Learning Markovian Homogenized Models in Viscoelasticity

Kaushik Bhattacharya\textsuperscript{1}, Burigede Liu\textsuperscript{2}, Andrew Stuart\textsuperscript{3}, Margaret Trautner\textsuperscript{4}

\textbf{Abstract:} Fully resolving dynamics of materials with rapidly-varying features involves expensive fine-scale computations which need to be conducted on macroscopic scales. The theory of homogenization provides an approach to derive effective macroscopic equations which eliminates the small scales by exploiting scale separation. An accurate homogenized model avoids the computationally-expensive task of numerically solving the underlying balance laws at a fine scale, thereby rendering a numerical solution of the balance laws more computationally tractable.

In complex settings, homogenization only defines the constitutive model implicitly, and machine learning can be used to learn the constitutive model explicitly from localized fine-scale simulations. In the case of one-dimensional viscoelasticity, the linearity of the model allows for a complete analysis. We establish that the homogenized constitutive model may be approximated by a recurrent neural network (RNN) that captures the memory. The memory is encapsulated in the evolution of an appropriate finite set of internal variables, discovered through the learning process and dependent on the history of the strain. Simulations are presented which validate the theory. Guidance for the learning of more complex models, such as arise in plasticity, by similar techniques, is given.

\textbf{1 Introduction}

Many problems in continuum mechanics lead to constitutive laws which are history-dependent. This property may be inherent to physics beneath the continuum scale (for example in plasticity \cite{1, 2}) or may arise from homogenization of rapidly varying continua \cite{3, 4} (for example in the Kelvin-Voigt (KV) model of viscoelasticity \cite{5}). When history-dependence is present, Markovian models that capture the history dependence are desirable for both interpretability and computability. In some cases theory may be used to justify Markovian models which capture this history-dependence, but in many cases data plays a central role in finding such models. The goal of this paper is to study data-driven methods to learn Markovian models for history-dependence and to provide theoretical underpinnings for understanding them.

The paper \cite{6} adopted a data-driven learning approach to uncovering history-dependent homogenized models arising in crystal plasticity. However, the resulting constitutive model is not causal and instead learns causality approximately from computations performed at the level of the cell problem. The paper \cite{7} introduces a different approach, learning \textit{causal} constitutive models of plasticity. In order to give rigorous underpinnings to the empirical results therein, the present work is devoted to studying the methodology from \cite{7} in the setting of linear one-dimensional viscoelasticity. Here we can use theoretical understanding to justify and validate the methodology; we show

\textsuperscript{1}Mechanical and Civil Engineering, California Institute of Technology, Pasadena, CA, USA (bhatta@caltech.edu)
\textsuperscript{2}Department of Engineering, University of Cambridge, Cambridge, UK (bl377@eng.cam.ac.uk)
\textsuperscript{3}Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA (astuart@caltech.edu)
\textsuperscript{4}Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA (trautner@caltech.edu)
that machine-learned homogenized models can accurately approximate the dynamics of multiscale models at much cheaper evaluation cost. We obtain insight into desirable choice of training data to learn the homogenized constitutive model, and we study the effect of the multiple local minimizers which appear in the underlying optimization problem. Furthermore, the rigorous underpinnings enable us to gain insight into how to test model hypotheses. We demonstrate that hypothesizing the correct model leads to robustness with respect to changes in time-discretization in the causal model: the model can be trained at one time-step and used at others; and the model can be trained with one time-integration method and used with others. In contrast, hypothesizing an incorrect model leads to intolerable sensitivity with respect to the time-step. Thus training at one time-step and testing at other levels of resolution provides a method for testing model form hypotheses. We work primarily with the one-dimensional KV model for viscoelasticity for which the constitutive model depends only on strain and strain rate. We will also briefly touch on the standard linear solid (SLS) model for which the constitutive relation depends only on the strain and the strain history; in so doing we show that the ideas presented extend beyond the specifics of the one-dimensional KV setting.

In Subsection 1.1 we describe the overarching mathematical framework adopted, and Subsection 1.2 contains a detailed literature review. This is followed, in Subsection 1.3, by a statement of our contributions and an overview of the paper. In Subsection 1.4 we summarize notation used throughout the remainder of the paper.

1.1 Set-up

Consider the problem of material response on an arbitrary spatial domain \( D \subset \mathbb{R}^d \) where the material properties vary rapidly within the domain. We denote by \( u_\epsilon \in \mathbb{R}^d \) the displacement, where \( \epsilon : 0 < \epsilon \ll 1 \) denotes the scale of the material fluctuations. Denote by \( T = (0, T) \) the time domain of interest. We consider continuum models which satisfy dynamical equations of the form

\[
\rho \frac{\partial^2}{\partial t^2} u_\epsilon = \nabla \cdot \sigma_\epsilon + f, \quad (x, t) \in D \times T, \tag{1.1a}
\]

\[
\sigma_\epsilon(x, t) = \Psi_\epsilon^\dagger \left( \nabla u_\epsilon(x, t), \partial_t \nabla u_\epsilon(x, t), \{\nabla u_\epsilon(x, \tau)\}_{\tau \in T} \right), \quad (x, t) \in D \times T, \tag{1.1b}
\]

\[
u_\epsilon = u^*, \quad \partial_t u_\epsilon = v^*, \quad (x, t) \in D \times \{0\}, \tag{1.1c}
\]

\[
u_\epsilon = 0, \quad x \in \partial D, \quad (x, t) \in \partial D \times T. \tag{1.1d}
\]

From these equations we seek \( u_\epsilon : D \times T \to \mathbb{R}^d \). Equation (1.1a) is the balance equation with inertia term \( \rho \frac{\partial^2}{\partial t^2} u_\epsilon \) for known parameter \( \rho \in \mathbb{R}^+ \), resultant stress term \( \nabla \cdot (\sigma_\epsilon) \) where \( \sigma_\epsilon \in \mathbb{R}^{d \times d} \) is the internal stress tensor, and known external forcing \( f \in \mathbb{R}^d \); equations (1.1c,1.1d) specify the initial and boundary data for the displacement. Equation (1.1b) is the constitutive law relating properties of the the strain \( \nabla u_\epsilon \) to the stress \( \sigma_\epsilon \) via map \( \Psi_\epsilon^\dagger \). In this paper we will consider this model with inertia (\( \rho > 0 \)) and without inertia (\( \rho \equiv 0 \)).

1.1.1 Constitutive Model

The constitutive model is defined by \( \Psi_\epsilon^\dagger : \mathbb{R}^{d \times d} \times \mathbb{R}^{d \times d} \times C(T; \mathbb{R}^{d \times d}) \times D \times T \to \mathbb{R}^{d \times d} \). It takes time \( t \) as input, which enables the stress at time \( t \) to be expressed only in terms of the strain history up to time \( t \), \( \{\nabla u_\epsilon(x, \tau)\}_{\tau = 0}^t \), and not on the future of the strain for \( \tau \in (t, T] \). It takes \( x \) as input to allow for material properties which depend on the rapidly varying \( x/\epsilon \); it is also possible to allow for
material properties which exhibit additional dependence on the slowly varying $x$, but we exclude this case for simplicity.

Such constitutive models include a variety of plastic, viscoelastic and viscoplastic materials. In this paper we focus mainly on the one-dimensional KV viscoelastic setting in order to highlight ideas; in this case $\Psi_\epsilon^\dagger$ is independent of the history of the strain. However, we will briefly demonstrate that similar concepts relating to the learning of constitutive models also apply to the case of an SLS, for which $\Psi_\epsilon^\dagger$ is independent of the strain rate but does depend on the history of the strain; the SLS contains the KV and Maxwell models of viscoelasticity as special cases. Furthermore, the paper [7] includes empirical evidence that similar concepts relating to the learning of constitutive models also apply in the setting of Maxwell models of plasticity.

1.1.2 Homogenized Constitutive Model

The goal of homogenization is to find constitutive models which eliminate small-scale dependence. To this end, we first discuss the form of a general homogenized problem: we seek the equation satisfied by $u_0$, an appropriate limit of $u_\epsilon$ as $\epsilon \to 0$. Then map $\Psi_\epsilon^\dagger$ defines the constitutive relationship in a homogenized model for $u_0$ of the form

$$\rho \partial_t^2 u_0 = \nabla \cdot \sigma_0 + f, \quad (x, t) \in D \times T,$$

$$\sigma_0(x, t) = \Psi_0^\dagger (\nabla u_0(x, t), \partial_t \nabla u_0(x, t), \{\nabla u_0(x, \tau)\}_{\tau \in T}, t), \quad (x, t) \in D \times T,$$

$$u_0 = u^*, \quad \partial_t u_0 = v^*, \quad (x, t) \in D \times \{0\},$$

$$u_0 = 0, \quad (x, t) \in \partial D \times T.$$

The key assumption about this homogenized model is that parameter $\epsilon$ no longer appears. Furthermore, since we assumed that the multiscale model material properties depend only on the rapidly varying scale $x/\epsilon$ and not on $x$, we have that $\Psi_\epsilon^\dagger$ is independent of $x$: $\Psi_\epsilon^\dagger : \mathbb{R}^{d \times d} \times \mathbb{R}^{d \times d} \times C(T; \mathbb{R}^{d \times d}) \times T \to \mathbb{R}^{d \times d}$.

If the homogenized model is identified correctly then dynamics under the multiscale model $\Psi_\epsilon^\dagger$, i.e. $u_\epsilon$, can approximated by dynamics under the homogenized model $\Psi_0^\dagger$, i.e. $u_0$. This, potentially, facilitates cheaper computations since length-scales of size $\epsilon$ need not be resolved. We observe, however, that for KV viscoelasticity, the homogenized model contains history-dependence (memory) even though the multiscale model does not. Markovian history-dependence is desirable for two primary reasons: first, Markovian models encode conceptual understanding, representing the history dependence in a compact, interpretable form; second, Markovian expression reduces computational cost from $O(|T|^2)$ in the general memory case to $O(|T|)$ in the Markovian case. In the general media setting, for a multitude of models in viscoelasticity, viscoplasticity, and plasticity, the homogenized model will depend on the memory in a non-Markovian manner. However, it is interesting to determine situations in which accurate Markovian approximations can be found.

1.1.3 Markovian Homogenized Constitutive Model

We will seek to identify internal variables $\xi$ and functions $F, G$ such that, for $B \in C(T; \mathbb{R}^{d \times d})$, $\Psi_0^\dagger$ can be approximated by

$$\Psi_0 (B(t), \partial_t B(t), \{B(\tau)\}_{\tau \in T}, t) = F (B(t), \partial_t B(t), \xi(t)), \quad (1.3a)$$

$$\partial_t \xi(t) = G (\xi(t), B(t)). \quad (1.3b)$$
Note that $\xi$ carries the history dependence on $B$ through its Markovian evolution. We assume that $\xi \in C(T;\mathbb{R}^r)$ for some integer $r$ and hence that $F : \mathbb{R}^{d \times d} \times \mathbb{R}^{d \times d} \times \mathbb{R}^r \to \mathbb{R}^{d \times d}$ and that $G : \mathbb{R}^r \times \mathbb{R}^{d \times d} \to \mathbb{R}^r$. In dimension $d > 1$ there will be further symmetries that should be built into the model, but as the concrete analysis in this paper is in dimension $d = 1$ we will not detail these symmetries here.

In general such a Markovian model can only approximate the true model, and the nature of the physics leading to a good approximation will depend on the specific continuum mechanics problem. To determine $F$ and $G$ in practice we will parameterize them as neural networks, which enables us to use general purpose optimization software to determine suitable values of the parameters. In doing so we identify an operator class $\Psi_0(\cdot ; \theta)$ and parameter space $\Theta$ such that, for some judiciously chosen $\theta^* \in \Theta$, $\Psi_0(\cdot ; \theta^*) \approx \Psi_0^\dagger \epsilon$.

In this paper we will concentrate on justifying a Markovian homogenized approximation in the context of one-dimensional KV viscoelasticity. Our justification will use theory that is specific to one dimensional linear viscoelasticity, and we demonstrate that the approach also works for the general SLS, which includes the KV model as a particular limit. Furthermore, the paper [7] contains evidence that the ideas we develop apply beyond the confines of one dimensional linear viscoelasticity and into nonlinear plasticity in higher spatial dimensions.

The specific property of one-dimensional viscoelasticity that we exploit to underpin our analysis (and which applies to the SLS and therefore also to the KV model) is that, for piecewise-constant media, the homogenized model has a memory term which can be represented in a Markovian way. Therefore, to justify our strategy of approximating by Markovian models we will: first, approximate the rapidly-varying medium by a piecewise-constant rapidly-varying medium; secondly, homogenize this model to find a Markovian description; and finally, demonstrate how the Markovian description can be learned from data at the level of the unit cell problem, using neural networks. For more general problems we anticipate a similar justification holding, but with different specifics leading to the existence of good approximate Markovian homogenized models. The benefit of the one-dimensional viscoelastic setting is that, through theory, we obtain underpinning insight into the conceptual approach more generally, and in particular for plasticity; this theory underpins the numerical experiments which follow.

1.2 Literature Review

The continuum assumption for physical materials approximates the inherently particulate nature of matter by a continuous medium and thus allows the use of partial differential equations to describe response dynamics. We refer the reader to [8, 9, 10] for a general background. In continuum mechanics, the governing equations are derived by combining universal balance laws of physics (balance of mass, momenta and energy) with a constitutive relation that describes the properties of the material being studied. This is typically specified as the relation between a dynamic quantity like stress or energy and kinematic quantities like strain and its history. The constitutive relation of many materials are history dependent, i.e., the state of stress at an instant depends on the history of deformation. It is common in continuum mechanics to incorporate this history dependence through the introduction of internal variables. We refer the reader to [11] for a systematic formulation of internal variable theories.

Of particular interest in this work are viscoelastic materials. We refer the reader to [12, 13] for
a general background. In viscoelastic materials, the state of stress at any instant depends on the strain and its history. There are various models where the stress depends only on strain and strain rate (Kelvin-Voigt), internal variables (standard linear solids), convolution kernels and fractional time derivatives.

While constitutive relations were traditionally determined empirically, more recently there has been a systematic attempt to understand them from more fundamental solids, and this has given rise to a rich activity in multiscale modeling of materials [14, 15, 16]. Materials are heterogeneous on various length (and time) scales, and it is common to use different theories to describe the behavior at different scales [17]. The goal of multiscale modeling of materials is to use this hierarchy of scales to understand the overall constitutive behavior at the scale of applications. The hierarchy of scales include a number of continuum scales. For example, a composite material is made of various distinct materials arranged at a scale that is small compared to the scale of application, but large enough compared to an atomistic/discrete scale that their behavior is adequately described by continuum mechanics. Or, for example, a polycrystal, is made of a large collection of grains (regions of identical anisotropic material but with differing orientation) that are small compared to the scale of application but large enough for a continuum theory. Homogenization theory leverages the assumption of the separation of scales to average out the effects of fine-scale material variations. To estimate macroscopic response of heterogeneous materials, asymptotic expansion of the displacement field yields a set of boundary value problems whose solution produces an approximation that does not depend on the microscale [3, 18]. The fundamentals of asymptotic homogenization theory are well-established [4, 19, 20]. Milton [21] provides a comprehensive survey of the effective or homogenized properties.

