(E,\nu), b(E,\nu)$ are bounded (see (4.20)), the dynamics of ξ_{pc} in (4.17) are continuous and hence $\overline{\sigma}_{\mathrm{pc}} \in C(\mathcal{T};\mathbb{R})$ as expected.

For any microstructural properties $E, \nu : \mathbb{T} \to \mathbb{R}$ define its *piecewise-constant discretization* $E_{\mathrm{pc}}, \nu_{\mathrm{pc}} : \mathbb{T} \to \mathbb{R}$ with $n+1$ pieces as

$$\begin{aligned} E_{\mathrm{pc}}(y) = E_{\mathrm{pc}_i} &:= (n+1)\int_{y_i}^{y_{i+1}} E(z)\,\mathrm{d}z, \\ \nu_{\mathrm{pc}}(y) = \nu_{\mathrm{pc}_i} &= (n+1)\int_{y_i}^{y_{i+1}} \nu(z)\,\mathrm{d}z, \quad y \in [y_i, y_{i+1}) \end{aligned} \tag{4.18}$$

for $i \in [n+1]$ where $y_i = \frac{i-1}{n+1}$. The coefficient functions of the ODE model (4.17) are

$$\begin{aligned} A : \mathcal{M}^B_{E_{\min},E_{\max}} \times \mathcal{M}^B_{\nu_{\min},\nu_{\max}} &\longrightarrow \mathbb{D}^n_+ \\ b : \mathcal{M}^B_{E_{\min},E_{\max}} \times \mathcal{M}^B_{\nu_{\min},\nu_{\max}} &\longrightarrow \mathbb{R}^n_+ \\ E'_{\mathrm{pc}}, \nu'_{\mathrm{pc}} : \mathcal{M}^B_{E_{\min},E_{\max}} \times \mathcal{M}^B_{\nu_{\min},\nu_{\max}} &\longrightarrow \mathbb{R}_+, \end{aligned} \tag{4.19}$$

defined in Proposition 3.4 and Lemma 3.5 as continuous functions of the piecewise constants $\{E_{\mathrm{pc}_i}, \nu_{\mathrm{pc}_i}\}_{i=1}^{n+1}$, and hence implicitly, as continuous functions of the original microstructure $(E,\nu) \in \mathcal{M}^B_{E_{\min},E_{\max}} \times \mathcal{M}^B_{\nu_{\min},\nu_{\max}}$. Since the diagonal of A is nonnegative, the dynamics of (4.15) are stable. In fact, by Lemma 3.6 the diagonal entries of the matrix A and the entries of the vector b are bounded by

$$\frac{E_{\min}}{\nu_{\max}} \le \mathrm{diag}(A) \le \frac{E_{\max}}{\nu_{\min}}, \quad 0 \le b \le \nu_{\max}\left(\frac{E_{\max}}{\nu_{\min}} - \frac{E_{\min}}{\nu_{\max}}\right)^2. \tag{4.20}$$

Note that if E, ν are piecewise-constant with $n+1$ equisized pieces, in other words if $E = E_{\mathrm{pc}}$ and $\nu = \nu_{\mathrm{pc}}$, then the piecewise constant and true constitutive laws agree exactly

$$\Psi^{\mathrm{pc}}(\overline{\epsilon}; E_{\mathrm{pc}}, \nu_{\mathrm{pc}}) = \Psi_0(\overline{\epsilon}; E_{\mathrm{pc}}, \nu_{\mathrm{pc}})$$

as a result of Proposition 3.4 and Lemma 3.5. For a general microstructure, however, Ψ^{pc} will only serve as an approximation to the true constitutive law Ψ_0 of the material.

The basic idea behind using Ψ^{pc} to approximate the constitutive law of any microstructure is as follows. The homogenized constitutive model for any reasonable choice of microstructural properties (E, ν) can be approximated by the constitutive model arising from making a piecewise-constant approximation $E_{\text{pc}}, \nu_{\text{pc}}$ of the material properties. In fact, we show that Ψ^{pc} approximates Ψ_0 *uniformly* over all microstructures $E, \nu \in \mathcal{M}^B_{E_{\min}, E_{\max}} \times \mathcal{M}^B_{\nu_{\min}, \nu_{\max}}$ to arbitrary accuracy $\mathfrak{e}$, for choice of n sufficiently large.

Proposition 4.5. *For any microstructures (E, ν) and average strain input $\overline{\epsilon}$ satisfying Assumptions 3.2 and any $\mathfrak{e} > 0$, there exists a differential equation strain-to-stress map Ψ^{pc} defined by* (4.14) *with dimension $n = n(\mathfrak{e})$ such that it uniformly approximates the true strain-to-stress map Ψ_0 defined by* (3.3) *of the homogenized material to accuracy $\mathfrak{e}$, in the following sense:*

$$\sup_{\substack{E \in \mathcal{M}^B_{E_{\min}, E_{\max}} \\ \nu \in \mathcal{M}^B_{\nu_{\min}, \nu_{\max}}}} \sup_{\overline{\epsilon} \in \mathcal{C}_{\overline{\epsilon}_{\max}, \dot{\overline{\epsilon}}_{\max}}} \|\Psi^{\text{pc}}(\overline{\epsilon}; E, \nu) - \Psi_0(\overline{\epsilon}; E, \nu)\|_{L^\infty} < \mathfrak{e}. \tag{4.21}$$

Proof. We begin by taking any $E \in \mathcal{M}^B_{E_{\min}, E_{\max}}$, $\nu \in \mathcal{M}^B_{\nu_{\min}, \nu_{\max}}$ and $\overline{\epsilon} \in \mathcal{C}_{\overline{\epsilon}_{\max}, \dot{\overline{\epsilon}}_{\max}}$. Assume that the differential equation model Ψ^{pc} has hidden state dimension $n = n(\mathfrak{e})$ where we will specify the dependence of n on $\mathfrak{e}$ below. Because $\Psi^{\text{pc}}(\overline{\epsilon}; E_{\text{pc}}, \nu_{\text{pc}}) = \Psi_0(\overline{\epsilon}; E_{\text{pc}}, \nu_{\text{pc}})$, using Corollary 4.3, we now have that

$$\|\Psi^{\text{pc}}(\overline{\epsilon}; E_{\text{pc}}, \nu_{\text{pc}}) - \Psi_0(\overline{\epsilon}; E, \nu)\|_{L^\infty} \leq C_1 \|\nu_{\text{pc}} - \nu\|_{L^2} + C_2 \|E_{\text{pc}} - E\|_{L^2} \tag{4.22}$$

where $C_1, C_2 > 0$ are constants that depend on $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ and $\overline{\epsilon}_{\max}, \dot{\overline{\epsilon}}_{\max}$. Note that E, E_{pc} and ν, ν_{pc} are bounded so in particular

$$|\nu_{\text{pc}}(x) - \nu(x)| \leq \nu_{\max} - \nu_{\min}, \quad |E_{\text{pc}}(x) - E(x)| \leq E_{\max} - E_{\min}$$

for almost every $x \in \Omega$. This implies, by L^1-L^∞ interpolation of L^2, that

$$\begin{aligned} \|\nu_{\text{pc}} - \nu\|_{L^2} &\leq (\nu_{\max} - \nu_{\min})^{\frac{1}{2}} \|\nu_{\text{pc}} - \nu\|_{L^1}^{\frac{1}{2}}, \\ \|E_{\text{pc}} - E\|_{L^2} &\leq (E_{\max} - E_{\min})^{\frac{1}{2}} \|E_{\text{pc}} - E\|_{L^1}^{\frac{1}{2}}. \end{aligned} \tag{4.23}$$

Combining (4.22) with (4.23) gives us

$$\|\Psi^{\text{pc}}(\overline{\epsilon}; E_{\text{pc}}, \nu_{\text{pc}}) - \Psi_0(\overline{\epsilon}; E, \nu)\|_{L^\infty} \leq C_1 (\nu_{\max} - \nu_{\min})^{\frac{1}{2}} \|\nu_{\text{pc}} - \nu\|_{L^1}^{\frac{1}{2}} + C_2 (E_{\max} - E_{\min})^{\frac{1}{2}} \|E_{\text{pc}} - E\|_{L^1}^{\frac{1}{2}}.$$

As proven in Lemma D.1, piecewise constant functions with $n+1$ pieces can uniformly approximate functions of bounded variation with total variation at most B with error

$$\begin{aligned} \|\nu_{\text{pc}} - \nu\|_{L^1} &\leq \frac{B}{n+1} < \frac{1}{\nu_{\max} - \nu_{\min}} \Big(\frac{\mathfrak{e}}{2C_1} \Big)^2, \\ \|E_{\text{pc}} - E\|_{L^1} &\leq \frac{B}{n+1} < \frac{1}{E_{\max} - E_{\min}} \Big(\frac{\mathfrak{e}}{2C_2} \Big)^2 \end{aligned}$$

assuming we set $n = n(\mathfrak{e}) = C' \frac{B}{\mathfrak{e}^2} - 1$. Here C' is a constant that depends on $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ and $\overline{\epsilon}_{\max}, \dot{\overline{\epsilon}}_{\max}$. Combining these results gives us that

$$\|\Psi^{\text{pc}}(\overline{\epsilon}; E, \nu) - \Psi_0(\overline{\epsilon}; E, \nu)\|_{L^\infty} < \mathfrak{e}. \tag{4.24}$$

Because we proved this statement for any $E \in \mathcal{M}^B_{E_{\min}, E_{\max}}$, $\nu \in \mathcal{M}^B_{\nu_{\min}, \nu_{\max}}$ and $\overline{\epsilon} \in \mathcal{C}_{\overline{\epsilon}_{\max}, \dot{\overline{\epsilon}}_{\max}}$, our bound holds uniformly over this class. ∎

The fact that $(E, \nu) \in \mathcal{M}^B_{E_{\min}, E_{\max}} \times \mathcal{M}^B_{\nu_{\min}, \nu_{\max}}$ are functions of bounded variation B is necessary to show that our approximation guarantees hold uniformly over this class of functions as shown in the proof of Proposition 4.5. In the next section, we show that the right-hand sides of the differential equations in (4.17) are well approximated by Fourier neural mapping architectures. This relies on approximation theory results requiring that the set of material parameters E, ν entering the theorem

statement is compact in $L^2(\Omega)$; this is ensured by working in a subset of $L^2(\Omega)$ in which we have bounded variation.

4.3. **Approximation Through Neural Operators**

In this section, we combine the results from the previous section on piecewise-constant approximation, with approximation guarantees of Fourier neural operators, to prove that the homogenized constitutive law Ψ_0 of a multiscale KV material from (3.3) is well approximated by an FNM–RNO model $\Psi^{\text{FNM–RNO}}$.

Recall from Definition 2.2 that an FNM–RNO model

$$\begin{aligned}\Psi^{\text{FNM–RNO}} : C^1(\mathcal{T};\mathbb{R}) \times L^2(\mathbb{T};\mathbb{R}^2) &\longrightarrow C(\mathcal{T};\mathbb{R}),\\ (\{\overline{\epsilon}(t)\}_{t\in\mathcal{T}}; E, \nu) &\longmapsto \{\overline{\sigma}_{\text{RNO}}(t)\}_{t\in\mathcal{T}}\end{aligned}$$

with a one-dimensional Kelvin–Voigt microstructure $M = (E, \nu) \in L^2(\mathbb{T};\mathbb{R}^2)$ is given by

$$\begin{aligned}\overline{\sigma}_{\text{RNO}}(t) &= F_{\text{FNM}}(\overline{\epsilon}(t), \dot{\overline{\epsilon}}(t), \xi_{\text{RNO}}(t); E, \nu)\\ \dot{\xi}_{\text{RNO}}(t) &= G_{\text{FNM}}(\overline{\epsilon}(t), \xi_{\text{RNO}}(t); E, \nu), \quad \xi_{\text{RNO}}(0) = 0,\end{aligned} \tag{4.25}$$

where F_{FNM} and G_{FNM} are two FNMs as defined in Definition C.1.

Note that the differential equation model Ψ^{pc} (4.15) is of the same form as the FNM–RNO model $\Psi^{\text{FNM–RNO}}$ (4.25), and the right-hand sides $F_{\text{pc}}, G_{\text{pc}}$ of this ODE model can be approximated by $F_{\text{FNM}}, G_{\text{FNM}}$ if designed appropriately. To show this formally, we use the universal approximation theorem for Fourier neural mappings, an important result in neural operator theory [29, 34, 35]. We rely on the compact embedding of the space of bounded BV functions in L^2, used in prior works on approximation theory of FNMs for elliptic homogenization problems [6, Lemma C.1]. We combine these results here to obtain the universal approximation result for our FNM architecture, proven in detail in Proposition D.3.

Proposition 4.6. *Under Assumptions 3.2, choosing any $\overline{\epsilon}_{\max}, \dot{\overline{\epsilon}}_{\max}, \xi_{\max} > 0$ and $\mathfrak{e}_F, \mathfrak{e}_G > 0$, there exist FNMs F_{FNM} and G_{FNM}, in* (2.3)*, that approximate F_{pc} and G_{pc} of* (4.17) *such that*

$$\sup_{\substack{|z_1|\le\overline{\epsilon}_{\max}, |z_2|\le\dot{\overline{\epsilon}}_{\max}, \|z_3\|_\infty\le\xi_{\max},\\ E\in\mathcal{M}^B_{E_{\min},E_{\max}}, \nu\in\mathcal{M}^B_{\nu_{\min},\nu_{\max}}}} |F_{\text{FNM}}(z_1, z_2, z_3; E, \nu) - F_{\text{pc}}(z_1, z_2, z_3; E, \nu)| < \mathfrak{e}_F \tag{4.26}$$

$$\sup_{\substack{|z_1|\le\overline{\epsilon}_{\max}, \|z_3\|_\infty\le\xi_{\max},\\ E\in\mathcal{M}^B_{E_{\min},E_{\max}}, \nu\in\mathcal{M}^B_{\nu_{\min},\nu_{\max}}}} \|G_{\text{FNM}}(z_1, z_3; E, \nu) - G_{\text{pc}}(z_1, z_3; E, \nu)\|_\infty < \mathfrak{e}_G. \tag{4.27}$$

Now we are ready to prove that for a general class of microstructures (E, ν) and strain histories $\overline{\epsilon}$, the true constitutive law Ψ_0 from (3.3) can be approximated by Ψ^{pc} which can in turn be approximated by $\Psi^{\text{FNM–RNO}}$.

Theorem 4.7 (FNM-RNO Universal Approximation)**.** *For any materials E, ν and strain $\overline{\epsilon}$ satisfying Assumptions 3.2 and any $\mathfrak{e} > 0$, there exist FNMs F_{FNM} and G_{FNM} such that the map*

$$\Psi^{\text{FNM–RNO}} : (\{\overline{\epsilon}(t)\}_{t\in\mathcal{T}}; E, \nu) \longmapsto \{\overline{\sigma}_{\text{RNO}}(t)\}_{t\in\mathcal{T}}$$

defined by (2.3) *uniformly approximates the map*

$$\Psi_0 : (\{\overline{\epsilon}(t)\}_{t\in\mathcal{T}}; E, \nu) \longmapsto \{\overline{\sigma}(t)\}_{t\in\mathcal{T}}$$

defined by (3.3) *to accuracy $\mathfrak{e}$, in the following sense:*

$$\sup_{\substack{E\in\mathcal{M}^B_{E_{\min},E_{\max}}\\ \nu\in\mathcal{M}^B_{\nu_{\min},\nu_{\max}}}} \sup_{\overline{\epsilon}\in\mathcal{C}_{\overline{\epsilon}_{\max},\dot{\overline{\epsilon}}_{\max}}} \|\Psi^{\text{FNM–RNO}}(\overline{\epsilon}; E, \nu) - \Psi_0(\overline{\epsilon}; E, \nu)\|_{L^\infty(\mathcal{T})} < \mathfrak{e}. \tag{4.28}$$

Proof. Assume throughout this proof that $E \in \mathcal{M}^B_{E_{\min},E_{\max}}$, $\nu \in \mathcal{M}^B_{\nu_{\min},\nu_{\max}}$ and $\overline{\epsilon} \in \mathcal{C}_{\overline{\epsilon}_{\max},\dot{\overline{\epsilon}}_{\max}}$. The bounds derived below hold uniformly over all functions in these classes. Denote the true strain-to-stress map defined by the cell problem (3.3) as $\Psi_0 : (\{\overline{\epsilon}(t)\}_{t\in\mathcal{T}}; E, \nu) \mapsto \{\overline{\sigma}(t)\}_{t\in\mathcal{T}}$. We know by Proposition 4.5 that

$$\|\Psi^{\mathrm{pc}}(\overline{\epsilon}; E, \nu) - \Psi_0(\overline{\epsilon}; E, \nu)\|_{L^\infty} = \|\overline{\sigma}_{\mathrm{pc}} - \overline{\sigma}\|_{L^\infty} < \frac{\mathfrak{e}}{2} \tag{4.29}$$

as long as the dimension $n = n(\mathfrak{e}/2)$ in (4.15) is taken sufficiently large. Now we study the error between $\Psi^{\text{FNM–RNO}}$ and Ψ^{pc}. To do this, we first need to show that the trajectories of the hidden variables ξ_{pc} in (4.15) and ξ_{RNO} in (4.25) stay in a bounded domain so that we can apply well-established FNM universal approximation results. First studying ξ_{pc} note that

$$\dot{\xi}_{\mathrm{pc}}(t) \leq -A(E,\nu)\xi_{\mathrm{pc}}(t) + b(E,\nu)\overline{\epsilon}_{\max}.$$

Since A is a diagonal matrix with strictly positive entries, using (4.20) we can apply Gronwall's inequality to each entry of ξ_{pc}, using that $\xi_{\mathrm{pc}}(0) = 0$, to write

$$\xi_{\mathrm{pc}}(t) \leq A(E,\nu)^{-1}b(E,\nu)\overline{\epsilon}_{\max} \leq \overline{\epsilon}_{\max}\frac{\nu_{\max}^2}{E_{\min}}\left(\frac{E_{\max}}{\nu_{\min}} - \frac{E_{\min}}{\nu_{\max}}\right)^2$$

where the bound above is interpreted element-wise. In the preceding inequality, we have used the lower and upper bounds on A and b, respectively, derived in Proposition 4.5. We can derive the same bound for $-\xi_{\mathrm{pc}}$ through a similar application of Gronwall's inequality, which proves that

$$\sup_{t\in\mathcal{T}} \|\xi_{\mathrm{pc}}(t)\|_\infty \leq \overline{\epsilon}_{\max}\frac{\nu_{\max}^2}{E_{\min}}\left(\frac{E_{\max}}{\nu_{\min}} - \frac{E_{\min}}{\nu_{\max}}\right)^2. \tag{4.30}$$

