Determining kernels in linear viscoelasticity

Barbara Kaltenbacher [a], Ustim Khristenko [b], Vanja Nikolić [c,*],
Mabel Lizzy Rajendran [b], Barbara Wohlmuth [b]

[a] *Department of Mathematics, Alpen-Adria-Universität Klagenfurt, Universitätsstraße 65–67, A-9020 Klagenfurt, Austria*
[b] *Department of Mathematics, Technical University of Munich, Boltzmannstraße 3, 85748 Garching, Germany*
[c] *Department of Mathematics, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands*

A R T I C L E   I N F O

A B S T R A C T

In this work, we investigate the inverse problem of determining the kernel functions that best describe the mechanical behavior of a complex medium modeled by a general nonlocal viscoelastic wave equation. To this end, we minimize a tracking-type data misfit function under this PDE constraint. We perform the well-posedness analysis of the state and adjoint problems and, using these results, rigorously derive the first-order sensitivities. Numerical experiments in a three-dimensional setting illustrate the method.

1. Introduction

When elastic waves propagate through complex media such as biological tissues, their amplitude is attenuated according to a frequency-dependent law. In recent years, it has become evident that these attenuation laws are more complicated than initially thought and that nonlocal wave equations are needed to model such behavior. Indeed, elastic wave equations with weakly singular kernels arise in various important medical and industrial applications of wave propagation in complex media. For example, frequency power laws with powers between zero and two are often encountered in shear-wave elastography; see [1–3].

In this work, we tackle the inverse problem of determining the kernel functions that best describe the mechanical behavior of a complex medium from over-specified data. To this end, we first derive and analyze a general nonlocal elastic wave model containing three kernels, cf. (2.2) below. By taking the Laplace transform, it can be seen that, in general, not all three kernels can be uniquely determined independently of each other. Therefore, we consider the inverse problem of determining two of these kernels (cf. (2.2) below) from over-specified data and study the PDE-constrained optimization problem resulting from the (optionally regularized) minimization of the data misfit. Using an adjoint-based calculation, we then rigorously derive the first-order sensitivities and develop a numerical kernel recovery method.

We note that some work in the direction of kernel identification can be found in the literature; in particular, we refer to [4–8] and the references contained therein. For results with Abel integral kernels, that is, models involving time-fractional derivatives, we refer to, e.g., [9–13].

The remaining of our exposition is organized as follows. In Section 2, we derive a general nonlocal elastodynamic wave equation and discuss the setup of the inverse problem. Section 3 is concerned with the well-posedness analysis of the state

---

* Corresponding author.
E-mail addresses: barbara.kaltenbacher@aau.at (B. Kaltenbacher), khristen@ma.tum.de (U. Khristenko), vanja.nikolic@ru.nl (V. Nikolić), rajendrm@ma.tum.de (M.L. Rajendran), wohlmuth@ma.tum.de (B. Wohlmuth).

problem. In Section 4, we derive and analyze the corresponding adjoint problem. Section 5 is focused on the estimation of the kernels from additional observations. Finally, in Section 6, we present numerical experiments that illustrate our theoretical results.

2. Modeling and problem setup

In this section, we provide the setup for the problem and derive the viscoelastic equation from the conservation law and the constitutive relation of stress and strain.

Consider a bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, with Lipschitz smooth boundary $\partial\Omega$ acted upon by force. Let $\mathbf{u} = \mathbf{u}(x, t)$ be the displacement field and $\rho = \rho(x)$ the density. The linearized strain due to the deformation is defined by

$$\varepsilon(\mathbf{u}) = \frac{1}{2}\nabla\mathbf{u} + \nabla\mathbf{u}^T.$$

The balance of momentum in the body is given by

$$\rho\mathbf{u}_{tt} = \operatorname{div}\boldsymbol{\sigma} + \mathbf{f}, \tag{2.1}$$

where $\boldsymbol{\sigma}$ denotes the stress tensor. In this paper we consider the following three- and two-kernel constitutive relations involving hydrostatic and deviatoric parts of the strain $\operatorname{tr}\varepsilon(\mathbf{u}) = \sum_{i=1}^{d}\varepsilon(\mathbf{u})_{ii}$, $\varepsilon_d(\mathbf{u}) = \varepsilon(\mathbf{u}) - \frac{1}{d}\mathbb{I}\operatorname{tr}\varepsilon(\mathbf{u})$, separately

$$\boldsymbol{\sigma} + (k_\sigma * \boldsymbol{\sigma})_t = \mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\operatorname{tr}\varepsilon} * \mathbb{I}\operatorname{tr}\varepsilon(\mathbf{u}_t), \tag{2.2}$$

and

$$\boldsymbol{\sigma} = \mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\operatorname{tr}\varepsilon} * \mathbb{I}\operatorname{tr}\varepsilon(\mathbf{u}_t). \tag{2.3}$$

We point to Remark 1 for their relation to existing models in the literature. The operator $*$ denotes the convolution on the positive half-line with respect to the time variable.

2.1. The viscoelastic wave model

The model for the wave propagation in complex media considered in this work is derived by coupling the conservation law with a more appropriate constitutive relation, as described above. The two hyperbolic models obtained by coupling (2.1) with two forms of constitutive relations, (2.2) and (2.3), are as follows.

- Model 1:

$$\rho\mathbf{u}_{tt} + \rho(k_\sigma * \mathbf{u}_{tt})_t - \operatorname{div}[\mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\operatorname{tr}\varepsilon} * \mathbb{I}\operatorname{tr}\varepsilon(\mathbf{u}_t)]$$
$$= \mathbf{f} + (k_\sigma * \mathbf{f})_t;$$

- Model 2:

$$\rho\mathbf{u}_{tt} - \operatorname{div}[\mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\operatorname{tr}\varepsilon} * \mathbb{I}\operatorname{tr}\varepsilon(\mathbf{u}_t)] = \mathbf{f}.$$

We can rewrite both of these equations as

$$\rho\mathbf{z}_{tt} - \operatorname{div}[\mathbb{C}\varepsilon(\mathbf{z}) + \mathbb{D} * \varepsilon(\mathbf{u}_t)] = \mathbf{g},$$

by introducing the auxiliary variable $\mathbf{z}$, tensor $\mathbb{D}$, and the forcing function $\mathbf{g}$:

$$\mathbf{z} = \mathbf{u} + k_\sigma * \mathbf{u}_t, \quad \mathbb{D} = \mathbb{A}k_\varepsilon - \mathbb{C}k_\sigma + \mathbb{I}k_{\operatorname{tr}\varepsilon}\operatorname{tr},$$
$$\mathbf{g} = \mathbf{f} + (k_\sigma * \mathbf{f})_t + \rho k_{\sigma t} \cdot \mathbf{u}_t(0), \tag{2.4}$$

where $k_\sigma = 0$ for Model 2. We note that the scalar counterpart of this model has been considered in the literature with fractional kernels; see, for example, [14]. We mention in passing that third-order in time models of viscoelasticity were introduced in [15] and have been recently a topic of extensive research, often referred to as the Moore–Gibson–Thompson viscoelasticity; see, for example, [16,17] and the references therein.

Assumptions on the medium parameters, kernels, and tensors. To allow for heterogeneous viscoelastic materials, we assume that $\rho \in L^\infty(\Omega)$ and that there exists $\underline{\rho} > 0$, such that

$$\rho(x) \geq \underline{\rho} > 0 \qquad \text{a.e. in } \Omega. \tag{2.5}$$

The fourth-order tensors $\mathbb{C}$ and $\mathbb{A}$ are assumed to be constant and symmetric,

$$\mathbb{C}, \; \mathbb{A} \in \text{Sym}^4(\mathbb{R}^d), \text{ and } \mathbb{C} \text{ is positive definite.} \tag{2.6}$$

Moreover, we assume that the kernel k_σ and tensor $\mathbb{D}$ according to (2.4) are such that the following non-negativity conditions hold:

$$\int_0^t (\mathbb{D} * \mathbf{y})(s) : (k_\sigma * \mathbf{y})_t(s)\, ds \geq 0; \tag{2.7}$$

$$\int_0^t (\mathbb{D} * \mathbf{y})(s) : \mathbf{y}(s)\, ds \geq \underline{\gamma} \|\mathbf{y}\|^2_{H^{-\delta}(0,t)} \tag{2.8}$$

for some $\underline{\gamma} > 0$, $\delta > 0$ and all $\mathbf{y} \in W^{1,1}(0, T, \text{Sym}^4(\mathbb{R}^d))$. Here

$$\text{Sym}^4(\mathbb{R}^d) = \{\mathbb{C} \in \mathbb{R}^{d \times d \times d \times d} : \mathbb{C}_{ijk\ell} = \mathbb{C}_{ij\ell k} = \mathbb{C}_{jik\ell} = \mathbb{C}_{k\ell ij}\},$$

$$\text{Sym}^2(\mathbb{R}^d) = \{\mathbf{y} \in \mathbb{R}^{d \times d} : \mathbf{y}_{ij} = \mathbf{y}_{ji}\},$$

and $H^{-\delta}(0,t)$ is the space of all functions whose extension by zero to $\mathbb{R}$ is in $H^{-\delta}(\mathbb{R})$. In the scalar case

$$\mathbb{D}(t) = \mathbb{M}m(t)$$

with $\mathbb{M} \in \text{Sym}^4(\mathbb{R}^d)$ positive semidefinite and using the $\mathbb{M}$ bilinear form

$$\langle \mathbf{w}, \mathbf{v} \rangle = \mathbb{M}\mathbf{w} : \mathbf{v}$$

as well as seminorm, condition (2.7) follows from the integrated version of Lemma B.1, provided $k_\sigma = k * m$ with $k \geq 0$ and $k' \leq 0$, using $\mathbf{w} = m * \mathbf{y}$, which satisfies $\mathbf{w}(0) = 0$ and assuming $\mathbf{w} \in H^1(0, T)$. In the same setting, assuming strict positivity of $\mathbb{M}$ and $\Re(\mathcal{F}m)(\omega) \geq \gamma(1 + \omega^2)^{-\delta/2}$, $\omega \in \mathbb{R}$, (where $\Re$ denotes the real part and $\mathcal{F}$ the Fourier transform), we can conclude from Lemma B.2 that (2.8) holds.

Remark 1. We comment on the relation of (2.2) and (2.3) to other models in the literature. Assuming the material is homogeneous and isotropic, a very general relation between deviatoric and hydrostatic parts of stress (σ_d, tr σ) and strain ($\varepsilon_d(\mathbf{u})$, tr $\varepsilon(\mathbf{u})$) is given by

$$\sigma_d + (k_\sigma * \sigma_d)_t = 2\mu\varepsilon_d(\mathbf{u}) + k_\varepsilon * \varepsilon_d(\mathbf{u}_t) \tag{2.9a}$$

$$\text{tr}\,\sigma + (k_{\text{tr}\sigma} * \text{tr}\,\sigma)_t = (2\mu + d\lambda)\text{tr}\,\varepsilon(\mathbf{u}) + k_{\text{tr}\varepsilon} * \text{tr}\,\varepsilon(\mathbf{u}_t), \tag{2.9b}$$

where μ, λ are the Lamé constants and $k_\sigma, k_{\text{tr}\sigma}, k_\varepsilon, k_{\text{tr}\varepsilon}$ are four scalar-valued kernels.