Homogenization in the context of viscoelasticity was initiated by Sanchez-Palencia ([18] Chapter 6), who pointed out that the homogenization of a Kelvin-Voigt model leads to a model with fading memory. Further discussion of homogenization theory in (thermo-)viscoelasticity can be found in Francfort and Suquet [5], and a detailed discussion of the overall behavior including memory in Brenner and Suquet [22]. A broader discussion of homogenization and memory effects can be found in Tartar [23]. It is now understood that homogenization of various constitutive models gives rise to memory.

As noted above, according to homogenization theory, the macroscopic behavior depends on the solution of a boundary value problem at the microscale. Evaluating the macroscopic behavior by the solution of a boundary value problem computationally leads to what has been called computational micromechanics [24]. These often involve periodic boundary conditions, and fast Fourier transform-based methods are widely used since Moulinec and Suquet [25] (see [26, 27] for recent summaries). While these enable us to compute the macroscopic response for a particular deformation history, one needs to repeat the calculation for all possible deformation histories.

Therefore, recent work in the mechanics literature addresses the issue of learning homogenized constitutive models from computational data [28, 29, 6, 7] or experimental data [30]. This learning problem requires determination of maps that take as inputs functions describing microstructural properties, and leads in to the topic of operator learning.

Neural networks on finite-dimensional spaces have proven effective at solving longstanding computational problems. A natural question which arises from this success is whether neural networks can be used to solve partial differential equations. From a theoretical standpoint, the classical notion
of neural networks as maps between finite-dimensional spaces is insufficient. Indeed, a differential operator is a map between infinite-dimensional spaces. Similar to the case of finite-dimensional maps, universal approximation results for nonlinear operator learning have been developed [31]. In one approach to operator learning, model reduction methods are applied to the operator itself. In this setting, a low-dimensional approximation space is assumed, and the operator approximation is constructed via regression using the latent basis [32, 33]. In a second method, classical dimension reduction maps are applied to the input and output spaces, and an approximation map is learned between these finite-dimensional spaces [34]. An important result of this method is its mesh-invariance property. Data for any operator-learning problem must consist of a finite discretization of the true input and output functions. One side effect of this practical fact is that some existing methods for operator learning depend critically on the choice of discretization. Mesh-invariant methods are desirable for both practical and theoretical purposes. Practically, it would be expensive to train a new model to accomodate a finer data resolution. From a theory standpoint, since the true map between infinite-dimensional spaces is inherently resolution-independent, the operator approximation ought to have this property as well. Mesh-invariance allows the operator approximation to be applied in the use of various numerical approximation methods for PDEs, which is of particular importance when the operator is being used as a surrogate model for the overall PDE system [35].

Surrogate modeling bypasses expensive simulation computations by replacing part of the PDE with a neural network. In several application settings, including fluid flows and solid mechanics computations, surrogate modeling has met empirical success in approximating the true solution [36, 37]. This work continues to use ideas from physics-informed machine learning; in [37] in particular, the differential operator is incorporated into the cost function via automatic differentiation. Other work in surrogate modeling for solid mechanics proposes a hierarchical network architecture that mimics the heterogeneous material structure to yield an approximation to the homogenized solution [38].

In this paper we use an RNN as a surrogate model for the constitutive relation on the microscale. The RNN can then be used to evaluate the forward dynamic response on the microscale cells, whose results are combined with traditional numerical approximation methods to yield the macroscale response. Furthermore, the RNN that we train at a particular time-discretization is also accurate when used at other time-discretizations, if the correct model form is proposed. However, the RNN is trained for a particular choice of material parameters, and the resulting model cannot be used at different material parameter values; simultaneously learning material parameter dependence is left for future work.

1.3 Our Contributions and Paper Overview

Our contributions are as follows:

1. We provide theoretical underpinnings for the discovery of Markovian homogenized models in viscoelasticity and plasticity; the methods use data generated by solving cell problems to learn constitutive models of the desired form.

2. We prove that in the one-dimensional Kelvin-Voigt (KV) setting, any solution of the multiscale problem can be approximated by solution of a homogenized problem with Markovian structure.
3. We prove that the constitutive model for this Markovian homogenized system can be approximated by a recurrent neural network (RNN) operator, learnt from data generated by solving the appropriate cell problem.

4. We provide simulations which numerically demonstrate the accuracy of the learned Markovian model.

5. We offer guidance for the application of this methodology, beyond the setting of one-dimensional viscoelasticity, into multi-dimensional plasticity.

In Section 2, we formulate the KV viscoelastic problem and its homogenized solution. In Section 3, we present our main theoretical results, addressing contributions 1., 2. and 3.; these are in the setting of one-dimensional KV viscoelasticity. We prove that solution of the multiscale problem can be approximated by solution of a homogenized Markovian memory-dependent model that does not depend on small scales, and we prove that an RNN can approximate the constitutive law for this homogenized problem. Section 4 contains numerical experiments which address contributions 4. and 5.; the start of that section details the findings.

1.4 Notation

We define notation that will be useful throughout the paper. Recall that $\mathcal{T} = (0, T)$ is the time domain of interest, and in the one-dimensional setting we let $\mathcal{D} = [0, L]$ be the spatial domain. Let $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote the standard inner product and induced norm operations on the Hilbert space $L^2(\mathcal{D}; \mathbb{R})$. Additionally, let $\| \cdot \|_\infty$ denote the $L^\infty(\mathcal{D}; \mathbb{R})$ norm.

It will also be convenient to define the $\xi$-dependent quadratic form

$$q_\xi(u, w) := \int_\mathcal{D} \xi(x) \frac{\partial u(x)}{\partial x} \frac{\partial w(x)}{\partial x} \, dx$$

for arbitrary $\xi \in L^\infty(\mathcal{D}; (0, \infty))$; furthermore we define

$$\xi^+ := \text{ess sup}_{x \in \mathcal{D}} \xi(x) < \infty$$

and

$$\xi^- := \text{ess inf}_{x \in \mathcal{D}} \xi(x) \geq 0.$$  

In this paper we always work with $\xi$ such that $\xi^- > 0$. Under these assumptions $q_\xi(\cdot, \cdot)$ defines an inner-product, and we can define the following norm

$$\|u\|_{H_0^1, \xi}^2 := q_\xi(u, u)$$

from it; note also that we may define a norm on $H_0^1(\mathcal{D}; \mathbb{R})$ by

$$\|u\|_{H_0^1}^2 := q_1(u, u),$$

where $1(\cdot)$ is the function in $L^\infty(\mathcal{D}; (0, \infty))$ taking value 1 in $\mathcal{D}$ a.e. The resulting norms are all equivalent on the space $H_0^1(\mathcal{D}; \mathbb{R})$; this is a consequence of the following lemma:
Lemma 1.1. For any $\xi_1, \xi_2 \in L^\infty(D; (0, \infty))$ satisfying properties (1.5) and (1.6), the norms $\|u\|_{H^1_0, \xi_1}$ and $\|u\|_{H^1_0, \xi_2}$ are equivalent in the sense that

$$\frac{\xi_2}{\xi_1} \|u\|^2_{H^1_0, \xi_1} \leq \|u\|^2_{H^1_0, \xi_2} \leq \frac{\xi_2^+}{\xi_1^-} \|u\|^2_{H^1_0, \xi_1}.$$ 

Proof. For $i = 1, 2$:

$$\xi_i^- \int_0^1 \left| \frac{\partial u}{\partial x} \right|^2 dx \leq \int_0^1 \xi_i(x) \left| \frac{\partial u}{\partial x} \right|^2 dx \leq \xi_i^+ \int_0^1 \left| \frac{\partial u}{\partial x} \right|^2 dx.$$ 

The result follows.

We denote by $V$ the Hilbert space $H^1_0(D; \mathbb{R})$ noting that, as a consequence of the preceding lemma, we may use $q_\xi(u, w)$ as the inner product on this space for any $\xi$ satisfying (1.5) and (1.6). We also define

$$Z = L^\infty(T; L^2(D; \mathbb{R})), \quad Z_2 = L^2(T; L^2(D; \mathbb{R}))$$

with norms

$$\|r\|_Z = \text{ess sup}_{t \in T} (\|r(\cdot, t)\|), \quad \|r\|_{Z_2} = \left( \int_0^T \|r(\cdot, t)\|^2 dt \right)^{\frac{1}{2}}.$$ 

We note that $Z$ is continuously embedded into $Z_2$.

2 One-Dimensional Kelvin-Voigt Viscoelasticity

This paper is focused on one-dimensional KV viscoelasticity because the model is amenable to rigorous analysis. The resulting analysis sheds light on the learning of constitutive models more generally. In Section 2.1 we present the equations for the model in an informal fashion and in a weak form suitable for analysis; and in Section 2.2 we homogenize the model and define the operator defining the effective constitutive model.

2.1 Governing Equations and Weak Form

The one-dimensional KV model for viscoelasticity postulates that stress is affine in the strain and strain rate, with affine transformation dependent on the (typically spatially varying) material properties. For a multiscale material varying with respect to $x/\epsilon$ we thus have the following definition of $\Psi_\epsilon$ from (1.1), in the one-dimensional KV model:

$$\sigma_\epsilon = E_\epsilon \partial_x u_\epsilon + \nu_\epsilon \partial^2_{xt} u_\epsilon$$

where $E_\epsilon(x) = E \left( \frac{x}{\epsilon} \right)$ and $\nu_\epsilon(x) = \nu \left( \frac{x}{\epsilon} \right)$ are rapidly-varying material elasticity and viscosity, respectively. Both $E$ and $\nu$ are assumed to be 1-periodic. Then equations (1.1) without momentum ($\rho \equiv 0$) become

$$-\partial_x \left( E_\epsilon \partial_x u_\epsilon + \nu_\epsilon \partial^2_{xt} u_\epsilon \right) = f, \quad (x, t) \in \partial D \times T,$$

$$u_\epsilon(0, x) = u^*, \quad \partial_t u_\epsilon(0, x) = v^*, \quad x \in \partial D,$$

$$u_\epsilon(x, t) = 0, \quad x \in \partial D.$$ (2.1a)
Any classical solution to equations (2.1) will also solve the corresponding weak form: find \( u_\epsilon \in \mathcal{C}(T; V) \) such that
\[
q_\nu(\partial_t u_\epsilon, \varphi) + q_E(u_\epsilon, \varphi) = \langle f, \varphi \rangle
\]
for all test functions \( \varphi \in V \).

### 2.2 Homogenization

In the inertia-free setting \( \rho = 0 \) we perform homogenization to eliminate the dependence on the small scale \( \epsilon \) in (2.1). First, we take the Laplace transform of (2.1), which gives, for Laplace parameter \( s \) and with the hat symbol denoting Laplace transform,
\[
-\partial_x ((E_\epsilon + \nu_\epsilon s) \partial_x \hat{u}_\epsilon) = \hat{f}, \quad x \in \mathcal{D},
\]
\[
\hat{u}_\epsilon(x, s) = 0, \quad x \in \partial \mathcal{D}.
\]
The initial condition \( u_\epsilon(0, x) = u^* \) is applied upon Laplace inversion. Since \( \epsilon \ll 1 \), we may apply standard techniques from multiscale analysis [3, 4] and seek a solution in the form
\[
\hat{u}_\epsilon = \hat{u}_0 + \epsilon \hat{u}_1 + \epsilon^2 \hat{u}_2 + \ldots.
\]

For convenience, define \( \hat{a}(s, y) = E(y) + \nu(y)s \). Note that \( \hat{a}(s, \cdot) \) is 1-periodic. The leading order term in our approximation, \( \hat{u}_0 \), solves the following uniformly elliptic PDE with Dirichlet boundary conditions:
\[
-\partial_x (\hat{a}_0(s) \partial_x \hat{u}_0) = \hat{f} \quad \text{for} \quad x \in \mathcal{D},
\]
\[
\hat{u}_0 = 0 \quad \text{for} \quad x \in \partial \mathcal{D}.
\]

Here the coefficient \( \hat{a}_0 \) is given by
\[
\hat{a}_0 = \int_0^1 (\hat{a}(y) + \hat{a}(y) \partial_y \chi(y)) \, dy
\]
and \( \chi(y) : [0, 1] \to \mathbb{R} \) satisfies the cell problem
\[
-\partial_y (\hat{a}(y) \partial_y \chi(y)) = \partial_y \hat{a}(y) \tag{2.4a}
\]
\[
\chi \text{ is 1-periodic,} \quad \int_0^1 \chi(y) dy = 0. \quad \tag{2.4b}
\]

Using this, the coefficient \( \hat{a}_0 \) can be computed explicitly as the harmonic average of the original coefficient \( \hat{a} \) [4][Subsection 12.6.1]:
\[
\hat{a}_0(s) = \langle \hat{a}(s)^{-1} \rangle^{-1} = \left( \int_0^1 \frac{dy}{s\nu(y) + E(y)} \right)^{-1}, \tag{2.5}
\]
where \( \langle \cdot \rangle \) denotes spatial averaging over the unit cell.

Equations (2.3) indicate that the homogenized map \( \Psi_0^\dagger \) appearing in (1.2) is, for one-dimensional linear viscoelasticity, defined from
\[
\Psi_0^\dagger(\partial_x u_0(x, t), \partial_{xt} u_0(x, t), \{\partial_x u_0(x, \tau)\}_{\tau \in \mathcal{T}}, t) = \mathcal{L}^{-1}\left( \hat{a}_0(s) \partial_x \hat{u}_0 \right); \tag{2.6}
\]
here $L^{-1}$ denotes the inverse Laplace transform. Note that (2.5) shows that $\hat{a}_0$ grows linearly in $s \to \infty$ and computing the constant term in a regular power series expansion at $s = \infty$ shows that we may write
\[\hat{a}_0 = \nu' s + E' + \hat{\kappa}(s)\]
where $\hat{\kappa}(s)$ decays to 0 as $s \to \infty$. Here
\[\nu' = \left(\int_0^1 \frac{1}{\nu(y)} dy\right)^{-1}, \quad E' = \left(\int_0^1 \frac{E(y)}{\nu(y)^2} dy\right) / \left(\int_0^1 \frac{1}{\nu(y)} dy\right)^2.\]
Details is presented in Appendix B.1. Laplace inversion of $\hat{a}_0(s)\partial_x \hat{u}_0$ then yields the conclusion that
\[\Psi^1_0(\partial_x u_0(t), \partial_x^2 u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t; \theta) = E'\partial_x u_0(t) + \nu'\partial_x^2 u_0(t) + \int_0^t \kappa(t - \tau) \partial_x u_0(\tau) \, d\tau.\] (2.7)

**Remark 2.1.** When $\rho = 0$, the homogenized solution provably approximates $u_\epsilon$ in the $\epsilon \to 0$ limit; see Theorem 3.7. However, although we derived it with inertia set to zero, the homogenized solution given by equation (2.8) is also valid when the inertia term $\rho \partial_t^2 u_\epsilon$ generates contributions which are $O(1)$ with respect to $\epsilon$.

The homogenized PDE for one-dimensional viscoelasticity follows by combining equations (1.2) with equation (2.7) to give
\[\rho \partial_t^2 u_0 = \nabla \cdot \sigma_0 + f, \quad (x, t) \in \mathcal{D} \times \mathcal{T},\] (2.8a)
\[\sigma_0(t) = E'\partial_x u_0(t) + \nu'\partial_x^2 u_0(t) + \int_0^t \kappa(t - \tau) \partial_x u_0(\tau) \, d\tau, \quad (x, t) \in \mathcal{D} \times \mathcal{T},\] (2.8b)
\[u_0 = u^*, \quad \partial_t u_0 = v^*, \quad (x, t) \in \mathcal{D} \times \{0\},\] (2.8c)
\[u_0 = 0, \quad (x, t) \in \partial \mathcal{D} \times \mathcal{T}.\] (2.8d)

The price paid for homogenization is dependence on the strain history. We will show in the next section, however, that we can approximate the general homogenized map with one in which the history-dependence is expressed in a Markovian manner.