The next step is to bound the norm of ξ_{RNO}, which we do by a bootstrap argument. First by FNM universal approximation Proposition 4.6, for any $\xi_{\max} > 0$ and $\mathfrak{e}_G > 0$ there exists an FNM G_{FNM} such that

$$\|G_{\mathrm{pc}}(\overline{\epsilon}, \xi; E, \nu) - G_{\mathrm{FNM}}(\overline{\epsilon}, \xi; E, \nu)\|_\infty < \mathfrak{e}_G$$

assuming that $\|\xi\|_\infty \leq \xi_{\max}$. For reasons evident below, we choose

$$\xi_{\max} = \overline{\epsilon}_{\max}\frac{\nu_{\max}^2}{E_{\min}}\left(\frac{E_{\max}}{\nu_{\min}} - \frac{E_{\min}}{\nu_{\max}}\right)^2 + \frac{\nu_{\max}}{E_{\min}}$$

and $\mathfrak{e}_G$ sufficiently smaller than one. Note that $\xi_{\mathrm{RNO}}(0) = 0$, so it is initialized in the set where this error bound holds, and its dynamics are continuous in time since the Fourier neural mapping G_{FNM} is Lipschitz in the variable ξ. We explicitly derive the form of this Lipschitz constant in Lemma D.5. Now we can write

$$\begin{aligned}\dot{\xi}_{\mathrm{RNO}}(t) &= [G_{\mathrm{FNM}}(\overline{\epsilon}, \xi_{\mathrm{RNO}}; E, \nu) - G_{\mathrm{pc}}(\overline{\epsilon}, \xi_{\mathrm{RNO}}; E, \nu)] + G_{\mathrm{pc}}(\overline{\epsilon}, \xi_{\mathrm{RNO}}; E, \nu) \\ &\leq -A(E,\nu)\xi_{\mathrm{RNO}} + (b(E,\nu)\overline{\epsilon}_{\max} + \mathbb{1}_n\mathfrak{e}_G)\end{aligned}$$

where the upper bound above should be interpreted in an element-wise sense. By the same application of Gronwall's inequality above, we obtain

$$\xi_{\mathrm{RNO}}(t) \leq A(E,\nu)^{-1}(b(E,\nu)\overline{\epsilon}_{\max} + \mathbb{1}_n\mathfrak{e}_G) \leq \overline{\epsilon}_{\max}\frac{\nu_{\max}^2}{E_{\min}}\left(\frac{E_{\max}}{\nu_{\min}} - \frac{E_{\min}}{\nu_{\max}}\right)^2 + \mathfrak{e}_G\frac{\nu_{\max}}{E_{\min}}.$$

Applying the same bound to $-\xi_{\mathrm{RNO}}$ we achieve

$$\sup_{t\in\mathcal{T}} \|\xi_{\mathrm{RNO}}(t)\|_\infty \leq \overline{\epsilon}_{\max}\frac{\nu_{\max}^2}{E_{\min}}\left(\frac{E_{\max}}{\nu_{\min}} - \frac{E_{\min}}{\nu_{\max}}\right)^2 + \mathfrak{e}_G\frac{\nu_{\max}}{E_{\min}} \tag{4.31}$$

proving that $\|\xi_{\mathrm{RNO}}(t)\|_\infty \leq \xi_{\max}$ for all time. To summarize, we have now shown that

$$\sup_{t\in\mathcal{T}} \|\xi_{\mathrm{pc}}(t)\|_\infty, \sup_{t\in\mathcal{T}} \|\xi_{\mathrm{RNO}}(t)\|_\infty < \xi_{\max} = \overline{\epsilon}_{\max}\frac{\nu_{\max}^2}{E_{\min}}\left(\frac{E_{\max}}{\nu_{\min}} - \frac{E_{\min}}{\nu_{\max}}\right)^2 + \frac{\nu_{\max}}{E_{\min}}. \tag{4.32}$$

We now bound the difference between the trajectories of $\xi_{\mathrm{pc}}, \xi_{\mathrm{RNO}}$ given that we have established that they stay in bounded sets where our FNM approximation guarantees apply. We write

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Big(\xi_{\mathrm{RNO}}(t) - \xi_{\mathrm{pc}}(t)\Big) &= [G_{\mathrm{FNM}}(\bar{\epsilon}, \xi_{\mathrm{RNO}}; E, \nu) - G_{\mathrm{pc}}(\bar{\epsilon}, \xi_{\mathrm{RNO}}; E, \nu)] \\ &\quad + [G_{\mathrm{pc}}(\bar{\epsilon}, \xi_{\mathrm{RNO}}; E, \nu) - G_{\mathrm{pc}}(\bar{\epsilon}, \xi_{\mathrm{pc}}; E, \nu)] \\ &\leq -A(E,\nu)\Big(\xi_{\mathrm{RNO}}(t) - \xi_{\mathrm{pc}}(t)\Big) + \mathfrak{e}_G \mathbb{1}_n\end{aligned}$$

where again the bound above should be interpreted element-wise. Gronwall's inequality now gives us

$$\xi_{\mathrm{RNO}}(t) - \xi_{\mathrm{pc}}(t) \leq A(E,\nu)^{-1}\mathbb{1}_n \mathfrak{e}_G \leq \mathbb{1}_n \frac{\nu_{\max}}{E_{\min}} \mathfrak{e}_G$$

The same upper bound can be proven for $\xi_{\mathrm{pc}} - \xi_{\mathrm{RNO}}$ hence showing that

$$\sup_{t\in\mathcal{T}} \|\xi_{\mathrm{RNO}}(t) - \xi_{\mathrm{pc}}(t)\|_\infty \leq \frac{\nu_{\max}}{E_{\min}} \mathfrak{e}_G. \tag{4.33}$$

Lastly, again invoking FNM universal approximation Proposition 4.6 there exists an FNM F_{FNM} such that

$$|F_{\mathrm{FNM}}(\bar{\epsilon}, \dot{\bar{\epsilon}}, \xi_{\mathrm{RNO}}; E, \nu) - F_{\mathrm{pc}}(\bar{\epsilon}, \dot{\bar{\epsilon}}, \xi_{\mathrm{RNO}}; E, \nu)| < \mathfrak{e}_F$$

for any small $\mathfrak{e}_F > 0$. Because F_{pc} is linear in ξ, we further have by the triangle inequality that

$$|F_{\mathrm{pc}}(\bar{\epsilon}, \dot{\bar{\epsilon}}, \xi_{\mathrm{RNO}}; E, \nu) - F_{\mathrm{pc}}(\bar{\epsilon}, \dot{\bar{\epsilon}}, \xi_{\mathrm{pc}}; E, \nu)| = |\langle \mathbb{1}_n, \xi_{\mathrm{RNO}} - \xi_{\mathrm{pc}}\rangle| \leq n(\mathfrak{e})\|\xi_{\mathrm{RNO}} - \xi_{\mathrm{pc}}\|_\infty \leq n(\mathfrak{e})\frac{\nu_{\max}}{E_{\min}}\mathfrak{e}_G$$

where in the last inequality we used the bound derived in (4.33). By a final application of the triangle inequality, we have

$$\begin{aligned}|\overline{\sigma}_{\mathrm{RNO}}(t) - \overline{\sigma}_{\mathrm{pc}}(t)| &\leq |F_{\mathrm{FNM}}(\bar{\epsilon}, \dot{\bar{\epsilon}}, \xi_{\mathrm{RNO}}; E, \nu) - F_{\mathrm{pc}}(\bar{\epsilon}, \dot{\bar{\epsilon}}, \xi_{\mathrm{RNO}}; E, \nu)| \\ &\quad + |F_{\mathrm{pc}}(\bar{\epsilon}, \dot{\bar{\epsilon}}, \xi_{\mathrm{RNO}}; E, \nu) - F_{\mathrm{pc}}(\bar{\epsilon}, \dot{\bar{\epsilon}}, \xi_{\mathrm{pc}}; E, \nu)| \\ &\leq \mathfrak{e}_F + n(\mathfrak{e})\frac{\nu_{\max}}{E_{\min}}\mathfrak{e}_G.\end{aligned}$$

which proves the bound

$$\|\Psi^{\mathrm{FNM\text{–}RNO}}(\bar{\epsilon}; E, \nu) - \Psi^{\mathrm{pc}}(\bar{\epsilon}; E, \nu)\|_{L^\infty} = \|\overline{\sigma}_{\mathrm{RNO}} - \overline{\sigma}_{\mathrm{pc}}\|_{L^\infty} < \frac{\mathfrak{e}}{2} \tag{4.34}$$

by choosing $\mathfrak{e}_F, \mathfrak{e}_G$ sufficiently small. Finally, by combining (4.29) and (4.34) through a triangle inequality we get the desired bound

$$\|\Psi^{\mathrm{FNM\text{–}RNO}}(\bar{\epsilon}; E, \nu) - \Psi_0(\bar{\epsilon}; E, \nu)\|_{L^\infty} < \mathfrak{e}. \tag{4.35}$$

■

This proves the main theoretical result of our paper, namely that the homogenized constitutive law of the one-dimensional Kelvin–Voigt model can be approximated by an FNM–RNO architecture uniformly over a large class of microstructures and strain inputs. The theorem justifies considering the FNM–RNO more generally, beyond the specifics of linear Kelvin–Voigt viscoelasticity, an avenue we pursue further in the next section on numerical experiments.

Remark 4.8. The assumption on the regularity of microstructure properties, i.e., in BV and L^∞, is physically benign. Commonly observed microstructures (e.g., laminates, fiber-reinforced composites, and polycrystalline materials) satisfy this assumption. Mathematically, showing boundedness in this class is generally easier than showing compactness in L^2, which is the requirement for universal approximation. For microstructures outside of the class $\mathsf{BV} \cap L^\infty$ (e.g., inclusions with fractal interfaces), one would need to show compactness in L^2 directly.

5. Numerical Experiments

In this section, we apply our proposed recurrent neural operator architecture to learn and deploy homogenized constitutive laws of viscoelastic and elasto-viscoplastic materials. We initially consider linear viscoelasticity with piecewise-constant microstructures of varying numbers of pieces, then design and study high-memory continuous microstructures. We first discuss, in Section 5.1, our data generation procedure for sampling these microstructures and our choice of strain trajectories. In Section 5.2, we give further details of our FNM–RNO architecture and model training. We then demonstrate that the ability of our neural operator to encode memory in the strain-to-stress relationship allows us to improve significantly over memory-less models. Our numerical results are shown in Section 5.3, where our architecture is tested on the multiscale Kelvin–Voigt cell problem and is then used within homogenized macroscale simulations. In Section 5.4, we show that the same model can be used to learn the constitutive law of elasto-viscoplastic materials. Taken together, the experiments demonstrate that our approach applies to different constitutive models and generalizes across a wide array of microstructures and strain inputs.

5.1. Dataset Generation

The dataset for our FNM–RNO architecture consists of n_s microstructures, averaged strain trajectories, and averaged stress trajectories $\{E^{(j)}, \nu^{(j)}, \overline{\epsilon}^{(j)}, \overline{\sigma}^{(j)}\}_{j=1}^{n_s}$. We consider two methods for producing joint samples of E and ν that yield piecewise-constant (PC) random microstructures and high-memory continuous (HMC) random microstructures. The strain trajectories $\overline{\epsilon}^{(j)}$ are independently sampled following the procedure in [7, 42, 65] which is detailed below. Given these samples, we solve the cell problem in (3.3) for the averaged stress $\overline{\sigma}^{(j)}$ on a uniform grid using linear Lagrange finite elements with 501 spatial degrees of freedom (DoFs) and a 4th order explicit Runge–Kutta method with $5,001$ temporal DoFs. The solver is implemented using the FEniCS [4, 45] and hIPPYlib [63] libraries.

We first detail the sampling procedures for piecewise-constant and high-memory continuous microstructures and then describe the construction of the average strains; taken together, these define the data sets used later for training and testing. Further testing of generalization with respect to strain trajectories is implicit in the macroscale calculations, also presented later, since these generate strains that are not in our training set.

Piecewise-Constant Random Microstructures. The piecewise constant random microstructure properties $(E,\ \nu)$ are generated to be spatially periodic and share the same set of jump discontinuities. The number of constant pieces L is selected uniformly at random from 5–20. Locations of the discontinuities are drawn at random from the finite set $\{0.02k\}_{k=0}^{50}$, with equal probability and with replacement; this leads to a minimum length of 0.02 for each piece. The values of E and ν in each piece are sampled from a uniform distribution on $[0.1, 1]^2$.

High-Memory Continuous Random Microstructures. We construct High-Memory Continuous (HMC) microstructure properties $(E,\ \nu)$ to again be spatially-periodic. They are designed by taking samples from a periodic random mean shift $m \in C_{\text{per}}(\Omega; \mathbb{R}^2)$ and a periodic centered Gaussian random function $g \in C_{\text{per}}(\Omega; \mathbb{R}^2)$,

$$\begin{bmatrix} E(y) \\ \nu(y) \end{bmatrix} = 0.45 \times \Big(\text{erf}\Big(m(y) + g(y)\Big) + 1\Big) + 0.1, \tag{5.1}$$

where erf is the error function that smoothly enforces E, ν to be bounded in $[0.1, 1]$.

We design $m : \Omega \to \mathbb{R}^2$ using a piecewise-constant random microstructure that has large contrasts between the magnitude of its first and second coordinates, hence resulting in a large contrast between

the elasticity E and viscosity ν. The choice of this mean function typically corresponds to viscoelastic materials with large memory kernels, as it leads to large exponential weights β_l, as given by (3.21), of the memory kernel derived in Lemma 3.5. Specifically, m is generated by sampling its two piecewise-constant pieces from a Gaussian mixture distribution with two equally-weighted modes centered at $[-1, 1]$ and $[1, -1]$ with a small covariance $0.06I$. Lastly, m is turned into a continuous function by applying a spatial Gaussian convolution with a standard deviation of 0.01 to each component.

The random perturbation $g : \Omega \to \mathbb{R}^2$ is sampled from a centered Gaussian distribution with a diagonal covariance matrix with entries $\rho^{(k)}{\sigma^{(k)}}^2(1 - {\rho^{(k)}}^2\partial_x^2)^{-2}$, where $k = 1, 2$ indicates the E or ν component and $\rho^{(k)}$ and $\sigma^{(k)}$ represent correlation length and pointwise standard deviation. The statistics $\rho^{(k)}, \sigma^{(k)}$ are sampled i.i.d. from a reciprocal distribution on $[0.01, 0.3]$ and a uniform distribution on $[0.1, 0.3]$ respectively.

Averaged Strain Trajectories. The averaged strain trajectories $\bar{\epsilon}$ are generated by first randomly picking a total number of time points $3 \le n \le 21$ with $0 = t_1 < \cdots < t_n = T = 1$, where the internal time points $\{t_k\}_{k=2}^{n-1}$ are uniformly randomly placed in $[0, 1]$. At each time point, we assign its averaged strain value $\bar{\epsilon}(t_k)$ by first randomly picking a sign $v_k \in \{-1, 1\}$ and then taking

$$\bar{\epsilon}(t_k) = \bar{\epsilon}(t_{k-1}) + (0.5 - \bar{\epsilon}(t_{k-1}))v_k\sqrt{t_k - t_{k-1}} \tag{5.2}$$

where we initialize $\bar{\epsilon}(0) = 0$. We use piecewise-cubic Hermite interpolating polynomials (PCHIP) to create the averaged strain trajectories from these points.

Samples of the piecewise-constant (PC) microstructure dataset and the high-memory continuous (HMC) microstructure dataset are provided in Figure 5.1. We visualize the averaged stress response with and without memory effects. The stress response without the memory effects is given by (3.12) with $K \equiv 0$. In Figure 5.1, the stress responses without memory for PC sample #2 and HMC sample #3 show large discrepancies in comparison to the stress response with memory, demonstrating the importance of modeling memory effects for those material and strain trajectory inputs.

5.2. **Architecture and Training**

We now describe the architecture of the FNM–RNO from Section 4.3 introduced in Definition 2.2, which is used in the following experiments. The internal variable size L of our RNO differential equation is varied through all values in the set $\{1, 3, 5, 10, 15, 20, 25\}$. We simplify the architecture by letting our vector-to-function map $\mathscr{D}$ and the vector lifting function S_v be the identity. We use spatial coordinates as an additional channel in our function input, alongside the microstructure properties (E, ν), as a form of positional encoding. Hence, the number of functional inputs into our architecture is $d_{\text{in}}^f = 3$. The function lifting layer leads to hidden channels of size $d_0 = 32$. We use 3 Fourier layers with the same hidden channel size $d_t = 32$ for $t = 0, 1, 2$. Each layer uses Gaussian error linear unit activations and 4 Fourier modes to parameterize all the convolution operators. The output dimension of the functional layers is $d_{\text{proj}}^{fv} = 64$.