Assuming $\varepsilon(\mathbf{u})(0) = 0$ and using Laplace transform,

$$(1 + \omega\widehat{k_\sigma})\widehat{\sigma}_d = (2\mu + \omega\widehat{k_\varepsilon})\varepsilon_d(\widehat{\mathbf{u}})$$

$$(1 + \omega\widehat{k_{\text{tr}\sigma}})\text{tr}\,\widehat{\sigma} = (2\mu + d\lambda + \widehat{k_{\text{tr}\varepsilon}})\text{tr}\,\varepsilon(\widehat{\mathbf{u}}),$$

as well as $\sigma = \sigma_d + \frac{1}{d}\mathbb{I}\text{tr}\,\sigma$, it is not difficult to see that (2.9a) and (2.9b) can be reduced to a form with three kernels non-uniquely. One such reduction is with the kernels $(k_\sigma, k_\varepsilon, k_{\text{tr}\varepsilon}^{(1)})$, where

$$k_{\text{tr}\varepsilon}^{(1)} = \mathcal{L}^{-1}\left(\frac{1}{d(1 + \omega\widehat{k_{\text{tr}\sigma}})}\left[(2\mu + d\lambda)(\widehat{k_\sigma} - \widehat{k_{\text{tr}\sigma}}) + \widehat{k_{\text{tr}\varepsilon}}(1 + \omega\widehat{k_\sigma})\right]\right),$$

where skipping the superscript $^{(1)}$ we obtain the form (2.2) in case of the special choice

$$\mathbb{C}_{ijk\ell} = \delta_{ik}\delta_{j\ell}(2\mu + \lambda\delta_{ij}\text{tr}), \quad \mathbb{A}_{ijk\ell} = \delta_{ik}\delta_{j\ell}. \tag{2.10}$$

Using the same technique once again, the constitutive relation (2.2) can be reduced to a form with only two kernels $(k_\varepsilon^{(2)}, k_{\text{tr}\varepsilon}^{(2)})$, where

$$k_\varepsilon^{(2)} = \mathcal{L}^{-1}\left(\frac{\widehat{k_\varepsilon} - 2\mu\widehat{k_\sigma}}{1 + \omega\widehat{k_\sigma}}\right), \; k_{\text{tr}\varepsilon}^{(2)} = \mathcal{L}^{-1}\left(\frac{\widehat{k_{\text{tr}\varepsilon}^{(1)}} - \lambda\widehat{k_\sigma}}{1 + \omega\widehat{k_\sigma}}\right),$$

which we again denote by $k_\varepsilon, k_{\text{tr}\varepsilon}$ and obtain the form (2.3). Existing works consider one of the above three forms (2.2), (2.3), or (2.9), as constituent relation between stress and strain in 3D with either fractional kernels

$$g_\alpha(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$$

or the finite sum of these fractional kernels. The constitutive relation of the form (2.9) with $k_\sigma = k_\varepsilon, k_{\mathrm{tr}\sigma} = k_{\mathrm{tr}\varepsilon}$ is used in, e.g., [18,19]. The form (2.2) with $k_\sigma = k_\varepsilon = k_{\mathrm{tr}\varepsilon}$ is studied in [20]. The two-kernel representation form (2.3) is employed in [21] with fractional kernels. The four-kernel form (2.9) with the kernels being a sum of fractional kernels is considered in [22].

2.2. The inverse problem

We next discuss the inverse problem of determining the kernels from the measurements of the displacement $\mathbf{u}$ on finite number of points on the surface of the boundary. We assume that the domain boundary $\partial\Omega$ is the disjoint union of $\Gamma_D \neq \emptyset$, where homogeneous Dirichlet conditions will be imposed, and Γ_N with Neumann boundary excitation.

We use Tikhonov regularization, (see, e.g., [23–25]) for recovering the unknown kernels from over-specified data and end up with a PDE-constrained optimization problem:

$$\min_{\vec{k},\mathbf{u}\in X\times\boldsymbol{U}} J(\vec{k},\mathbf{u}), \tag{2.11}$$

where $\vec{k} = (k_\sigma, k_\varepsilon, k_{\mathrm{tr}\varepsilon})$, such that

$$\begin{cases} \rho\mathbf{z}_{tt} - \mathrm{div}\,[\mathbb{C}\varepsilon(\mathbf{z}) + \mathbb{D} * \varepsilon(\mathbf{u}_t)] = \mathbf{g}, & \text{in } \Omega \times (0,T), \\ \mathbf{z} = \mathbf{u} + k_\sigma * \mathbf{u}_t, & \text{in } \Omega \times (0,T), \\ \mathbf{u} = 0 & \text{on } \Gamma_D \times (0,T), \\ \sigma \cdot n = \mathbf{h} & \text{on } \Gamma_N \times (0,T), \\ (\mathbf{u}, \mathbf{u}_t)|_{t=0} = (\mathbf{u}_0, \mathbf{u}_1) & \text{in } \Omega, \end{cases} \tag{2.12}$$

is satisfied in a weak sense, with $\mathbb{D}$ and $\mathbf{g}$ defined in (2.4). The regularity assumptions on the Neumann data h and initial conditions $(\mathbf{u}_0, \mathbf{u}_1)$ needed for the state problem to be well-posed will be specified in the upcoming Propositions 3.1 and 3.2.

There exists a large variety of optimization applications in the context of viscoelasticity. For this reason, we here consider a general cost function J. In Section 5 below, we will specify some particular choices that can be made in the context of parameter identification problems arising from the estimation of the kernels from additional measurements.

2.3. Notation and theoretical preliminaries

Before proceeding with the analysis, we shortly introduce the function spaces and analytical techniques which will be used in the following sections.

We recall that we have assumed $\Omega \subset \mathbb{R}^d, d \in \{1, 2, 3\}$ to be an open, connected, and bounded set with Lipschitz regular boundary $\partial\Omega$. Furthermore, throughout the paper, we assume $\partial\Omega$ to be a disjoint union of Γ_D and Γ_N with Γ_D nonempty.

As usual, we equip the Sobolev and Lebesgue spaces $W^{k,p}(\Omega)$ (special case of $p = 2$, $H^k(\Omega)$) and $L^p(\Omega)$ on Ω with the norms $\|\cdot\|_{W^{k,p}(\Omega)}$ and $\|\cdot\|_{L^p(\Omega)}$; their vector-valued variants are denoted by $W^{k,p}(\Omega)^d$ and $L^p(\Omega)^d$.

For notational brevity, we introduce the Sobolev space that incorporates homogeneous Dirichlet boundary conditions:

$$\boldsymbol{H}_D^1(\Omega) = \{\mathbf{w} \in H^1(\Omega)^d : \mathbf{w} = 0 \text{ on } \Gamma_D\}.$$

We often use $x \lesssim y$ to denote $x \leq Cy$, where $C > 0$ is a generic constant that may depend on the final time. Throughout the paper $\langle\cdot,\cdot\rangle$ denotes the duality pairing between $(\boldsymbol{H}_D^1(\Omega))^*$ and $\boldsymbol{H}_D^1(\Omega)$, and $(\cdot,\cdot)_\Omega$ is the L^2-product on Ω for scalars, vectors and tensors. Furthermore, $(\cdot,\cdot)_{\Gamma_N}$ denotes the scalar product on $L^2(\Gamma_N)$.

Auxiliary results. We recall the useful Leibniz integral rule, integration by parts, and the transposition identity:

$$(k * \mathbf{w})_t(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t k(s)\mathbf{w}(t-s)\,\mathrm{d}s = (k * \mathbf{w}_t)(t) + k(t)\mathbf{w}(0), \tag{2.13}$$

$$k \in L^1(0,T), \quad \mathbf{w} \in W^{1,1}(0,T;X) \subseteq C(0,T;X),$$

$$\int_0^T \mathbf{w}_t(t)\mathbf{q}(T-t)\,\mathrm{d}t = \int_0^T \mathbf{w}(t)\mathbf{q}_t(T-t)\,\mathrm{d}t + \mathbf{w}(T)\mathbf{q}(0) - \mathbf{w}(0)\mathbf{q}(T), \tag{2.14}$$

$$\mathbf{q}, \mathbf{w} \in W^{1,1}(0,T;X),$$

$$\int_0^T (k * \mathbf{w})(t)\mathbf{q}(T - t)\,\mathrm{d}t = \int_0^T \mathbf{w}(t)(k * \mathbf{q})(T - t)\,\mathrm{d}t, \tag{2.15}$$

$$k \in L^1(0, T), \quad \mathbf{q},\ \mathbf{w} \in L^2(0, T; X)$$

for some Banach space X. The last identity is obtained by changing the order of integration as follows:

$$\int_0^T (k * \mathbf{w})(t)\mathbf{q}(T - t)\,\mathrm{d}t = \int_0^T \int_0^t k(t - s)\mathbf{w}(s)\,\mathrm{d}s\,\mathbf{q}(T - t)\,\mathrm{d}t$$

$$= \int_0^T \mathbf{w}(s) \int_s^T k(t - s)\mathbf{q}(T - t)\,\mathrm{d}t\,\mathrm{d}s = \int_0^T \mathbf{w}(s) \int_0^{T-s} k(T - s - r)\mathbf{q}(r)\,\mathrm{d}r\,\mathrm{d}s$$

$$= \int_0^T \mathbf{w}(T - \ell) \int_0^\ell k(\ell - r)\mathbf{q}(r)\,\mathrm{d}r\,\mathrm{d}\ell = \int_0^T \mathbf{w}(T - \ell)(k * \mathbf{q})(\ell)\,\mathrm{d}\ell$$

$$= \int_0^T \mathbf{w}(t)(k * \mathbf{q})(T - t)\,\mathrm{d}t.$$

3. Analysis of a nonlocal viscoelastic equation

In this section, we analyze the state problem (2.12) associated with our inverse problem, which can be written in terms of $\mathbf{u}$ only as follows:

$$\begin{cases} \rho\mathbf{u}_{tt} + \rho(k_\sigma * \mathbf{u}_{tt})_t \\ \quad - \operatorname{div}[\mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\operatorname{tr}\varepsilon} * \operatorname{tr}\varepsilon(\mathbf{u}_t)\mathbb{I}] = \mathbf{g} & \text{in } \Omega \times (0, T), \\ \mathbf{u} = 0 & \text{on } \Gamma_{\mathrm{D}} \times (0, T), \\ [\mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\operatorname{tr}\varepsilon} * \operatorname{tr}\varepsilon(\mathbf{u}_t)\mathbb{I}] \cdot n = \mathbf{h} + (k_\sigma * \mathbf{h})_t & \text{on } \Gamma_{\mathrm{N}} \times (0, T), \\ (\mathbf{u}, \mathbf{u}_t)|_{t=0} = (\mathbf{u}_0, \mathbf{u}_1) & \text{in } \Omega, \end{cases} \tag{3.1}$$

with $\mathbf{g}$ defined in (2.4), where the Neumann boundary condition on Γ_{N} results from the traction condition $\boldsymbol{\sigma} \cdot n = \mathbf{h}$.

The well-posedness analysis follows in spirit of the arguments in [20,19], where the main novelty arises from handling the k_σ term, which is not present in these references as a separate kernel, possibly different from the other kernels. In particular, we need to distinguish two cases in our well-posedness analysis based on whether the kernel k_σ is singular or not. This condition will influence the regularity of $\mathbf{u}$.

Proposition 3.1. *Let assumptions* (2.5) *and* (2.6) *hold. Given* $T > 0$, *let* $k_\sigma \in W^{1,1}(0, T)$ *with*

$$k_\sigma(t) \geq \underline{k} > -1,$$

k_σ *monotonically decreasing, and* k_ε, $k_{\operatorname{tr}\varepsilon} \in L^1(0, T)$. *Let also* $\mathbb{D} \in W^{1,1}(0, T; Sym^4(\mathbb{R}^d))$ *and the coercivity assumptions* (2.7)–(2.8) *on the involved kernels hold for all* $t \in [0, T]$. *Consider problem* (3.1) *with initial conditions*

$$(\mathbf{u}_0, \mathbf{u}_1) \in \boldsymbol{H}^1_{\mathrm{D}}(\Omega) \times L^2(\Omega)^d.$$

Further, assume that $\mathbf{g} \in L^1(0, T; L^2(\Omega)^d)$ *and*

$$\mathbf{h} + (k_\sigma * \mathbf{h})_t \in W^{1,1}(0, T; H^{-1/2}(\Gamma_{\mathrm{N}})^d).$$

Then there exists a unique

$$\mathbf{u} \in \boldsymbol{U} = \left\{ \mathbf{u} \in L^\infty(0, T; \boldsymbol{H}^1_{\mathrm{D}}(\Omega)) : \mathbf{u}_t \in L^\infty(0, T; L^2(\Omega)^d),\ \rho\mathbf{u}_{tt} \in L^1(0, T; (\boldsymbol{H}^1_{\mathrm{D}}(\Omega))^*) \right\},$$

such that

$$\langle \rho\mathbf{u}_{tt} + \rho(k_\sigma * \mathbf{u}_{tt})_t, \mathbf{v} \rangle + (\mathbb{C}\varepsilon(\mathbf{u}), \varepsilon(\mathbf{v}))_\Omega$$
$$+ (k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t), \varepsilon(\mathbf{v}))_\Omega + (k_{\operatorname{tr}\varepsilon} * \operatorname{tr}\varepsilon(\mathbf{u}_t)\mathbb{I}, \varepsilon(\mathbf{v}))_\Omega \tag{3.2}$$
$$= (\mathbf{g}, \mathbf{v})_\Omega + (\mathbf{h} + (k_\sigma * \mathbf{h})_t, \mathbf{v})_{\Gamma_{\mathrm{N}}},$$

for all $\mathbf{v} \in \boldsymbol{H}_{\mathrm{D}}^1(\Omega)^d$ a.e. in time, with $(\mathbf{u}, \mathbf{u}_t)|_{t=0} = (\mathbf{u}_0, \mathbf{u}_1)$. Furthermore, the solution satisfies the estimate

$$\|\mathbf{u}_t\|_{L^\infty(0,T;L^2(\Omega)^d)}^2 + \|\mathbf{u}\|_{L^\infty(0,T;H^1(\Omega)^d)}^2 + \underline{\gamma}\|\varepsilon(\mathbf{u}_t)\|_{H^{-\delta}(0,t;L^2(\Omega)^{d2})}^2$$
$$\lesssim \|\mathbf{u}_0\|_{H^1(\Omega)^d}^2 + \|\mathbf{u}_1\|_{L^2(\Omega)^d}^2 + \|\mathbf{g}\|_{L^1(0,T;L^2(\Omega)^d)}^2 + \|\mathbf{h} + (k_\sigma * \mathbf{h})_t\|_{W^{1,1}(0,t;H^{-1/2}(\Gamma_{\mathrm{N}})^d)}^2.$$

If additionally $\mathbf{u}_1 = 0$ and $\mathbf{g} \in L^\infty(0, T; (\boldsymbol{H}_{\mathrm{D}}^1(\Omega))^*)$, then $\rho\mathbf{u}_{tt} \in L^\infty(0, T; (\boldsymbol{H}_{\mathrm{D}}^1(\Omega))^*)$.