### 3 Main Theorems: Statement and Interpretation

In this section we present theoretical results of three types. Firstly, in Subsection 3.1, we show that the solution $u_\epsilon$ to equation (2.1) is Lipschitz when viewed as a mapping from the unit cell material properties $E(\cdot), \nu(\cdot)$ in $L^\infty$ into $Z$; hence, an $O(\delta)$ approximation of $E, \nu$ by piecewise-constant functions leads to an $O(\delta)$ approximation of $u_\epsilon$. Second, in Subsection 3.2, we demonstrate that the homogenized model based on piecewise-constant material properties can be represented in a Markovian fashion by introducing *internal variables*; hence, combining with the first point, we have a mechanism to approximate $u_\epsilon$ by solution a Markovian homogenized model. Third, in Subsection 3.3, we show the existence of neural networks which provide arbitrarily good approximation of the constitutive law arising in the Markovian homogenized model; this suggests a model class within which to learn homogenized, Markovian constitutive models from data. Subsection 3.4 establishes our framework for the optimization methods used to learn such constitutive models; this framework is employed in the subsequent Section 4.
Assumptions 3.1. We will make the following assumptions on $E$, $\nu$, and $f$ throughout:

1. $f \in L^2(\mathcal{D}; \mathbb{R})$ for all $t \in \mathcal{T}$; thus $\|f\|_Z < \infty$;
2. $E^+, \nu^+ < \infty$, and $E^-, \nu^- > 0$.

Note that $E^+ = E_\epsilon^+$ and $\nu^+ = \nu_\epsilon^+$, so we will drop the $\epsilon$ superscript in this notation.

3.1 Approximation by Piecewise Constant Material

Consider (2.1) with continuous material properties $E$ and $\nu$. We show in Theorem 3.4 that we can approximate the solution $u_\epsilon$ to this system by a solution $u_{\epsilon}^{PC}$ which solves (2.1) with suitable piecewise-constant material properties $E^{PC}$ and $\nu^{PC}$, in such a way that $u_\epsilon$ and $u_{\epsilon}^{PC}$ are close. To this end we make precise the definition of piecewise-constant material properties.

Definition 3.2 (Piecewise Constant). A material is piecewise constant on the unit cell with $L$ pieces if the elasticity function $E(y)$ and the viscosity function $\nu(y)$ both take constant values on $L$ intervals $[0, a_1), [a_1, a_2), \ldots, [a_{L-1}, 1]$. In particular, $E(y)$ and $\nu(y)$ have discontinuities only at the same $L-1$ points in the unit cell. We use terminology $L-$piecewise constant to specify the number of pieces.

Remark 3.3. The situation in which $E(y)$ and $\nu(y)$ have discontinuities at different values of $y \in (0, 1)$ can be reduced to the case in Definition 3.2 by increasing the value of $L$.

Theorem 3.4 (Piecewise-Constant Approximation). Let $E$ and $\nu$ be piecewise continuous functions, with a finite number of discontinuities, satisfying Assumptions 3.1; let $u_{\epsilon}$ be the corresponding solution to (2.1). Then, for any $\delta > 0$, there exist piecewise constant $E^{PC}$ and $\nu^{PC}$ (in the sense of Definition 3.2) such that solution $u_{\epsilon}^{PC}$ of equations (2.1) with these material properties satisfies

$$\|u_{\epsilon}^{PC} - u_{\epsilon}\|_Z < \delta.$$ 

Note that Theorem 3.4 is stated in the setting of no inertia. The proof depends on the following lemma; proof of both the theorem and the lemma may be found in Appendix A.1. We observe that, since the Lipschitz result is in the $L^\infty-$norm with respect to the material properties, it holds with constant $C$ independent of $\epsilon$, in the case of interest where the material properties vary rapidly on scale $\epsilon$.

Lemma 3.5 (Lipschitz Solution). Let $u_i$ be the solution to

$$-\partial_x \left( E_i \partial_x u_i + \nu_i \partial_x^2 u_i \right) = f, \quad (x, t) \in \partial \mathcal{D} \times \mathcal{T}, \quad (3.1)$$
$$u_i(x, t) = u^*, \quad (x, t) \in \mathcal{D} \times \{0\}, \quad (3.2)$$
$$u_i(x, t) = 0, \quad (x, t) \in \partial \mathcal{D} \times \mathcal{T}, \quad (3.3)$$

associated with material properties $E_i$, $\nu_i$, for $i \in \{1, 2\}$, and forcing $f$, all satisfying the Assumptions 3.1. Then

$$\|u_1 - u_2\|_Z \leq C (\|\nu_1 - \nu_2\|_\infty + \|E_1 - E_2\|_\infty)$$

for some constant $C \in \mathbb{R}^+$ dependent on $f, E_1^+, E_1^-, \nu_1^+, \nu_1^-$ and $L$ and independent of $\epsilon$. 

11
3.2 Homogenization for Piecewise Constant Material

We show in Theorem 3.6 that for piecewise-constant material properties $E(\cdot)$ and $\nu(\cdot)$, the homogenized map $\Psi_0^\dagger$ given in (2.7) can be written explicitly with a finite number of parameters, and in particular the memory is expressible in a Markovian form. This Markovian form implicitly defines a finite number of internal variables.

**Theorem 3.6 (Existence of Exact Parametrization).** Let $\Psi_0^\dagger$ be the map from strain history to stress in the homogenized model, as defined by equation (2.7), in a piecewise constant material with $L' + 1$ pieces. Define $\Psi_0^{PC}: \mathbb{R}^2 \times C(\mathcal{T}; \mathbb{R}) \times \mathcal{T} \times \Theta \rightarrow \mathbb{R}$ by

$$
\Psi_0^{PC}(\partial_x u_0(t), \partial_{xt}^2 u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t; \theta) = E_0 \partial_x u_0(t) + \nu_0 \partial_{xt}^2 u_0(t) - \sum_{\ell=1}^{L_0} \xi_\ell(t),
$$

$$
\partial_t \xi_\ell(t) = \beta_{t}\partial_x u_0(t) - \alpha_\ell \xi_\ell(t), \quad \xi_\ell(0) = 0, \quad \ell \in \{1, \ldots, L_0\},
$$

with parameter space

$$
\Theta = \left( E_0 \in \mathbb{R}_+, \nu_0 \in \mathbb{R}_+, L_0 \in \mathbb{Z}_+, \alpha_0 \in \mathbb{R}_{L_0}^+, \beta_0 \in \mathbb{R}^{L_0} \right).
$$

Then, under Assumptions 3.1, there exists $\theta^* \in \Theta$ with $(E_0, \nu_0, L_0, \alpha_0, \beta_0) = (E', \nu', L', \alpha, \beta)$ such that

$$
\Psi_0^\dagger(\partial_x u_0(t), \partial_{xt}^2 u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t) = \Psi_0^{PC}(\partial_x u_0(t), \partial_{xt}^2 u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t; \theta^*)
$$

for all $u_0 \in C^2(\mathcal{D} \times \mathcal{T}; \mathbb{R})$ and $t \in \mathcal{T}$.

The proof of the above theorem may be found in Appendix A.2. Note that the model in equations (3.4) is Markovian. Furthermore, although the model in (3.4) requires an input of $t$ for evaluation, the spatial variable $x$ only enters implicitly through the local values of $\partial_x u_0$ and $\partial_{xt}^2 u_0$; the model acts pointwise in space. Thus we have not included $x$ explicitly in the theorem statement, for economy of notation. In what follows it is useful to define $u_0^{PC}$ to be the solution to the following system defined with constitutive model $\Psi_0^{PC}$:

$$
\rho \partial_t^2 u_0^{PC} - \partial_x \sigma_0 = f, \quad (x, t) \in \mathcal{D} \times \mathcal{T}
$$

$$
\sigma_0(t) = \Psi_0^{PC}\left(\partial_x u_0^{PC}(t), \partial_{xt}^2 u_0^{PC}(t), \{\partial_x u_0^{PC}(\tau)\}_{\tau \in \mathcal{T}}, t\right), \quad (x, t) \in \mathcal{D} \times \mathcal{T}
$$

$$
u_u |_{t=0} = u^*, \quad \partial_t u_0^{PC} |_{t=0} = v^*, \quad (x, t) \in \mathcal{D} \times \{0\}
$$

$$
\left. u_0^{PC} \right|_{t=0} = 0, \quad (x, t) \in \partial \mathcal{D} \times \mathcal{T}.
$$

Using a homogenization theorem, together with approximation by piecewise constant material properties, we now show that $u_\epsilon$ can be approximated by $u_0^{PC}$; this will follow from the inequality

$$
\|u_\epsilon - u_0^{PC}\|_{L_2} \leq \|u_\epsilon - u_\epsilon^{PC}\|_{L_2} + \|u_\epsilon^{PC} - u_0^{PC}\|_{L_2}.
$$

The first term on the righthand-side may be controlled using Theorem 3.4. The fact that dynamics under constitutive law $\Psi_0^\dagger$ converge to those under $\Psi_0^\dagger$ as $\epsilon \rightarrow 0$ may be used to control the second term; this fact is a consequence of the following theorem:
Theorem 3.7. Under Assumptions 3.1, the solution $u_\epsilon$ to equations (2.1) converges weakly to $u_0$, the solution to equations (2.8) with $\rho = 0$, in $W^{1,2}(\mathcal{T}; V)$. Thus, for any $\eta > 0$ there exists $\epsilon_{\text{crit}} > 0$ such that for all $\epsilon \in (0, \epsilon_{\text{crit}})$,
\[\|u_\epsilon - u_0\|_{Z_2} < \eta.\]  
(3.7)

Proof. Since $f \in Z$, continuous embedding gives $f \in Z_2$. Applying Theorem 3.1[5] (noting that the work in that paper is set in dimension $d = 3$, but is readily extended to dimension $d = 1$) establishes weak convergence of $u_\epsilon$ to $u_0$ in $W^{1,2}(\mathcal{T}; V)$. Hence strong convergence in $Z_2$ follows, by compact embedding of $W^{1,2}(\mathcal{T}; V)$ into $Z_2$. \qed

The following corollary is a consequence of Theorem 3.7.

Corollary 3.8. Under Assumptions 3.1 and assuming $E, \nu$ are piecewise constant, the solution $u_{\epsilon}^{\text{PC}}$ to equations (2.1) converges weakly to $u_0^{\text{PC}}$, the solution to equations (3.6) with $\rho = 0$, in $W^{1,2}(\mathcal{T}; V)$. Thus, for any $\eta > 0$ there exists $\epsilon_{\text{crit}} > 0$ such that for all $\epsilon \in (0, \epsilon_{\text{crit}})$,
\[\|u_{\epsilon}^{\text{PC}} - u_0^{\text{PC}}\|_{Z_2} < \eta.\]  
(3.8)

Combining this result with that of Theorem 3.4, noting continuous embedding of $Z$ into $Z_2$, allows us to approximate $u_\epsilon$ by $u_0^{\text{PC}}$.

Corollary 3.9. Let $E$ and $\nu$ be piecewise-continuous functions, with a finite number of discontinuities satisfying, along with $f$, Assumptions 3.1; let $u_\epsilon$ be the corresponding solution to (2.1). Then for any $\eta > 0$, there exists $L_{\text{crit}}$ and $\epsilon_{\text{crit}}$ with the property that for all $L \geq L_{\text{crit}}$ there are $L$–piecewise-constant $E^{\text{PC}}$ and $\nu^{\text{PC}}$ such that for all $\epsilon \in (0, \epsilon_{\text{crit}})$, the solution to $u_0^{\text{PC}}$ to (3.6) with $\rho = 0$ satisfies
\[\|u_\epsilon - u_0^{\text{PC}}\|_{Z_2} < \eta.\]  
(3.9)

3.3 Neural Network Approximation of Constitutive Model

For the specific KV model in dimension $d = 1$ we know the postulated form of $\Psi_0^{\text{PC}}$ and can in principle use this directly as a constitutive model. However, in more complex problems we do not know the constitutive model analytically, and it is then desirable to learn it from data from within an expressive model class. To this end we demonstrate that $\Psi_0^{\text{PC}}$ can be approximated by an operator $\Psi_0^{\text{RNN}}$ which has a similar form to that defined by equations (3.4) but in which the right hand sides of those equations are represented by neural networks, leading to a recurrent neural network structure.

Recall the definitions of $(E', \nu', L')$ and $\theta^*$ in Theorem 3.6. We first define the linear functions $\mathcal{F}^{\text{PC}} : \mathbb{R} \times \mathbb{R} \times \mathbb{R}' \to \mathbb{R}$ and $\mathcal{G}^{\text{PC}} : \mathbb{R}' \times \mathbb{R} \to \mathbb{R}$ by
\[\mathcal{F}^{\text{PC}}(b, c, r) = E'b + \nu'c - \langle 1, r \rangle\]  
(3.10a)
\[\mathcal{G}^{\text{PC}}(r, b) = -Ar + \beta b\]  
(3.10b)
where $A = \text{diag}(\alpha_i) \in \mathbb{R}' \times \mathbb{R}'$ and $\beta = \{\beta_1, \ldots, \beta_\ell\} \in \mathbb{R}'$. We then have

$\Psi_0^{\text{PC}}(\partial_x u_0(t), \partial_{xx} u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t; \theta^*) = \mathcal{F}^{\text{PC}}(\partial_x u_0(t), \partial_{xx} u_0(t), \xi(t)), \xi(t) = \mathcal{G}^{\text{PC}}(\xi(t), \partial_x u_0(t)), \xi(0) = 0$  
(3.11a)
(3.11b)
as in Theorem 3.6.

We seek to approximate this map by \( \Psi_0^{RNN} \) defined by replacing the linear functions \( \mathcal{F}^{PC} \) and \( \mathcal{G}^{PC} \) by neural networks \( \mathcal{F}^{RNN} : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^L \to \mathbb{R} \) and \( \mathcal{G}^{RNN} : \mathbb{R}^L \times \mathbb{R} \to \mathbb{R} \) to obtain
\[
\Psi_0^{RNN}(\partial_x u_0(t), \partial^2_{xt} u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t) = \mathcal{F}^{RNN}(\partial_x u_0(t), \partial^2_{xt} u_0(t), \xi(t))
\]
\[
\xi(t) = \mathcal{G}^{RNN}(\xi(t), \partial_x u_0(t)), \quad \xi(0) = 0.
\]

Let \( R > 0 \) and define the bounded set \( Z_R = \{w : \mathbb{R}^+ \to \mathbb{R} | \sup_{t \in \mathcal{T}} |w(t)| \leq R\} \).