We consider the loss function given by a squared relative L^2 error with a penalty term. Let $\overline{\sigma}_{\text{RNO}}^{(j)}$ denote the FNM–RNO prediction of the averaged stress trajectories for the data sample $E^{(j)}, \nu^{(j)}, \overline{\epsilon}^{(j)}$ and $\overline{\sigma}^{(j)}$. The prediction depends on the learnable parameters of F_{FNM} and G_{FNM}, and these parameters are found by minimizing the following loss function:

$$\text{Loss}(\{\overline{\sigma}^{(j)}\}_{j=1}^{n_{\text{train}}}, \{\overline{\sigma}_{\text{RNO}}^{(j)}\}_{j=1}^{n_{\text{train}}}) = \frac{1}{n_{\text{train}}}\sum_{j=1}^{n_{\text{train}}}\Big(\mathcal{E}(\overline{\sigma}^{(j)}, \overline{\sigma}_{\text{RNO}}^{(j)})^2 + \|G_{\text{FNM}}(0, 0; E^{(j)}, \nu^{(j)})\|^2\Big), \tag{5.3}$$

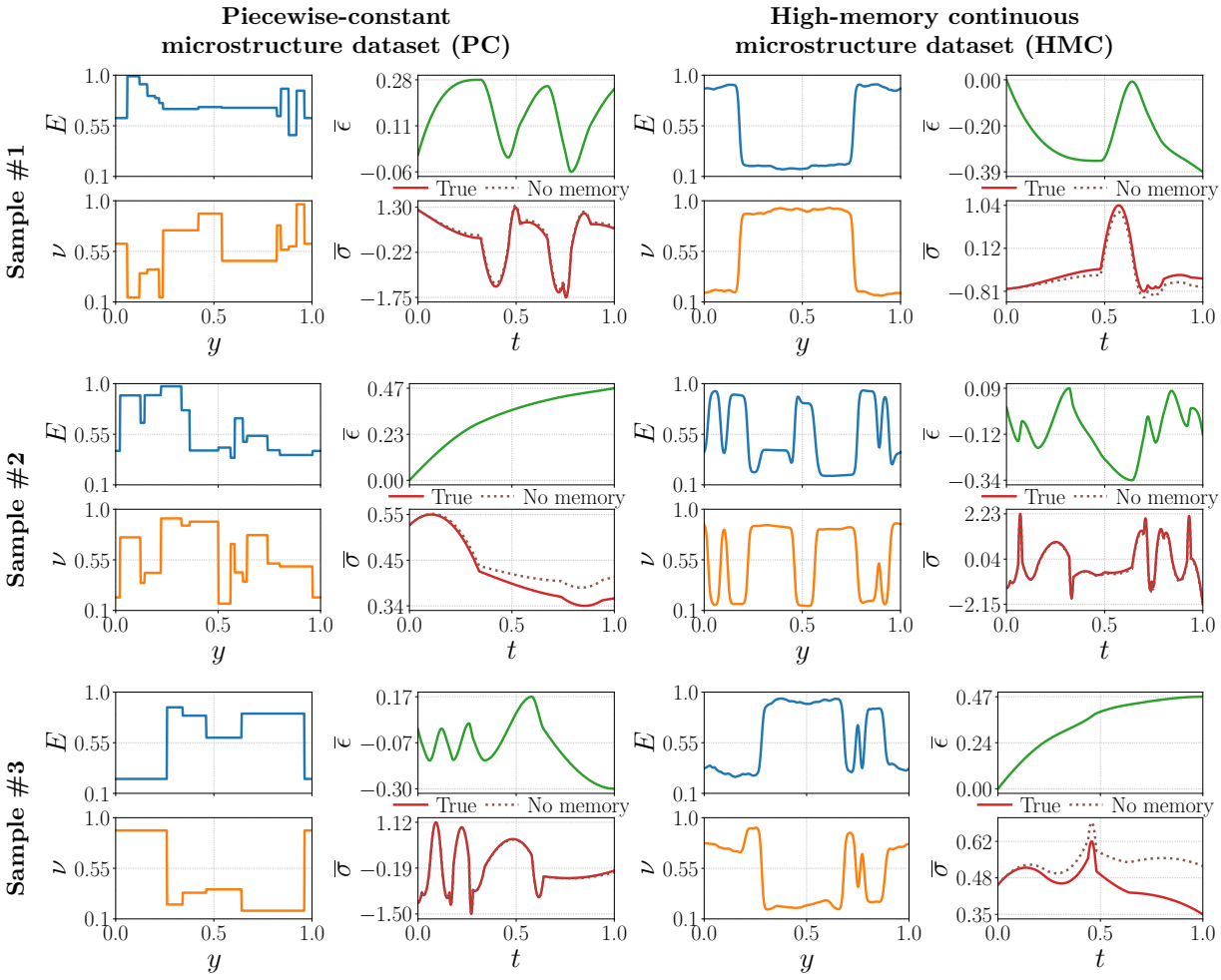


Figure 5.1. Visualization of samples from the two datasets: piecewise-constant microstructure (PC) and high-memory continuous microstructure (HMC); see Section 5.1. Each dataset consists of microstructure samples $(E^{(j)}, \nu^{(j)})$, averaged strain trajectory samples $\overline{\epsilon}^{(j)}$, and the averaged stress trajectory samples $\overline{\sigma}^{(j)}$. We visualize the averaged stress response with (*solid lines*) and without (*dotted lines*) memory effects.

where $\mathcal{E}$ returns the relative L^2 error in the average stress

$$\mathcal{E}(\overline{\sigma}^{\dagger}, \overline{\sigma}) = \left(\frac{\int_0^T \left| \overline{\sigma}^{\dagger}(t) - \overline{\sigma}(t) \right|^2 \mathrm{d}t}{\int_0^T \left| \overline{\sigma}^{\dagger}(t) \right|^2 \mathrm{d}t} \right)^{1/2}. \tag{5.4}$$

where our stress trajectories are simulated up to time $T = 1$. The penalty term is included in the loss function because we find that the physical constraint $G_{\mathrm{FNM}}(0, 0; E^{(j)}, \nu^{(j)}) = 0$ is usually not learned without the penalty, which leads to a large error in averaged stress predictions when the material has no deformation history. See a detailed discussion on the penalty term in Appendix E.

Remark 5.1 (On the Mismatch in Norms)**.** Our universal approximation result in Theorem 4.7 is established in the L^∞ norm. The training of our model is based on the relative L^2 error (5.4), which assuming the stress dynamics is of constant order, is equivalent to the time-averaged L^2 norm

$\frac{1}{T^{\frac{1}{2}}}\|\cdot\|_{L^2}$. Since the time-averaged L^2 norm is dominated by the L^∞ norm, any model achieving uniform accuracy guarantees small relative L^2 error. The training process, therefore, searches for these accurate models using the relative L^2 metric as a computationally tractable proxy, though attaining them is not guaranteed in practice due to the complexity of stochastic, nonlinear optimization.

We use $n_{\text{train}} = 2,049$ samples from the PC dataset to train 7 FNM–RNOs with varying numbers of the internal variables L, with resolutions of the training data reduced to 251 spatial and 501 temporal DoFs. The PC dataset is used for training as it leads to better model generalization. We use the HMC dataset for testing only. To evaluate the loss function, we use the forward Euler scheme to estimate $\sigma^{(j)}_{\text{RNO}}$ and the trapezoidal rule to estimate the temporal integration. We use the Adam optimizer with a learning rate of 10^{-3}, a batch size of 32, a total of 500 epochs, and cosine annealing of the learning rate that decays to 10^{-5}. The training is implemented in PyTorch [53].

5.3. **Numerical Results**

High Generalization Accuracy. We evaluate the trained FNM–RNOs on 2,500 testing samples from both the PC and HMC datasets. In Figure 5.2, we present the distributions of the relative L^2 error for the FNM–RNO stress response alongside the error for the linear stress response excluding memory effects. Additionally, we visualize the testing samples with the largest and median errors for the FNM–RNO using five internal variables in Figure 5.3.

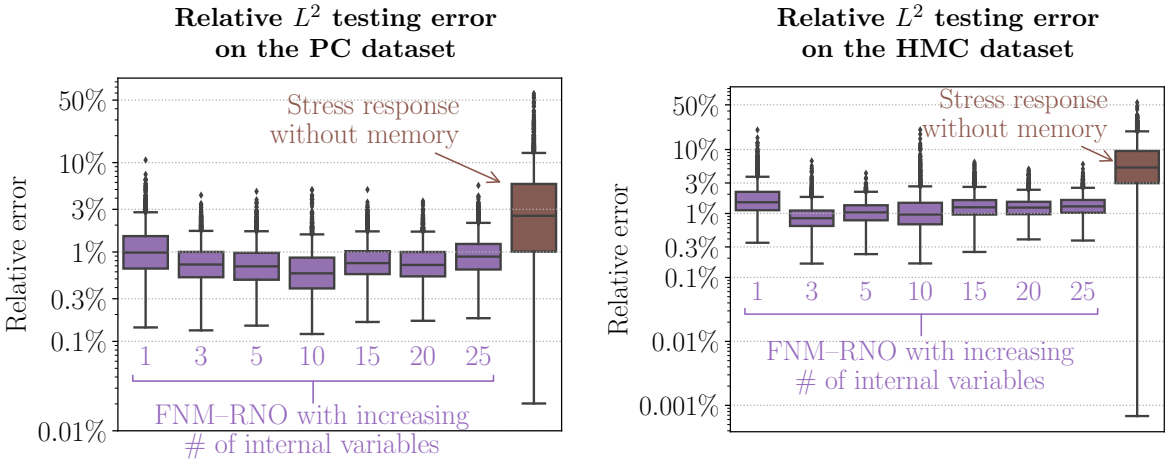


Figure 5.2. The distributions of the relative L^2 error on 2,500 testing samples from the PC dataset (*left*) and the HMC dataset (*right*). We visualize errors in FNM–RNO predictions, where the trained FNM–RNOs have varying numbers of internal variables. We also visualize the distribution of error given by the linear stress response without memory effects, where the response function is obtained using (3.12) with $K \equiv 0$.

The FNM–RNOs achieve consistently low relative L^2 testing errors on both datasets, with mean errors of 0.7%–1.2% for the PC dataset and 0.9%–1.9% for the HMC dataset. On average, the FNM–RNO stress response considerably outperforms the linear stress response without memory effects, which exhibits mean errors of 4.5% and 7.3% for the PC and HMC datasets, respectively. For testing samples with large FNM–RNO stress response errors, the stress response without memory typically shows a much greater discrepancy from the true stress response; see, e.g., Figure 5.3.

Furthermore, the generalization accuracy of FNM–RNO improves with the inclusion of additional internal variables; however, no significant improvement beyond three internal variables is observed. We

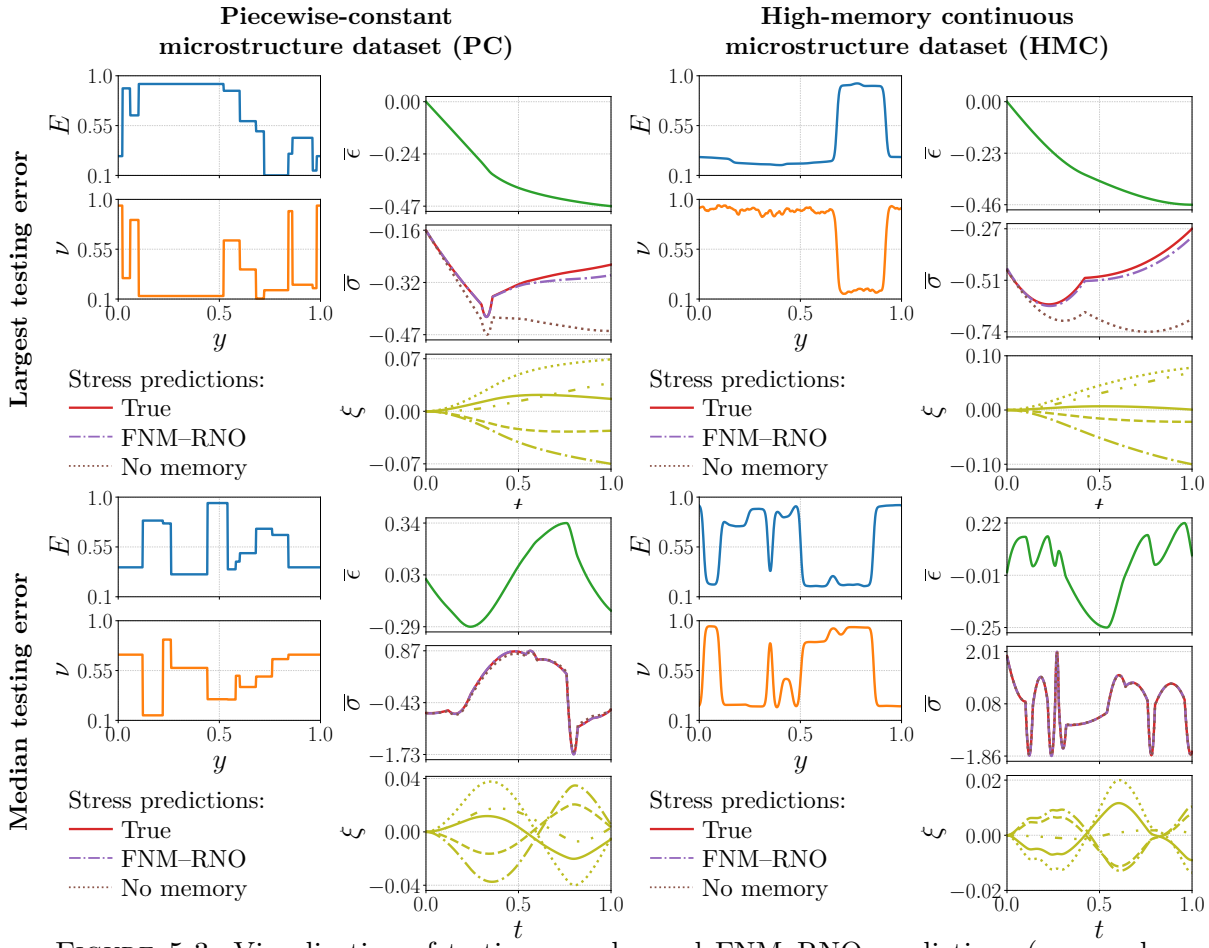


Figure 5.3. Visualization of testing samples and FNM–RNO predictions (averaged stress $\overline{\sigma}$ and internal variables ξ) with the largest and median relative L^2 testing error. We consider the trained FNM–RNO with 5 internal variables. We visualize the averaged stress response with (*solid lines*) and without (*dotted lines*) memory effects along with the FNM–RNO prediction (*dash-dot lines*).

note that the analytical form of the RNO for the PC data set requires at least 20 internal variables, as shown in Proposition 3.4, whereas the FNM–RNO architecture does not discover this, likely because training is harder with more internal variables.

Finally, we present a sample-complexity study in Figure 5.4. The results show that the FNM–RNO continues to improve in generalization accuracy as the number of training samples increases.

Discretization Agnostic. The FNM–RNO constitutive model can be trained on data with one set of spatial and temporal resolutions and used to predict at another. To illustrate this property, we evaluate the trained FNM–RNO model on testing samples with varying spatial and temporal resolutions, and the resulting mean relative L^2 testing errors are visualized in Figure 5.5. Testing samples with different resolutions for the HMC materials and the averaged stress trajectories are generated using linear interpolation.

The results show that the accuracy of FNM–RNO predictions is relatively sensitive to changes in temporal resolution, primarily due to truncation errors in estimating the evolution of internal variables. For the PC dataset, the testing error shows less sensitivity to changes in spatial resolution, since the

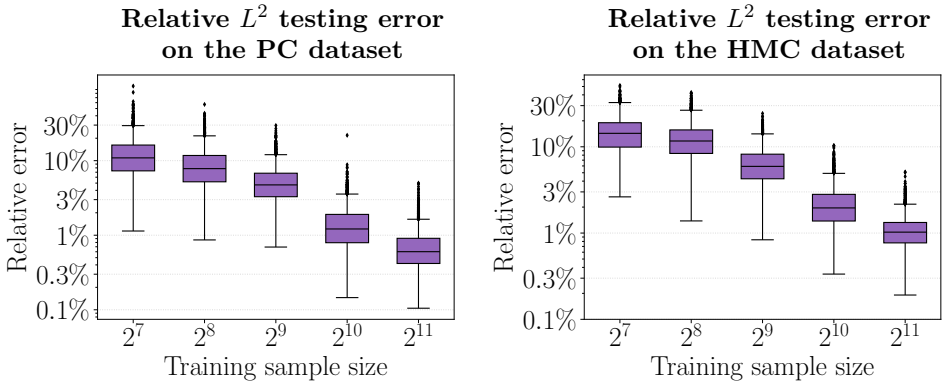


FIGURE 5.4. The sample complexity of the FNM–RNO trained on the PC dataset. We show the distributions of the relative L^2 testing errors as a function of the number of training samples. The relative errors are computed using 2,500 testing samples from the PC dataset (*left*) and the HMC dataset (*right*). The FNM–RNO uses five internal variables.

discontinuity points are located on a low-resolution grid. In contrast, for the HMC dataset, the testing error increases with decreased spatial resolution when the temporal resolution is high. Overall, the FNM–RNO can predict stress response on finer spatial and temporal resolutions than those used for training without significant deterioration in accuracy.

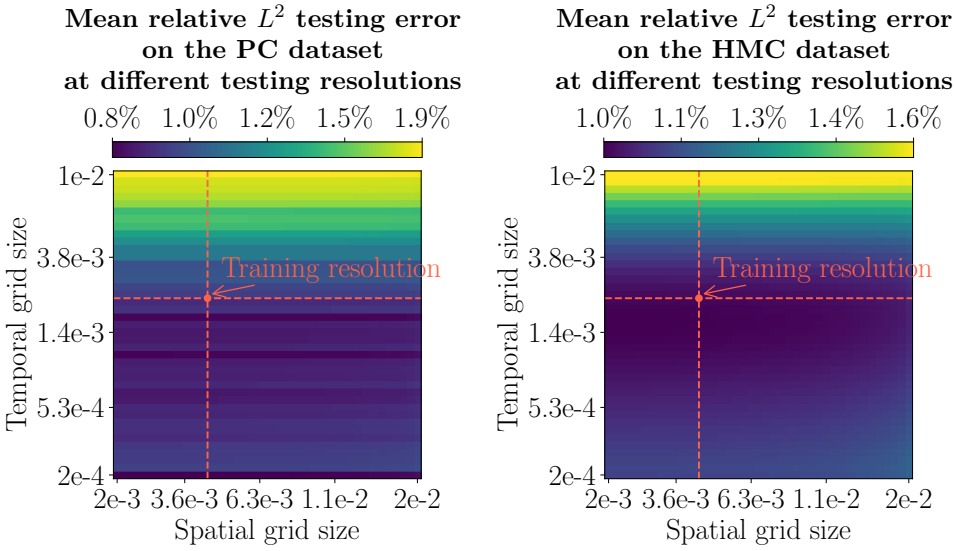


FIGURE 5.5. The mean relative L^2 error on the PC (*left*) and HMC (*right*) testing samples at different testing resolutions for the trained FNM–RNO with 5 internal variables.

Deployment in Macroscale Simulations. We consider a macroscale problem on domain $\mathcal{D}\times\mathcal{T} = [0,1]\times[0,1]$ and employ a body forcing term $b(x,t) = 100\sin(8\pi(x+t))$. This leads to a stress evolution given by $\sigma(x,t) = -800\cos(8\pi(x+t))$. Our goal is to solve for the displacement $u(x,t)$ with the

boundary conditions $u(0,t) = u(1,t) = 0$ for all $t \in [0,1]$ and the initial condition $u(x,0) = 0$ for all $x \in [0,1]$. We use four different constitutive models to solve for the displacement:

(i) Homogenized stress response using a memory kernel form as in (3.12),

(ii) Multiscale stress response with $\varepsilon^{-1} = 5, 10, 20, 40$ or 80,

(iii) Homogenized stress response without memory effects; $K \equiv 0$ in (3.12),

(iv) FNM–RNO stress response with 5 internal variables.