Proof. The proof follows by employing a Galerkin discretization in space based on the eigenvectors $\{\boldsymbol{\phi}_j\}_{j\geq 1}$, where

$$a(\boldsymbol{\phi}_j, \mathbf{v}) = \tilde{\lambda}_j(\rho\boldsymbol{\phi}_j, \mathbf{v})_\Omega, \quad \forall\mathbf{v} \in \boldsymbol{H}_{\mathrm{D}}^1(\Omega)$$

with the bilinear form

$$a(\boldsymbol{\phi}_j, \mathbf{v}) = \int_\Omega \mathbb{C}\varepsilon(\boldsymbol{\phi}_j) : \varepsilon(\mathbf{v})\,\mathrm{d}x;$$

cf. [20, Theorem 4.1]. The basis can be chosen as orthonormal in the weighted Lebesgue space $L_\rho^2(\Omega)^d$, as proven in [20, Lemma 4.1]. Denote $V_n = \mathrm{span}\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n\}$. We seek an approximate solution in the form

$$\mathbf{u}^n = \sum_{j=1}^n \xi_j(t)\boldsymbol{\phi}_j(x)$$

with $\mathbf{u}^n(0)$, $\mathbf{u}_t^n(0)$ chosen as $L_\rho^2(\Omega)^d$ projections of $\mathbf{u}_0$, $\mathbf{u}_1$ onto V_n. The semi-discrete problem can then be rewritten as the system

$$\begin{aligned}
&\xi_{i,tt}(t) + (k_\sigma * \xi_{i,tt})_t(t) + \tilde{\lambda}_i\xi_i(t) \\
&\quad + \sum_{j=1}^n (k_\varepsilon * \xi_{j,t})(t)(\mathbb{A}\varepsilon(\boldsymbol{\phi}_j), \varepsilon(\boldsymbol{\phi}_i))_\Omega + \sum_{j=1}^n (k_{\mathrm{tr}\,\varepsilon} * \xi_{j,t})(t)(\mathrm{tr}\,\varepsilon(\boldsymbol{\phi}_j), \mathrm{tr}\,\varepsilon(\boldsymbol{\phi}_i))_\Omega \\
&= (\mathbf{g}(t), \boldsymbol{\phi}_i)_\Omega + ((\mathbf{h} + (k_\sigma * \mathbf{h})_t)(t), \boldsymbol{\phi}_i)_{\Gamma_{\mathrm{N}}}
\end{aligned} \tag{3.3}$$

for $i = 1, \ldots, n$ and $t \in [0, T]$. In vector notation $\boldsymbol{\xi} = [\xi_1 \ \ldots \ \xi_n]^T$, we have

$$\boldsymbol{\xi}_{tt} + (k_\sigma * \boldsymbol{\xi}_{tt})_t + \tilde{\boldsymbol{\lambda}}\boldsymbol{\xi} + \mathbf{A}k_\varepsilon * \boldsymbol{\xi}_t + \boldsymbol{\mu}k_{\mathrm{tr}\,\varepsilon} * \boldsymbol{\xi}_t = \boldsymbol{b}, \tag{3.4}$$

with the right-hand side vector given by

$$\boldsymbol{b}(t) = (\mathbf{g}(t), \boldsymbol{\phi}_i)_\Omega + ((\mathbf{h} + (k_\sigma * \mathbf{h})_t)(t), \boldsymbol{\phi}_i)_{\Gamma_{\mathrm{N}}}.$$

Further, we have introduced $\tilde{\boldsymbol{\lambda}} = [\tilde{\lambda}_1 \ \ldots \ \tilde{\lambda}_n]$, $\mathbf{A} = [A_{ij}]_{n \times n}$, and $\boldsymbol{\mu} = [\mu_{ij}]_{n \times n}$ with

$$A_{ij} = (\mathbb{A}\varepsilon(\boldsymbol{\phi}_j), \varepsilon(\boldsymbol{\phi}_i))_\Omega, \quad \mu_{ij} = (\mathrm{tr}\,\varepsilon(\boldsymbol{\phi}_j), \mathrm{tr}\,\varepsilon(\boldsymbol{\phi}_i))_\Omega.$$

Note that we have $k_\sigma \in W^{1,1}(0, T) \hookrightarrow C[0, T]$. Denote $\boldsymbol{\chi} = \boldsymbol{\xi}_{tt}$. Then

$$(k_\sigma * \boldsymbol{\xi}_{tt})_t = k_{\sigma t} * \boldsymbol{\xi}_{tt} + k_\sigma(0)\boldsymbol{\xi}_{tt}$$

with $k_{\sigma t} \in L^1(0, T)$. We have

$$\boldsymbol{\xi}_t = 1 * \boldsymbol{\chi} + \boldsymbol{\xi}_1, \quad \boldsymbol{\xi} = 1 * 1 * \boldsymbol{\chi} + t\boldsymbol{\xi}_1 + \boldsymbol{\xi}_0,$$

where $\boldsymbol{\xi}_0$ and $\boldsymbol{\xi}_1$ are the vectors of coordinates of $\mathbf{u}^n(0)$ and $\mathbf{u}_t^n(0)$, respectively, in the basis. Then since we have assumed that $k_\sigma(0) > -1$, we can rewrite equation (3.4) as

$$\begin{aligned}
&(1 + k_\sigma(0))\boldsymbol{\chi} + k_{\sigma t} * \boldsymbol{\chi} + \tilde{\boldsymbol{\lambda}}(1 * 1) * \boldsymbol{\chi} + \mathbf{A}(k_\varepsilon * 1) * \boldsymbol{\chi} + \boldsymbol{\mu}(k_{\mathrm{tr}\,\varepsilon} * 1) * \boldsymbol{\chi} \\
&= \boldsymbol{b} - \tilde{\boldsymbol{\lambda}}(t\boldsymbol{\xi}_1 + \boldsymbol{\xi}_0) - \mathbf{A}k_\varepsilon * \boldsymbol{\xi}_1 - \boldsymbol{\mu}k_{\mathrm{tr}\,\varepsilon} * \boldsymbol{\xi}_1.
\end{aligned}$$

We can then see this problem as a system of Volterra integral equations of the second kind

$$(1 + k_\sigma(0))\boldsymbol{\chi} + \boldsymbol{K} * \boldsymbol{\chi} = \tilde{\boldsymbol{b}}(t) \tag{3.5}$$

with the kernel

$$K = k_{\sigma t} + \tilde{\lambda}(1 * 1) + \mathbf{A}(k_\varepsilon * 1) + \boldsymbol{\mu}(k_{\mathrm{tr}\,\varepsilon} * 1)$$

and the right-hand side

$$\tilde{\boldsymbol{b}}(t) = \boldsymbol{b} - \tilde{\lambda}(t\boldsymbol{\xi}_1 + \boldsymbol{\xi}_0) - \mathbf{A}k_\varepsilon * \boldsymbol{\xi}_1 - \boldsymbol{\mu}k_{\mathrm{tr}\,\varepsilon} * \boldsymbol{\xi}_1.$$

According to existence theory for the Volterra integral equations of the second kind (see, e.g. [26, Ch. 2, Theorem 3.5]), there exists a unique $\chi \in L^1(0, T)$, which solves (3.5). Combined with the imposed initial conditions, we recover a unique $\boldsymbol{\xi} \in W^{2,1}(0, T)$ and thus $\mathbf{u}^n \in W^{2,1}(0, T; \boldsymbol{H}_{\mathrm{D}}^1(\Omega))$. Note that then the combined function $\mathbf{z} = k_\sigma * \mathbf{u}_t^n + \mathbf{u}^n$ due to (2.13) and the identities

$$\mathbf{z}_t^n = k_{\sigma t} * \mathbf{u}_t^n + (1 + k_\sigma(0))\mathbf{u}_t^n, \tag{3.6}$$

$$\mathbf{z}_{tt}^n = k_{\sigma t} * \mathbf{u}_{tt}^n + k_{\sigma t} * \mathbf{u}_t^n(0) + (1 + k_\sigma(0))\mathbf{u}_{tt}^n \tag{3.7}$$

has the regularity $\mathbf{z}^n \in W^{2,1}(0, T; \boldsymbol{H}_{\mathrm{D}}^1(\Omega))$ as well, which we will use in the energy analysis below. For ease of notation, we omit the superscript n in the semi-discrete solution from now on.

Energy analysis. Testing the semi-discrete problem (3.8) given in terms of $\mathbf{z} = k_\sigma * \mathbf{u}_t + \mathbf{u}$:

$$(\rho\mathbf{z}_{tt}, \mathbf{v})_\Omega + (\mathbb{C}\varepsilon(\mathbf{z}), \varepsilon(\mathbf{v}))_\Omega + (\mathbb{D} * \varepsilon(\mathbf{u}_t), \varepsilon(\mathbf{v}))_\Omega$$
$$= (\mathbf{g}, \mathbf{v})_\Omega + (\mathbf{h} + (k_\sigma * \mathbf{h})_t, \mathbf{v})_{\Gamma_\mathrm{N}}, \tag{3.8}$$

with $\mathbf{z}_t(t) = (k_\sigma * \mathbf{u}_t + \mathbf{u})_t(t) \in V_n$ and integrating in space and time leads at first to the following identity:

$$\frac{1}{2}\|\rho^{1/2}\mathbf{z}_t(t)\|_{L^2(\Omega)^d}^2 + \frac{1}{2}\|\mathbb{C}^{1/2}\varepsilon(\mathbf{z}(t))\|_{L^2(\Omega)^{d^2}}^2 + \int_0^t (\mathbb{D} * \varepsilon(\mathbf{u}_t), \varepsilon(\mathbf{z}_t))_\Omega \, \mathrm{d}s$$

$$= \frac{1}{2}\|\rho^{1/2}\mathbf{z}_t(0)\|_{L^2(\Omega)^d}^2 + \frac{1}{2}\|\mathbb{C}^{1/2}\varepsilon(\mathbf{z}(0))\|_{L^2(\Omega)^{d^2}}^2 + \int_0^t (\mathbf{g}, \mathbf{z}_t)_\Omega \, \mathrm{d}s \tag{3.9}$$

$$+ \int_0^t (\mathbf{h} + (k_\sigma * \mathbf{h})_t, \mathbf{z}_t)_{\Gamma_\mathrm{N}} \, \mathrm{d}s.$$

By Hölder's inequality and Young's inequality $xy \leq x^2 + \frac{1}{4}y^2$, we have

$$\int_0^t (\mathbf{g}, \mathbf{z}_t)_\Omega \, \mathrm{d}s \leq \left(\int_0^t \|\rho^{-1/2}\mathbf{g}\|_{L^2(\Omega)^d} \, \mathrm{d}s\right)^2 + \frac{1}{4}\sup_{s\in(0,t)} \|\rho^{1/2}\mathbf{z}_t(s)\|_{L^2(\Omega)^d}^2.$$