**Theorem 3.10** (RNN Approximation). Consider \( \Psi_0^{PC} \) defined as by equations (3.10), (3.11). Assume that there exist \( \rho > 0 \) and \( 0 \leq B < \infty \) such that \( \rho < \min_{\ell} |\alpha_\ell| \) and \( \max_{\ell} |\beta_\ell| \leq B \). Then, under Assumptions 3.1, for every \( \eta > 0 \) there exists \( \Psi_0^{RNN} \) of the form (3.12) such that
\[
\sup_{t \in \mathcal{T}, b, c \in Z_R} |\Psi_0^{PC}(b(t), c(t), \{b(\tau)\}_{\tau \in \mathcal{T}}, t; \theta^*) - \Psi_0^{RNN}(b(t), c(t), \{b(\tau)\}_{\tau \in \mathcal{T}}, t)| < \eta.
\]

The proof of Theorem 3.10 can be found in Appendix A.3.

Note that \( \Psi_0^{RNN} \) both avoids dependence on the fine-scale \( \epsilon \) and is Markovian. The non-homogenized map \( \Psi_0^\epsilon \) is local in time while the homogenized map \( \Psi_0^{RNN} \) is non-local in time and depends on the strain history. Let \( u_0^{RNN} \) be the solution to the following system with constitutive model \( \Psi_0^{RNN} \):
\[
\rho \partial^2_t u_0^{RNN} - \partial_x \sigma_0 = f, \quad (x, t) \in \mathcal{D} \times \mathcal{T}
\]
\[
\sigma_0(t) = \Psi_0^{RNN}(\partial_x u_0^{RNN}(t), \partial^2_{xt} u_0^{RNN}(t), \{\partial_x u_0^{RNN}(\tau)\}_{\tau \in \mathcal{T}}, t), \quad (x, t) \in \mathcal{D} \times \mathcal{T}
\]
\[
u_0^{RNN}|_{t=0} = u^*, \quad \partial_t u_0^{RNN}|_{t=0} = v^*, \quad (x, t) \in \mathcal{D} \times \{0\}
\]
\[
u_0^{RNN} = 0, \quad (x, t) \in \partial\mathcal{D} \times \mathcal{T}.
\]

Ideally we would like an approximation result bounding \( \|u_\epsilon - u_0^{RNN}\|_{\mathcal{Z}_2} \), the difference between solution of the multiscale problem (2.1) and the Markovian RNN model (3.13), in the case \( \rho = 0 \). Using Corollary 3.9 shows that this would follow from a bound on \( \|u_0^{PC} - u_0^{RNN}\|_{\mathcal{Z}} \), where \( u_0^{PC} \) solves (3.6), in the case \( \rho = 0 \). We note, however, that although Theorem 3.10 gives us an approximation result between \( \Psi_0^{PC} \) and \( \Psi_0^{RNN} \), proving that \( u_0^{PC} \) and \( u_0^{RNN} \) are close requires developing new theory for the fully nonlinear PDE for \( u_0^{RNN} \); developing such a theory is beyond the scope of this work. Developing such a theory is difficult for two primary reasons: (i) the monotonicity property of \( \Psi_0^{RNN} \) with respect to strain rate is hard to establish globally, for a trained model; (ii) the functions \( \mathcal{F}^{RNN}, \mathcal{G}^{RNN} \) may not be differentiable. As a result, existence and uniqueness of \( u_0^{RNN} \) remains unproven; however, numerical experiments in Section 4 indicate that in practice, \( u_0^{RNN} \) does approximate \( u_\epsilon \) well.

**Remark 3.11.**

- Monotonicity of \( \Psi_0^{RNN} \) with respect to strain rate is a particular issue when \( \rho = 0 \) (no inertia) as in this case it is needed to define an (implicit) equation for \( \partial_t u_0 \) to determine the dynamics. It is for this reason that our experiments will all be conducted with \( \rho > 0 \), obviating the need for the determination of an (implicit) equation for \( \partial_t u_0 \). However this leads to the issue that the homogenized equation is only valid for a subset of initial conditions, in the inertial setting \( \rho > 0 \); see Remark 2.1.
In practice we will train $\Psi^\text{RNN}_0$ using data for the inputs $b, c$ which are obtained from a random function, or set of realizations of random functions. The choice of the probability measure from which this training data is drawn will affect the performance of the learned model when $\Psi^\text{RNN}_0$ is evaluated at $b = \partial_x u_0^\text{RNN}$ and $c = \partial^2_{xx} u_0^\text{RNN}$, for displacement $u$ generated by the model given by (3.13).

### 3.4 RNN Optimization

In the following section, we present numerical results using a trained RNN as a surrogate model: an efficient approximation of the complex microscale dynamics; in this section we discuss the problem of finding such an RNN. To learn the RNN operator approximation, we are given data

$$\{(\partial_x u_0)_n, (\partial^2_{xx} u_0)_n, (\sigma_0)_n\}_{n=1}^N$$

where the suffix $n$ denotes the $n^{\text{th}}$ strain, strain rate, and stress trajectories over the entire time interval $\mathcal{T}$. Each strain trajectory $(\partial_x u_0)_n$ is drawn i.i.d. from a measure $\mu$ on $\mathcal{C}(\mathcal{T}; \mathbb{R})$. There is no need to generate training data on the same time interval $\mathcal{T}$ as the macroscale model; we do so for simplicity.

The data for the homogenized constitutive model is given by

$$\sigma_0(t) = \Psi^\dagger_0(\partial_x u_0(t), \partial^2_{xx} u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t),$$

defined via solution of the cell-problem (2.4); but it may also be obtained as the solution to a forced boundary problem on the microscale, as stated in the following lemma.

**Lemma 3.12.** Let $u$ solve the equations

\begin{align}
\partial_y \sigma(y,t) &= 0, \quad (y,t) \in \mathcal{D} \times \mathcal{T} \\
\sigma(y,t) &= E(y)\partial_y u(y,t) + \nu(y)\partial_y u(y,t), \quad (y,t) \in \mathcal{D} \times \mathcal{T} \\
u(0,t) &= 0, \quad u(1,t) = b(t), \quad (y,t) \in \partial \mathcal{D} \times \mathcal{T} \\
u(y,0) &= 0, \quad y \in \mathcal{D}.
\end{align}

Then

$$\{\sigma(t)\}_{t=0}^T = \Psi^\dagger_0(b(t), \partial_t b(t), \{b(t)\}_{t=0}^T)$$

where $\Psi^\dagger_0$ is the map defined in (2.6).

The proof can be found in Appendix B.2 and justifies the application of data resulting from this problem to the homogenized model. In the following, we denote by $(\tilde{\sigma}_n)_n$ and $\tilde{\xi}_n$ the output of $\mathcal{F}^\text{RNN}$ and $\mathcal{G}^\text{RNN}$ on data point $n$:

$$(\tilde{\sigma}_n)_n(t) = \mathcal{F}^\text{RNN}\left((\partial_x u_0)_n(t), (\partial^2_{xx} u_0)_n(t), \tilde{\xi}_n(t)\right)$$

$$\tilde{\xi}_n(t) = \mathcal{G}^\text{RNN}\left((\partial_x u_0)_n(t), \tilde{\xi}_n(t)\right), \quad \tilde{\xi}_n(0) = 0.$$
Accessible Loss Function:

\[ L_1(\{\sigma_0\}_{n=1}^N, \{\hat{\sigma}_0\}_{n=1}^N) = \frac{1}{N} \sum_{n=1}^{N} \frac{\| (\sigma_0)_n - (\hat{\sigma}_0)_n \|_{L^2(T;\mathbb{R})}}{\| (\sigma_0)_n \|_{L^2(T;\mathbb{R})}}. \]  

(3.16)

Remark 3.13. To test robustness of our conclusions, we also employed relative and absolute \( L^2 \) squared loss functions. In doing so we did not observe significant differences in the predictive accuracy of the resulting models.

In the case of a material that is 2-piecewise-constant on the microscale, we can explicitly write down the analytic form of the solution, and thus can also know the values of the hidden variable \( \{\xi_n\}_{n=1}^N \) and its derivative \( \{\dot{\xi}_n\}_{n=1}^N \), for each data trajectory as expressed in equation (3.11). It is intuitive that training an RNN on an extended data set which includes the hidden variable should be easier than using the original data set (3.14). In order to deepen our understanding of the training process we will include training in a 2-piecewise-constant which uses this hidden data, motivating the following loss function. Since, in general, the hidden variable is inaccessible in the data, we refer to the resulting loss as the inaccessible relative loss function:

Inaccessible Loss Function:

\[ L_2(\{(\sigma_0)_n\}_{n=1}^N, \{(\hat{\sigma}_0)_n\}_{n=1}^N, \{\dot{\xi}_n\}_{n=1}^N, \{\dot{\hat{\xi}}_n\}_{n=1}^N) = \frac{1}{N} \sum_{n=1}^{N} \left( \frac{\| (\sigma_0)_n - (\hat{\sigma}_0)_n \|_{L^2(T;\mathbb{R})}}{\| (\sigma_0)_n \|_{L^2(T;\mathbb{R})}} + \frac{\| \dot{\xi}_n - \dot{\hat{\xi}}_n \|_{L^2(T;\mathbb{R})}}{\| \dot{\xi}_n \|_{L^2(T;\mathbb{R})}} \right). \]  

(3.17)

This inaccessible loss function helps identify the RNN whose existence is proved in previous sections.

4 Numerical Results

The numerical results make the following contributions, all of which guide the use of machine-learned constitutive models within complex nonlinear problems beyond the one-dimensional linear viscoelastic setting considered in this work.

I. Machine-Learned Constitutive Models. We can find RNNs that yield low-error simulations when used as a surrogate model in the macroscopic system (1.2),(1.3) to approximate the multiscale system (1.1), in the one dimensional KV setting with inertia. We also discuss how in some material parameter settings, inertial effects lead to higher error in the homogenized approximation.

II. Choice of Training Data. We describe our choice of data sampling distribution \( \mu \) and show that it exhibits desirable properties.

III. Effect of Non-Convex Optimization When the inaccessible loss function is used for training, the trained RNN exhibits desirable properties of (approximate) linearity in its arguments in the domain of interest, as is proved for the homogenized constitutive model (3.10) for piecewise-constant materials. When using the accessible loss function, the trained RNN may perform well as a surrogate model without exhibiting linearity in the equation for evolution of the hidden variables. This is attributable to the existence of local minimizers of the loss function and highlights the need for caution in training constitutive models.
IV. Model Choice. The correct choice of architecture for the RNN leads to discretization-robustness in time: a model learned with one choice of time discretization \( dt \) performs well when tested on another \( dt \); this is not true for poor model choices. Discretization-robustness can thus be used as a guide to model choice.

In Subsection 4.1 we demonstrate that in appropriate settings the solution \( u^\text{RNN}_0 \) obtained under the dynamics of a trained RNN approximates the true solution \( u_\epsilon \) well when used in the macroscopic setting; furthermore, this RNN is shown to exhibit linearity in its arguments in the domain of interest. We also discuss the error arising from inertial effects. In Subsection 4.2 we discuss the performance of an RNN learned using the accessible loss function. In Subsection 4.3 we discuss the discretization-robustness property of the RNN and the choice of data sampling distribution \( \mu \).

4.1 RNN as Surrogate Model

In this subsection we discuss two RNNs: RNN “A” trained using only the inaccessible loss function in equation (3.17) and RNN “B” trained with the standard loss function in equation (3.16), but initialized at parameters obtained via training with the inaccessible loss function. Descriptions of these RNNs, and others we will introduce in subsequent subsections, may be found in Table 1. For details on RNN training, see Appendix C.1.

In the first surrogate model experiment, we subjected the material to sinusoidal boundary forcing of \( b(t) = 0.1 \sin(2\pi t) \) starting from 0 initial displacement and velocity. As a ground-truth comparison, we used a traditional finite element solver with periodic domain of width 0.04, spatial resolution of \( h = 0.005 \), and time discretization \( dt = 0.1h^2 \); we refer to this solution as \( u_\epsilon \) and name it FEM. In contrast, the RNN-based macroscale computation employs spatial resolution of \( h_{\text{cell}} = 0.04 \), with a time discretization of \( dt = 0.4h_{\text{cell}}^2 \); for economy of notation, this solution is denoted \( u^\text{RNN}_0 \) and named RNN. We also compared the results to the displacement obtained using as macroscale constitutive model the analytic solution to the cell problem. To make comparisons we use the relative error given by

\[
e(u_0, u_\epsilon)(t) = \frac{\|u_\epsilon(t) - u^\text{RNN}_0(t)\|_{L^2(D;R)}}{\|u_\epsilon(t)\|_{L^2(D;R)} + 0.01}.
\]

The relative error plots for RNNs “A” and “B” are shown in Figures 1a and 1b. In a second experiment, we subjected the material to integrated Brownian motion forcing starting from null initial conditions. The FEM solver uses the same discretizations as in the sinusoidal forcing experiment, and the RNN spatial discretization was \( h_{\text{cell}} = 0.05 \) with time discretization of \( dt = 0.4h_{\text{cell}}^2 \). The results for RNNs “A” and “B” are shown in Figures 2a and 2b.

| RNN | Description |
|-----|-------------|
| A   | Trained on 2-piecewise-constant media only with inaccessible loss function (3.17) |
| B   | Trained on 2-piecewise-constant media; initialized at solution found with inaccessible loss function then trained with accessible loss function (3.16) |
| C   | Trained on 2-piecewise-constant media only with accessible loss function |
| D   | Trained on continuous media with accessible loss function |

Table 1: RNN Descriptions
Both sets of experiments show that the RNN-based macroscopic models accurately reproduce the microscale FEM simulation at far lower computational cost. The RNN-based results have some errors in comparison with the microscale simulation, but the errors are of the same order of magnitude as the errors arising when the exact homogenized constitutive model is used. The initial error between the analytic solution and the FEM solution is due to inertial effects discussed in Remark 2.1. The inertial errors become more significant with the ratio between $E$ and $\nu$ varies more across the interval. Relative error results for an RNN trained on a material with more inertial effects is shown in Figure 3.

Both RNNs also exhibit the desirable property of linearity in the inputs in an appropriate domain as discussed in Section 3.4; this is presented in Figure 4.

![Figure 1: Analytic cell and RNN relative error versus FEM solution using sinusoidal forcing; this supports Numerical Experiments, conclusion I.](image)

![Figure 2: Analytic cell and RNN relative error versus FEM solution using integrated Brownian motion forcing; this supports Numerical Experiments, conclusion I.](image)
Figure 3: Relative error of RNN trained on material parameters with higher inertial effects in response to sinusoidal and integrated Brownian motion forcing; this demonstrates Numerical Experiments, conclusion I.