We expect that the multiscale solution converges to the homogenized solution in L^2 as $\varepsilon \to 0$ for each microstructure, whereas the FNM–RNO stress response and the stress response without memory lead to biased macroscale solutions relative to the homogenized stress response. The multiscale stress response (ii) and the memory-free approach (iii) serve as simpler comparative baselines for FNM-RNO.

The macroscale solvers have the following specifications. All problems are solved on a uniform grid using linear Lagrange finite elements. The multiscale and homogenized solutions have $500/\varepsilon$ and 4×10^5 spatial DoFs and use an explicit Euler method for time stepping. The FNM–RNO solutions have 250 spatial DoFs and employ an operator-splitting method that treats the internal variables explicitly and the displacements implicitly, with L-BFGS used to solve the nonlinear equations. All solutions have 501 temporal DoFs.

In Figure 5.6, we visualize the distributions of the relative L^2 error in macroscale solutions using 800 microstructure samples from the HMC dataset, where the macroscale solutions obtained by the homogenized stress response are used as the reference. The multiscale solutions linearly converge to the homogenized solutions as $\varepsilon \to 0$, and the FNM–RNO solutions have an error distribution similar to that arising from the multiscale solutions at $\varepsilon^{-1} = 20$. The stress response without memory leads to a higher average macroscale solution error than FNM–RNO and a lower average than multiscale simulation with $\varepsilon^{-1} = 10$. In Figure 5.7, we visualize a microstructure and its macroscale solutions. This sample corresponds to the median of the FNM–RNO macroscale solution error distribution shown in Figure 5.6. We also visualize the pointwise absolute error of macroscale solutions relative to the homogenized stress response. The multiscale stress response does not exhibit noticeable accumulation of error in time. The FNM–RNO solution accumulates errors over time, but the overall error is less than that from linear stress response without memory effects.

Remark 5.2 (On Computational Costs)**.** For the specific macroscale problems we considered, the FNM–RNO solutions are slower to compute than the multiscale solutions at a similar accuracy ($\varepsilon = 20$). The reasons are as follows. First, the solution of a linear time-evolving PDE can be accelerated substantially by amortizing the cost of solving the same linear systems for time-stepping, e.g., by reusing matrix factorizations. Note that this advantage is unavailable for nonlinear constitutive models. Second, we consider macroscale problems with no inertia, which is consistent with our theoretical settings but requires the displacement field to be computed iteratively at each time step. We expect our learned model to outperform the multiscale solvers in 2D and 3D for nonlinear constitutive models.

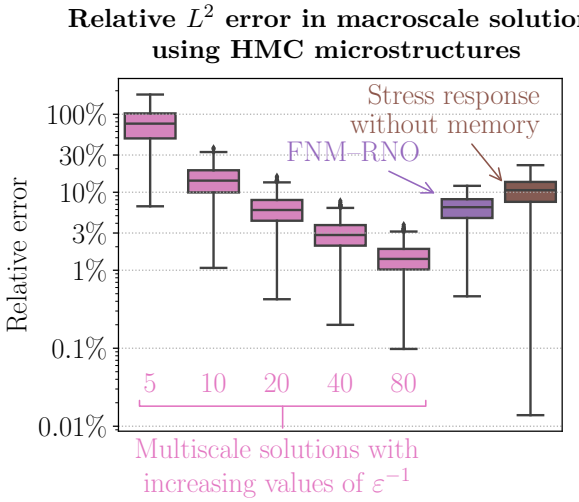


FIGURE 5.6. The distributions of the relative L^2 error in macroscale solutions obtained by different constitutive models using 800 microstructure samples from the HMC dataset. The error is computed relative to the macroscale solutions obtained by the homogenized stress response using memory kernels.

5.4. **Application to Elasto-Viscoplasticity**

We consider elasto-viscoplastic composites in one dimension. The cell problem is given by

$$
\begin{aligned}
&-\partial_y \sigma(y,t) = 0, && (y,t) \in \Omega \times \mathcal{T}, && (5.5a)\\
&\sigma(y,t) = E(y)(\partial_y u(y,t) - \epsilon_p(y,t)), && (y,t) \in \Omega \times \mathcal{T}, && (5.5b)\\
&\partial_t \epsilon_p(y,t) = \dot{\epsilon}_{p0}(y)\,\mathrm{sign}(\sigma(y,t))\left(\frac{|\sigma(y,t)|}{\sigma_Y(y)}\right)^{n(y)}, && (y,t) \in \Omega \times \mathcal{T}, && (5.5c)\\
&u(0,t) = 0,\ u(1,t) = \bar{\epsilon}(t), && t \in \mathcal{T}, && (5.5d)\\
&u(y,0) = 0,\ \epsilon_p(y,0) = 0, && y \in \Omega. && (5.5e)
\end{aligned}
$$

Here ϵ_p is the plastic strain, E is the Young's modulus, $\dot{\epsilon}_{p0}$ is the rate constant, σ_Y is the yield stress, and n is the rate exponent. We assume that these four microstructural properties $(E, \dot{\epsilon}_{p0}, \sigma_Y, n)$ vary spatially in the unit cell. Our goal is to learn the constitutive model $\{\{\bar{\epsilon}(t)\}_{t\in\mathcal{T}}, E, \dot{\epsilon}_{p0}, \sigma_Y, n\} \mapsto \{\bar{\sigma}(t)\}_{t\in\mathcal{T}}$, where $\bar{\epsilon}(t) = \int_\Omega \partial_y u(y,t)\,\mathrm{d}y$ and $\bar{\sigma}(t) = \int_\Omega \sigma(y,t)\,\mathrm{d}y$. We highlight the fact that the constitutive model can be expressed using the averaged plastic strain $\bar{\epsilon}_p(t) = \int_\Omega \epsilon_p(y,t)\,\mathrm{d}y$ as an internal variable; see [42, Eq. 11]

We generate two datasets following a strategy similar to that described in Section 5.1. The piecewise-constant random microstructure (PC-EVP) uses a uniform distribution on $[1,10]\times[0.5,2.0]\times[0.1,1.0]\times[1,20]$, independently drawn in each of the four components of the materials property vector $(E^{(j)}, \dot{\epsilon}_{p0}^{(j)}, \sigma_Y^{(j)}, n^{(j)})$, for each piece with label j, and drawn i.i.d. with respect to j. Continuous random microstructure (C-EVP) uses the spatially smooth piecewise-constant random microstructure as the random mean function. The sampling procedure for the mean is similar to HMC microstructures in Section 5.1, except that the values taken in each piece for the four microstructural properties are independently distributed.

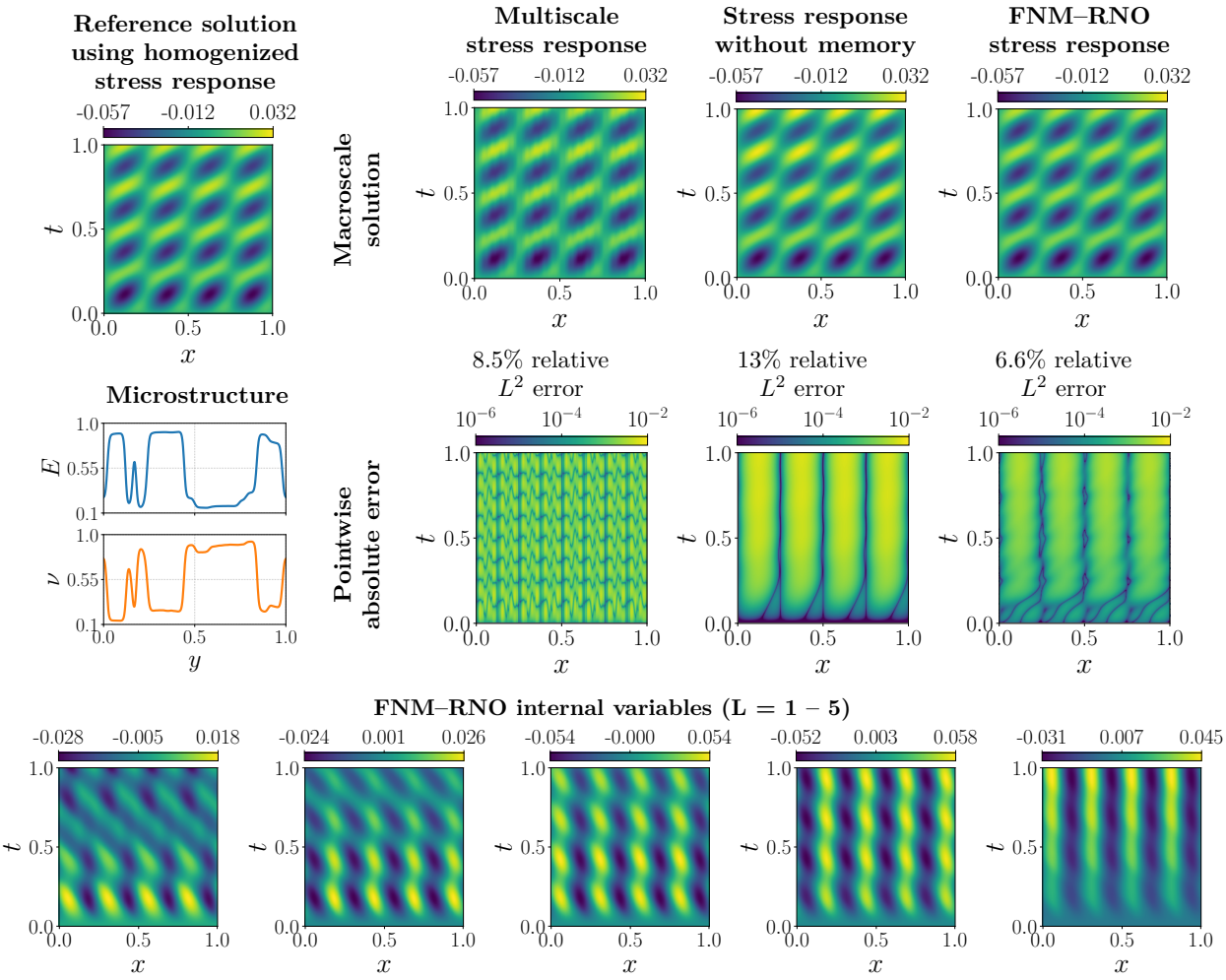


Figure 5.7. Visualization of macroscale solutions using different constitutive models for an HMC microstructure sample. We visualize the multiscale solution with $\varepsilon^{-1} = 20$. This sample corresponds to the median of the FNM–RNO macroscale solution relative error distribution in Figure 5.6.

We consider an FNM–RNO architecture given by

$$\overline{\sigma}(t) = F_{\text{FNM}}(\overline{\epsilon}(t), \xi(t); E, \dot{\epsilon}_{p0}, \sigma_Y, n) \tag{5.6a}$$

$$\dot{\xi}(t) = G_{\text{FNM}}(\overline{\epsilon}(t), \xi(t); E, \dot{\epsilon}_{p0}, \sigma_Y, n) \tag{5.6b}$$

$$\xi(0) = 0. \tag{5.6c}$$

Note that, comparing with (2.4), we have suppressed dependence on $\dot{\overline{\epsilon}}(t)$ in F_{FNM}, motivated by the analysis in [42]. The same work motivates an expectation that the internal variable should be scalar. Indeed we expect it to follow $\bar{\epsilon}_p(t) \approx c(\bar{\epsilon}, E, \dot{\epsilon}_{p0}, \sigma_Y, n)\xi(t)$, where c is a scalar-valued function that can be found numerically for each set of materials and averaged strain trajectory. We use a similar FNM architecture and training procedure as in Section 5.2, except that: (i) 2 Fourier modes are used to parameterize all the convolution operators; and (ii) the penalty term in the loss function is not included.

We evaluate the trained FNM–RNO on 2,500 testing samples from the PC-EVP and C-EVP datasets. The distributions of the relative L^2 testing error are shown in Figure 5.8. For the PC-EVP dataset, the FNM–RNO achieves mean relative L^2 errors of 3.4% in predicting the averaged stress and 1.4% in predicting the averaged plastic strain up to a constant. For the C-EVP dataset, the FNM–RNO achieves mean relative L^2 errors of 2.8% in predicting the averaged stress and 1.4% in predicting the averaged plastic strain up to a constant. In Figure 5.9, we visualize the testing samples in the PC-EVP and C-EVP datasets with the largest and median relative L^2 error in the averaged stress.

Finally, we present a sample-complexity study in Figure 5.10. The results show that the FNM–RNO continues to improve in generalization accuracy as the number of training samples increases.

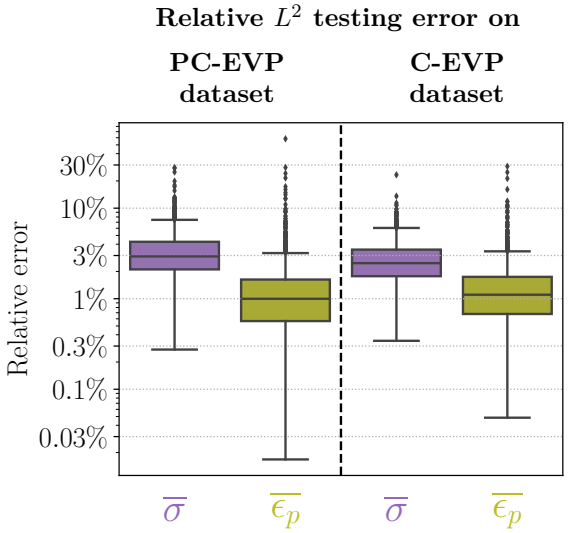


FIGURE 5.8. The distributions of the relative L^2 error on 2,500 testing samples from the PC-EVP dataset (*left*) and the C-EVP dataset (*right*) for elasto-viscoplasticity. We consider the error in the trained FNM–RNO predictions of (i) the averaged stress $\overline{\sigma}$ and (ii) the averaged plastic strain $\overline{\epsilon}_p$ up to a multiplicative constant via the internal variable ξ.

6. Conclusion and Outlook

In this paper, we present a novel Markovian recurrent and Fourier neural operator architecture capable of predicting the memory-dependent constitutive laws of homogenized multiscale materials over a wide array of microstructures. Our architecture is designed as a neural differential equation with a Fourier neural mapping on the right-hand side, making it agnostic to the level of discretization or sampling of the microstructure. Guided by the theory of homogenization in the one-dimensional Kelvin–Voigt model, we derive Lipschitz properties of the cell problem and, to our knowledge, present the first universal approximation theorem for a data-driven model that predicts a viscoelastic constitutive law as a function of the microstructure. This theorem is established for a wide class of one-dimensional spatially varying microstructural properties with uniform boundedness, uniform coercivity, and uniformly bounded variation, as well as for a wide class of strain trajectories with uniformly bounded magnitude and rate. Our numerical experiments confirm, in the context of the multiscale Kelvin–Voigt linear viscoelastic model, that the neural operator accurately predicts the homogenized dynamics of

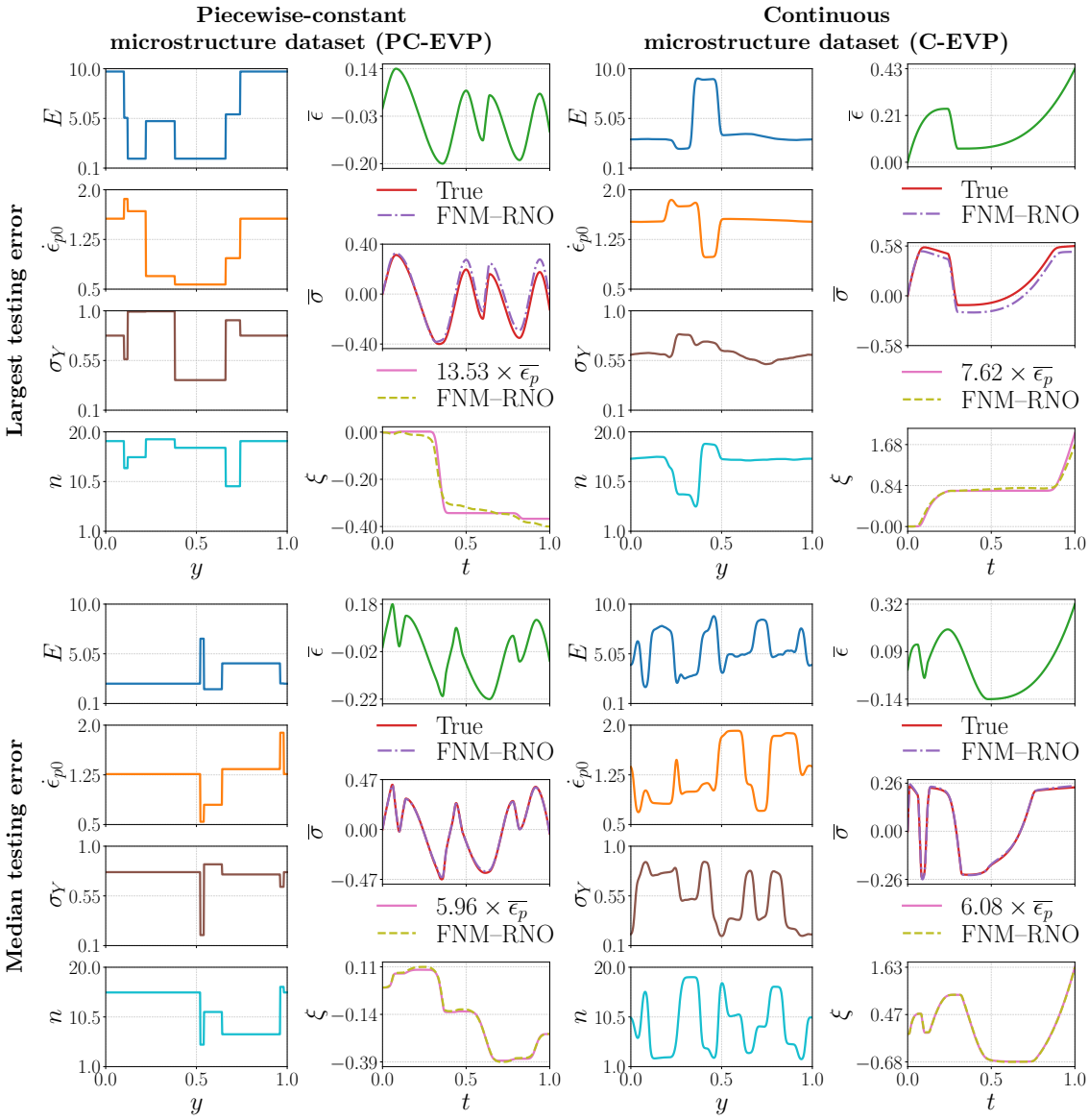


Figure 5.9. Visualization of testing samples and FNM–RNO predictions (averaged stress $\overline{\sigma}$ and internal variables ξ) with the largest and median relative L^2 error in stress response for elasto-viscoplasticity. We visualize the internal variable ξ (*dashed line*) along with the plastic strain $\overline{\epsilon}_p$ scaled by a constant (*solid line*), where the constant is determined through minimizing the L^2 distance between the internal variable and the scaled averaged plastic strain.

the material and generalizes from training on piecewise-constant microstructures to testing on continuous microstructures. We also show that our architecture can be utilized beyond the specific case of one-dimensional linear viscoelasticity where the theory directly applies: we demonstrate its efficacy in learning a nonlinear constitutive model for homogenized viscoplastic materials.