Employing integration by parts with respect to time yields

$$\int_0^t (\mathbf{h} + (k_\sigma * \mathbf{h})_t, \mathbf{z}_t)_{\Gamma_\mathrm{N}} \, \mathrm{d}s = (\mathbf{h} + (k_\sigma * \mathbf{h})_t, \mathbf{z})_{\Gamma_\mathrm{N}}\Big|_0^t - \int_0^t (\mathbf{h}_t + (k_\sigma * \mathbf{h})_{tt}, \mathbf{z})_{\Gamma_\mathrm{N}} \, \mathrm{d}s.$$

Thus,

$$\int_0^t (\mathbf{h} + (k_\sigma * \mathbf{h})_t, \mathbf{z}_t)_{\Gamma_\mathrm{N}} \, \mathrm{d}s$$
$$\leq \|(\mathbf{h} + (k_\sigma * \mathbf{h})_t)(t)\|_{H^{-1/2}(\Gamma_\mathrm{N})^d}\|\mathbf{z}(t)\|_{H^{1/2}(\Gamma_\mathrm{N})^d} + \|\mathbf{h}(0)\|_{H^{-1/2}(\Gamma_\mathrm{N})^d}\|\mathbf{z}(0)\|_{H^{1/2}(\Gamma_\mathrm{N})^d}$$
$$- \|\mathbf{h}_t + (k_\sigma * \mathbf{h})_{tt}\|_{L^1(0,t;H^{-1/2}(\Gamma_\mathrm{N})^d)}\|\mathbf{z}\|_{L^\infty(0,t;H^{1/2}(\Gamma_\mathrm{N})^d)}.$$

By recalling the definition of $\mathbf{z}$, we further have

$$\int_0^t (\mathbb{D} * \varepsilon(\mathbf{u}_t), \varepsilon(\mathbf{z}))_\Omega \, \mathrm{d}s = \int_0^t (\mathbb{D} * \varepsilon(\mathbf{u}_t), (k_\sigma * \varepsilon(\mathbf{u}_t))_t + \varepsilon(\mathbf{u}_t))_\Omega \, \mathrm{d}s.$$

We can employ our assumption (2.8) on the kernels, which guarantees that

$$\int\limits_0^t (\mathbb{D} * \varepsilon(\mathbf{u}_t), \varepsilon(\mathbf{u}_t))_\Omega \, \mathrm{d}s \geq \underline{\gamma} \, \|\varepsilon(\mathbf{u}_t)\|^2_{H^{-\delta}(0,t;L^2(\Omega)^{d^2})}.$$

We can further use assumption (2.7) to conclude that

$$\int\limits_0^t (\mathbb{D} * \varepsilon(\mathbf{u}_t), (k_\sigma * \varepsilon(\mathbf{u}_t))_t)_\Omega \, \mathrm{d}s \geq 0.$$

Thus, from (3.9), by relying on the assumptions on ρ and $\mathbb{C}$, we have the estimate

$$\|\mathbf{z}_t(t)\|^2_{L^2(\Omega)^d} + \|\mathbf{z}(t)\|^2_{H^1(\Omega)^d} + \underline{\gamma}\|\varepsilon(\mathbf{u}_t)\|^2_{H^{-\delta}(0,t;L^2(\Omega)^{d^2})}$$
$$\lesssim \|\mathbf{z}_t(0)\|^2_{L^2(\Omega)^d} + \|\mathbf{z}(0)\|^2_{H^1(\Omega)^d} + \|\mathbf{g}\|^2_{L^1(0,T;L^2(\Omega)^d)} \qquad (3.10)$$
$$+ \|\mathbf{h} + (k_\sigma * \mathbf{h})_t\|^2_{W^{1,1}(0,t;H^{-1/2}(\Gamma_N)^d)}$$

at first in a discrete setting, which can be extended to the continuous setting via standard (weak) compactness arguments.

Applying the $L^\infty(0,T;L^2(\Omega)^d)$ norm to (3.6) with monotonically decreasing kernel k_σ and using the estimate

$$\|k_{\sigma t} * \mathbf{u}_t\|_{L^\infty(0,T;L^2(\Omega)^d)} \qquad\qquad\qquad (3.11)$$
$$\leq \|k_{\sigma t}\|_{L^1(0,T)}\|\mathbf{u}_t\|_{L^\infty(0,T;L^2(\Omega)^d)} = (k_\sigma(0) - k_\sigma(T))\|\mathbf{u}_t\|_{L^\infty(0,T;L^2(\Omega)^d)}$$

under the assumption $k_\sigma(t) \geq \underline{k} > -1$ we get

$$(1 + \underline{k})\|\mathbf{u}_t\|_{L^\infty(0,T;L^2(\Omega))} \leq \|\mathbf{z}_t\|_{L^\infty(0,T;L^2(\Omega))}$$

Likewise, taking the $L^\infty(0,T;\boldsymbol{H}^1_D(\Omega)^d)$ norm of $\mathbf{z} = k_\sigma * \mathbf{u}_t + \mathbf{u} = k_{\sigma t} * \mathbf{u} + k_\sigma(0)\mathbf{u} - k_\sigma\mathbf{u}(0)$, we conclude

$$(1 + \underline{k})\|\mathbf{u}\|_{L^\infty(0,T;\boldsymbol{H}^1_D(\Omega)^d)} \leq \|\mathbf{z}_t\|_{L^\infty(0,T;\boldsymbol{H}^1_D(\Omega)^d)} + \|k_\sigma\|_{L^\infty(0,T)}\|\mathbf{u}_0\|_{\boldsymbol{H}^1_D(\Omega)^d}.$$

Since $\mathbb{D} \in W^{1,1}(0,T;\mathrm{Sym}^4(\mathbb{R}^d))$, we can obtain an additional estimate on $\mathbf{z}_{tt}$ as follows. For $\mathbf{v} \in \boldsymbol{H}^1_D(\Omega)$ with $\|\mathbf{v}\|_{\boldsymbol{H}^1_D(\Omega)} \leq 1$, we decompose it into $\mathbf{v} = \mathbf{v}^1 + \mathbf{v}^2$, where $\mathbf{v}^1 \in \mathrm{span}\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n\}$ and $\mathbf{v}^2 \in \mathrm{span}\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n\}^\perp$; see, e.g., [27, Ch. 7.2] for similar arguments. Then $\|\mathbf{v}^1\|_{\boldsymbol{H}^1_D(\Omega)} \leq 1$ and

$$(\rho\mathbf{z}_{tt}, \mathbf{v})_\Omega = (\rho\mathbf{z}_{tt}, \mathbf{v}^1)_\Omega$$
$$= -(\mathbb{C}\varepsilon(\mathbf{z}), \varepsilon(\mathbf{v}^1))_\Omega - (\mathbb{D} * \varepsilon(\mathbf{u}_t), \varepsilon(\mathbf{v}^1))_\Omega + (\mathbf{g}, \mathbf{v}^1)_\Omega + (\mathbf{h} + (k_\sigma * \mathbf{h})_t, \mathbf{v}^1)_{\Gamma_N}.$$

From here, using the identity

$$(\mathbb{D} * \varepsilon(\mathbf{u}_t))(t) = (\mathbb{D}_t * \varepsilon(\mathbf{u}))(t) + \mathbb{D}(0)\varepsilon(\mathbf{u}(t)) - \mathbb{D}(t)\varepsilon(\mathbf{u}(0))$$

we have

$$\|\rho\mathbf{z}_{tt}\|_{L^p(0,T;(\boldsymbol{H}^1_D(\Omega))^*)} \lesssim \|\mathbb{C}\varepsilon(\mathbf{z})\|_{L^p(0,T;L^2(\Omega)^{d^2})} + \|\mathbb{D}_t * \varepsilon(\mathbf{u}) + \mathbb{D}(0)\varepsilon(\mathbf{u})\|_{L^p(0,T;L^2(\Omega)^{d^2})}$$
$$+ \|\mathbf{g}\|_{L^p(0,T;(\boldsymbol{H}^1_D(\Omega))^*)} + \|\mathbf{h} + (k_\sigma * \mathbf{h})_t\|_{L^p(0,T;H^{-1/2}(\Gamma_N)^d)}$$
$$+ \|\mathbb{D}\|_{L^p(0,T)}\|\varepsilon(\mathbf{u}(0))\|_{L^2(\Omega)^{d^2}}.$$

With $p = 1$ we get $\rho\mathbf{z}_{tt} \in L^1(0,T;(\boldsymbol{H}^1_D(\Omega))^*)$. If additionally $\mathbf{g} \in L^\infty(0,T;(\boldsymbol{H}^1_D(\Omega))^*)$ and $\mathbf{u}_t(t = 0) = 0$, then we can also use $p = \infty$.

Similarly to (3.11), by taking the $L^p(0,T;(\boldsymbol{H}^1_D(\Omega))^*)$ norm of (3.7), we conclude the same regularity for $\rho\mathbf{u}_{tt}$ as for $\rho\mathbf{z}_{tt}$.

Usual compactness arguments then lead to a solution of the state problem, where the initial conditions $\mathbf{u}(0) = \mathbf{u}_0$ and $\mathbf{u}_t(0) = \mathbf{u}_1$ are satisfied as equalities in $C_w([0,T];H^1_D(\Omega))$ and $C_w([0,T];L^2_\rho(\Omega)^d)$, respectively.

Uniqueness follows by proving that the only solution of the homogeneous problem

$$\langle\rho\mathbf{z}_{tt}, \mathbf{v}\rangle + (\mathbb{C}\varepsilon(\mathbf{z}), \varepsilon(\mathbf{v}))_\Omega + (\mathbb{D} * \varepsilon(\mathbf{u}_t), \varepsilon(\mathbf{v}))_\Omega = 0 \qquad (3.12)$$

with zero initial and boundary data, is $\mathbf{u} = 0$. To prove uniqueness, we cannot test the above weak form with $\mathbf{z}_t$ as in the semi-discrete setting due to its insufficient spatial regularity. Instead, following, e.g., [27, Ch. 7.2], we fix $0 \leq s \leq T$ and take

$$\mathbf{v}(t) = \begin{cases} \int\limits_t^s \mathbf{z}(\tau) \, \mathrm{d}\tau & \text{if } 0 \leq t \leq s, \\ 0 & \text{if } s \leq t \leq T, \end{cases}$$

which satisfies $\mathbf{v}(t) \in \boldsymbol{H}_{\mathrm{D}}^1(\Omega)$ for each $0 \leq t \leq T$. Furthermore,

$$\mathbf{v}_t = -\mathbf{z}, \quad 0 \leq t \leq s.$$

Since $\mathbf{v}(s) = 0$ and (with zero initial data) $\mathbf{z}_t(0) = 0$, taking this test function in (3.12) yields

$$\int\limits_0^s \left(\langle \rho \mathbf{z}_t, \mathbf{z} \rangle - (\mathbb{C}\varepsilon(\mathbf{v}_t), \varepsilon(\mathbf{v}))_\Omega + (\mathbb{D} * \varepsilon(\mathbf{u}))_t, \varepsilon(\mathbf{v}))_\Omega \right) \mathrm{d}t = 0,$$

where we have also used that $\mathbb{D} * \varepsilon(\mathbf{u}_t) = (\mathbb{D} * \varepsilon(\mathbf{u}))_t - \mathbb{D}(t)\varepsilon(\mathbf{u}(0)) = (\mathbb{D} * \varepsilon(\mathbf{u}))_t$. Note that

$$\int\limits_0^s -(\mathbb{C}\varepsilon(\mathbf{v}_t), \varepsilon(\mathbf{v}))_\Omega = -\frac{1}{2}(\mathbb{C}\varepsilon(\mathbf{v}), \varepsilon(\mathbf{v}))_\Omega \Big|_0^s$$

$$= \frac{1}{2}(\mathbb{C}\varepsilon(\mathbf{v}(0)), \varepsilon(\mathbf{v}(0)))_\Omega$$

since $\mathbf{v}(s) = 0$. Integration by parts with respect to time thus yields

$$\frac{1}{2}\|\rho^{1/2}\mathbf{z}(s)\|_{L^2(\Omega)^d}^2 + \frac{1}{2}\|\mathbb{C}^{1/2}\varepsilon(\mathbf{v}(0))\|_{L^2(\Omega)^{d2}}^2 + \int\limits_0^s (\mathbb{D} * \varepsilon(\mathbf{u}), \varepsilon(\mathbf{z}))_\Omega \, \mathrm{d}t = 0. \tag{3.13}$$

Due to our assumptions on $\mathbb{D}$, we know that

$$\int\limits_0^s (\mathbb{D} * \varepsilon(\mathbf{u}), \varepsilon(\mathbf{z}))_\Omega \, \mathrm{d}t = \int\limits_0^s (\mathbb{D} * \varepsilon(\mathbf{u}), k_\sigma * \varepsilon(\mathbf{u}_t) + \varepsilon(\mathbf{u}))_\Omega \, \mathrm{d}t$$

$$= \int\limits_0^s (\mathbb{D} * \varepsilon(\mathbf{u}), (k_\sigma * \varepsilon(\mathbf{u}))_t + \varepsilon(\mathbf{u}))_\Omega \, \mathrm{d}t \geq 0.$$

Thus, from (3.13), we conclude that $\mathbf{z} = 0$ and thus

$$k_\sigma * \mathbf{u}_t + \mathbf{u} = k_{\sigma t} * \mathbf{u} + k_\sigma(0)\mathbf{u} + \mathbf{u} = 0.$$

Together with $\mathbf{u}|_{t=0} = 0$, this implies that $\mathbf{u} = 0$. $\square$

We next consider the well-posedness of the state problem for a singular k_σ. Note that the analysis will lead to the same estimate (3.10) of the combined quantity $\mathbf{z}$, but the resulting estimate for $\mathbf{u}$ crucially depends on whether k_σ is singular or not. We also point out that even for a singular kernel $k_\sigma(t) \sim t^{-\gamma}$ as $t \to 0$, where $(k_\sigma * \mathbf{u}_{tt})_t$ behaves similarly to a Riemann–Liouville derivative of order γ of $\mathbf{u}_{tt}$, no initial condition on $\mathbf{u}_{tt}$ is imposed, which is in line with the theory of Riemann–Liouville fractional ODEs; see, e.g., [28,29], where initial conditions are imposed on the fractional derivatives up to one order less than the order of the fractional ODE, which is $\gamma < 1$ here resulting in no initial conditions.