Figure 4: RNN outputs versus the truth (dashed) for each of the three candidate RNNs. The columns correspond to RNNS “A,” “B,” and “C” respectively. The first row shows the strain-stress dependence for five fixed strain rate inputs. The second row shows the strain rate-stress dependence for five fixed strain inputs. The third row shows the $\xi$, stress relationships for hidden variable $\xi$ for five fixed strain inputs. The fourth row shows the strain, $\dot{\xi}$ relationship for five different fixed values of $\xi$. Finally, the fifth row shows the $\xi$, $\dot{\xi}$ relationship for five fixed strain inputs. This supports Numerical Experiments, conclusion III.
4.2 RNN Trained with Standard Loss

We also trained a third RNN, denoted RNN “C,” using only the accessible loss function. Details of training may be found in Appendix C.1. The performance of this RNN as a surrogate model in the macroscale simulation experiments may be seen in Figures 1c and 2c. While this RNN performed well as a surrogate model, indeed is comparable in errors to those of RNNs “A” and “B”, it does not exhibit a close linear match to the known analytic expression for $G$, as shown in Figure 4. In the figure, all three RNNs approximate the linear structure of $F$ well; the difficulty is in obtaining the correct linear dependence in the hidden variable rate, $\dot{\xi}$. Interestingly, by changing the material parameter $\nu_2$ from 0.2 to 2, training via the method of RNN “C” with only the accessible loss function yields an RNN that matches the true linear dependence in $G$ very well. However, in this parameter regime, inertial effects perturb the simulations on the macroscale to an unacceptable degree, meaning that the homogenization theory that we use as benchmark is not valid, and so we avoid this regime. The inability of RNN “C” to capture the exact linear dependence in $G$ is unsurprising; indeed, had we guaranteed convergence to the optimal function for any choice of material parameters, we would have entirely sidestepped the problem of high-dimensional optimization inherent to machine learning. In the case of continuous material properties, we do not have a known analytic solution to the microscale problem and thus do not have access to the hidden variable $\xi$ in the train and test data; in this case, we may only use the accessible loss function. We trained RNNs, denoted RNN type “D” on continuous media with different numbers of hidden variables $\xi$ and used the trained RNNs as surrogate models in the macroscale system subjected to boundary forcing. Training details may be found in Appendix C.2. The relative error of RNN “D” for the sinusoidal and Brownian motion forcing experiments described previously is shown in Figure 5. We note that similar error is found with all dimensions of the hidden variable, suggesting that 1 hidden variable suffices in this case; the fact that the error does not decrease suggests that the error we see is primarily from homogenization rather than piecewise-constant approximation.
4.3 Time Discretization and RNN Training

Discretization-robustness is a desirable feature of an RNN surrogate model. To test robustness to time discretization we work in the piecewise-constant media setup leading to RNNs “A”, “B” and “C”. We evaluated the test error when employing each of the three RNNs using values of timestep $dt$ different from those used in the training. Additionally, to demonstrate the value of postulating the correct model form, we trained three additional RNNs via the same methods as described in Subsection 4.2 but without strain rate dependence. Figure 6 shows that all three RNNs trained with strain rate as an input parameter were more robust to changes in time discretization than their non-strain-rate counterparts.

![Figure 6: Time discretization error for RNNs “A,” “B,” and “C.” This supports Numerical Experiments, conclusion IV.](image)

To generate the training strain, we sampled trajectories as follows: first, we randomly partitioned the time interval $T$ into 10 pieces; second, at each point between these time intervals, we generated a value of strain via a balanced random walk from the previous value scaled by the length of the time interval; third, we used a piecewise cubic hermite interpolating polynomial (pchip) function to interpolate between these values of strain. This choice of distribution has the desirable property that it generates data with a variety of strain/strain-rate pairings evenly dispersed throughout the domain of interest rather than introducing large correlations between the two. A scatterplot of the associated values is shown in Figure 7.

![Figure 7: Strain and strain rate distributions in the training data generated by $\mu$; this supports the choice of training data discussed in Numerical Experiments, conclusion II.](image)

5 Conclusions

In this paper, we develop theory underlying the learning of Markovian models for history-dependent constitutive laws. The theory presented applies to the one-dimensional KV case, but the underlying ideas extend to more complex systems. In [7], numerical experiments suggest that the method-
ology can be useful in the case of plasticity. Conclusions drawn from our numerical experiments, underpinned by the theory of this paper and enumerated at the start of Section 4, provide useful guidance for these more complex nonlinear models in higher spatial dimensions.

Several research directions are suggested by this work. Firstly, when inertial effects are significant, the homogenization theory used in this paper, and underlying the computational work in [7], is not valid; extending the methodology to this setting would be useful. Second, the development of theoretical guidance and methodology for choice of training measure \( \mu \) will be very important in this field. Third, as is the case with most machine learning applications, convergence of the RNN to the globally optimal solution is not guaranteed; considering learning techniques from reservoir computing could be useful to alleviate this issue as they lead to convex quadratic optimization problems.

Acknowledgements The work of KB, BL and AMS was sponsored by the Army Research Laboratory, United States, and was accomplished under Cooperative Agreement Number W911NF-12-2-0022. MT is funded by the Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC002111. The authors are grateful to Matt Levine for helpful discussions about training RNN models and to Pierre Suquet for helpful discussions about homogenization theory.

References

[1] Juan C Simo. Numerical analysis and simulation of plasticity. *Handbook of numerical analysis*, 6:183–499, 1998.
[2] Juan C Simo and Thomas JR Hughes. *Computational inelasticity*, volume 7. Springer Science & Business Media, 2006.
[3] Alain Bensoussan, Jacques-Louis Lions, and George Papanicolaou. *Asymptotic analysis for periodic structures*, volume 374. American Mathematical Soc., 2011.
[4] Grigoris Pavliotis and Andrew Stuart. *Multiscale methods: averaging and homogenization*. Springer Science & Business Media, 2008.
[5] Gilles A Francfort and Pierre M Suquet. Homogenization and mechanical dissipation in thermoviscoelasticity. *Archive for Rational Mechanics and Analysis*, 96(3):265–293, 1986.
[6] Burigede Liu, Nikola Kovachki, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, Andrew M Stuart, and Kaushik Bhattacharya. A learning-based multiscale method and its application to inelastic impact problems. *Journal of the Mechanics and Physics of Solids*, 158:104668, 2022.
[7] Burigede Liu, Margaret Trautner, Andrew Stuart, and Kaushik Bhattacharya. Learning macroscopic internal variables and history dependence from microscopic models. *In preparation*, 2022.
[8] Morton E Gurtin, Eliot Fried, and Lallit Anand. *The Mechanics and Thermodynamics of Continua*. Cambridge University Press, Cambridge, 2010.
[9] A.J.M. Spencer. *Continuum Mechanics*. Longman Group U.K., Essex, 1980.

[10] Oscar Gonzalez and Andrew M Stuart. *A first course in continuum mechanics*, volume 42. Cambridge University Press, 2008.

[11] James R Rice. Inelastic constitutive relations for solids: an internal-variable theory and its application to metal plasticity. *Journal of the Mechanics and Physics of Solids*, 19(6):433–455, 1971.

[12] Richard C. Christensen. *Theory of Viscoelasticity*. Academic Press, 1982.

[13] Roderic S. Lakes. *Viscoelastic Solids*. C.R.C Press, 1998.

[14] Jacob Fish. *Multiscale methods: bridging the scales in science and engineering*. Oxford University Press on Demand, 2010.

[15] Ellad B. Tadmor. *Modeling Materials: Continuum, Atomistic and Multiscale Techniques*. Cambridge University Press, 2012.

[16] Erik Van Der Giessen, Peter A Schultz, Nicolas Bertin, Vasily V Bulatov, Wei Cai, Gábor Csányi, Stephen M Foiles, Marc GD Geers, Carlos González, Markus Hütter, et al. Roadmap on multiscale materials modeling. *Modelling and Simulation in Materials Science and Engineering*, 28(4):043001, 2020.

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

[18] Enrique Sánchez-Palencia. Non-homogeneous media and vibration theory. *Lecture notes in physics*, 127, 1980.

[19] Doina Cioranescu and Patrizia Donato. *An Introduction to Homogenization*. Oxford University Press, Oxford, 1999.

[20] Grégoire Allaire. Homogenization and two-scale convergence. *SIAM Journal on Mathematical Analysis*, 23(6):1482–1518, 1992.

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

[22] Renald Brenner and Pierre Suquet. Overall response of viscoelastic composites and polycrystals: exact asymptotic relations and approximate estimates. *International Journal of Solids and Structures*, 50(10):1824–1838, 2013.

[23] Luc Tartar. Memory effects and homogenization. *Archive for Rational Mechanics and Analysis*, 111:121–133, 1990.

[24] T I Zohdi and P Wriggers. *Introduction to computational micromechanics*. Springer Verlag, 2005.

[25] Hervé Moulinec and Pierre Suquet. A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Computer methods in Applied Mechanics and Engineering*, 157(1-2):69–94, 1998.
[26] N. Mishra, J. Vondřejc, and J. Zeman. A comparative study on low-memory iterative solvers for fft-based homogenization of periodic media. *Journal of Computational Physics*, 321:151–168, 2016.

[27] H. Moulinec, P. Suquet, and G. Milton. Convergence of iterative methods based on neumann series for composite materials: Theory and practice. *International Journal of Numerical Methods in Engineering*, 114:1103–1130, 2018.

[28] M Mozaffar, R Bostanabad, W Chen, K Ehmann, Jian Cao, and MA Bessa. Deep learning predicts path-dependent plasticity. *Proceedings of the National Academy of Sciences*, 116(52):26414–26420, 2019.

[29] L. Wu, N.G. Kilingar, and L. Noels. A recurrent neural network-accelerated multi-scale model for elasto-plastic heterogeneous materials subjected to random cyclic and non-proportional loading paths. *Computer Methods in Applied Mechanics and Engineering*, 369:113234, 2020.

[30] Faisal As'ad, Philip Avery, and Charbel Farhat. A mechanics-informed artificial neural network approach in data-driven constitutive modeling. In *AIAA Scitech 2022 Forum*, page 0100, 2022.

[31] Tianping Chen and Hong Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995.

[32] Benjamin Peherstorfer and Karen Willcox. Data-driven operator inference for nonintrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering*, 306:196–215, 2016.

[33] Elizabeth Qian, Boris Kramer, Benjamin Peherstorfer, and Karen Willcox. Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems. *Physica D: Nonlinear Phenomena*, 406:132401, 2020.

[34] Kaushik Bhattacharya, Bamdad Hosseini, Nikola B Kovachki, and Andrew M Stuart. Model reduction and neural networks for parametric pdes. *SMAI Journal of Computational Mathematics*, 7:121–157., 2021.

[35] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*, 2020.

[36] Luning Sun, Han Gao, Shaowu Pan, and Jian-Xun Wang. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. *Computer Methods in Applied Mechanics and Engineering*, 361:112732, 2020.

[37] Ehsan Haghhighat, Maziar Raissi, Adrian Moure, Hector Gomez, and Ruben Juanes. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 379:113741, 2021.

[38] Zeliang Liu, CT Wu, and M391298807188754 Koishi. A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. *Computer Methods in Applied Mechanics and Engineering*, 345:1138–1168, 2019.
Appendices

A Proofs

A.1 Proof of Theorem 3.4

The proof of Lemma 3.5, which underlies the proof of Theorem 3.4, uses the following two propositions

Proposition A.1. Under Assumptions 3.1, for all solutions $u$ of equation (2.2) the following bounds hold for some constant $C_1$:

1. $\sup_{t \in T} \|u\|^2_{H_0^1, \nu} \leq \|u|_{t=0}\|^2_{H_0^1, \nu} + \left(\frac{\nu^+}{E^-}\right)^2 \frac{1}{\nu} C_1^2 \|f\|^2_Z$

2. $\sup_{t \in T} \|u\|^2_{H_0^1, E} \leq \frac{E^-}{\nu^-} \|u|_{t=0}\|^2_{H_0^1, \nu} + \left(\frac{\nu^+}{E^-}\right)^2 \frac{E^-}{(\nu^-)^2} C_1^2 \|f\|^2_Z$

3. $\|\partial_t u\|^2_{H_0^1, \nu} \leq \frac{C_1 \|f\|^2_Z}{\nu^-} + \frac{E^-}{\nu^-} \|u\|^2_{H_0^1, E}$, for all $t \in T$.

Proof. To show the first bound, let $\varphi = u$ in equation (2.2). We have

\[ q_\nu(\partial_t u, u) + q_E(u, u) = \langle f, u \rangle \]

so that

\[ \frac{1}{2} \frac{d}{dt} \|u\|^2_{H_0^1, \nu} + \|u\|^2_{H_0^1, E} \leq \|f\|_{H^{-1}} \|u\|_{H_0^1} \leq C_1 \|f\| \|u\|_{H_0^1} \]

for some constant $C_1$, by compact embedding. Then, using Lemma 1.1,

\[ \frac{1}{2} \frac{d}{dt} \|u\|^2_{H_0^1, \nu} + \frac{E^-}{\nu^-} \|u\|^2_{H_0^1, \nu} \leq \frac{C_1}{2\delta^2} \|f\|^2 + \frac{\delta^2}{2\nu^-} \|u\|^2_{H_0^1, \nu} \]

for any $\delta > 0$ by Young’s Inequality. Letting $\delta^2 = \frac{E^-}{\nu^-}$, we have

\[ \frac{d}{dt} \|u\|^2_{H_0^1, \nu} + \frac{E^-}{\nu^-} \|u\|^2_{H_0^1, \nu} \leq \frac{C_1^2 \nu^+}{E^- \nu^-} \|f\|^2_Z \]

Finally, Gronwall’s inequality yields

\[ \sup_{t \in T} \|u\|^2_{H_0^1, \nu} \leq \|u|_{t=0}\|^2_{H_0^1, \nu} + \left(\frac{\nu^+}{E^-}\right)^2 \frac{C_1^2}{\nu^-} \|f\|^2 \]

The second bound follows from Lemma 1.1. For the third bound, let $\varphi = \partial_t u$ in equation (2.2). Then

\[ q_\nu(\partial_t u, \partial_t u) + q_E(u, \partial_t u) = \langle f, \partial_t u \rangle \]
so that, again using Lemma 1.1, and using the Poincaré inequality,
\[ \| \partial_t u \|_{H^1_{0, \nu}}^2 \leq \frac{C_1}{\nu} \| f \| \| \partial_t u \|_{H^1_{0, \nu}} + \frac{E^+}{\nu} \| u \|_{H^1_{0, E}} \| \partial_t u \|_{H^1_{0, \nu}} \]
and
\[ \| \partial_t u \|_{H^1_{0, \nu}} \leq \frac{C_1}{\nu} \| f \| + \frac{E^+}{\nu} \| u \|_{H^1_{0, E}} \]
Then
\[ \frac{1}{2} \frac{d}{dt} \|\gamma\|^2_{H^{1,\nu}_0} + \frac{E^-}{\nu^+} \|\gamma\|^2_{H^{1,\nu}_0} \leq \frac{1}{2\delta^2} \|g\|^2 + \frac{\delta^2}{2\nu^-} \|\gamma\|^2_{H^{1,\nu}_0} \]
for any \( \delta > 0 \) by Young’s Inequality. Letting \( \delta = \frac{E^-}{\nu^+} \), we have
\[ \frac{d}{dt} \|\gamma\|^2_{H^{1,\nu}_0} + \frac{E^-}{\nu^+} \|\gamma\|^2_{H^{1,\nu}_0} \leq \frac{\nu^+}{E^-\nu^-} \|g\|^2. \]
Finally, since \( \gamma(0) = 0 \), Gronwall’s inequality yields
\[ \sup_{t \in T} \|\gamma\|^2_{H^{1,\nu}_0} \leq \left( \frac{\nu^+}{E^-} \right)^2 \frac{1}{\nu^-} \|g\|^2. \]
The second bound follows from Lemma 1.1. For the third bound, let \( \varphi = \partial_t \gamma \) in equation (A.1). Then
\[ q_v(\partial_t \gamma, \partial_t \gamma) + q_E(\gamma, \partial_t \gamma) = \langle g, \partial^2_t \gamma \rangle \]
so that, again using Lemma 1.1,
\[ \|\partial_t \gamma\|^2_{H^{1,\nu}_0} \leq \frac{1}{\nu^-} \|g\| \|\partial_t \gamma\|_{H^{1,\nu}_0} + \frac{E^+}{\nu^+} \|\gamma\|_{H^{1,E}_0} \|\partial_t \gamma\|_{H^{1,\nu}_0} \]
and
\[ \|\partial_t \gamma\|_{H^{1,\nu}_0} \leq \frac{1}{\nu^-} \|g\| + \frac{E^+}{\nu^+} \|\gamma\|_{H^{1,E}_0}. \]

To prove the Lipschitz property of the solution in Theorem 3.4, we will need the following lemma.