While our framework serves as the first step towards learning memory- and microstructure-dependent constitutive models with theoretical certification, several avenues for improvements and

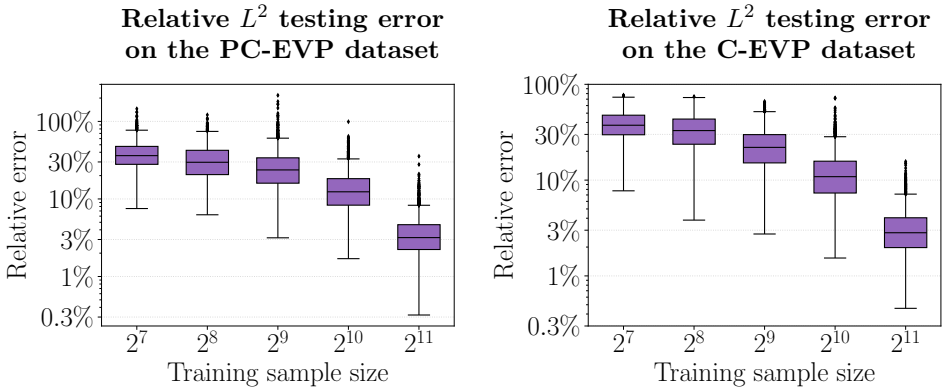


Figure 5.10. The sample complexity of the FNM–RNO trained on the PC-EVP dataset. We show the distributions of the relative L^2 testing errors as a function of the number of training samples. The relative errors are computed using 2,500 testing samples from the PC-EVP dataset (*left*) and the C-EVP dataset (*right*).

extensions remain. We first discuss possible theoretical extensions of our universal approximation theory, followed by several important directions for further numerical experimentation.

Extending the Universal Approximation Result. The universal approximation theory developed in this paper was limited to the two-scale Kelvin–Voigt model in one dimension. We remark that all results presented in Section 4 on the Lipschitz continuity of the cell problem and the universal approximation of the FNM-RNO architecture can be extended to higher dimensions, with the added difficulty of working with symmetrized strain and stress tensors (e.g., using Korn's inequality) and requiring standard coercivity conditions on the elasticity and viscosity tensors. Our results could also be applied directly to one-dimensional Maxwell materials, with the modification that the homogenized constitutive law would not depend on strain rate. The main complexity in generalizing our universal approximation theory to higher dimensions lies in the analysis of Section 3, which relied on the analytical form of the homogenized constitutive law, and specifically that a *dense* class of microstructures (piecewise constant) had homogenized constitutive laws that could be written as finite-dimensional ODEs. We relied on the fact that the ODE coefficients depend continuously on the microstructure, making their approximation by neural networks possible. In two and three dimensions, laminar materials, dual-multiphase composites, and Hashin–Shtrikman spherical assemblies are large classes of microstructures with closed-form homogenization, but they do not approximate the space of all microstructures. Hence, our analysis in Section 3 could be used to prove universal approximation theory for these special classes of microstructures, but a general result would not be readily attainable. A more direct route would require building on the operator theoretic arguments of [21, 61] to prove that the homogenized constitutive law of any coercive and bounded variation microstructure is governed by an integral equation whose coefficients and memory kernel are continuous in the microstructure. This line of proof could extend our theoretical results to higher dimensions.

Multiscale Material Models Beyond 1D. For future work, we aim to apply Markovian recurrent and Fourier neural operators to learn homogenized constitutive laws for 2D and 3D viscoelasticity and elasto-viscoplasticity in more practical, realistic settings. In particular, the computational efficiency of macroscale solvers using learned models remains to be thoroughly tested. Additionally, advanced training techniques, such as adjoint training [12] and tensor factorization [33], are necessary in 2D and 3D due to the high memory burden.

Thermodynamic Consistency and Stability. In the current model construction, thermodynamic consistency and stability are not strictly enforced. Since the training datasets are generated from homogenized responses of thermodynamically consistent, stable models, the learned models can be expected to closely follow these properties for loading paths and microstructures well represented in the training datasets. Our numerical results provide empirical evidence that the learned model satisfies these properties beyond the training datasets that consist of piecewise-constant microstructures. For viscoelasticity, macroscale simulations with continuous microstructures were solved robustly without numerical instability. For elasto-viscoplasticity, we observe that the internal variable is identifiable with high accuracy up to a scalar factor for continuous microstructures, consistent with the expected invariance of a thermodynamically consistent model. Nevertheless, these results do not rigorously validate the thermodynamic consistency or stability of the learned model. Enforcing them reliably (e.g., through the energetic formulation [20, 46]) remains an important direction for future work.

Memory Representations. Lastly, an important open mathematical and experimental problem [52] is to determine in what cases memory and fractional derivatives are necessary to model homogenized constitutive laws in higher dimensions. More generally, learning microstructure-dependent constitutive models now allows us to investigate which microstructural properties give rise to distinct features of homogenized constitutive laws.

Appendix A. Equivalence between Cell Problems

Here, we show that our original cell problem (3.3) can be derived from the classical cell problem of a viscoelastic material [7, 54] with a periodic boundary when the microstructural properties E and ν are periodic functions. The classical periodic cell problem is given by

$$-\partial_y\big((E(y)+s\nu(y))\partial_y\chi(y)\big) = \partial_y(E(y)+s\nu(y)), \qquad y\in\Omega, \tag{A.1a}$$

$$\chi \text{ is 1-periodic}, \quad \int_\Omega \chi(y)\,\mathrm{d}y = 0. \tag{A.1b}$$

where $s\in\mathbb{R}$ is a Laplace variable which the solution χ will depend on. Hence, we can write the solution to this cell problem more explicitly as $\chi(y,s)$. Now take any forcing trajectory $\bar\epsilon(t)$ and take its Laplace transform $\widehat{\bar\epsilon}(s)$. Then we can define $\widehat u(y,s)=\widehat{\bar\epsilon}(s)(\chi(y,s)+y)$ and rewrite the cell problem above as

$$-\partial_y\widehat\sigma(y,s) = 0, \qquad y\in\Omega, \tag{A.2a}$$

$$\widehat\sigma(y,s) = (E(y)+s\nu(y))\partial_y\widehat u(y,s), \qquad y\in\Omega, \tag{A.2b}$$

$$\widehat u(1,s) = \widehat u(0,s)+\widehat{\bar\epsilon}(s), \tag{A.2c}$$

$$\partial_y\widehat u(1,s) = \partial_y\widehat u(0,s), \tag{A.2d}$$

$$\int_\Omega \widehat u(y,s)\,\mathrm{d}y = \frac{\widehat{\bar\epsilon}(s)}{2}. \tag{A.2e}$$

Now converting back into the time domain from the Laplace domain for $t\in\mathcal{T}:=[0,T]$ we have

$$-\partial_y\sigma(y,t) = 0, \qquad (y,t)\in\Omega\times\mathcal{T}, \tag{A.3a}$$

$$\sigma(y,t) = E(y)\partial_y u(y,t)+\nu(y)\partial_{yt}u(y,t), \qquad (y,t)\in\Omega\times\mathcal{T}, \tag{A.3b}$$

$$u(1,t) = u(0,t)+\bar\epsilon(t), \qquad t\in\mathcal{T}, \tag{A.3c}$$

$$\partial_y u(1,t) = \partial_y u(0,t), \qquad t\in\mathcal{T}, \tag{A.3d}$$

$$\int_\Omega u(y,t)\,\mathrm{d}y = \frac{\bar\epsilon(t)}{2}, \qquad t\in\mathcal{T}. \tag{A.3e}$$

Omitting temporarily the last integral constraint, the solution $u(y,t)$ of the PDE above still remains a valid solution for $u(y,t)+c(t)$ for any trajectory $c(t)\in\mathbb{R}$. Hence, we can shift our solution $u(y,t)\to u(y,t)-u(0,t)$ and it will now satisfy the Dirichlet problem

$$\begin{aligned}
-\partial_y\sigma(y,t) &= 0, && (y,t)\in\Omega\times\mathcal{T}, && \text{(A.4a)}\\
\sigma(y,t) &= E(y)\partial_y u(y,t)+\nu(y)\partial_{yt}u(y,t), && (y,t)\in\Omega\times\mathcal{T}, && \text{(A.4b)}\\
u(0,t) &= 0, \quad u(1,t)=\bar{\epsilon}(t), && t\in\mathcal{T}, && \text{(A.4c)}\\
\partial_y u(1,t) &= \partial_y u(0,t), && t\in\mathcal{T}, && \text{(A.4d)}\\
u(y,0) &= 0, && y\in\Omega && \text{(A.4e)}
\end{aligned}$$

which is precisely the cell problem (3.3) we began our discussion from in Section 3.1.

Appendix B. Lipschitz Continuity Proofs

This section proves several bounds on the solution of the cell problem which we rely on in the proofs of Lemmas 4.1 and 4.2. The proof of these bounds follow the proof of Lemma 3.5 in [7] with the difference that these bounds depend on the minimum and maximum value of the cell problem microstructure $E_{\min}, E_{\max}, \nu_{\min}, \nu_{\max}$ *as well as* the maximum absolute value of the strain input and its derivative $\bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}$. This allows us to uniformly control the variation in cell problem solutions across a class of microstructures and strain inputs. Also, our result derives Lipschitz bounds on the solution of the cell problem, whereas Lemma 3.5 in [7] works at the level of the original multiscale problem, which does not suit our purposes here.

Proposition B.1. *Under Assumptions 3.2, for the solutions p of the periodic cell problem* (3.5)*, the following bounds hold*

(a) $\sup_{t\in\mathcal{T}}\|p\|_{H^1_0,\nu}\le\dfrac{\nu_{\max}(\dot{\bar{\epsilon}}_{\max}\nu_{\max}+\bar{\epsilon}_{\max}E_{\max})}{E_{\min}\sqrt{\nu_{\min}}}.$

(b) $\sup_{t\in\mathcal{T}}\|\partial_t p\|_{H^1_0,\nu}\le\dfrac{(E_{\max}\nu_{\max}+E_{\min}\nu_{\min})(\dot{\bar{\epsilon}}_{\max}\nu_{\max}+\bar{\epsilon}_{\max}E_{\max})}{E_{\min}\nu_{\min}^{\frac{3}{2}}}.$

Proof. We show the first bound by choosing a test function $\varphi=p$ and writing the weak form of our periodic cell problem (3.11) as

$$q_\nu(\partial_t p,p)+q_E(p,p)=-\dot{\bar{\epsilon}}(t)\langle\nu,\partial_y p\rangle-\bar{\epsilon}(t)\langle E,\partial_y p\rangle \tag{B.1}$$

and hence by Cauchy–Schwarz and the definition of the weighted H^1_0 norm in (3.8) we get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|p\|^2_{H^1_0,\nu}+\|p\|^2_{H^1_0,E}\le\Big(|\dot{\bar{\epsilon}}(t)|\|\nu\|_{L^2}+|\bar{\epsilon}(t)|\|E\|_{L^2}\Big)\|p\|_{H^1_0}\le C\|p\|_{H^1_0}$$

for the constant $C=\dot{\bar{\epsilon}}_{\max}\nu_{\max}+\bar{\epsilon}_{\max}E_{\max}$. Applying Lemma 3.1 we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\|p\|^2_{H^1_0,\nu}+2\frac{E_{\min}}{\nu_{\max}}\|p\|^2_{H^1_0,\nu}\le 2\frac{C}{\sqrt{\nu_{\min}}}\|p\|_{H^1_0,\nu}$$

which by Young's inequality for $\delta>0$ gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\|p\|^2_{H^1_0,\nu}+2\frac{E_{\min}}{\nu_{\max}}\|p\|^2_{H^1_0,\nu}\le\frac{C^2}{\nu_{\min}\delta^2}+\delta^2\|p\|^2_{H^1_0,\nu}.$$

Setting $\delta^2=\frac{E_{\min}}{\nu_{\max}}$ gives us

$$\frac{\mathrm{d}}{\mathrm{d}t}\|p\|^2_{H^1_0,\nu}+\frac{E_{\min}}{\nu_{\max}}\|p\|^2_{H^1_0,\nu}\le\frac{C^2\nu_{\max}}{E_{\min}\nu_{\min}}$$

which by Gronwall's inequality yields

$$\sup_{t\in\mathcal{T}} \|p\|^2_{H^1_0,\nu} \leq \left(\frac{\nu_{\max}}{E_{\min}}\right)^2 \frac{C^2}{\nu_{\min}} = \left(\frac{\nu_{\max}(\dot{\bar{\epsilon}}_{\max}\nu_{\max} + \bar{\epsilon}_{\max}E_{\max})}{E_{\min}\sqrt{\nu_{\min}}}\right)^2, \tag{B.2}$$

so the first bound is proved.
To prove the second bound, we take a test function $\varphi = \partial_t p \in H^1_0(\Omega)$ and write the weak form of the periodic cell problem

$$q_\nu(\partial_t p, \partial_t p) + q_E(\partial_t p, p) = -\dot{\bar{\epsilon}}(t)\langle \nu, \partial_{yt} p\rangle - \bar{\epsilon}(t)\langle E, \partial_{yt} p\rangle \tag{B.3}$$

which by Cauchy–Schwarz gives us

$$\|\partial_t p\|^2_{H^1_0,\nu} \leq \|p\|_{H^1_0,E}\|\partial_t p\|_{H^1_0,E} + C\|\partial_t p\|_{H^1_0}$$

for the same constant $C = \dot{\bar{\epsilon}}_{\max}\nu_{\max} + \bar{\epsilon}_{\max}E_{\max}$. Then by Lemma 3.1 we get

$$\|\partial_t p\|^2_{H^1_0,\nu} \leq \left(\frac{E_{\max}}{\nu_{\min}}\|p\|_{H^1_0,\nu} + \frac{C}{\sqrt{\nu_{\min}}}\right)\|\partial_t p\|_{H^1_0,\nu}$$

and therefore using our first bound on $\|p\|^2_{H^1_0,\nu}$ in (B.2) we get

$$\|\partial_t p\|_{H^1_0,\nu} \leq (\dot{\bar{\epsilon}}_{\max}\nu_{\max} + \bar{\epsilon}_{\max}E_{\max})\left(\frac{E_{\max}\nu_{\max}}{E_{\min}\nu_{\min}^{3/2}} + \frac{1}{\sqrt{\nu_{\min}}}\right). \qquad \blacksquare$$

Now recall that a solution $u(y,t)$ to the original cell problem (3.3) can be decomposed into its periodic and nonperiodic parts as $u(y,t) = p(y,t) + \bar{\epsilon}(t)y$ where p is a solution to the periodic cell problem (3.5). This allows us to bound by the triangle inequality and Lemma 3.1

$$\begin{aligned}
\sup_{t\in\mathcal{T}} \|u\|_{H^1_0,\nu} &\leq \sup_{t\in\mathcal{T}} \|p\|_{H^1_0,\nu} + \nu_{\max}\sup_{t\in\mathcal{T}}|\bar{\epsilon}(t)|,\\
\sup_{t\in\mathcal{T}} \|\partial_t u\|_{H^1_0,\nu} &\leq \sup_{t\in\mathcal{T}} \|\partial_t p\|_{H^1_0,\nu} + \nu_{\max}\sup_{t\in\mathcal{T}}|\dot{\bar{\epsilon}}(t)|.
\end{aligned} \tag{B.4}$$

Combining these bounds with Proposition B.1 immediately leads to the following corollary.