Since k_σ is singular, we can now introduce

$$\ell \in L^1(0, T) \text{ such that } \ell * k_\sigma = 1. \tag{3.14}$$

The fact that $\ell \in L^1(0, T)$ is due to Tauberian Theorems [30, Theorems 2, 3, Chapter XIII.5]

$$k_\sigma(t) \sim t^{-\gamma} \text{ as } t \to 0 \Leftrightarrow \widehat{k_\sigma}(s) \sim s^{\gamma-1} \text{ as } s \to \infty \Leftrightarrow$$

$$\widehat{\ell}(s) = \frac{\widehat{1}}{\widehat{k_\sigma}(s)} = \frac{1}{s\widehat{k_\sigma}(s)} \sim s^{-\gamma} \text{ as } s \to \infty \Leftrightarrow \ell(t) \sim t^{\gamma-1} \text{ as } t \to 0.$$

Additionally we will assume

$$\|\ell\|_{L^1(0,T)} < 1 \tag{3.15}$$

and

$$k_\sigma^\gamma(t) := k_\sigma(t)t^\gamma \text{ satisfies } k_\sigma^\gamma(0) \neq 0, \quad k_\sigma^\gamma \in C[0, T] \cap W^{1,\infty}(0, T) \tag{3.16}$$

Proposition 3.2. *Given $T > 0$, let assumptions* (2.5), (2.6) *and assumptions* (2.7)–(2.8) *on the involved kernels hold. Assume that $k_\sigma \in L^1(0, T)$ is singular; that is, there exists $\gamma \in (0, 1]$, such that*

$$k_\sigma(t) \sim t^{-\gamma} \text{ as } t \to 0$$

with additionally (3.16) *being satisfied and* (3.15) *holding for ℓ as in* (3.14). *Furthermore, assume that k_ε, $k_{\text{tr}\,\varepsilon} \in L^1(0, T)$. Consider problem* (3.1) *with the initial conditions*

$$(\mathbf{u}, \mathbf{u}_t)|_{t=0} = (\mathbf{u}_0, \mathbf{u}_1) \in \boldsymbol{H}_{\mathrm{D}}^1(\Omega) \times \{0\}.$$

Further, assume that the source term and boundary data satisfy the assumptions of Proposition 3.1. *Then there exists a unique*

$$\mathbf{u} \in \boldsymbol{U} = \left\{ \mathbf{u} \in L^\infty(0, T; \boldsymbol{H}_{\mathrm{D}}^1(\Omega)) : \mathbf{u}_t \in H^{\gamma-1}(0, T; \boldsymbol{H}_{\mathrm{D}}^1(\Omega)), \quad \mathbf{u}_{tt} \in H^{\gamma-1}(0, T; L^2(\Omega)^d) \right\},$$

such that (3.2) *holds, with $(\mathbf{u}, \mathbf{u}_t)|_{t=0} = (\mathbf{u}_0, \mathbf{u}_1)$. The solution satisfies the following estimate:*

$$\|\mathbf{u}_t\|_{H^{\gamma-1}(0,T;\boldsymbol{H}_{\mathrm{D}}^1(\Omega))}^2 + \|\mathbf{u}_{tt}\|_{H^{\gamma-1}(0,T;L^2(\Omega)^d)}^2 + \underline{\gamma}\|\varepsilon(\mathbf{u}_t)\|_{H^{-\delta}(0,t;L^2(\Omega)^{d^2})}^2$$
$$\lesssim \|\mathbf{u}_0\|_{\boldsymbol{H}_{\mathrm{D}}^1(\Omega)}^2 + \|\mathbf{u}_1\|_{L^2(\Omega)^d}^2 + \|\mathbf{g}\|_{L^1(0,T;L^2(\Omega)^d)}^2 + \|\mathbf{h} + (k_\sigma * \mathbf{h})_t\|_{W^{1,1}(0,t;H^{-1/2}(\Gamma_{\mathrm{N}})^d)}^2.$$

Proof. The proof in the singular case follows via a Galerkin approximation similarly to before. However, to prove the existence of the solution of the semi-discrete problem (3.3), we use the kernel $\ell \in L^1(0, T)$ as defined in (3.14) We then set

$$\widehat{\boldsymbol{\chi}} = (k_\sigma * \boldsymbol{\xi}_{tt})_t, \tag{3.17}$$

as the new unknown. From (3.17), we have

$$\boldsymbol{\xi}_{tt}(t) = (\ell * \widehat{\boldsymbol{\chi}})(t)$$

Then

$$\boldsymbol{\xi}_t = 1 * \boldsymbol{\xi}_{tt} + \boldsymbol{\xi}_1 = 1 * \ell * \widehat{\boldsymbol{\chi}} + \boldsymbol{\xi}_1,$$
$$\boldsymbol{\xi} = 1 * \boldsymbol{\xi}_t + \boldsymbol{\xi}_0 = 1 * 1 * \ell * \widehat{\boldsymbol{\chi}} + \boldsymbol{\xi}_1 t + \boldsymbol{\xi}_0$$

with $\boldsymbol{\xi}_0$ and $\boldsymbol{\xi}_1$, being the coordinates of $\mathbf{u}^n(0)$ and $\mathbf{u}_t^n(0)$ in the basis, respectively. Equation (3.4) can thus be rewritten as the following equation for $\widehat{\boldsymbol{\chi}}$:

$$\ell * \widehat{\boldsymbol{\chi}} + \widehat{\boldsymbol{\chi}} + \tilde{\lambda}(1 * 1 * \ell) * \widehat{\boldsymbol{\chi}} + \mathbf{A}(1 * k_\varepsilon * \ell) * \widehat{\boldsymbol{\chi}} + \boldsymbol{\mu}(1 * k_{\text{tr}\,\varepsilon} * \ell) * \widehat{\boldsymbol{\chi}}$$
$$= \boldsymbol{b} - \tilde{\lambda}(\boldsymbol{\xi}_1 t + \boldsymbol{\xi}_0) - \mathbf{A}k_\varepsilon(t)\boldsymbol{\xi}_1 - \boldsymbol{\mu}k_{\text{tr}\,\varepsilon}(t)\boldsymbol{\xi}_1.$$

We can see this problem as a system of Volterra integral equations of second kind

$$\widehat{\boldsymbol{\chi}} + \boldsymbol{K} * \widehat{\boldsymbol{\chi}} = \tilde{\boldsymbol{b}}$$

with the kernel

$$\boldsymbol{K} = \ell + \tilde{\lambda}(1 * 1 * \ell) + \mathbf{A}(1 * k_\varepsilon * \ell) + \boldsymbol{\mu}(1 * k_{\text{tr}\,\varepsilon} * \ell)$$

and the right-hand side

$$\tilde{\boldsymbol{b}}(t) = \boldsymbol{b} - \tilde{\lambda}(\boldsymbol{\xi}_1 t + \boldsymbol{\xi}_0) - \mathbf{A}k_\varepsilon(t)\boldsymbol{\xi}_1 - \boldsymbol{\mu}k_{\text{tr}\,\varepsilon}(t)\boldsymbol{\xi}_1.$$

Similarly to the proof of Proposition 3.1, existence theory for the Volterra integral equations of the second kind yields a unique $\widehat{\boldsymbol{\chi}} \in L^1(0, T)$ and thus $\mathbf{z}, \mathbf{u} \in W^{2,1}(0, T; \boldsymbol{H}_{\mathrm{D}}^1(\Omega))$.

Energy analysis. We can derive the uniform estimate (3.10) for $\mathbf{z}$ as before. In order to conclude additional regularity of $\mathbf{u}$, we use the fact that the auxiliary function $\mathbf{v} = k_\sigma * \mathbf{u}_t$, which vanishes at $t = 0$, satisfies, along with its derivatives, the Volterra integral equations

$$\mathbf{v} + \ell * \mathbf{v} = \mathbf{z} - \mathbf{u}_0, \quad \mathbf{v}_t + \ell * \mathbf{v}_t = \mathbf{z}_t, \quad \mathbf{v}_{tt} + \ell * \mathbf{v}_{tt} = \mathbf{z}_{tt} + \ell(t)\mathbf{v}_t(0).$$

We then take the $L^\infty(0, T; \boldsymbol{H}_{\mathrm{D}}^1(\Omega))$, $L^\infty(0, T; L^2(\Omega)^d)$, and $L^1(0, T; (\boldsymbol{H}_{\mathrm{D}}^1(\Omega))^*)$ norms, respectively and use the estimate

$$\|\ell * \mathbf{w}\|_{L^\infty(0,T;X)} \le \|\ell\|_{L^1(0,T)} \|\mathbf{w}\|_{L^\infty(0,T;X)}.$$

Assuming $\|\ell\|_{L^1(0,T)} < 1$, we can thus estimate

$$
\begin{aligned}
(1 - \|\ell\|_{L^1(0,T)})\|\mathbf{v}\|_{L^\infty(0,T;\mathbf{H}_D^1(\Omega))} &\leq \|\mathbf{z}\|_{L^\infty(0,T;\mathbf{H}_D^1(\Omega))} + \|\mathbf{u}_0\|_{\mathbf{H}_D^1(\Omega)^d}, \\
(1 - \|\ell\|_{L^1(0,T)})\|\mathbf{v}_t\|_{L^\infty(0,T;L^2(\Omega)^d)} &\leq \|\mathbf{z}_t\|_{L^\infty(0,T;L^2(\Omega)^d)}, \\
(1 - \|\ell\|_{L^1(0,T)})\|\mathbf{v}_{tt}\|_{L^1(0,T;L^2(\Omega)^d)} &\leq \|\mathbf{z}_{tt}\|_{L^1(0,T;(\mathbf{H}_D^1(\Omega))^*)} + \|\ell\|_{L^1(0,T)}\|\mathbf{v}_t(0)\|_{(\mathbf{H}_D^1(\Omega))^*} \\
&\leq \|\mathbf{z}_{tt}\|_{L^1(0,T;(\mathbf{H}_D^1(\Omega))^*)} + \|\mathbf{z}_t(0)\|_{(\mathbf{H}_D^1(\Omega))^*}
\end{aligned}
$$

To extract additional temporal regularity of $\mathbf{u}$ from these bounds, we consider $k_\sigma * \mathbf{u}_t = \mathbf{v}$ as an Abel integral equation for $\mathbf{u}_t$ and specify the assumption $k_\sigma \sim t^{-\gamma}$ for $t \to 0$ to (3.16). Then from, e.g., [31, Theorem 1] we can conclude

$$
\begin{aligned}
&\|\mathbf{u}_t\|^2_{H^{\gamma-1}(0,T;\mathbf{H}_D^1(\Omega))} + \|\mathbf{u}_{tt}\|^2_{H^{\gamma-1}(0,T;L^2(\Omega)^d)} \\
&\lesssim \|\mathbf{z}_t(0)\|^2_{L^2(\Omega)^d} + \|\mathbf{z}(0)\|^2_{H^1(\Omega)^d} + \|\mathbf{g}\|^2_{L^1(0,T;L^2(\Omega)^d)} + \|\mathbf{h} + (k_\sigma * \mathbf{h})_t\|^2_{W^{1,1}(0,t;H^{-1/2}(\Gamma_N)^d)}.
\end{aligned}
$$

The rest of the arguments follow as in Proposition 3.1. □

Remark 2. Note that the use of a Hilbert space argument like [31, Theorem 1] in the last part of the proof of Proposition 3.2 leads to a loss as compared to some possibly most sharp result, and also the assumption of differentiability of k_σ^γ can probably be relaxed.