**Lemma 3.5 (Lipschitz Solution).** Let \( u_i \) be the solution to
\[ -\partial_x \left( E_i \partial_x u_i + \nu_i \partial^2_x u_i \right) = f, \quad (x, t) \in \partial D \times T, \] (3.1)
\[ u_i(x, t) = u^*, \quad (x, t) \in D \times \{0\}, \] (3.2)
\[ u_i(x, t) = 0, \quad (x, t) \in \partial D \times T, \] (3.3)
associated with material properties \( E_i, \nu_i \), for \( i \in \{1, 2\} \), and forcing \( f \), all satisfying the Assumptions 3.1. Then
\[ \|u_1 - u_2\| \leq C (\|\nu_1 - \nu_2\|_{\infty} + \|E_1 - E_2\|_{\infty}) \]
for some constant \( C \in \mathbb{R}^+ \) dependent on \( f, E^+_i, E^-_i, \nu^+_i, \nu^-_i \) and \( L \) and independent of \( \epsilon \).

**Proof.** Let \( \gamma \) and \( g \) be as defined before and after equation A.1. Then, by the result of Proposition A.2,
\[ \sup_{t \in T} \|\gamma\|_{H^{1}_0} \leq \frac{1}{\nu^-} \sup_{t \in T} \|\gamma\|^2_{H^{1,\nu}_0} \leq \left( \frac{\nu^+}{E^-\nu^-} \right)^2 \|g\|^2. \]
To bound the RHS:

\[
\|g\|_Z = \| (\Delta E) \partial_x u_2 + (\Delta \nu) \partial_{t,x}^2 u_2 \|_Z \\
\leq \| (\Delta E) \partial_x u_2 \|_Z + \| (\Delta \nu) \partial_{t,x}^2 u_2 \|_Z \\
\leq \| \Delta E \|_{\infty} \| \partial_x u_2 \|_Z + \| \Delta \nu \|_{\infty} \| \partial_{t,x}^2 u_2 \|_Z \\
\leq \sup_{t \in T} \| u_2 \|_{H^1} \| \Delta E \|_{\infty} + \sup_{t \in T} \| \partial_t u_2 \|_{H^1_{\mu}} \| \Delta \nu \|_{\infty} \\
\leq \frac{1}{(\nu^-)^2} \left( \sup_{t \in T} \| u_2 \|_{H^1_{\mu}} \| \Delta E \|_{\infty} + \sup_{t \in T} \| \partial_t u_2 \|_{H^1_{\mu}} \| \Delta \nu \|_{\infty} \right)
\]

To bound \( \| \partial_x u_2 \|_Z \) and \( \| \partial_{t,x}^2 u_2 \| \), note that any solution \( u_2 \) will satisfy equation (2.2) for \((u, E, \nu) \mapsto (u^2, E^2, \nu^2)\). The analysis of Proposition A.1 yields

\[
\sup_{t \in T} \| u_2 \|_{H^1_{\mu}} \leq \| u \|_{t=0} \| H^1_{\mu} + \left( \frac{\nu^+}{E^-} \right) \frac{C_1}{(\nu^-)^{1/2}} \| f \|_Z
\]

and

\[
\sup_{t \in T} \| \partial_t u_2 \|_{H^1_{\mu}} \leq \frac{C_1}{\nu^-} \| f \|_Z + \left( \frac{E^+}{\nu^-} \right)^{3/2} \| u \|_{t=0} \| H^1_{\mu} + \left( \frac{E^+}{\nu^-} \right)^{3/2} \frac{\nu^+}{E^-} C_1 \| f \|_Z.
\]

By the Poincaré inequality, \( \| \| \leq C_p \sup_{t \in T} \| \gamma \|_{H^1_{\mu}} \) for some constant \( C_p \) and setting

\[
C = C_p \left( \frac{\nu^+}{E^- (\nu^-)^{3/2}} \right) \max \left\{ \| u \|_{t=0} \| H^1_{\mu} + \left( \frac{\nu^+}{E^-} \right)^{1/2} \| f \|_Z, \right. \right. \\
\left. \left. \frac{C_1}{\nu^-} \| f \|_Z + \left( \frac{E^+}{\nu^-} \right)^{3/2} \| u \|_{t=0} \| H^1_{\mu} + \left( \frac{E^+}{\nu^-} \right)^{3/2} \frac{\nu^+}{E^-} C_1 \| f \|_Z \right\}
\]

gives the result.

Now we can prove the piecewise constant approximation theorem.

**Theorem 3.4 (Piecewise-Constant Approximation).** Let \( E \) and \( \nu \) be piecewise continuous functions, with a finite number of discontinuities, satisfying Assumptions 3.1; let \( u_\epsilon \) be the corresponding solution to (2.1). Then, for any \( \delta > 0 \), there exist piecewise constant \( E^{PC} \) and \( \nu^{PC} \) (in the sense of Definition 3.2) such that solution \( u_\epsilon^{PC} \) of equations (2.1) with these material properties satisfies

\[
\| u_\epsilon^{PC} - u_\epsilon \|_Z < \delta.
\]
Proof. Let $\mathcal{A}_E$ and $\mathcal{A}_\nu$ be the finite sets of discontinuities of $E$ and $\nu$, respectively, and let $\mathcal{A} = \mathcal{A}_E \cup \mathcal{A}_\nu$ with elements $a_1, a_2, \ldots, a_K$. Partition the interval $\mathcal{D}$ into intervals $D_1 = (a_0, a_1)$, $D_2 = (a_1, a_2)$, ..., $D_K = [a_{K-1}, a_K)$ such that $\bigcup_{k=1}^{K} D_k = \mathcal{D}$ and $\bigcap_{k=1}^{K} D_k = \emptyset$. Let $B_{k,\delta} = \{b_k^{\delta}, b_1^\delta, \ldots, b_N^{\delta}\}$ be a uniform partition of $D_k$ such that $b_i^\delta - b_{i-1}^\delta = \delta$. Furthermore, define $E^{\delta}_{\epsilon}$ and $\nu^{\delta}_{\epsilon}$ via

$$E^{\delta}_{\epsilon}(x) = \sum_{k=1}^{K} \sum_{n=1}^{N} I_{x \in (b_{n-1}^\delta, b_n^\delta]} E \left( \frac{1}{2} b_n^{\delta} + \frac{1}{2} b_{n-1}^{\delta} \right)$$

$$\nu^{\delta}_{\epsilon}(x) = \sum_{k=1}^{K} \sum_{n=1}^{N} I_{x \in (b_{n-1}^\delta, b_n^\delta]} \nu \left( \frac{1}{2} b_n^{\delta} + \frac{1}{2} b_{n-1}^{\delta} \right)$$

for $x \in \mathcal{D}$, noting that $E^{\delta}_{\epsilon}$ and $\nu^{\delta}_{\epsilon}$ are piecewise constant with $KN(\delta)$ pieces.

$E_{\epsilon}$ and $\nu_{\epsilon}$ are continuous on each interval $D_k$, so for each $\delta > 0$, there exists a mesh width $\delta$ such that with partitions $\{B_{k,\delta}\}_{k=1}^{K}$

$$\sup_{x \in (b_{n-1}^\delta, b_n^\delta]} \| E_{\epsilon} \left( \frac{1}{2} b_n^{\delta} + \frac{1}{2} b_{n-1}^{\delta} \right) - E_{\epsilon}(x) \| < \delta'$$

$$\sup_{x \in (b_{n-1}^\delta, b_n^\delta]} \| \nu_{\epsilon} \left( \frac{1}{2} b_n^{\delta} + \frac{1}{2} b_{n-1}^{\delta} \right) - \nu_{\epsilon}(x) \| < \delta'$$

for all $n \in \{1, \ldots, N(\delta)\}$. Thus, $\| E^{\delta}_{\epsilon} - E \|_\infty < \delta'$ and $\| \nu^{\delta}_{\epsilon} - \nu \|_\infty < \delta'$. Since $\delta'$ was arbitrary, we can pick $\delta' < \frac{\eta}{C_1}$ where $C_1$ is as in Lemma 3.5, and the theorem follows by use of the same lemma. \hfill \Box

A.2 Proof of Theorem 3.6

We will need the following lemma:

**Lemma A.3 (Existence of Exact Parametrization).** For a piecewise constant material with $L' + 1$ pieces and under Assumptions 3.1, $a_0$ in equation (2.5) can be written exactly as

$$\hat{a}_0(s) = E' + \nu' s - \sum_{\ell=1}^{L'} \frac{\beta_\ell}{s + \alpha_\ell}$$

where $E', \nu', \beta_\ell \in \mathbb{R}$ and $\alpha_\ell \in \mathbb{R}_+$ for all $\ell \in [L']$.

**Proof.** Let $E(y)$ and $\nu(y)$ have $L' + 1$ constant pieces of lengths $\{d_\ell\}_{\ell=1}^{L'+1}$, each associated to values $\{E_\ell\}_{\ell=1}^{L'+1}$ and $\{\nu_\ell\}_{\ell=1}^{L'+1}$ of $E$ and $\nu$. Then equation (2.5), rewritten here for convenience

$$\hat{a}_0(s) = \left( \int_0^1 \frac{dy}{sv(y) + E(y)} \right)^{-1}$$

29
positive coefficients. We now show that all roots of $Q(s)$ for some constants $E$, $C$, and $\nu$. Therefore, there exists a decomposition

$$\frac{P(s)}{Q(s)} = E' + \nu's - \frac{C(s)}{Q(s)}$$

(A.5)

where $P(s)$ is a polynomial of degree $L' + 1$ and $Q(s)$ a polynomial of degree $L'$. Therefore, there exists a decomposition

$$\frac{P(s)}{Q(s)} = E' + \nu's - \frac{C(s)}{Q(s)}$$

for some constants $E'$ and $\nu'$ and polynomial $C(s)$ of degree $L' - 1$.

Let $-\alpha_1, \ldots, -\alpha_{L'}$ be the roots of $Q(s)$. Then $C(s) = \sum_{\ell=1}^{L'} \beta_\ell$ for some constants $\beta_\ell \in \mathbb{C}$ by partial fraction decomposition. We wish to show that $\Re(\alpha_\ell) > 0$ for all roots $-\alpha_\ell$ of $Q(s)$ so that we can take the inverse Laplace transform. Furthermore, we wish to show that, in fact, $-\alpha_\ell \in \mathbb{R}$ for all roots $-\alpha_\ell$ so that $\beta_\ell \in \mathbb{R}$ as well. Since $E_j$ and $\nu_j$ are positive for all $j \in [L' + 1]$, it is clear that if a root $-\alpha_\ell$ is real, then it cannot be positive since $Q(s) = \sum_{\ell=1}^{L' + 1} d_\ell \prod_{j \neq \ell} (E_j + \nu_j s)$ has all positive coefficients. We now show that all roots of $Q(s)$ are real. Suppose $a + bi$ is a root of $Q(s)$. Then

$$Q(a + bi) = \sum_{\ell=1}^{L' + 1} d_\ell \prod_{j \neq \ell} (E_j + \nu_j (a + bi))$$

$$= \left[ \prod_{j=1}^{L' + 1} (E_j + \nu_j (a + bi)) \right] \cdot \sum_{\ell=1}^{L' + 1} \frac{d_\ell}{E_\ell + \nu_\ell (a + bi)}$$

$$= \left[ \prod_{j=1}^{L' + 1} (E_j + \nu_j (a + bi)) \right] \cdot \sum_{\ell=1}^{L' + 1} \left( \frac{d_\ell (E_\ell + \nu_\ell a)}{(E_\ell + \nu_\ell a)^2 + (\nu_\ell b)^2} - \frac{d_\ell (\nu_\ell b)}{(E_\ell + \nu_\ell a)^2 + (\nu_\ell b)^2} \right).$$

The term $\prod_{j=1}^{L' + 1} (E_j + \nu_j (a + bi))$ is a nonzero constant for $b \neq 0$ since $E_j, \nu_j \in \mathbb{R}_+$. Therefore, for $Q(a + bi) = 0$, both the real and imaginary components of the sum on the right must total 0. However, since $d_\ell, \nu_\ell$, and the denominator term $(E_\ell + \nu_\ell a)^2 + (\nu_\ell b)^2$ are all positive as well, $b$ must equal 0 to make $\Im[Q(a + bi)] = 0$. Therefore, all roots of $Q(s)$ are in $\mathbb{R}_-$. Returning to the decomposition, we now have

$$\tilde{a}_0(s) = E' + \nu's - \sum_{\ell=1}^{L'} \frac{\beta_\ell}{s + \alpha_\ell}$$

(A.6)

where $\beta_\ell \in \mathbb{R}$ and $\alpha_\ell \in \mathbb{R}_+$ for all $\ell \in [L']$.

Now we may prove the theorem.
Theorem 3.6 (Existence of Exact Parametrization). Let $\Psi_0^{\dagger}$ be the map from strain history to stress in the homogenized model, as defined by equation (2.7), in a piecewise constant material with $L' + 1$ pieces. Define $\Psi_0^{PC}: \mathbb{R}^2 \times C(\overline{T}; \mathbb{R}) \times \mathcal{T} \times \Theta \to \mathbb{R}$ by

$$
\Psi_0^{PC}(\partial_x u_0(t), \partial^2_{xt} u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t; \theta) = E_0 \partial_x u_0(t) + \nu_0 \partial^2_{xt} u_0(t) - \sum_{\ell=1}^{L_0} \xi_\ell(t),
$$

$$
\partial_t \xi_\ell(t) = \beta_\ell \partial_x u_0(t) - \alpha_\ell \xi_\ell(t), \xi_\ell(0) = 0, \quad \ell \in \{1, \ldots, L_0\},
$$

with parameter space

$$
\Theta = \left(E_0 \in \mathbb{R}_+, \nu_0 \in \mathbb{R}_+, L_0 \in \mathbb{Z}_+, \alpha_0 \in \mathbb{R}_{+}^{L_0}, \beta_0 \in \mathbb{R}_{+}^{L_0} \right).
$$

Then, under Assumptions 3.1, there exists $\theta^* \in \Theta$ with $(E_0, \nu_0, L_0, \alpha_0, \beta_0) = (E', \nu', L', \alpha, \beta)$ such that

$$
\Psi_0^{\dagger}(\partial_x u_0(t), \partial^2_{xx} u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t) = \Psi_0^{PC}(\partial_x u_0(t), \partial^2_{xt} u_0(t), \{\partial_x u_0(\tau)\}_{\tau \in \mathcal{T}}, t; \theta^*)
$$

for all $u_0 \in C^2(\overline{\mathcal{T}} \times \mathcal{T}; \mathbb{R})$ and $t \in \mathcal{T}$.