Corollary B.2. *Under Assumptions 3.2, for the solution u of the cell problem* (3.3)*, the following bounds hold*

(a) $\sup_{t\in\mathcal{T}} \|u\|_{H^1_0,\nu} \leq \dfrac{\nu_{\max}(\dot{\bar{\epsilon}}_{\max}\nu_{\max} + \bar{\epsilon}_{\max}E_{\max})}{E_{\min}\nu_{\min}} + \nu_{\max}\bar{\epsilon}_{\max}$

(b) $\sup_{t\in\mathcal{T}} \|\partial_t u\|_{H^1_0,\nu} \leq \dfrac{(E_{\max}\nu_{\max}+E_{\min}\nu_{\min})(\dot{\bar{\epsilon}}_{\max}\nu_{\max}+\bar{\epsilon}_{\max}E_{\max})}{E_{\min}\nu_{\min}^{\frac{3}{2}}} + \nu_{\max}\dot{\bar{\epsilon}}_{\max}.$

Proposition B.3. *Under Assumptions 3.2, for all solutions γ of* (4.4)*, the following bounds hold*

(a) $\sup_{t\in\mathcal{T}} \|\gamma\|_{H^1_0,\nu_1} \leq \dfrac{\nu_{\max}}{E_{\min}\sqrt{\nu_{\min}}}\|g\|_{\mathcal{Z}}$

(b) $\sup_{t\in\mathcal{T}} \|\partial_t\gamma\|_{H^1_0,\nu_1} \leq \dfrac{E_{\max}\nu_{\max} + E_{\min}\nu_{\min}}{E_{\min}\nu_{\min}^{\frac{3}{2}}}\|g\|_{\mathcal{Z}}.$

Proof. Choosing the test function $\varphi = \gamma$ which by definition is zero on the boundary $\partial\Omega$, we can write the weak form of the PDE (4.4) as

$$q_{\nu_1}(\partial_t\gamma, \gamma) + q_{E_1}(\gamma, \gamma) = -\langle g, \partial_y\gamma\rangle.$$

Now using Cauchy–Schwarz we can write

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\gamma\|^2_{H^1_0,\nu_1} + \|\gamma\|^2_{H^1_0,E_1} \leq \|g\|_{\mathcal{Z}}\|\gamma\|_{H^1_0}$$

and applying Lemma 3.1 gives us

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\gamma\|^2_{H^1_0,\nu_1} + \frac{E_{\min}}{\nu_{\max}}\|\gamma\|^2_{H^1_0,\nu_1} \le \frac{1}{\sqrt{\nu_{\min}}}\|g\|_{\mathcal{Z}}\|\gamma\|_{H^1_0,\nu_1}$$

and applying Young's inequality for any $\delta > 0$ gives us

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\gamma\|^2_{H^1_0,\nu_1} + \frac{E_{\min}}{\nu_{\max}}\|\gamma\|^2_{H^1_0,\nu_1} \le \frac{1}{2\delta^2}\|g\|^2_{\mathcal{Z}} + \frac{\delta^2}{2\nu_{\min}}\|\gamma\|^2_{H^1_0,\nu_1}.$$

Now setting $\delta^2 = \frac{E_{\min}\nu_{\min}}{\nu_{\max}}$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\gamma\|^2_{H^1_0,\nu_1} + \frac{E_{\min}}{\nu_{\max}}\|\gamma\|^2_{H^1_0,\nu_1} \le \frac{\nu_{\max}}{E_{\min}\nu_{\min}}\|g\|^2_{\mathcal{Z}}$$

Note that $\gamma(y,0) = 0$ since $u_1(y,0) = u_2(y,0) = 0$. Hence, by Gronwall's inequality we get that

$$\sup_{t\in\mathcal{T}}\|\gamma\|^2_{H^1_0,\nu_1} \le \Big(\frac{\nu_{\max}}{E_{\min}}\Big)^2\frac{1}{\nu_{\min}}\|g\|^2_{\mathcal{Z}} \tag{B.5}$$

which proves the first bound.

To prove the second bound, we substitute $\varphi = \partial_t\gamma$ into the weak form of the PDE to get

$$q_{\nu_1}(\partial_t\gamma, \partial_t\gamma) + q_{E_1}(\gamma, \partial_t\gamma) = -\langle g, \partial_{yt}\gamma\rangle.$$

Now rearranging terms and using Cauchy–Schwarz we can write

$$\|\partial_t\gamma\|^2_{H^1_0,\nu_1} \le \|g\|\|\partial_t\gamma\|_{H^1_0} + \|\gamma\|_{H^1_0,E_1}\|\partial_t\gamma\|_{H^1_0,E_1}.$$

Applying Lemma 3.1 gives us

$$\|\partial_t\gamma\|^2_{H^1_0,\nu_1} \le \frac{1}{\sqrt{\nu_{\min}}}\|g\|\|\partial_t\gamma\|_{H^1_0,\nu_1} + \frac{E_{\max}}{\nu_{\min}}\|\gamma\|_{H^1_0,\nu_1}\|\partial_t\gamma\|_{H^1_0,\nu_1}.$$

Dividing by $\|\partial_t\gamma\|_{H^1_0,\nu_1}$ on both sides we get

$$\|\partial_t\gamma\|_{H^1_0,\nu_1} \le \frac{1}{\sqrt{\nu_{\min}}}\|g\| + \frac{E_{\max}}{\nu_{\min}}\|\gamma\|_{H^1_0,\nu_1}$$

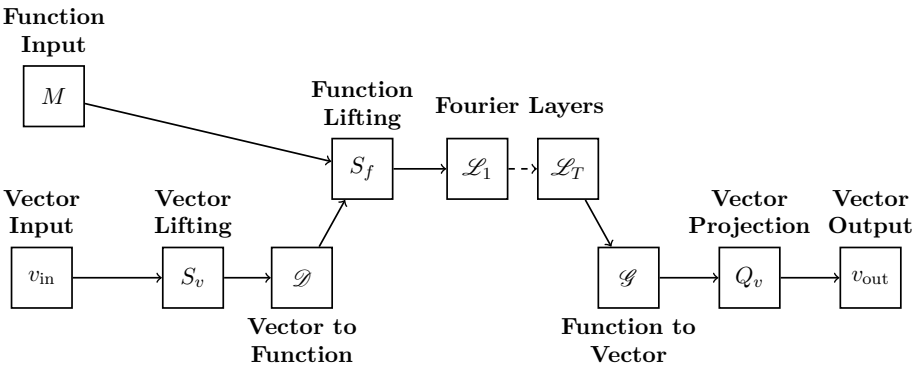
and applying our first bound in (B.5) gives us

$$\|\partial_t\gamma\|_{H^1_0,\nu_1} \le \Big(\frac{E_{\max}\nu_{\max}}{E_{\min}\nu_{\min}^{\frac{3}{2}}} + \frac{1}{\sqrt{\nu_{\min}}}\Big)\|g\|_{\mathcal{Z}}. \tag{B.6}$$

■

Appendix C. Definition of Neural Operator

Here we give the full detailed definition of the Fourier Neural Mapping (FNM). In the following definition, $\psi_k = e^{2\pi\mathrm{i}\langle k,\cdot\rangle_{\mathbb{R}^d}}$ are the complex Fourier basis elements of $L^2(\mathbb{T}^d;\mathbb{C})$.



Definition C.1 (Fourier Neural Mapping (FNM))**.** Let the function input $M \in L^2(\mathbb{T}^d; \mathbb{R}^{d_M})$. Define the vector input $v_{\text{in}} \in \mathbb{R}^{d^v_{\text{in}}}$ and vector output $v_{\text{out}} \in \mathbb{R}^{d^v_{\text{out}}}$. Let $x \in \mathbb{T}^d$ and define the following layers:

$$\begin{aligned}
&\text{(Vector Lifting)} && S_v : \mathbb{R}^{d^v_{\text{in}}} \to \mathbb{R}^{d^v_{\text{lift}}} \\
&\text{(Vector to Function)} && \mathscr{D} : \mathbb{R}^{d^v_{\text{lift}}} \to L^2(\mathbb{T}^d; \mathbb{R}^{d^{vf}_{\text{lift}}}) \\
& && z \mapsto \mathscr{D}z = \kappa_v(\,\cdot\,)z \\
& && z \mapsto \mathscr{D}z = \left\{ \textstyle\sum_{k\in\mathbb{Z}^d} \left(P_v^{(k)} z\right)_j \psi_k \right\}_{j \in [d^{vf}_{\text{lift}}]} \\
&\text{(Function Lifting)} && S_f : L^2(\mathbb{T}^d; \mathbb{R}^{d_M + d^{vf}_{\text{lift}}}) \to L^2(\mathbb{T}^d; \mathbb{R}^{d_0}) \\
&\text{(Fourier)} && \mathscr{L}_t : L^2(\mathbb{T}^d; \mathbb{R}^{d_{t-1}}) \to L^2(\mathbb{T}^d; \mathbb{R}^{d_t}),\ t \in [T], \\
& && (\mathscr{L}_t(u))(x) = \sigma\big(W_t u(x) + (\mathcal{K}_t u)(x) + b_t\big), \\
&\text{(Function to Vector)} && \mathscr{G} : L^2(\mathbb{T}^d; \mathbb{R}^{d_T}) \to \mathbb{R}^{d^{fv}_{\text{proj}}} \\
& && h \mapsto \mathscr{G}h = \textstyle\int_{\mathbb{T}^d} \kappa_f(x) h(x)\,\mathrm{d}x \\
& && h \mapsto \mathscr{G}h = \left\{ \textstyle\sum_{k\in\mathbb{Z}^d} \left(\sum_{j=1}^{d_T} (P_f^{(k)})_{\ell j} \langle \psi_k, h_j \rangle_{L^2(\mathbb{T}^d;\mathbb{C})} \right) \right\}_{\ell \in [d^{fv}_{\text{proj}}]} \\
&\text{(Vector Projection)} && Q_v : \mathbb{R}^{d^{fv}_{\text{proj}}} \to \mathbb{R}^{d^v_{\text{out}}}.
\end{aligned}$$

The convolution operator is given, for $u : \mathbb{T}^d \to \mathbb{R}^{d_{t-1}}$ and $x \in \mathbb{T}^d$, by

$$(\mathcal{K}_t u)(x) = \left\{ \sum_{k\in\mathbb{Z}^d} \left(\sum_{j=1}^{d_{t-1}} (P_t^{(k)})_{\ell j} \langle \psi_k, u_j \rangle_{L^2(\mathbb{T}^d;\mathbb{C})} \right) \psi_k(x) \right\}_{\ell \in [d_t]} \in \mathbb{R}^{d_t}. \tag{C.1}$$

For given layer index t and wave vector $k \in \mathbb{Z}^d$, the matrix $P_t^{(k)} \in \mathbb{C}^{d_t \times d_{t-1}}$ comprises learnable parameters of the integral operator $\mathcal{K}_t$; furthermore, $W_t \in \mathbb{R}^{d_t \times d_{t-1}}$ is a weights matrix, $b_t \in \mathbb{R}^{d_t}$ is a bias vector, both learnable. For given wave vector $k \in \mathbb{Z}^d$, $P_v^{(k)} \in \mathbb{C}^{d^{vf}_{\text{lift}} \times d^v_{\text{lift}}}$ are the learnable parameters of the vector to function map $\mathscr{D}$, and $P_f^{(k)} \in \mathbb{C}^{d^{fv}_{\text{proj}} \times d_T}$ are the learnable parameters of the function to vector map $\mathscr{G}$. The vector lifting and projection layers, S_v and Q_v, are either shallow neural networks or linear maps, and hence also contain learnable weight and bias parameters. Finally, the function lifting layer S_f is applied pointwise in $\mathbb{T}^d-$a.e. and is also defined by either a shallow neural network or a linear map containing learnable weight and bias parameters.

Remark C.2. Note that the function $\kappa_f : \mathbb{T}^d \to \mathbb{R}^{d^{fv}_{\text{proj}} \times d_T}$ is parametrized in the Fourier domain, where the coefficients $P_f^{(k)}$ correspond to the Fourier coefficients of κ_f. Similarly, the function $\kappa_v : \mathbb{T}^d \to \mathbb{R}^{d^{vf}_{\text{lift}} \times d^v_{\text{lift}}}$ for the vector to function layer $\mathscr{D}$ is parameterized in the Fourier domain such that $P_v^{(k)}$ correspond to the Fourier coefficients of κ_v.

Using the definition of the Fourier Neural Mapping above and comparing to (2.5), we know the input and output dimensionalities for F_{FNM} are $d^v_{\text{in}} = 2d^2 + L, d^v_{\text{out}} = d^2$ and for G_{FNM} are $d^v_{\text{in}} = d^2 + L, d^v_{\text{out}} = d^2$. The input dimensionality d_M of the microstructural properties M depends on the material model as discussed in Remark 2.3.

Appendix D. Approximation Proofs

This lemma is used to show, in one dimension, that functions that are integrable and of bounded variation are approximable by piecewise constants. This is a classical and well-known result which

we derive here for convenience in the one dimensional setting of our theory, and can be extended to higher dimensions by using the equivalence between functions of bounded variation and Lipschitz-1 functions [6, Lemma C.1].

Lemma D.1. *For the domain $\Omega = [0,1]$ take any integrable function of bounded variation $f \in L^1(\Omega) \cap \mathsf{BV}(\Omega)$. Then defining a grid $y_i = \frac{i}{M}$ for $i \in \{0, \dots, M\}$ there exists a piecewise constant function $f_{\text{pc}} : [0,1] \to \mathbb{R}$ with M pieces given by*

$$f_{\text{pc}}(y) = f_{\text{pc}}^i := M \int_{y_i}^{y_{i+1}} f(z)\,\mathrm{d}z, \quad y \in [y_i, y_{i+1}] \tag{D.1}$$

satisfying the approximation bound

$$\|f - f_{\text{pc}}\|_{L^1} \le \frac{|f|_{\mathsf{BV}}}{M}. \tag{D.2}$$

Proof. We begin by noting that

$$\|f - f_{\text{pc}}\|_{L^1} = \sum_{i=0}^{M-1} \int_{y_i}^{y_{i+1}} |f(z) - f_{\text{pc}}^i|\,\mathrm{d}z.$$

Studying one of the terms in the sum we can bound

$$\begin{aligned}
\sup_{y \in [y_i, y_{i+1}]} |f(y) - f_{\text{pc}}^i| &= \sup_{y \in [y_i, y_{i+1}]} M \Big| \int_{y_i}^{y_{i+1}} (f(y) - f(z))\,\mathrm{d}z \Big| \\
&\le \sup_{y \in [y_i, y_{i+1}]} M \int_{y_i}^{y_{i+1}} |f(y) - f(z)|\,\mathrm{d}z \\
&\le \sup_{y, y' \in [y_i, y_{i+1}]} |f(y) - f(y')|.
\end{aligned}$$

Now suppose that for each of the M intervals $[y_i, y_{i+1}]$ we choose two points $y, y' \in [y_i, y_{i+1}]$ and form a partition from the union of all of these points. This partition now has $2M$ points if we also include the endpoints $y = 0, 1$. This implies that

$$\begin{aligned}
\|f - f_{\text{pc}}\|_{L^1} &\le \frac{1}{M} \sum_{i=0}^{M-1} \sup_{y, y' \in [y_i, y_{i+1}]} |f(y) - f(y')| \\
&\le \frac{1}{M} \sup \Big\{ \sum_{i=0}^{2M-1} |f(z_{i+1}) - f(z_i)| \;\Big|\; 0 = z_0 < z_1 < \ldots < z_{2M} = 1 \Big\} \\
&\le \frac{|f|_{\mathsf{BV}}}{M}
\end{aligned}$$

where the last line is given by the definition of the total variation norm in (1.1). ∎

The following two propositions are critical to the RNO approximation result. The first is a general universal approximation result for FNMs with both finite and infinite-dimensional inputs, and the second applies this general result to the constitutive map of interest in this work.

Proposition D.2. *Consider a bounded set $K_v \subset \mathbb{R}^{d_{\text{in}}^v}$ and a compact set $K_f \subset L^2(\mathbb{T}^d; \mathbb{R}^{d_{\text{in}}^f})$. Let $\Phi^\dagger : K_v \times K_f \to \mathbb{R}^{d_{\text{out}}^v}$ be continuous. Then for any $\mathfrak{e} > 0$, there exists an* FNM *$\Phi : K_v \times K_f \to \mathbb{R}^{d_{\text{out}}^v}$ of the form*

$$\Phi = Q_v \circ \mathscr{G} \circ \mathscr{L}_T \circ \cdots \circ \mathscr{L}_1 \circ S_f \circ (\mathscr{D} \circ S_v, I_{d_{\text{in}}^f}), \tag{D.3}$$

for some $T \in \mathbb{N}_{>0}$ where S_v acts on the input from K_v and $I_{d_{\text{in}}^f}$ acts on the input from K_f, such that

$$\sup_{u \in K_v \times K_f} \|\Phi^\dagger(u) - \Phi(u)\| \le \mathfrak{e}. \tag{D.4}$$

The layers in (D.3) *take the form of their homonymous counterparts in Definition C.1.*

This proposition closely follows those of Theorems 3.2 and 3.3 in [29], but we state a proof here for completeness. We adapt the proofs of these theorems to our setting because our FNM architecture is novel in the sense that it accepts vector and function inputs jointly as opposed to purely vector or purely function inputs discussed in [29]. The workhorse of this theoretical result is the Dugundji extension theorem that allows us to extend an operator acting on a compact subset of L^2 to all of L^2.

Proof. Let $\mathbb{1}(x) = 1$ be the constant function and define the vector to function map $\mathsf{L} : \mathbb{R}^{d_{\text{in}}^v} \to L^2(\mathbb{T}^d; \mathbb{R}^{d_{\text{in}}^v})$ given by $\mathsf{L}(z) \mapsto z\mathbb{1}$. Note that clearly $\|\mathsf{L}z\|_{L^2(\mathbb{T}^d;\mathbb{R}^{d_{\text{in}}^v})} = \|z\|$, so L is continuous. Let $K_{vf} := \{\mathsf{L}z : z \in K_v\}$ and note that this set is compact since K_v is compact and L is continuous. Define $K = K_{vf} \times K_f \subset L^2(\mathbb{T}^d; \mathbb{R}^{d_{\text{in}}^v + d_{\text{in}}^f})$ by $K_{vf} \times K_f$ which is also a compact set. Define $\Phi_{fv}^\dagger : K \to \mathbb{R}^{d_{\text{out}}^v}$ by $\mathsf{L}z \times f \mapsto \Phi^\dagger(z, f)$.

Lastly, define $\Phi_{ff}^\dagger : K \to L^2(\mathbb{T}^d; \mathbb{R}^{d_{\text{out}}^v})$ by the map $z \mapsto \Phi_{fv}^\dagger(z)\mathbb{1}$. We first show that $\Phi_{ff}^\dagger$ is continuous.

$$\begin{aligned}
\|\Phi_{ff}^\dagger((Lz, f))\|_{\mathcal{Y}}^2 &= \int_{\mathbb{T}^d} \|\Phi_{ff}^\dagger((Lz, f))\|_{\mathbb{R}^{d_{\text{out}}^v}}^2 \, \mathrm{d}x \\
&= \int_{\mathbb{T}^d} \|\Phi^\dagger(z, f)\mathbb{1}\|_{\mathbb{R}^{d_{\text{out}}^v}}^2 \, \mathrm{d}x \\
&= \|\Phi^\dagger(z, f)\|_{\mathbb{R}^{d_{\text{out}}^v}}^2 \\
&= \left\| \Phi^\dagger\left(\int_{\mathbb{T}^d} Lz \ \mathrm{d}x, f \right) \right\|_{\mathbb{R}^{d_{\text{out}}^v}}^2 .
\end{aligned}$$

Since the averaging operator $\int_{\mathbb{T}^d} \cdot \, \mathrm{d}x$ over the torus is continuous and $\Phi^\dagger$ is also continuous, $\Phi_{ff}^\dagger$ is continuous from K to $L^2(\mathbb{T}^d; \mathbb{R}^{d_{\text{out}}^v})$. By the Dugundji extension theorem, there exists a continuous operator $\widetilde{\Phi}^\dagger : L^2(\mathbb{T}^d; \mathbb{R}^{d_{\text{in}}^v + d_{\text{in}}^f}) \to L^2(\mathbb{T}^d; \mathbb{R}^{d_{\text{out}}^v})$ such that $\widetilde{\Phi}^\dagger(u) = \Phi_{ff}^\dagger(u)$ for every $u \in K$. By Theorem 9 of [34], for any $\mathfrak{e}$, there exists an FNO $\widehat{\Phi}$ of the form

$$\widehat{\Phi} = \widehat{Q}_f \circ \widehat{\mathscr{L}_T} \circ \cdots \circ \widehat{\mathscr{L}_1} \circ \widehat{S}_f \tag{D.5}$$

such that

$$\sup_{a \in K} \|\widetilde{\Phi}^\dagger(a) - \widehat{\Phi}(a)\|_{L^2(\mathbb{T}^d;\mathbb{R}^{d_{\text{out}}^v})} = \sup_{a \in K} \|\Phi_{ff}^\dagger(a) - \widehat{\Phi}(a)\|_{L^2(\mathbb{T}^d;\mathbb{R}^{d_{\text{out}}^v})} < \mathfrak{e}. \tag{D.6}$$

Specifically, in (D.5) we have that $\widehat{S}_f : \mathbb{R}^{d_{\text{in}}^v + d_{\text{in}}^f} \to \mathbb{R}^{d_0}$, and $\widehat{Q}_f : \mathbb{R}^{d_T} \to \mathbb{R}^{d_{\text{out}}^v}$.