We note that higher spatial regularity of the solution can be obtained in case $\Gamma_N = \emptyset$; a discussion of this case is included in Appendix A.

4. The adjoint problem

In order to allow for efficient gradient computation in the iterative solution of the optimization problem for (2.11), (2.12), we next wish to derive the adjoint problem. Recall that we are considering the optimization problem

$$
\min_{\vec{k},\mathbf{u} \in X \times \mathbf{U}} J(\vec{k}, \mathbf{u}) \text{ such that } \mathbf{u} \text{ solves (3.1)},
$$

where $\vec{k} = (k_\sigma, k_\varepsilon, k_{\mathrm{tr}\varepsilon})$ and $\mathbf{U}$ is the solution space for $\mathbf{u}$; cf. Propositions 3.1 and 3.2, where

$$
\mathbf{g} = \mathbf{f} + (k_\sigma * \mathbf{f})_t.
$$

In here, the cost function J is a general Fréchet differentiable mapping from $X \times \mathbf{U}$ to $\mathbb{R}$. In Section 5 it will be specified for the purpose of parameter identification, but more generally one could think of other criteria to be optimized for, such as certain properties of the state $\mathbf{u}$ or the kernels k_σ, k_ε, and $k_{\mathrm{tr}\varepsilon}$. The Lagrange function reads

$$
\begin{aligned}
\mathcal{L}(\vec{k}, \mathbf{u}, \mathbf{p}) = &J(\vec{k}, \mathbf{u}) \\
&+ \int_0^T \int_\Omega \Big(\big((\rho\mathbf{u}_{tt} + \rho(k_\sigma * \mathbf{u}_{tt})_t - \mathbf{f} - (k_\sigma * \mathbf{f})_t) \cdot \mathbf{p} + (\mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t)) : \varepsilon(\mathbf{p}) \\
&\qquad\qquad + k_{\mathrm{tr}\varepsilon} * \mathrm{tr}\varepsilon(\mathbf{u}_t)) \,\mathrm{tr}\varepsilon(\mathbf{p}) \big) \Big) \,\mathrm{d}x - \int_{\Gamma_N} (\mathbf{h} + (k_\sigma * \mathbf{h})_t) \cdot \mathbf{p} \,\mathrm{d}S \Big) \,\mathrm{d}t.
\end{aligned}
$$

In order to derive first order optimality conditions by setting all derivatives of $\mathcal{L}$ to zero, we use a solution space that incorporates not only the homogeneous boundary conditions but also the initial data, by writing $\mathbf{u} = \{\mathbf{u}_0 + t\mathbf{u}_1\} + \tilde{\mathbf{u}} \in \{\mathbf{u}_0 + t\mathbf{u}_1\} + \tilde{\mathbf{U}}$ with

$$
\tilde{\mathbf{U}} \subseteq \{\mathbf{u} \in \mathbf{U} : (\mathbf{u}, \mathbf{u}_t)|_{t=0} = (0, 0)\}.
$$

Formally writing the first order optimality conditions

$$
\frac{\partial \mathcal{L}}{\partial k_\sigma} = 0, \quad \frac{\partial \mathcal{L}}{\partial k_\varepsilon} = 0, \quad \frac{\partial \mathcal{L}}{\partial k_{\mathrm{tr}\varepsilon}} = 0, \quad \frac{\partial \mathcal{L}}{\partial \tilde{\mathbf{u}}} = 0, \quad \frac{\partial \mathcal{L}}{\partial \mathbf{p}} = 0,
$$

at $(\vec{k}, \mathbf{u}, \mathbf{p}) = (\vec{k}, \{\mathbf{u}_0 + t\mathbf{u}_1\} + \tilde{\mathbf{u}}, \mathbf{p})$, the fifth and fourth condition yield the state and adjoint equation, respectively. We will now focus on the second to last condition, that is, the adjoint equation:

$$0 = \frac{\partial \mathcal{L}}{\partial \mathbf{u}}(\vec{k}, \mathbf{u}, \mathbf{p})\mathbf{v}$$

$$= \frac{\partial \mathcal{J}}{\partial \mathbf{u}}(\vec{k}, \mathbf{u})\mathbf{v} + \int_0^T \int_\Omega \Big(\big(\rho \mathbf{v}_{tt} + \rho(k_\sigma * \mathbf{v}_{tt})_t \big) \cdot \mathbf{p} + \big(\mathbb{C}\varepsilon(\mathbf{v}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{v}_t) \big) : \varepsilon(\mathbf{p}) \tag{4.1}$$

$$+ k_{\mathrm{tr}\,\varepsilon} * \mathrm{tr}\varepsilon(\mathbf{u}_t)) \, \mathrm{tr}\varepsilon(\mathbf{p})) \Big) \Big) \, \mathrm{d}x\mathrm{d}t$$

for all $\mathbf{v} \in \tilde{\boldsymbol{U}}$ which is supposed to uniquely determine the adjoint state $\mathbf{p} \in \tilde{\boldsymbol{P}}$ with the test space $\tilde{\boldsymbol{P}} \subseteq L^2(0, T; L^2(\Omega)^d)$ yet to be determined. Knowing that the adjoint state $\mathbf{p}$ will solve a backwards in time PDE with end conditions at T, we make a timeflip right away and write $\mathbf{p}(t) = \overline{\mathbf{p}}(T - t)$. This allows us to use the elementary integration by parts and transposition identities (2.14) and (2.15). This method yields

$$\int_0^T \int_\Omega \rho \mathbf{v}_{tt}(t) \cdot \overline{\mathbf{p}}(T - t) \, \mathrm{d}x\mathrm{d}t = \int_0^T \int_\Omega \rho \mathbf{v}_t(t) \cdot \overline{\mathbf{p}}_t(T - t) \, \mathrm{d}x\mathrm{d}t + \int_\Omega \rho \mathbf{v}_t(T) \cdot \overline{\mathbf{p}}(0) \, \mathrm{d}x$$

$$= \int_0^T \int_\Omega \rho \mathbf{v}(t) \cdot \overline{\mathbf{p}}_{tt}(T - t) \, \mathrm{d}x\mathrm{d}t + \int_\Omega \rho \Big(\mathbf{v}(T) \cdot \overline{\mathbf{p}}_t(0) + \mathbf{v}_t(T) \cdot \overline{\mathbf{p}}(0) \Big) \, \mathrm{d}x$$

since $\mathbf{v}_t(0) = 0$, $\mathbf{v}(0) = 0$,

$$\int_0^T \int_\Omega \rho(k_\sigma * \mathbf{v}_{tt})_t(t) \cdot \overline{\mathbf{p}}(T - t) \, \mathrm{d}x\mathrm{d}t$$

$$= \int_0^T \int_\Omega \rho(k_\sigma * \mathbf{v}_{tt})(t) \cdot \overline{\mathbf{p}}_t(T - t) \, \mathrm{d}x\mathrm{d}t + \int_\Omega \rho(k_\sigma * \mathbf{v}_{tt})(T) \cdot \overline{\mathbf{p}}(0) \, \mathrm{d}x$$

$$= \int_0^T \int_\Omega \rho \mathbf{v}_{tt}(t) \cdot (k_\sigma * \overline{\mathbf{p}}_t)(T - t) \, \mathrm{d}x\mathrm{d}t + \int_\Omega \rho(k_\sigma * \mathbf{v}_{tt})(T) \cdot \overline{\mathbf{p}}(0) \, \mathrm{d}x$$

$$= \int_0^T \int_\Omega \rho \mathbf{v}_t(t) \cdot (k_\sigma * \overline{\mathbf{p}}_t)_t(T - t) \, \mathrm{d}x\mathrm{d}t + \int_\Omega \rho(k_\sigma * \mathbf{v}_{tt})(T) \cdot \overline{\mathbf{p}}(0) \, \mathrm{d}x$$

$$= \int_0^T \int_\Omega \rho \mathbf{v}(t) \cdot (k_\sigma * \overline{\mathbf{p}}_t)_{tt}(T - t) \, \mathrm{d}x\mathrm{d}t$$
$$+ \int_\Omega \rho \Big(\mathbf{v}(T) \cdot (k_\sigma * \overline{\mathbf{p}}_t)_t(0) + (k_\sigma * \mathbf{v}_{tt})(T) \cdot \overline{\mathbf{p}}(0) \Big) \, \mathrm{d}x,$$

due to the fact that $(k_\sigma * \mathbf{v}_{tt})(0) = 0$, $(k_\sigma * \overline{\mathbf{p}}_t)(0) = 0$, $\mathbf{v}_t(0) = 0$, and $\mathbf{v}(0) = 0$, where

$$(k_\sigma * \overline{\mathbf{p}}_t)_t(0) = \lim_{t \to 0+} k_\sigma(t)\overline{\mathbf{p}}_t(0).$$

Therefore, in case of a singularity in k_σ at $t = 0$, we need to impose $\overline{\mathbf{p}}_t(0) = 0$ and

$$\int_0^T \int_\Omega (k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{v}_t))(t) : \varepsilon(\overline{\mathbf{p}}(T - t)) \, \mathrm{d}x\mathrm{d}t = \int_0^T \int_\Omega \mathbb{A}\varepsilon(\mathbf{v}_t(t)) : (k_\varepsilon * \varepsilon(\overline{\mathbf{p}}))(T - t) \, \mathrm{d}x\mathrm{d}t$$

$$= \int_0^T \int_\Omega \varepsilon(\mathbf{v}(t)) : (\mathbb{A}k_\varepsilon * \varepsilon(\overline{\mathbf{p}}_t))(T - t) \, \mathrm{d}x\mathrm{d}t,$$

where we have used $(\mathbb{A}k_\varepsilon * \varepsilon(\overline{\mathbf{p}}_t))(0) = 0$ and $\varepsilon(\mathbf{v}(0)) = 0$.

Furthermore, we define $\nabla_u J(\vec{k}, \mathbf{u}) \in \tilde{\boldsymbol{U}}^*$ by the variational equation

$$\int_0^T \int_\Omega \nabla_u J(\vec{k}, \mathbf{u})(x, t) \cdot \mathbf{v}(x, t)\, dx dt = \frac{\partial \mathcal{J}}{\partial \mathbf{u}}(\vec{k}, \mathbf{u})\mathbf{v} \quad \text{for all } \mathbf{v} \in \tilde{\boldsymbol{U}}.$$

Altogether, (4.1) becomes

$$
\begin{aligned}
0 = \int_0^T \int_\Omega &\Big(\mathbf{v}(t) \cdot \big(\nabla_u J(\vec{k}, \mathbf{u})(t) + \rho \overline{\mathbf{p}}_{tt}(T - t) + \rho(k_\sigma * \overline{\mathbf{p}}_t)_{tt}(T - t)\big) \\
&+ \varepsilon(\mathbf{v}(t)) : \big(\mathbb{C}\varepsilon(\overline{\mathbf{p}}(T - t)) + (\mathbb{A}k_\varepsilon * \varepsilon(\overline{\mathbf{p}}_t))(T - t)\big) \\
&+ \mathrm{tr}\varepsilon(\mathbf{v}(t))\,(k_{\mathrm{tr}\varepsilon} * \mathrm{tr}\varepsilon(\overline{\mathbf{p}}_t))(T - t)\Big)\, dx dt \\
+ \int_\Omega \rho\Big(\mathbf{v}(T) \cdot &\big(1 + \lim_{t\to 0+} k_\sigma(t)\big)\overline{\mathbf{p}}_t(0) + \big(\mathbf{v}_t(T) + (k_\sigma * \mathbf{v}_{tt})(T)\big) \cdot \overline{\mathbf{p}}(0)\Big)\, dx.
\end{aligned}
$$

Since $\mathbf{v}$ is arbitrary, this implies that $\overline{\mathbf{p}}$ solves (in the weak sense) the following adjoint problem:

$$
\begin{cases}
\rho \overline{\mathbf{p}}_{tt} + \rho(k_\sigma * \overline{\mathbf{p}}_t)_{tt} - \mathrm{div}[\mathbb{C}\varepsilon(\overline{\mathbf{p}}) + k_\varepsilon * \mathbb{A}\varepsilon(\overline{\mathbf{p}}_t) + k_{\mathrm{tr}\varepsilon} * \mathrm{tr}\varepsilon(\overline{\mathbf{p}}_t)\mathbb{I}] & \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = -\nabla_u \overline{J}(\vec{k}, \mathbf{u}), & \text{in } \Omega \times (0, T), \\
\overline{\mathbf{p}} = 0 & \text{on } \Gamma_{\mathrm{D}} \times (0, T), \\
[\mathbb{C}\varepsilon(\overline{\mathbf{p}}) + k_\varepsilon * \mathbb{A}\varepsilon(\overline{\mathbf{p}}_t) + k_{\mathrm{tr}\varepsilon} * \mathrm{tr}\varepsilon(\overline{\mathbf{p}}_t)\mathbb{I}] \cdot n = 0 & \text{on } \Gamma_{\mathrm{N}} \times (0, T), \\
(\overline{\mathbf{p}}, \overline{\mathbf{p}}_t)|_{t=0} = (0, 0) & \text{in } \Omega,
\end{cases}
\tag{4.2}
$$

with $\overline{J}(\vec{k}, \mathbf{u})(t) = J(\vec{k}, \mathbf{u})(T - t)$.