Proof. By Lemma A.3, we have that

$$
\hat{\sigma}_0 = \hat{a}_0(s) \partial_x \hat{u}_0 = \left(E' + \nu' s - \sum_{\ell=1}^{L'} \frac{\beta_\ell}{s + \alpha_\ell} \right) \partial_x \hat{u}_0,
$$

where $\beta_\ell \in \mathbb{R}$ and $\alpha_\ell \in \mathbb{R}_+$ for all $\ell \in [L']$. Taking an inverse Laplace transform, we get

$$
\sigma_0(t) = E' \partial_x u_0(t) + \nu' \partial_t \partial_x u_0(t) - \sum_{\ell=1}^{L'} \beta_\ell \int_0^t \partial_x u_0(\tau) \exp[-\alpha_\ell (t - \tau)] \, d\tau.
$$

(A.7)

The above may be reexpressed as equations (3.4) with a choice of parameters $\theta = (E', \nu', L', \alpha, \beta)$ and auxiliary variable $\xi$.

A.3 RNN Approximation Theorem 3.10 Proof

In this subsection we use $| \cdot |$, $\| \cdot \|$ to denote modulus and Euclidean norm, respectively, and $\langle \cdot, \cdot \rangle$ to denote Euclidean inner-product. This overlap with the notation from Subsection 1.4 should not lead to any confusion as it is confined to this subsection.

To prove Theorem 3.10 we first study the simple case where $\mathcal{F}^{PC}, \mathcal{G}^{PC}$ are uniformly approximated across all inputs; subsequently we will use this to establish Theorem 3.10 as stated.

Assumptions A.4. For any $\delta > 0$, there exist $\mathcal{F}^{RNN}$ and $\mathcal{G}^{RNN}$ such that

$$
\sup_{z \in \mathbb{R}^{1+L'}} | \mathcal{F}^{PC}(z) - \mathcal{F}^{RNN}(z) | \leq \delta
$$

$$
\sup_{z \in \mathbb{R}^{1+L'}} \| \mathcal{G}^{PC}(z) - \mathcal{G}^{RNN}(z) \| \leq \delta
$$

31
Proposition A.5. Under Assumptions A.4, if \( \{\alpha_t\} \) in equations (3.10) are bounded such that 0 < \( \rho < \alpha_t \) for some \( \rho \) for all \( t \in [L'] \), then for any \( \eta > 0 \), there exists a map \( \tilde{\Psi}_0^{RNN} \) defined as in equations (3.12) such that for \( \tilde{\Psi}_0^{PC} \) defined in equations (3.11), for any \( t \in \mathbb{R}_+^+ \) and functions \( b, c : \mathbb{R}_+^+ \rightarrow \mathbb{R} \),

\[
|\tilde{\Psi}_0^{PC}(b(t), c(t), \{b(\tau)\}_{\tau \in \mathcal{T}}, t; \theta^*) - \tilde{\Psi}_0^{RNN}(b(t), c(t), \{b(\tau)\}_{\tau \in \mathcal{T}}, t)| \leq \eta.
\]

Proof. Note that the main difficulty in this proof results from the fact that \( \mathcal{F}^{RNN} \) and \( \mathcal{F}^{PC} \) act on different hidden variables \( \xi \), which we will denote \( \xi^{RNN} \) and \( \xi^{PC} \), and whose first order time derivatives are given by \( \mathcal{G}^{RNN} \) and \( \mathcal{G}^{PC} \) respectively. We write

\[
\begin{align*}
|\tilde{\Psi}_0^{PC}(b(t), c(t), \{b(\tau)\}_{\tau \in \mathcal{T}}, t; \theta^*) - \tilde{\Psi}_0^{RNN}(b(t), c(t), \{b(\tau)\}_{\tau \in \mathcal{T}}, t)| &= |\mathcal{F}^{PC}(b(t), c(t), \xi^{PC}(t)) - \mathcal{F}^{RNN}(b(t), c(t), \xi^{RNN}(t))| \\
&\leq |\mathcal{F}^{PC}(b(t), c(t), \xi^{RNN}(t)) - \mathcal{F}^{RNN}(b(t), c(t), \xi^{RNN}(t))| \\
&+ |\mathcal{F}^{PC}(b(t), c(t), \xi^{PC}(t)) - \mathcal{F}^{PC}(b(t), c(t), \xi^{RNN}(t))| \\
&\leq \delta + |\mathcal{F}^{PC}(b(t), c(t), \xi^{PC}(t)) - \mathcal{F}^{PC}(b(t), c(t), \xi^{RNN}(t))|
\end{align*}
\]

by Assumptions A.4 since \( \mathcal{F}^{PC} \) and \( \mathcal{F}^{RNN} \) share arguments in the first term. To bound the second term,

\[
|\mathcal{F}^{PC}(b(t), c(t), \xi^{PC}(t)) - \mathcal{F}^{PC}(b(t), c(t), \xi^{RNN}(t))| = |\langle 1, \xi^{PC}(t) - \xi^{RNN}(t) \rangle| \leq \sqrt{L'}\|\xi^{PC}(t) - \xi^{RNN}(t)\|
\]

using the known form of \( \mathcal{F}^{PC} \) where \( \| \cdot \| \) is the Euclidean norm in \( \mathbb{R}^L' \).

Let \( e_\xi(t) = \xi^{PC}(t) - \xi^{RNN}(t) \). Note that \( \xi^{PC}(0) = \xi^{RNN}(0) = 0 \), so \( e_\xi(0) = 0 \). We wish to bound \( \|e_\xi(t)\| \). To do so, we first bound \( \|e_\xi(t)\| \), where \( \dot{e}_\xi(t) = \frac{d}{dt}e_\xi(t) \):

\[
\begin{align*}
\dot{e}_\xi(t) &= \dot{\xi}^{PC}(t) - \dot{\xi}^{RNN}(t) \\
&= \mathcal{G}^{PC}(\xi^{PC}(t), b(t)) - \mathcal{G}^{PC}(\xi^{RNN}(t), b(t)) - \mathcal{G}^{RNN}(\xi^{RNN}(t), b(t)) + \mathcal{G}^{PC}(\xi^{RNN}(t), b(t)) + q(t)
\end{align*}
\]

where we have defined \( q(t) = \mathcal{G}^{PC}(\xi^{RNN}(t), b(t)) - \mathcal{G}^{RNN}(\xi^{RNN}(t), b(t)) \) and \( \|q(t)\| \leq \delta \) by Assumptions A.4. Now note that \( \dot{e}_\xi(t) = -Ae_\xi(t) + q(t) \) by the form of \( \mathcal{G}^{PC} \), so we can bound

\[
\begin{align*}
\frac{1}{2} \frac{d}{dt}\|e_\xi(t)\|^2 &= \langle \dot{e}_\xi(t), e_\xi(t) \rangle = -\langle e_\xi(t), Ae_\xi(t) \rangle + \langle q(t), e_\xi(t) \rangle \\
&\leq -\alpha_{\min}\|e_\xi(t)\|^2 + \left\langle \frac{1}{\alpha_{\min}^{1/2}} q(t), e_\xi(t)^{1/2} \right\rangle \\
&\leq -\alpha_{\min}\|e_\xi(t)\|^2 + \frac{1}{2\alpha_{\min}}\|q(t)\|^2 + \frac{\alpha_{\min}}{2}\|e_\xi(t)\|^2 \\
\frac{d}{dt}\|e_\xi(t)\|^2 &\leq -\alpha_{\min}\|e_\xi(t)\|^2 + \frac{\delta^2}{\alpha_{\min}}
\end{align*}
\]

by Young’s inequality. Then by Gronwall’s inequality

\[
\|e_\xi(t)\|^2 \leq \frac{\delta^2}{\alpha_{\min}} (1 - e^{-\alpha_{\min} t})
\]  

(A.8)
For any \( t, b, c \) and \( G \) assumption in the proof of Proposition A.5 because the function arguments of the theorem by showing that, for inputs bounded via Assumptions A.6. We will show that sup \( \| \xi^{PC}(t) - \xi^{RNN}(t) \| \leq \frac{\sqrt{T}R}{\rho} \).

Since by Assumptions A.4, \( \delta \) is arbitrarily small, the theorem result is shown. \( \square \)

Although we did not need to restrict the inputs \( t, b, c \) and \( G \) in Proposition A.5 to compact sets in order to prove it, we will argue that the statement holds under weaker assumptions if the inputs are also bounded. The following weaker assumptions follow from the Universal Approximation Theorem for RNNs[39].

**Assumptions A.6.** If \( D_1 \in \mathbb{R}^{2+L'} \) and \( D_2 \in \mathbb{R}^{1+L'} \) are compact sets, then for any \( \delta > 0 \), there exist \( \mathcal{F}^{RNN} \) and \( \mathcal{G}^{RNN} \) such that

\[
\sup_{z \in D_1} |\mathcal{F}^{PC}(z) - \mathcal{F}^{RNN}(z)| \leq \delta \\
\sup_{z \in D_2} \|\mathcal{G}^{PC}(z) - \mathcal{G}^{RNN}(z)\| \leq \delta.
\]

**Theorem 3.10** (RNN Approximation). Consider \( \Psi_0^{PC} \) defined as by equations (3.10), (3.11). Assume that there exist \( \rho > 0 \) and \( 0 \leq B < \infty \) such that \( \rho < \min_{\ell} |\alpha_{\ell}| \) and \( \max_{\ell} |\beta_{\ell}| \leq B \). Then, under Assumptions 3.1, for every \( \eta > 0 \) there exists \( \Psi_0^{RNN} \) of the form (3.12) such that

\[
\sup_{t \in \mathcal{T}, b, c \in \mathbb{Z}_R} |\Psi_0^{PC}(b(t), c(t), \{b(\tau)\}_{\tau \in \mathcal{T}}, t; \theta^*) - \Psi_0^{RNN}(b(t), c(t), \{b(\tau)\}_{\tau \in \mathcal{T}}, t)| < \eta.
\]

**Proof.** Notice first that Assumptions A.6 are a weaker version of Assumptions A.4. We will prove the theorem by showing that, for inputs bounded via \( t \in \mathcal{T} \) and \( b, c \in \mathbb{Z}_R \), we never need the stronger assumption in the proof of Proposition A.5 because the function arguments of \( \mathcal{F}^{PC}, \mathcal{F}^{RNN}, \mathcal{G}^{PC}, \) and \( \mathcal{G}^{RNN} \) never leave a compact set. First we show that \( \sup_{t \in \mathcal{T}} \|\xi^{PC}(t)\| \leq R_3 \) for some \( R_3 > 0 \). For any \( \ell \in \{1, \ldots, L'\} \), we have

\[
\dot{\xi}^{PC}(t) = \beta_b(t) - \alpha_{\ell}\xi^{PC}_{\ell}(t) \\
|\xi^{PC}(t)| \leq e^{-\alpha_{\ell}t} \beta_\ell \left( \sup_{t \in \mathcal{T}} |b(t)| \right) \int_0^t e^{\alpha_{\ell}t'} dt' \\
\leq e^{-\alpha_{\ell}t} \beta_\ell \left( \sup_{t \in \mathcal{T}} |b(t)| \right) \frac{1}{\alpha_{\ell}} e^{\alpha_{\ell}t} \\
\sup_{t \in \mathcal{T}} |\xi^{PC}(t)| \leq \frac{B}{\rho} R,
\]

so that \( \sup_{t \in \mathcal{T}} \|\xi^{PC}(t)\| \leq \frac{\sqrt{TB}R}{\rho} \). Let \( R_3 = \frac{\sqrt{TB}R}{\rho} \). Define \( R_4 = \max\{R_3 + \frac{\delta}{\rho}, R\} \) for \( \delta \) in Assumptions A.6. We will show that \( \sup_{t \in \mathcal{T}} \|\xi^{RNN}(t)\| \leq R_4 \) for \( \xi^{RNN} \) defined by \( \xi^{RNN}(t) \) in equations (3.12). Then the proof of Proposition A.5 will apply for bounded \( t, b, c \) with the weaker assumptions since all inputs to \( \mathcal{F}^{PC}, \mathcal{F}^{RNN}, \mathcal{G}^{PC}, \) and \( \mathcal{G}^{RNN} \): \( b(t), c(t), \xi^{PC}(t), \) and \( \xi^{RNN}(t) \), will remain in a compact set for \( t \in \mathcal{T} \).
Suppose for the sake of contradiction that there exists a time $t' \in \mathcal{T}$ at which $\|\xi^{RNN}(t')\| > R_4$. Then there exists a time $T' < t' < T$ and $\epsilon > 0$ such that for $t \in [0, T']$, $\|\xi^{RNN}(t)\| \leq R_4$, for $t \in (T', T' + \epsilon)$, $\|\xi^{RNN}(t)\| > R_4$, and $\|\xi^{RNN}(T')\| = R_4$ by continuity. In other words, $T'$ is the time at which $\xi^{RNN}$ first crosses the $R_4$ radius. Then
\[
\|e_\xi(T')\| := \|\xi^{RNN}(T') - \xi^{PC}(T')\| \geq R_4 - R_3 \geq \frac{\delta}{\rho}
\]
by triangle inequality. Since $\|\xi^{RNN}(t)\| \leq R_4$ for $t \in [0, T']$, the bound on $\|e_\xi(t)\|$ in equation (A.8) in the proof of Proposition A.5 applies on the interval $t \in [0, T']$ under the weaker assumptions A.6, and $\|e_\xi(T')\| < \frac{\delta}{\rho}$. This is a contradiction. Therefore, $\sup_{t \in \mathcal{T}} \|\xi^{RNN}(t)\| \leq R_4$, and the proof of Proposition A.5 holds with the weaker Assumptions A.6 for bounded inputs $t \in \mathcal{T}$ and $b, c \in \mathbb{Z}_R$. 

The bounds on $\alpha$ and $\beta$ required in Theorem 3.10 are justified because for known material properties $E$ and $\nu$, $\alpha$ and $\beta$ are determined and finite-dimensional, so they have maximum and minimum values.

B Special Case Solutions

B.1 Laplace Transform Limit

Here we derive the form of $\Psi_0^\dagger$ in equation (2.7) via a power series expansion of the Laplace transform at $s = \infty$. Starting from the definition in equation (2.6):
\[
\Psi_0^\dagger = \mathcal{L}^{-1}\left(\left(\int_0^1 \frac{dy}{sv(y) + E(y)}\right)^{-1} \partial_x \hat{u}_0\right).
\]

For $s \gg 1$, we have that
\[
\left(\int_0^1 \frac{dy}{sv(y) + E(y)}\right)^{-1} \approx \left(\int_0^1 \frac{dy}{sv(y)}\right)^{-1} = s \left(\frac{dy}{\nu(y)}\right)^{-1}.
\]

Setting $\nu' = \left(\frac{dy}{\nu(y)}\right)^{-1}$, we now subtract out the linear dependence on $s$ and let $z = \frac{1}{s}$. We define
\[
F(z) = \hat{a}_0(s) - \nu' s \bigg|_{s = z^{-1}}
\]
to obtain

\[ F(z) = \hat{a}_0(z^{-1}) - \nu' z^{-1} \]
\[ = \left( \int_0^1 \frac{z \, dy}{\nu(y) + zE(y)} \right)^{-1} - \left( \int_0^1 \frac{z \, dy}{\nu(y)} \right)^{-1} \]
\[ = \int_0^1 \left( \frac{z}{\nu(y) + zE(y)} \right) \, dy 
\quad - \left( \int_0^1 \frac{z \, dy}{\nu(y)} \right)^{-1} \]
\[ = z^2 \int_0^1 \left( \frac{1}{\nu(y) + zE(y)} \right) \, dy 
\quad - \left( \int_0^1 \frac{dy}{\nu(y)} \right)^{-1} \]
\[ = \int_0^1 \left( \frac{z \, E(y)}{\nu(y) + zE(y)} \right) \, dy 
\quad \left( \int_0^1 \frac{dy}{\nu(y)} \right)^{-1} \]
\[ = \int_0^1 E(y) \nu(y)^2 \frac{dy}{\nu(y)} \]
\[ = \int_0^1 E(y) \nu^2(y) \frac{dy}{\nu(y)} \]
\[ = \int_0^1 \frac{E(y)}{\nu(y)} \frac{dy}{\nu(y)} \]

Since \( \inf_{y \in (0,1)} \nu(y) > 0 \),

\[ \lim_{z \to 0} F(z) = \int_0^1 E(y) \nu(y) \frac{dy}{\nu(y)} =: E' \]

From this same computation, we see that for \( \hat{a}_0(s) = s\nu' + E' + \kappa(s) \), the contribution \( \kappa(s) \) consists of lower order terms in \( s \) and is such that \( \lim_{s \to \infty} \kappa(s) = 0 \). Using the fact that the inverse Laplace transform of a product (if it exists) is a convolution, we justify the form of the integral term in equation (2.7).