Next we define the following FNM architecture

$$\Phi = Q_v \circ \mathscr{G} \circ \mathscr{L}_T \circ \ldots \mathscr{L}_1 \circ S_f \circ (\mathscr{D} \circ S_v, I_{d_{\text{in}}^f}) \tag{D.7}$$

where we will define its layers accordingly to be equivalent to the FNO architecture in (D.5). The layers are given by

$$\begin{aligned}
S_f &= \widehat{S}_f \\
S_v &= I_{d_{\text{in}}^v} \\
z \longmapsto \mathscr{D}z &= \mathbb{1}z \\
u \longmapsto \mathscr{L}_1 u(x) &= \sigma\left(W_1 u(x) + \sum_{k\in\mathbb{Z}^d}\left(\sum_{j=1}^{d_0}\left(P_1^{(k)}\right)_j \langle \psi_k, u_j\rangle_{L^2(\mathbb{T}^d;\mathbb{C})}\right)\psi_k(x) + b_1\right) \\
&\text{where } W_1 = \widehat{W}_1,\ P_1^{(k)} = \widehat{P}_1^{(k)} \text{ so in other words} \\
\mathscr{L}_t &= \widehat{\mathscr{L}}_t,\ t = \{1,\dots,T\} \\
z \longmapsto \mathscr{G}z &= \int_{\mathbb{T}^d} \kappa_f(x) z(x)\,\mathrm{d}x \text{ where } \kappa_f = \widehat{Q}_f \mathbb{1} \\
Q_v &= I_{d_{\text{out}}^v}.
\end{aligned}$$

In the preceding display, $\widehat{W}_1$ and $\widehat{P}_1$ are the associated coefficients in $\widehat{\mathscr{L}_1}$ of the FNO in (D.5). One can check by this construction that $\int_{\mathbb{T}^d} \widehat{\Phi}(Lv, f)\,\mathrm{d}x = \Phi(v, f)$. Finally, this allows us to show that

$$\begin{aligned}
&\sup_{v,f\in K_v\times K_f} \|\Phi^\dagger(v,f) - \Phi(v,f)\|_{\mathbb{R}^{d_{\text{out}}^v}} \\
&\quad\le \sup_{v,f\in K_v\times K_f} \left\|\Phi^\dagger(v,f) - \int_{\mathbb{T}^d} \Phi_{ff}^\dagger(Lv,f)\,\mathrm{d}x\right\|_{\mathbb{R}^{d_{\text{out}}^v}} \\
&\qquad + \left\|\int_{\mathbb{T}^d}\left(\Phi_{ff}^\dagger(Lv,f) - \widehat{\Phi}(Lv,f)\right)\mathrm{d}x\right\|_{\mathbb{R}^{d_{\text{out}}^v}} \\
&\qquad + \left\|\int_{\mathbb{T}^d} \widehat{\Phi}(Lv,f)\,\mathrm{d}x - \Phi(v,f)\right\|_{\mathbb{R}^{d_{\text{out}}^v}} \\
&\quad\le \mathfrak{e}.
\end{aligned}$$

The transition from the second to third line holds because $\int_{\mathbb{T}^d} \Phi_{ff}^\dagger(Lv,f)\,\mathrm{d}x = \Phi^\dagger(v,f)$, by the approximation result in (D.6), and by the fact that $\int_{\mathbb{T}^d} \widehat{\Phi}(Lv,f)\,\mathrm{d}x = \Phi(v,f)$. Since $\mathfrak{e}$ was arbitrary, the lemma is proven. ∎

Proposition D.3. *Under Assumptions 3.2, for any $\bar{\epsilon}_{\max}, \dot{\bar{\epsilon}}_{\max}, \xi_{\max} > 0$ and $\mathfrak{e}_F, \mathfrak{e}_G > 0$, there exist FNMs F_{FNM} and G_{FNM}, in* (2.3)*, that approximate F_{pc} and G_{pc} of* (4.17) *such that*

$$\sup_{\substack{|z_1|\le\bar{\epsilon}_{\max}, |z_2|\le\dot{\bar{\epsilon}}_{\max}, \|z_3\|_\infty\le\xi_{\max},\\ E\in\mathcal{M}^B_{E_{\min},E_{\max}}, \nu\in\mathcal{M}^B_{\nu_{\min},\nu_{\max}}}} |F_{\text{FNM}}(z_1,z_2,z_3;E,\nu) - F_{\text{pc}}(z_1,z_2,z_3;E,\nu)| < \mathfrak{e}_F \tag{D.8}$$

$$\sup_{\substack{|z_1|\le\bar{\epsilon}_{\max}, \|z_3\|_\infty\le\xi_{\max},\\ E\in\mathcal{M}^B_{E_{\min},E_{\max}}, \nu\in\mathcal{M}^B_{\nu_{\min},\nu_{\max}}}} \|G_{\text{FNM}}(z_1,z_3;E,\nu) - G_{\text{pc}}(z_1,z_3;E,\nu)\| < \mathfrak{e}_G. \tag{D.9}$$

Proof. The proof is a simple application of Proposition D.2. The function inputs E and ν are on the same domain $\mathbb{T}^d$ and may have their outputs concatenated to form a single function input set K_f consistent with the statement of Proposition D.2. Since $\mathcal{M}^B_{E_{\min},E_{\max}}$ and $\mathcal{M}^B_{\nu_{\min},\nu_{\max}}$ are compact in L^2 due to the embedding $\mathsf{BV}(\mathbb{T}) \cap L^\infty(\mathbb{T}) \hookrightarrow L^2(\mathbb{T})$, the set of input functions is a compact set. For a proof of this embedding result, see [6, Lemma C.1]. Similarly, since all the finite inputs are bounded, their product set is also a compact set K_v consistent with Proposition D.2. Applying the lemma gives the result for both F_{FNM} and G_{FNM} to arbitrary accuracy $\mathfrak{e}_F$ and $\mathfrak{e}_G$. ∎

The following assumptions are necessary to derive the Lipschitz constant of the FNM in Lemma D.5.

Assumptions D.4. *We assume*

(1) The activation σ is $B-$Lipschitz.

(2) $\|P_t\|_\infty$ and $\|W_t\|_\infty$ are bounded.

(3) $\left(\sum_{k\in\mathbb{Z}^d}\left\|P_v^{(k)}\right\|_F^2\right)^{\frac{1}{2}}$ and $\left(\sum_{k\in\mathbb{Z}^d}\left\|P_f^{(k)}\right\|_F^2\right)^{\frac{1}{2}}$ are bounded.

(4) S_f, S_v, Q_f and Q_v are feedforward neural nets with activation σ, bounded network weights, fixed maximum width, and fixed number of layers.

Lemma D.5. *Under Assumptions D.4 and using notation from Definition C.1, a Fourier Neural Mapping Φ of the form*

$$Q_v \circ \mathscr{G} \circ \mathscr{L}_T \circ \cdots \circ \mathscr{L}_1 \circ S_f \circ (\mathscr{D} \circ S_v, I_{d_{\text{in}}^f})$$

is Lipschitz in the vector inputs: there exists some $C > 0$ such that for vector inputs v_1 and v_2 and function input f,

$$\|\Phi(v_1, f) - \Phi(v_2, f)\|_2 \leq C\|v_1 - v_2\|_2.$$

Proof.

(i) Claim: The Fourier layers $\mathscr{L}_t : L^2(\mathbb{T}^d; \mathbb{R}^{d_{t-1}}) \to L^2(\mathbb{T}^d; \mathbb{R}^{d_t})$ for $t \in [T]$ are Lipschitz

Take any $u_1, u_2 \in L^2(\mathbb{T}^d; \mathbb{R}^{d_{t-1}})$ and write

$$\begin{aligned}
\|\mathscr{L}_t(u_1) - \mathscr{L}_t(u_2)\|_{L^2(\mathbb{T}^d;\mathbb{R}^{d_t})} &= \left\|\sigma\Big(W_t u_1 + (\mathcal{K}_t u_1) + b_t\Big) - \sigma\Big(W_t u_2 + (\mathcal{K}_t u_2) + b_t\Big)\right\|_{L^2} \\
&\leq B\left\|W_t(u_1 - u_2) + (\mathcal{K}_t(u_1 - u_2))\right\|_{L^2} \\
&\leq B\|W_t\|_\infty\sqrt{d_{t-1}d_t}\|u_1 - u_2\|_{L^2} + \|\mathcal{K}_t(u_1 - u_2)\|_{L^2}.
\end{aligned}$$

Now to bound the second term above we let $u = u_1 - u_2$ and write

$$\begin{aligned}
\|\mathcal{K}_t u\|^2_{L^2(\mathbb{T}^d;\mathbb{R}^{d_t})} &= \sum_{l=1}^{d_t}\left\|\sum_{k\in\mathbb{Z}^d}\left(\sum_{j=1}^{d_{t-1}}(P_t^{(k)})_{\ell j}\langle\psi_k, u_j\rangle_{L^2(\mathbb{T}^d;\mathbb{C})}\right)\psi_k(x)\right\|^2_{L^2(\mathbb{T}^d)} \\
&= \sum_{l=1}^{d_t}\sum_{k\in\mathbb{Z}^d}\left(\sum_{j=1}^{d_{t-1}}(P_t^{(k)})_{\ell j}\langle\psi_k, u_j\rangle_{L^2(\mathbb{T}^d;\mathbb{C})}\right)^2 \\
&\leq \|P_t\|^2_\infty d_{t-1}d_t\sum_{k\in\mathbb{Z}^d}\sum_{j=1}^{d_{t-1}}\langle\psi_k, u_j\rangle^2_{L^2(\mathbb{T}^d;\mathbb{C})} \\
&= \|P_t\|^2_\infty d_{t-1}d_t\|u\|^2_{L^2(\mathbb{T}^d;\mathbb{R}^{d_{t-1}})}.
\end{aligned}$$

Combining everything together we get

$$\|\mathscr{L}_t(u_1) - \mathscr{L}_t(u_2)\|_{L^2(\mathbb{T}^d;\mathbb{R}^{d_t})} \leq (B\|W_t\|_\infty + \|P_t\|_\infty)\sqrt{d_{t-1}d_t} \times \|u_1 - u_2\|_{L^2(\mathbb{T}^d;\mathbb{R}^{d_{t-1}})} \tag{D.10}$$

(ii) Claim: The lifting and projection layers S_v, S_f, and Q_v are Lipschitz.

Each of these layers is a feed-forward neural network with Lipschitz activation functions and hence it is Lipschitz by composition.

(iii) The vector to function layer $\mathscr{D} : \mathbb{R}^{d_{\text{lift}}^v} \to L^2(\mathbb{T}^d; \mathbb{R}^{d_{\text{lift}}^{vf}})$ is Lipschitz.

For any $z \in \mathbb{R}^{d_{\text{lift}}^v}$ we write out

$$\begin{aligned}\|\mathscr{D}z\|^2_{L^2(\mathbb{T}^d;\mathbb{R}^{d_{\text{lift}}^{vf}})} &= \sum_{j=1}^{d_{\text{lift}}^{vf}} \left\| \sum_{k\in\mathbb{Z}^d} \left(P_v^{(k)} z\right)_j \psi_k \right\|^2_{L^2(\mathbb{T}^d)} = \sum_{j=1}^{d_{\text{lift}}^{vf}} \sum_{k\in\mathbb{Z}^d} \left| \left(P_v^{(k)} z\right)_j \right|^2 \\ &= \sum_{k\in\mathbb{Z}^d} \left\| P_v^{(k)} z \right\|_2^2 \le \left(\sum_{k\in\mathbb{Z}^d} \left\| P_v^{(k)} \right\|_F^2 \right) \|z\|_2^2\end{aligned}$$

where the second line follows by an application of Cauchy–Schwarz. So for any $z_1, z_2 \in \mathbb{R}^{d_{\text{lift}}^v}$ we have that

$$\|\mathscr{D}(z_1 - z_2)\|_{L^2(\mathbb{T}^d;\mathbb{R}^{d_{\text{lift}}^{vf}})} \le \left(\sum_{k\in\mathbb{Z}^d} \left\| P_v^{(k)} \right\|_F^2 \right)^{\frac{1}{2}} \|z_1 - z_2\|_2. \tag{D.11}$$

(iv) The function to vector layer $\mathscr{G} : L^2(\mathbb{T}^d; \mathbb{R}^{d_T}) \to \mathbb{R}^{d_{\text{proj}}^{fv}}$ is Lipschitz.

For any $h \in L^2(\mathbb{T}^d; \mathbb{R}^{d_T})$ we can write

$$\begin{aligned}\|\mathscr{G}h\|_{\ell^2} \le \|\mathscr{G}h\|_{\ell^1} &= \sum_{l=1}^{d_{\text{proj}}^{fv}} \left| \sum_{k\in\mathbb{Z}^d} \sum_{j=1}^{d_T} (P_f^{(k)})_{\ell j} \langle \psi_k, h_j \rangle_{L^2(\mathbb{T}^d;\mathbb{C})} \right| \\ &\le \sum_{l=1}^{d_{\text{proj}}^{fv}} \sum_{j=1}^{d_T} \left| \sum_{k\in\mathbb{Z}^d} (P_f^{(k)})_{\ell j} \langle \psi_k, h_j \rangle_{L^2(\mathbb{T}^d;\mathbb{C})} \right| \\ &\le \sum_{l=1}^{d_{\text{proj}}^{fv}} \sum_{j=1}^{d_T} \left(\sum_{k\in\mathbb{Z}^d} |(P_f^{(k)})_{\ell j}|^2 \right)^{\frac{1}{2}} \left(\sum_{k\in\mathbb{Z}^d} |\langle \psi_k, h_j \rangle_{L^2(\mathbb{T}^d;\mathbb{C})}|^2 \right)^{\frac{1}{2}} \\ &= \sum_{l=1}^{d_{\text{proj}}^{fv}} \sum_{j=1}^{d_T} \left(\sum_{k\in\mathbb{Z}^d} |(P_f^{(k)})_{\ell j}|^2 \right)^{\frac{1}{2}} \|h_j\|_{L^2(\mathbb{T}^d;\mathbb{C})} \\ &\le \sum_{l=1}^{d_{\text{proj}}^{fv}} \left(\sum_{j=1}^{d_T} \sum_{k\in\mathbb{Z}^d} |(P_f^{(k)})_{\ell j}|^2 \right)^{\frac{1}{2}} \|h\|_{L^2(\mathbb{T}^d;\mathbb{R}^{d_T})} \\ &\le \sqrt{d_{\text{proj}}^{fv}} \left(\sum_{k\in\mathbb{Z}^d} \|P_f^{(k)}\|_F^2 \right)^{\frac{1}{2}} \|h\|_{L^2(\mathbb{T}^d;\mathbb{R}^{d_T})}\end{aligned}$$

where the third, fifth, and sixth lines above follow from an application of Cauchy–Schwarz. ■

Appendix E. Loss Function Penalty Term for Viscoelastic FNM–RNO

To demonstrate the necessity of the penalty term for the constraint $\|G_{\text{FNM}}(0, 0; E, \nu)\| = 0$, we train FNM–RNO with varying numbers of internal variables with and without the penalty term in the loss function (5.3). The FNM–RNO trained without penalty achieves a slightly smaller relative L^2 testing error on both datasets, while the relative L^∞ testing error is much larger. In Figure E.1, we visualize the distributions of their relative L^∞ testing error on the PC and HMC datasets. For the PC dataset,

the relative L^∞ prediction error of the FNM–RNO trained without penalty is, on average, much larger than that of the linear stress response without memory effects. For the HMC dataset, this discrepancy is less pronounced. On the other hand, the FNM–RNO trained with penalty typically achieves smaller relative L^∞ testing errors on average compared to linear stress response without memory effects, except when the number of internal variables is large.

In Figure E.2, we visualize the FNM–RNO predictions at the sample in the PC testing datasets with large relative L^∞ error when the FNM–RNO uses 5 internal variables. When the FNM–RNO is trained without penalty, its averaged stress prediction at the initial time has a large error, while the prediction by the FNM–RNO trained with penalty has no visible error. Rates of change of the internal variables are large at the initial time when trained without penalty, and the constraint $\|G_{\text{FNM}}(0,0;E,\nu)\| = 0$ is violated. This behavior is consistent when testing on the HMC dataset and when different FNM–RNO architectures are used (e.g., different numbers of internal variables, channels, and Fourier modes).