This way, we can not only characterize a minimizer of (2.11), (2.12), but also compute the gradient of the reduced cost function

$$j(\vec{k}) = J(\vec{k}, \mathbf{u}(\vec{k})), \quad \text{where } \mathbf{u}(\vec{k}) \text{ solves (2.12)}$$

by using the fact that $j(\vec{k}) = \mathcal{L}(\vec{k}, \mathbf{u}(\vec{k}), \mathbf{p}(\vec{k}))$. By means of the Chain Rule, we further have

$$
\begin{aligned}
\frac{\partial j}{\partial k_\sigma}(\vec{k})h &= \frac{\partial \mathcal{L}}{\partial k_\sigma}(\vec{k}, \mathbf{u}(\vec{k}), \mathbf{p}(\vec{k}))h \\
&\quad + \frac{\partial \mathcal{L}}{\partial \mathbf{u}}(\vec{k}, \mathbf{u}(\vec{k}), \mathbf{p}(\vec{k}))\frac{\partial \mathbf{u}}{\partial k_\sigma}(\vec{k})h + \frac{\partial \mathcal{L}}{\partial \mathbf{p}}(\vec{k}, \mathbf{u}(\vec{k}), \mathbf{p}(\vec{k}))\frac{\partial \mathbf{p}}{\partial k_\sigma}(\vec{k})h \\
&= \frac{\partial \mathcal{L}}{\partial k_\sigma}(\vec{k}, \mathbf{u}(\vec{k}), \mathbf{p}(\vec{k}))h.
\end{aligned}
\tag{4.3}
$$

Here we have used the fact that $\mathbf{u}(\vec{k})$ and $\mathbf{p}(\vec{k})$ solve the state and adjoint equations, respectively. Analogously for $\frac{\partial j}{\partial k_\varepsilon}$, $\frac{\partial j}{\partial k_{\mathrm{tr}\varepsilon}}$, we obtain

$$\frac{\partial j}{\partial k_\varepsilon}(\vec{k})h = \frac{\partial \mathcal{L}}{\partial k_\varepsilon}(\vec{k}, \mathbf{u}(\vec{k}), \mathbf{p}(\vec{k}))h, \qquad \frac{\partial j}{\partial k_{\mathrm{tr}\varepsilon}}(\vec{k})h = \frac{\partial \mathcal{L}}{\partial k_{\mathrm{tr}\varepsilon}}(\vec{k}, \mathbf{u}(\vec{k}), \mathbf{p}(\vec{k}))h.$$

Well-posedness of the adjoint problem (4.2) follows from Proposition 3.2 in case of a singular kernel k_σ, provided $\nabla_u J(\vec{k}, \mathbf{u}) \in L^1(0, T; L^2(\Omega)^d)$.

5. Estimation of the kernels from additional observations

In this section we turn to the inverse problem of parametrizing the model, that is, of determining the kernels. For this purpose we will need additional observations of the state $\mathbf{u}$. Since these are given in addition to the initial and boundary conditions, they are usually denoted as overspecified data.

By taking Laplace transforms, we have shown in Remark 1 that at least in the isotropic case (2.10) the two- and three-kernel formulations are equivalent. Therefore, without loss of generality, concerning parameter identification we will focus on the two-kernel model (formally setting $k_\sigma = 0$):

$$
\begin{cases}
\rho \mathbf{u}_{tt} - \operatorname{div}[\mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\mathrm{tr}\varepsilon} * \operatorname{tr}\varepsilon(\mathbf{u}_t)\mathbb{I}] = \mathbf{f}, & \text{in } \Omega \times (0, T), \\
\mathbf{u} = 0 & \text{on } \Gamma_{\mathrm{D}} \times (0, T), \\
[\mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\mathrm{tr}\varepsilon} * \operatorname{tr}\varepsilon(\mathbf{u}_t)\mathbb{I}] \cdot n = \mathbf{h} & \text{on } \Gamma_{\mathrm{N}} \times (0, T), \\
(\mathbf{u}, \mathbf{u}_t)|_{t=0} = (0, 0) & \text{in } \Omega.
\end{cases}
\tag{5.1}
$$

Applying Tikhonov regularization for the inverse problem kernel identification problem mentioned above, we end up with a PDE-constrained optimization problem of the form (2.11) under the constraint (5.1).

After having fixed the PDE constraint in (5.1), we will now define the cost function J in such a way that it takes into account the observations as well as some additional regularization that might be needed in order to solve this infinite-dimensional parameter identification problem. Measurements may be taken for multiple excitations, for example a pulling and a shearing experiment in order to distinguish between k_ε and $k_{\mathrm{tr}\varepsilon}$. To this end, note that the derivation from Section 4 easily extends to the case of multiple PDE constraints (5.1) that arise from considering states resulting from multiple different excitations $\mathbf{f}^n$, $n = 1, \dots N$. The corresponding overposed data are measurements $\mathbf{u}_i^{n,\mathrm{meas}}(t)$, $t \in (0, T)$ of state $\mathbf{u}^n$ at I given points $x_i^n \in \overline{\Omega}$ (typically at the boundary), $i \in \{1, \dots, I\}$. In case of limited regularity of $\mathbf{u}^n$ – note that according to Propositions 3.1, 3.2, we have $\mathbf{u}^n \in L^\infty(0, T; \boldsymbol{H}_{\mathrm{D}}^1(\Omega))$ – point evaluations in space cannot be justified. We therefore consider a variant that locally averages over a boundary patch $\Gamma_i^n \subseteq \partial\Omega$ concentrated at x_i^n and possibly weighted with some $L^\infty(\Gamma_i^n)$ function η_i^n, see (5.3) below.

Resulting examples of cost functions (abbreviating $\vec{k} = (k_\varepsilon, k_{\mathrm{tr}\varepsilon})$) are then

$$
\begin{aligned}
&J(\vec{k}, \mathbf{u}^1, \dots, \mathbf{u}^N) \\
&= \frac{1}{2} \sum_{n=1}^N \sum_{i=1}^I \int_0^T |\mathbf{u}^n(x_i^n, t) - \mathbf{u}_i^{n,\mathrm{meas}}(t)|^2 \, dt + \gamma_\epsilon \mathcal{R}_\epsilon(k_\varepsilon) + \gamma_{\mathrm{tr}\varepsilon} \mathcal{R}_{\mathrm{tr}\varepsilon}(k_{\mathrm{tr}\varepsilon})
\end{aligned}
\tag{5.2}
$$

for point measurements or

$$
\begin{aligned}
&J(\vec{k}, \mathbf{u}^1, \dots, \mathbf{u}^N) \\
&= \frac{1}{2} \sum_{n=1}^N \sum_{i=1}^I \int_0^T | \int_{\Gamma_i^n} \eta_i^n(x)\mathbf{u}^n(x, t) \, ds(x) - \mathbf{u}_i^{\mathrm{meas}}(t)|^2 \, dt + \gamma_\epsilon \mathcal{R}_\epsilon(k_\varepsilon) \\
&\qquad\qquad\qquad\qquad\qquad + \gamma_{\mathrm{tr}\varepsilon} \mathcal{R}_{\mathrm{tr}\varepsilon}(k_{\mathrm{tr}\varepsilon}),
\end{aligned}
\tag{5.3}
$$

for locally averaged observations, where $\mathcal{R}_\epsilon$, $\mathcal{R}_{\mathrm{tr}\varepsilon}$ are regularization functionals and γ_ϵ, $\gamma_{\mathrm{tr}\varepsilon} > 0$ are the corresponding regularization parameters. Before specifying these, we point out that we can already now compute the derivatives of J with respect to the state as needed for defining the adjoint state according to (4.2). For J as in (5.2), we have

$$
\nabla_u J(\vec{k}, \mathbf{u}^1, \dots, \mathbf{u}^N)(x, t) = \sum_{i=1}^I \sum_{n=1}^N (\mathbf{u}^n(x_i, t) - \mathbf{u}_i^{\mathrm{meas}}(t))\delta(x - x_i),
$$

and for J as in (5.3), we obtain

$$
\nabla_u J(\vec{k}, \mathbf{u}^1, \dots, \mathbf{u}^N)(x, t) = \sum_{i=1}^I \sum_{n=1}^N (\int_{\Gamma_i^n} \eta_i^n(x)\mathbf{u}^n(x, t) \, ds(x) - \mathbf{u}_i^{\mathrm{meas}}(t))\eta_i^n(x)\delta_{\Gamma_i^n}(x),
$$

where for any $f \in W^{1,1}(\Omega)$, $\int_\Omega \delta_{\Gamma_i^n}(x) f(x) \, dx = \int_{\Gamma_i^n}(x) f(x) \, ds(x)$; see [32, Chapter 15] for well-definedness of the trace.

Definition of the regularization terms $\gamma_\epsilon \mathcal{R}_\epsilon(k_\varepsilon)$, $\gamma_{\mathrm{tr}\varepsilon} \mathcal{R}_{\mathrm{tr}\varepsilon}(k_{\mathrm{tr}\varepsilon})$ ist closely related to the choice of function spaces for the kernels. A canonical choice of the space for the kernels is $L^p(0, T)$, with $p > 1$ close to one, since this space is still reflexive and also applicable to singular kernels as in Proposition 3.2. This choice allows forward solutions that are regular enough to admit locally averaging measurements according to (5.3) (but not point measurements as in (5.2) – for this we would need $\mathbf{u} \in L^2(0, T; H^s(\Omega))$ with s large enough so that $H^s(\Omega)$ embeds continuously into $C(\Omega)$). Thus, well-definedness of the forward operator

$$
F : L^1(0, T)^2 \to L^2(0, T; \mathbb{R}^{NI}), \quad \vec{k} \mapsto (t \mapsto (\int_\Omega \eta_i^n(x)\mathbf{u}^n(x, t))_{i \in \{1, \dots, I\}, n \in \{1, \dots, N\}}))
$$

then allows us to apply regularization theory in (reflexive) Banach spaces (see, e.g., [33–35] and the references therein) with quite general regularization functionals $\mathcal{R}_\epsilon$, $\mathcal{R}_{\mathrm{tr}\varepsilon}$. A Hilbert space setting is most convenient for defining a gradient, though.

Thus in place of $L^p(0,T)$, we alternatively consider a weighted L^2 space $X_\epsilon = L^2_{w_\epsilon}(0,T)$ with a weight function w_ϵ that vanishes at $t = 0$ (at a certain rate as $t \to 0$) if k_ϵ is expected to have a singularity that is stronger than $t^{-1/2}$. The simplest example of a regularization term is then just

$$\mathcal{R}_\epsilon(k_\epsilon) = \frac{1}{2}\|k_\epsilon\|^2_{X_\epsilon}, \tag{5.4}$$

but one might also make more sophisticated choices, such as

$$\mathcal{R}_\epsilon(k_\epsilon) = \inf_{\alpha_m \in (0,1), a_m \in \mathbb{R}} \left(\frac{1}{2}\|k_\epsilon - \sum_{m=1}^M a_m \frac{t^{-\alpha_m}}{\Gamma(1-\alpha_m)}\|^2_X + \beta \sum_{m=1}^M |a_m| \right)$$

to promote closeness to a multi-term fractional derivative kernel with as few components as possible. As a matter of fact, in our numerical experiments we did not need any regularization term to be added as the problem appears to be only mildly ill-posed and so the regularization induced by discretization of the kernel was sufficient to deal with even high noise levels in the data.