B.2 Forced Boundary Problem

Lemma 3.12. Let \( u \) solve the equations

\[ \partial_y \sigma(y, t) = 0, \quad (y, t) \in \mathcal{D} \times \mathcal{T} \]  
\[ \sigma(y, t) = E(y) \partial_y u(y, t) + \nu(y) \partial_y u(y, t), \quad (y, t) \in \mathcal{D} \times \mathcal{T} \]  
\[ u(0, t) = 0, \quad u(1, t) = b(t), \quad (y, t) \in \partial \mathcal{D} \times \mathcal{T} \]  
\[ u(y, 0) = 0, \quad y \in \mathcal{D}. \]

Then

\[ \{ \sigma(t) \}_{t=0}^T = \Psi_0 \left\{ b(t), \partial_t b(t), \{ b(t) \}_{t=0}^T \right\} \]

where \( \Psi_0 \) is the map defined in (2.6).

Proof. Taking the Laplace transform of (3.15) yields

\[ \hat{\sigma}(s) = (E(y) + \nu(y)s) \partial_y \hat{u}(y, s). \]

Spatially averaging and noting that \( b(t) = \langle \partial_y u(y, t) \rangle \), we have

\[ \hat{b}(s) = \int_0^1 \frac{dy}{(E + s\nu)(y)} \hat{\sigma}(s). \]  

(B.1)

35
Then \( \hat{\sigma}(s) = \left( \int_{0}^{1} \frac{dy}{(E + s\nu)(y)} \right)^{-1} \tilde{b}(s) \), which is equivalent to \( \hat{\sigma}(s) = \tilde{a}_0(s)\tilde{b}(s) \) using equation (2.5). The definition of \( \Psi^\dagger \) in equation (2.6) completes the proof. 

Lemma 3.12 justifies the use of data arising from the system (3.15) to train the map \( \Psi_0 \).

C Macroscale Numeric Comparisons

C.1 RNN Training and Testing: Piecewise-Constant Case

We trained three RNNs using the same dataset for the setting of a 2-piecewise constant material with material parameters \( E_1 = 1, E_2 = 3, \nu_1 = 0.1 \), and \( \nu_2 = 0.2 \). The data was generated using a forward Euler method with time discretization \( dt = 0.001 \) up to time \( T = 4 \) on the known analytic solution for the 2-piecewise-constant cell problem. Denote the data by \( \{(\partial_x u_0)_n, (\sigma_0)_n\}_{n=1}^N \) as discussed in Section 3.4. We repeat the two loss functions here:

**Accessible Loss Function:**

\[
L_1(\{(\sigma_0)_n\}_{n=1}^N, \{\tilde{\sigma}_0\}_n) = \frac{1}{N} \sum_{n=1}^{N} \| (\sigma_0)_n - (\tilde{\sigma}_0)_n \| / \| (\sigma_0)_n \|
\]

**Inaccessible Loss Function:**

\[
L_2(\{(\sigma_0)_n\}_{n=1}^N, \{\tilde{\sigma}_0\}_n, \{\xi\}_n, \{\tilde{\xi}_n\}_n) = \frac{1}{N} \sum_{n=1}^{N} \left( \| (\sigma_0)_n - (\tilde{\sigma}_0)_n \| / \| (\sigma_0)_n \| + \| (\xi)_n - (\tilde{\xi})_n \| / \| (\xi)_n \|_{L^2(D,\mathbb{R})} \right)
\]

For each of the following RNNs, the architecture for \( F_{\text{RNN}} \) and \( G_{\text{RNN}} \) consists of three internal layers of SeLU units of 100 nodes separated by linear layers, all followed by a final linear layer. The SeLU function is applied element-wise as

\[
\text{SELU}(x) = s(\max(0, x) + \min(0, \alpha(\exp(x) - 1)))
\]

where \( \alpha = 1.67326 \) and \( s = 1.05070 \). We trained three different RNNs on the same dataset in the following manner:

- **RNN “A”:** Using only the inaccessible loss function \( L_2 \), we trained on \( N = 400 \) data points with subsampled time discretization of \( dt = 0.004 \) up to \( T = 4 \) for 1500 epochs with a batch size of 50.

- **RNN “B”:** First we used the inaccessible loss function \( L_2 \) to train on \( N = 200 \) data points with subsampled time discretization of \( dt = 0.004 \) up to \( T = 2 \) for 1500 epochs with a batch size of 40. Then we initialized a new RNN at the parameters of this RNN and trained with the accessible loss function \( L_1 \) for 1000 epochs on 200 data with batch size of 40.

\[5\]https://pytorch.org/docs/stable/generated/torch.nn.SELU.html
- **RNN “C”**: Using only the accessible loss function $L_1$, we trained on $N = 500$ data points with subsampled time discretization of $dt = 0.004$ up to $T = 4$ for 3000 epochs with a batch size of 50.

The train and test errors are shown for the three RNNs in Figure 8.

![Figure 8: Train and test error for the three RNNs.](image)

(a) RNN “A” trained using inaccessible loss function  
(b) RNN “B” initialized at inaccessible loss solution  
(c) RNN “C” trained using only standard loss function

C.2 RNN Training and Testing: Continuous Case

We trained several RNNs on data $\{(\partial_x u_0)_n, (\sigma_0)_n\}_{n=1}^N$ for continuous material parameters given by $E(y) = 2 + \tanh\left(\frac{y-0.5}{0.2}\right)$ and $\nu(y) = 0.5 + 0.1 \tanh\left(\frac{y-0.5}{0.2}\right)$. Each of the four RNNs had 1, 2, 5, and 10 hidden variables ($L_0$, or the dimension of $\xi$) respectively. The data was generated by solving the cell problem using a finite difference method with 200 spatial nodes and $dt = h^2$ where $h$ is the spatial discretization. The RNN was trained for 3000 epochs on 500 data. The macroscale simulations were performed with a spacial discretization of $h_{cell} = 0.05$ and a time discretization of $dt = 0.4h_{cell}^2$. They were compared to an FEM solution computed with a spacial discretization of $h = 0.004$ with a material period of 0.04 and time discretization of $dt = 0.1h^2$.

D One-Dimensional Standard Linear Solid

In this section we address the model of the one-dimensional Maxwell version of the SLS, whose constitutive law depends only on the strain and strain history. The analysis for the SLS model demonstrates that the ideas presented for the KV model extend beyond that particular setting. In Section D.1, we present the governing equations, and in Section D.2 we homogenize the system.

D.1 Governing Equations

In the setting without inertia, the displacement $u_\epsilon$, strain $e_\epsilon$, and inelastic strain $e_\epsilon^p$ are related by

\[
\begin{align*}
-\partial_x \sigma_\epsilon &= f, \quad \text{(D.1a)} \\
\epsilon_\epsilon &= \partial_x u, \quad \text{(D.1b)} \\
\sigma_\epsilon &= E_{1,\epsilon} e_\epsilon + E_{2,\epsilon} (e_\epsilon - e_\epsilon^p), \quad \text{(D.1c)} \\
\partial_t e_\epsilon^p &= \frac{E_{2,\epsilon}}{\nu_\epsilon} (e_\epsilon - e_\epsilon^p); \quad \text{(D.1d)}
\end{align*}
\]
where \( f : \mathcal{D} \times \mathcal{T} \to \mathbb{R} \) is a known forcing, and we impose initial condition \( u(x,0) = 0 \) and boundary conditions \( u(x,t) = 0 \) for \( x \in \partial \mathcal{D} \). We seek a solution \( u : \mathcal{D} \times \mathcal{T} \to \mathbb{R} \). Once more we have small scale dependence in the material properties through \( \epsilon \): we have \( E_{i,\epsilon}(x) = E_i \left( \frac{x}{\epsilon} \right) \) for \( i = 1, 2 \) and \( \nu_\epsilon(x) = \nu \left( \frac{x}{\epsilon} \right) \) for \( 0 < \epsilon \ll 1 \).

## D.2 Homogenization

First, we take the Laplace transform of equation (D.1) and combine the transformed expressions of equations (D.1c) and (D.1d) to arrive at

\[
\hat{\sigma}_\epsilon = E_{1,\epsilon} \hat{\epsilon} + E_{2,\epsilon} s \left( s + \frac{E_{2,\epsilon}}{\nu_\epsilon} \right)^{-1} \hat{\epsilon}_\epsilon \quad (D.2)
\]

Letting \( \hat{a}(s) = E_{1,\epsilon} + E_{2,\epsilon} s \left( s + \frac{E_{2,\epsilon}}{\nu_\epsilon} \right)^{-1} \), the homogenization theory of Section 3.2 applies, and we can use the harmonic averaging expression in equation (2.5) to write

\[
\hat{a}_0(s) = \langle (a(s))^{-1} \rangle^{-1} = \left( \int_0^1 \frac{s + \frac{E_2}{\nu}}{(E_1 + E_2)s + \frac{E_1 E_2}{\nu}} dy \right)^{-1} \quad (D.3)
\]

where the homogenized solution \( u_0 \) solves system 1.2 with \( \Psi_0^\dagger \) is defined as

\[
\Psi_0^\dagger = \mathcal{L}^{-1} [\hat{a}_0(s) \partial_x \hat{u}_0], \quad (D.4)
\]

analogous to the KV case. However, in the case of piecewise-constant \( E_1, E_2 \) and \( \nu \) the inverse Laplace transform yields a different form in the SLS case:

\[
\Psi_0^{PC}(\partial_x u_0, t; \theta) = E' \partial_x u_0(t) - \sum_{\ell=1}^L \xi_\ell(t) \quad (D.5a)
\]

\[
\partial_t \xi_\ell(t) = \beta_\ell \partial_x u_0(t) - \alpha_\ell \xi_\ell(t), \quad \ell \in \{1, \ldots, L\} \quad (D.5b)
\]

for a material with \( L \) pieces. Note that this model does not have dependence on the strain rate, but it has one more internal variable than the piecewise-constant case for the KV model. The value of \( E' \) follows from taking the limit \( s \to \infty \) and is given by

\[
E' = \left( \int_0^1 \frac{1}{(E_1 + E_2) dy} \right)^{-1}.
\]

Full derivation may be found in Appendix D.3.

## D.3 SLS Derivation

Here we show that the SLS model has one more internal variable in the piecewise-constant case than the KV model does. This is the analog of Theorem 3.6 for the SLS model. Starting from
equation (D.3) for \( \hat{a}_0(s) \):

\[
\hat{a}_0(s) = (\hat{a}(s)^{-1})^{-1} = \left( \int_0^1 \frac{s + \frac{E_2(y)}{\nu}}{(E_1(y) + E_2(y))s + \frac{E_1(y)E_2(y)}{\nu(y)}} \, dy \right)^{-1} = \left( \sum_{i=1}^L \frac{(s + \frac{E_2,i}{\nu,i})d_i}{(E_1,i + E_2,i)s + \frac{E_1,iE_2,i}{\nu,i}} \right)^{-1}
\]

for \( L \)-piecewise-constant \( E_1, E_2, \) and \( \nu \) with pieces of length \( d_i \). Let \( c_i = \frac{E_2,i}{\nu,i}, k_i = E_1,i + E_2,i \) and \( p_i = \frac{E_1,iE_2,i}{\nu,i} \). Continuing,

\[
= \left( \sum_{i=1}^L \frac{(s + c_i)d_i}{k_is + p_i} \right)^{-1} = \frac{\prod_{i=1}^L (k_is + p_i)}{\sum_{i=1}^L d_i(s + c_i)\prod_{j \neq i}(k_js + p_j)} := \frac{P(s)}{Q(s)}
\]

Note that both \( P(s) \) and \( Q(s) \) have degree \( L \). There is a unique constant \( E' \) such that

\[
\frac{P(s)}{Q(s)} = E' + \frac{C(s)}{Q(s)}
\]

where \( C(s) \) has degree \( L \). Then \( \frac{C(s)}{Q(s)} \) decomposes uniquely as \( \sum_{i=1}^L \frac{\kappa_i}{s + \omega_i} \). Note that this is one more pole than the decomposition for the KV model in Theorem 3.6 has. We will now show that roots of \( Q \) are real and negative, which will lead to the expression in equation (D.5). First notice that if the roots of \( Q(s) \) are real, then they must be negative since \( k_i, c_i, d_i, \) and \( p_i \) are strictly positive for all \( i \in [L] \). Suppose for the sake of contradiction that \( Q(s) \) has a root with a nonzero imaginary component: \( s = a + bi \) where \( b \neq 0 \). Then

\[
Q(a + bi) = \sum_{\ell=1}^L d_\ell(a + bi + c_\ell) \prod_{j \neq \ell}(k_j(a + bi) + p_j)
\]

\[
= \left( \prod_{j}(k_j(a + bi) + p_j) \right) \sum_{\ell=1}^L \frac{d_\ell(a + bi + c_\ell)}{k_\ell(a + bi) + p_\ell}
\]

\[
= \left( \prod_{j}(k_j(a + bi) + p_j) \right) \sum_{\ell=1}^L \left( \frac{d_\ell a + d_\ell c_\ell + d_\ell bi}{k_\ell a + p_\ell + k_\ell b_i} \right) \left( \frac{k_\ell a + p_\ell - k_\ell b_i}{k_\ell a + p_\ell - k_\ell b_i} \right)
\]

\[
= \left( \prod_{j}(k_j(a + bi) + p_j) \right) \times \sum_{\ell=1}^L \frac{d_\ell}{(k_\ell a + p_\ell)^2 + (k_\ell b_i)^2} \left[ ((a + c_\ell)(k_\ell a + p_\ell) + k_\ell b_i^2) + (\mp k_\ell b c_\ell + b p_\ell) i \right]
\]
If $a+bi$ is a root of $Q$, then we need $b \sum_{\ell=1}^{L} \frac{d_{\ell}}{(k_{\ell} a + p_{\ell})^{2} + (k_{\ell} b)^{2}} (-k_{\ell} c_{\ell} + p_{\ell}) = 0$. Notice that $-k_{\ell} c_{\ell} + p_{\ell} = -\frac{E_{2,\ell}}{v_{\ell}}$, which is strictly negative, so for $b \neq 0$, $\Im(Q(a+bi)) < 0$, which is a contradiction. Therefore, $b = 0$, and all the roots of $Q$ are real and negative. Inverting the Laplace transform, we arrive at equation (D.5).