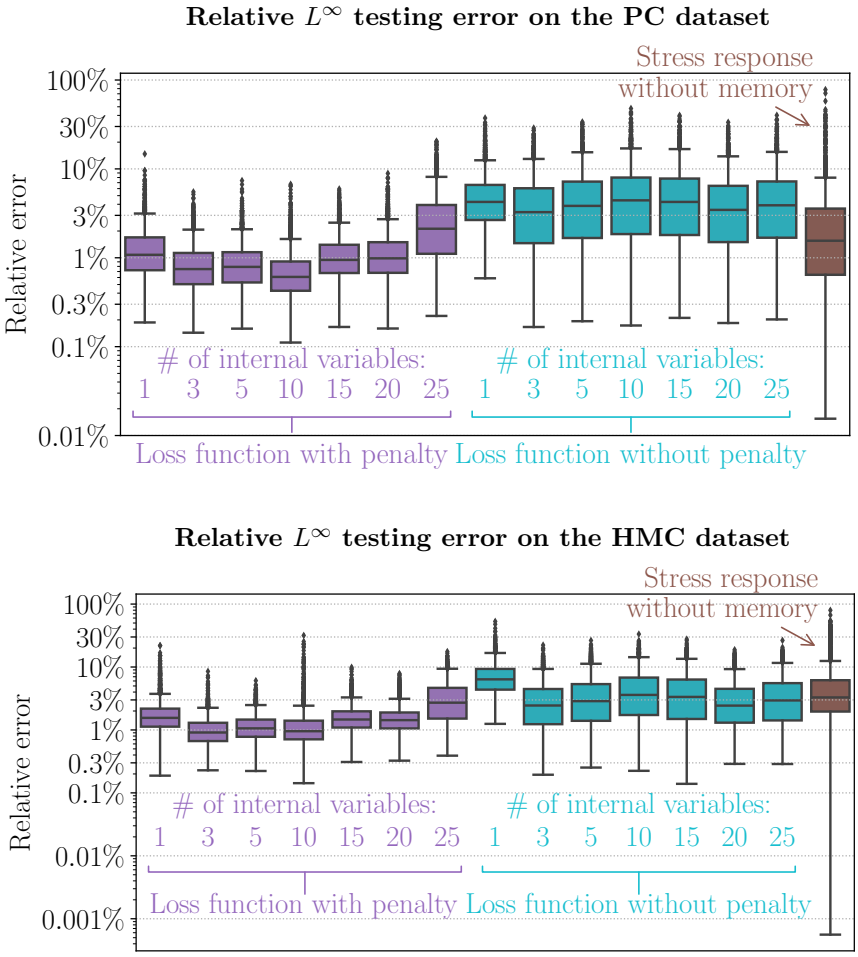


Figure E.1. The distributions of the relative L^∞ error on 2,500 testing samples from the PC dataset (*top*) and the HMC dataset (*bottom*). We visualize the errors in the FNM–RNO predictions where FNM–RNOs (i) are trained with or without the penalty term in (5.3), and (ii) have a varying number of internal variables. We also visualize the distribution of error given by the linear stress response without memory effects, where the response function is obtained using (3.12) with $K(t) = 0$ for all $t \in [0, 1]$.

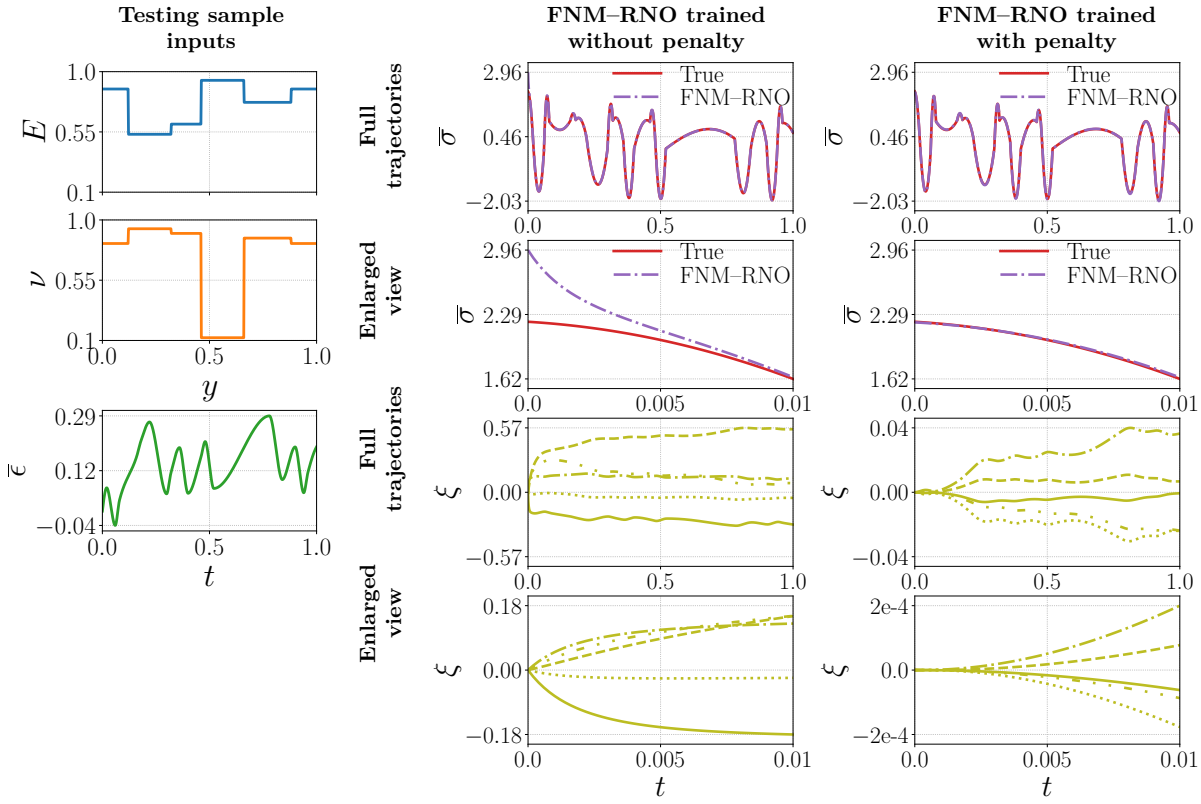


FIGURE E.2. Visualization of FNM–RNO predictions at the sample from the PC testing dataset with the largest relative L^∞ testing error (5 internal variables and without penalty in Figure E.1). We compare the predictions by FNM–RNOs trained using the loss function in (5.3) without the penalty term and with the penalty term. For the averaged stress and internal variable predictions, we show their full trajectories in $t \in [0, 1]$ and enlarged views in $t \in [0, 0.01]$.

References

[1] Assyr Abdulle, E. Weinan, Björn Engquist, and Eric Vanden-Eijnden. The heterogeneous multiscale method. *Acta Numer.*, 21:1–87, 2012.

[2] Andrew Akerson, Aakila Rajan, and Kaushik Bhattacharya. Learning constitutive relations from experiments: 1. PDE constrained optimization. `https://arxiv.org/abs/2412.02864`, 2024.

[3] Grégoire Allaire. Homogenization and two-scale convergence. *SIAM J. Math. Anal.*, 23(6):1482–1518, 1992.

[4] Martin S. Alnæs, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, Marie E. Rognes, and Garth Wells. The FEniCS Project Version 1.5. *Arch. Numer. Soft.*, 3(100):9–23, 2015.

[5] Alain Bensoussan, Jacques-Louis Lions, and George Papanicolaou. *Asymptotic analysis for periodic structures*. American Mathematical Society, 2011.

[6] Kaushik Bhattacharya, Nikola Kovachki, Aakila Rajan, Andrew Stuart, and Margaret Trautner. Learning homogenization for elliptic operators. *SIAM J. Numer. Anal.*, 62(4):1844–1873, 2024.

[7] Kaushik Bhattacharya, Burigede Liu, Andrew Stuart, and Margaret Trautner. Learning Markovian homogenized models in viscoelasticity. *Multiscale Model. Simul.*, 21(2):641–679, 2023.

[8] E. W. Billington, A. Tate, and W. O. Williams. *The Physics of Deformation and Flow*. American Institute of Physics, 1982.

[9] Dana Bishara, Yuxi Xie, Wing Kam Liu, and Shaofan Li. A state-of-the-art review on machine learning-based multiscale modeling, simulation, homogenization and design of materials. *Arch. Comput. Methods Eng.*, 30(1):191–222, 2023.

[10] Xavier Blanc and Claude Le Bris. *Homogenization Theory for Multiscale Problems: An Introduction*, volume 21 of *MS&A. Modeling, Simulation and Applications*. Springer, 2023.

[11] Renald Brenner and Pierre Suquet. Overall response of viscoelastic composites and polycrystals: exact asymptotic relations and approximate estimates. *Int. J. Solids Struct.*, 50(10):1824–1838, 2013.

[12] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, volume 31 of *Advances in Neural Information Processing Systems*, pages 6572–6583. Curran Associates, Inc., 2018.

[13] Doina Cioranescu and Patrizia Donato. *An introduction to homogenization*. Oxford University Press, 1999.

[14] Bernard D. Coleman and Walter Noll. Foundations of linear viscoelasticity. *Rev. Mod. Phys.*, 33(2):239–249, 1961.

[15] Bernard D. Coleman and Walter Noll. Recent results in the continuum theory of viscoelastic fluids. *Ann. N.Y. Acad. Sci.*, 89(4):672–714, 1961.

[16] David Darrow and George Stepaniants. A Spectral Theory of Scalar Volterra Equations. `https://arxiv.org/abs/2503.06957`, 2025.

[17] Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, volume 32 of *Advances in Neural Information Processing Systems*, pages 3140–3150. Curran Associates, Inc., 2019.

[18] Robert Eggersmann, Trenton Kirchdoerfer, Stefanie Reese, Laurent Stainier, and Michael Ortiz. Model-free data-driven inelasticity. *Comput. Methods Appl. Mech. Eng.*, 350:81–99, 2019.

[19] J. D. Ferry. *Viscoelastic Properties of Polymers*. John Wiley & Sons, 1980.

[20] Moritz Flaschel, Paul Steinmann, Laura De Lorenzis, and Ellen Kuhl. Convex neural networks learn generalized standard material models. *J. Mech. Phys. Solids*, 200: article no. 106103, 2025.

[21] Gilles A. Francfort and Pierre Suquet. Homogenization and mechanical dissipation in thermoviscoelasticity. *Arch. Ration. Mech. Anal.*, 96(3):265–293, 1986.

[22] Jan N. Fuhg, Govinda Anantha Padmanabha, Nikolaos Bouklas, Bahador Bahmani, WaiChing Sun, Nikolaos N. Vlassis, Moritz Flaschel, Pietro Carrara, and Laura De Lorenzis. A review on data-driven constitutive laws for solids. *Arch. Comput. Methods Eng.*, 32:1841–1883, 2025.

[23] Jan N. Fuhg and Nikolaos Bouklas. On physics-informed data-driven isotropic and anisotropic constitutive models through probabilistic machine learning and space-filling sampling. *Comput. Methods Appl. Mech. Eng.*, 394: article no. 114915 (27 pages), 2022.

[24] F. Ghavamian and A. Simone. Accelerating multiscale finite element simulations of history-dependent materials using a recurrent neural network. *Comput. Methods Appl. Mech. Eng.*, 357: article no. 112594 (23 pages), 2019.

[25] Bernhard Gross. *Mathematical structure of the theories of viscoelasticity*. 1968.

[26] JoséMiranda Guedes and Noboru Kikuchi. Preprocessing and postprocessing for materials based on the homogenization method with adaptive finite element methods. *Comput. Methods Appl. Mech. Eng.*, 83(2):143–198, 1990.

[27] Ehsan Haghighat, Sahar Abouali, and Reza Vaziri. Constitutive model characterization and discovery using physics-informed deep learning. *Eng. Appl. Artif. Intell.*, 120: article no. 105828, 2023.

[28] Mark F. Horstemeyer and Douglas J. Bammann. Historical review of internal state variable theory for inelasticity. *Int. J. Plast.*, 26(9):1310–1334, 2010.

[29] Daniel Zhenghu Huang, Nicholas H. Nelsen, and Margaret Trautner. An operator learning perspective on parameter-to-observable maps. *Found. Data Sci.*, 7(1):163–225, 2025.

[30] Reese E. Jones, Ari L. Frankel, and K. L. Johnson. A neural ordinary differential equation framework for modeling inelastic stress response via internal state variables. *J. Mach. Learn. Model. Comput.*, 3(3):1–35, 2022.

[31] Mina Karimi and Kaushik Bhattacharya. A learning-based multiscale model for reactive flow in porous media. *Water Resour. Res.*, 60(9): article no. e2023WR036303, 2024.

[32] Jeong Hee Kim, Daejong Yang, and Seungman Park. Experimental validation for the interconversion between generalized Kelvin–Voigt and Maxwell models using human skin tissues. *J. Biomech.*, 162: article no. 111908, 2024.

[33] Jean Kossaifi, Nikola Kovachki, Kamyar Azizzadenesheli, and Anima Anandkumar. Multi-Grid Tensorized Fourier Neural Operator for High-Resolution PDEs. `https://arxiv.org/abs/2310.00120`, 2023.

[34] Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On universal approximation and error bounds for Fourier neural operators. *J. Mach. Learn. Res.*, 22: article no. 290 (76 pages), 2021.

[35] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *J. Mach. Learn. Res.*, 24: article no. 89 (97 pages), 2023.

[36] Sergei Mikhailovich Kozlov. Averaging of random operators. *Sb. Math.*, 37(2):167–180, 1980.

[37] Michael A. Kraus, Miriam Schuster, Johannes Kuntsche, Geralt Siebert, and Jens Schneider. Parameter identification methods for visco-and hyperelastic material models. *Glass Struct. Eng.*, 2(2):147–167, 2017.

[38] Noël Lahellec, Renaud Masson, and Pierre Suquet. Effective thermodynamic potentials and internal variables: Particulate thermoviscoelastic composites. *J. Mech. Phys. Solids*, 193: article no. 105891 (29 pages), 2024.

[39] Noël Lahellec, Renaud Masson, and Pierre Suquet. Effective thermodynamic potentials and internal variables: linear viscoelastic composites. *J. Mech. Phys. Solids*, 188: article no. 105649 (25 pages), 2024.

[40] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. `https://arxiv.org/abs/2010.08895`, 2020.

[41] Burigede Liu, Nikola Kovachki, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, Andrew Stuart, and Kaushik Bhattacharya. A learning-based multiscale method and its application to inelastic impact problems. *J. Mech. Phys. Solids*, 158: article no. 104668 (16 pages), 2022.

[42] Burigede Liu, Eric Ocegueda, Margaret Trautner, Andrew Stuart, and Kaushik Bhattacharya. Learning macroscopic internal variables and history dependence from microscopic models. *J. Mech. Phys. Solids*, 178: article no. 105329 (21 pages), 2023.

[43] Xin Liu, Su Tian, Fei Tao, and Wenbin Yu. A review of artificial neural networks in the constitutive modeling of composite materials. *Compos. B. Eng.*, 224: article no. 109152, 2021.

[44] Zeliang Liu, C. T. Wu, and Masataka Koishi. A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. *Comput. Methods Appl. Mech. Eng.*, 345:1138–1168, 2019.

[45] Anders Logg, Kent-Andre Mardal, and Garth Wells, editors. *Automated Solution of Differential Equations by the Finite Element Method*, volume 84 of *Lecture Notes in Computational Science and Engineering*. Springer, 2012.

[46] Alexander Mielke and Aida M. Timofte. Two-Scale Homogenization for Evolutionary Variational Inequalities via the Energetic Formulation. *SIAM J. Math. Anal.*, 39(2):642–668, 2007.

[47] Graeme W. Milton. *The Theory of Composites*. Cambridge University Press, 2002.

[48] Nachiketa Mishra, Jaroslav Vondřejc, and Jan Zeman. A comparative study on low-memory iterative solvers for FFT-based homogenization of periodic media. *J. Comput. Phys.*, 321:151–168, 2016.

[49] Hervé Moulinec and Pierre Suquet. A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Comput. Methods Appl. Mech. Eng.*, 157(1-2):69–94, 1998.

[50] M. Mozaffar, R. Bostanabad, W. Chen, K. Ehmann, Jian Cao, and M. A. Bessa. Deep learning predicts path-dependent plasticity. *Proc. Natl. Acad. Sci. USA*, 116(52):26414–26420, 2019.

[51] A. Nikonov, Arthur Russell Davies, and Igor Emri. The determination of creep and relaxation functions from a single experiment. *J. Rheol.*, 49(6):1193–1211, 2005.

[52] Martin Ostoja-Starzewski and Jun Zhang. Does a fractal microstructure require a fractional viscoelastic model? *Fractal Fract.*, 2(1): article no. 12, 2018.

[53] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. *PyTorch: an imperative style, high-performance deep learning library*, volume 32 of *Advances in Neural Information Processing Systems*, pages 8026–8037. Curran Associates, Inc., 2019.

[54] Grigoris Pavliotis and Andrew Stuart. *Multiscale methods: averaging and homogenization*. Springer, 2008.

[55] Rob Phillips and Phillips Rob. *Crystals, defects and microstructures: modeling across scales*. Cambridge University Press, 2001.

[56] James R. Rice. Inelastic constitutive relations for solids: an internal-variable theory and its application to metal plasticity. *J. Mech. Phys. Solids*, 19(6):433–455, 1971.

[57] Enrique Sánchez-Palencia. *Non-homogeneous media and vibration theory*, volume 127 of *Lecture Notes in Physics*. Springer, 1980.

[58] A. Serra-Aguila, J. M. Puigoriol-Forcada, G. Reyes, and J. Menacho. Viscoelastic models revisited: characteristics and interconversion formulas for generalized Kelvin–Voigt and Maxwell models. *Acta Mech. Sin.*, 35:1191–1209, 2019.

[59] Sachin Shanbhag. A computer program for interconversion between creep compliance and stress relaxation. *J. Rheol.*, 67(5):965–975, 2023.

[60] Pierre Suquet. Elements of homogenization for inelastic solid mechanics. In *Homogenization techniques for composite media*, volume 272 of *Lecture Notes in Physics*, pages 193–278. Springer, 1987.

[61] Luc Tartar. Memory effects and homogenization. In *Mechanics and Thermodynamics of Continua: A Collection of Papers Dedicated to BD Coleman on His Sixtieth Birthday*, pages 537–549. Springer, 1991.

[62] Nicholas W. Tschoegl. *The phenomenological theory of linear viscoelastic behavior: an introduction*. Springer, 2012.

[63] Umberto Villa, Noemi Petra, and Omar Ghattas. Hippylib: An Extensible Software Framework for Large-Scale Inverse Problems Governed by PDEs: Part I: Deterministic Inversion and Linearized Bayesian Inference. *ACM Trans. Math. Softw.*, 47(2): article no. 16 (34 pages), 2021.

[64] E. Weinan. *Principles of multiscale modeling*. Cambridge University Press, 2011.

[65] Yupeng Zhang and Kaushik Bhattacharya. Iterated learning and multiscale modeling of history-dependent architectured metamaterials. `https://arxiv.org/abs/2402.12674`, 2024.

[66] Tarek I. Zohdi and Peter Wriggers. *An introduction to computational micromechanics*. Springer, 2008.