Starting from the expression (4.3), we can then define the gradient of the reduced cost function by using the inner product $\langle k_1, k_2 \rangle_{X_\epsilon} = \int_0^T w_\epsilon(t) k_1(t) k_2(t)\, dt$ and the variational equation

$$\int_0^T w_\epsilon(t) \nabla_{k_\epsilon} j(\vec{k})(t) h(t)\, dt = \frac{\partial j}{\partial k_\epsilon}(\vec{k}) h \quad \text{for all } h \in X.$$

Due to the fact that

$$\frac{\partial \mathcal{L}}{\partial k_\epsilon}(\vec{k}, \mathbf{u}^1(\vec{k}), \ldots, \mathbf{u}^N(\vec{k}), \mathbf{p}^1(\vec{k}), \ldots, \mathbf{p}^N(\vec{k})) h$$

$$= \sum_{n=1}^N \int_0^T \int_\Omega (\mathbb{A}\varepsilon(\mathbf{u}^n_t) * \varepsilon(\overline{\mathbf{p}}^n))(T-t)\, dx dt + \gamma_\epsilon \mathcal{R}'_\epsilon(k_\epsilon) h$$

$$= \sum_{n=1}^N \int_0^T h(s) \left(\int_\Omega \rho\tau(\mathbf{u}^n_{tt} * \overline{\mathbf{p}}^n_t)(T-s)\, dx + \gamma_\epsilon w_\epsilon(s) k_\epsilon(s) \right) ds,$$

we have

$$\nabla_{k_\epsilon} j(\vec{k})(t) = \gamma_\epsilon \nabla\mathcal{R}_\epsilon(k_\epsilon)(t) + \frac{1}{w_\epsilon(t)} \sum_{n=1}^N \int_\Omega (\mathbb{A}\varepsilon(\mathbf{u}^n_t) * \varepsilon(\overline{\mathbf{p}}^n))(T-t)\, dx,$$

where the function $\nabla\mathcal{R}_\epsilon(k_\epsilon)(t)$ is determined by the variational equation

$$\int_0^T w_\epsilon(t) \nabla\mathcal{R}_\epsilon(k_\epsilon)(t) h(t)\, dt = \mathcal{R}'_\epsilon(k_\epsilon) h \quad \text{for all } h \in L^2_{w_\epsilon}(0,T).$$

For example, in case (5.4), $\mathcal{R}'_\epsilon(k_\epsilon) h = \langle k_\epsilon, h \rangle_{X_\epsilon}$ and $\nabla\mathcal{R}_\epsilon(k_\epsilon)(t) = k_\epsilon(t)$.

Note that in case of a strongly singular kernel k_ε, also the gradient will contain a singularity via the $\dfrac{1}{w_\epsilon(t)}$ term.

Likewise, we obtain

$$\nabla_{k_{\mathrm{tr}\varepsilon}} j(\vec{k})(t) = \gamma_{\mathrm{tr}\varepsilon} \nabla\mathcal{R}_{\mathrm{tr}\varepsilon}(k_{\mathrm{tr}\varepsilon})(t) + \frac{1}{w_{\mathrm{tr}\varepsilon}(t)} \sum_{n=1}^N \int_\Omega (\mathrm{tr}\varepsilon(\mathbf{u}^n_t) * \mathrm{tr}\varepsilon(\overline{\mathbf{p}}^n))(T-t)\, dx.$$

Remark 3 *(On uniqueness).* To discuss uniqueness of identification of two kernels from two additional observations, we consider the equivalent reformulation in terms of the deviatoric and hydrostatic parts of the strain

$$\rho\mathbf{u}_{tt} - \mathrm{div}[\mathbb{C}\varepsilon_d(\mathbf{u}) + \mathbf{y}\mathrm{tr}\varepsilon(\mathbf{u}_t) + k_\varepsilon * \mathbb{A}\varepsilon_d(\mathbf{u}_t) + \widetilde{k_{\mathrm{tr}\varepsilon}} * \mathrm{tr}\varepsilon(\mathbf{u}_t)] = \mathbf{f}$$

with $\mathbf{y} = \frac{1}{d}\mathbb{C}\mathbb{I}$, $\widetilde{k_{\mathrm{tr}\varepsilon}}(t) = k_{\mathrm{tr}\varepsilon}(t)\mathbb{I} + k_\varepsilon(t)\mathbb{A}\mathbb{I} \in \mathrm{Sym}^2(\mathbb{R}^d)$ of the PDE in (5.1)

$$\rho\mathbf{u}_{tt} - \mathrm{div}[\mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\mathrm{tr}\varepsilon} * \mathrm{tr}\varepsilon(\mathbf{u}_t)\mathbb{I}] = \mathbf{f}$$

and focus on the case of $\mathbb{A}$ and $\widetilde{k_{\mathrm{tr}\,\varepsilon}}$ being scalar multiples of $\mathbb{C}$ and $\mathbf{y}$, respectively. An example for this is the isotropic case

$$(\mathbb{C})_{ijkl} = 2\mu\delta_{ik}\delta_{j\ell}, \quad \mathbf{y}_{ij} = (\lambda + \frac{2\mu}{d})\delta_{ij}, \quad (\mathbb{A})_{ijkl} = \delta_{ik}\delta_{j\ell}. \tag{5.5}$$

Additionally we assume to have a separable source term so that we arrive at the form

$$\begin{cases} \rho\mathbf{u}_{tt} - \mathrm{div}[\mathbb{C}_1\varepsilon(\mathbf{u}) + \mathbb{C}_2\varepsilon(\mathbf{u}) \\ \qquad + k_1 * \mathbb{C}_1\varepsilon(\mathbf{u}_t) + k_2 * \mathbb{C}_2\varepsilon(\mathbf{u}_t)] = \ell(t)\mathbf{f}(x) & \text{in } \Omega \times (0, T), \\ \mathbf{u} = 0 & \text{on } \Gamma_{\mathrm{D}} \times (0, T), \\ [\mathbb{C}_1\varepsilon(\mathbf{u}) + \mathbb{C}_2\varepsilon(\mathbf{u}) + k_1 * \mathbb{C}_1\varepsilon(\mathbf{u}_t) + k_2 * \mathbb{C}_2\varepsilon(\mathbf{u}_t)] \cdot n = 0 & \text{on } \Gamma_{\mathrm{N}} \times (0, T), \\ (\mathbf{u}, \mathbf{u}_t)|_{t=0} = (0, 0) & \text{in } \Omega. \end{cases} \tag{5.6}$$

Moreover, we consider an idealized setting in which we can excite the system via two different eigenpairs (λ_1, φ_1), (λ_2, φ_2) of the operators $-\frac{1}{\rho}\mathrm{div}[\mathbb{C}_1\varepsilon]$ and $-\frac{1}{\rho}\mathrm{div}[\mathbb{C}_2\varepsilon]$, (equipped with mixed Dirichlet-Neumann boundary conditions) that are selfadjoint and positive with respect to the weighted L^2 inner product $(\mathbf{v}_1, \mathbf{v}_2) = \int_\Omega \rho\mathbf{v}_1 \cdot \mathbf{v}_2\, dx$. We assume the eigenfunctions to be orthogonal to each other as well as normalized with respect to this inner product and take measurements of the displacements resulting from excitations $\mathbf{f}_i = f_i\rho\varphi_i$, $i = \{1, 2\}$

$$y^i(t) = B\mathbf{u}^i(t) \quad t \in (0, T) \quad i \in \{1, 2\},$$

e.g., $B\mathbf{v} = \mathbf{v}(x_0)$ for some boundary point x_0, or $B\mathbf{v} = \frac{1}{|\Gamma_0|}\int_{\Gamma_0} \mathbf{v}(x)\, dx$ for some boundary patch $\Gamma_0 \subseteq \partial\Omega$). Then testing (5.6) with φ_k, after integration by parts with respect to space and using the eigenvalue equation, as well as orthogonality $(\varphi_1, \varphi_2) = 0$ yields

$$u^i_{tt} + \lambda_i u^i + \lambda_i k_i * u^i_t = \ell(t)f_i \quad t \in (0, T) \quad i \in \{1, 2\}$$

for $u^i(t) = (\mathbf{u}(t), \varphi_i)$, and due to linearity we have $\mathbf{u}^{(i)}(t, x) = u^i(t)\varphi_i(x)$. Assuming that $T = \infty$, we can apply the Laplace transform to obtain

$$\mathcal{L}y^i(s) = \frac{f_i\, B\varphi_i\, \mathcal{L}\ell(s)}{s^2 + \lambda_i\, s\, \mathcal{L}k_i(s) + \lambda_i} \quad s \in \mathbb{C},$$

from which we further have

$$\mathcal{L}k_i(s) = \frac{1}{\lambda_i\, s}\left(\frac{f_i\, B\varphi_i\, \mathcal{L}\ell(s)}{\mathcal{L}y^i(s)} - s^2 - \lambda_i\right) \quad s \in \mathbb{C}.$$

By injectivity of the Laplace transform, this yields uniqueness of k_i, $i = \{1, 2\}$.

Of course, this setting is heavily idealized, but can be expected to be achieved – at least approximately – in the isotropic case (5.5) when excitation is done via longitudinal and shear forces, respectively. This is the setup we consider in our numerical experiments.

6. Numerical simulations

Let us consider a 3-dimensional rectangular viscoelastic beam of the size $1 \times 0.1 \times 0.04$. The beam is clamped on the left end and excited on the right during the time interval $[0, t_{load}]$ with $t_{load} = 0.8$ and then released. The final time is $T = 4$. In our experiments, we consider two types of load: bending in (x_1, x_2) plane and uniaxial extension (x_1-direction), respectively,

$$\mathbf{B}(t) = [0, \; B_0\, t/t_{load}, \; 0], \tag{6.1}$$

$$\mathbf{T}(t) = [T_0\, t/t_{load}, \; 0, \; 0], \tag{6.2}$$

where the magnitudes are $B_0 = 1$ and $T_0 = 100$. The Young's modulus is 10^3 and the Poisson ratio is 0.3. There is no external volumetric force. The beam is at rest at the initial time. Thus, the dynamical system is given by the balance equation (2.1) and the constitutive law (2.3):

$$\begin{cases} \rho\mathbf{u}_{tt} = \mathrm{div}\boldsymbol{\sigma}, \\ \boldsymbol{\sigma} = \mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \varepsilon_d(\mathbf{u}_t) + k_{\mathrm{tr}\,\varepsilon} * \mathbb{I}\mathrm{tr}\,\varepsilon(\mathbf{u}_t), \\ \mathbf{u}|_{x_1=0} = 0, \qquad \boldsymbol{\sigma} \cdot \mathbf{n}|_{x_1=1} = \mathbf{B} \text{ or } \mathbf{T}, \\ (\mathbf{u}, \mathbf{u}_t, \boldsymbol{\sigma}(\mathbf{u}))|_{t=0} = (0, 0, 0). \end{cases}$$
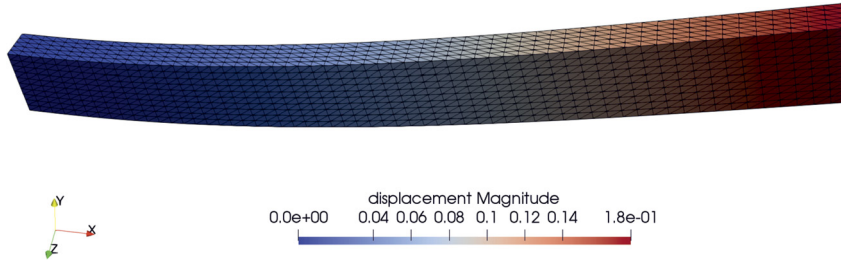
Fig. 1. 3D beam mesh. Color denotes the displacement magnitude at $t = t_{load}$ computed using the target kernel. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

We employ the Newmark method ($\beta = 0.25$, $\gamma = 0.5$) for time-stepping in combination with a sum-of-exponentials kernel representation; see, e.g., [36,37,47]. More precisely, we are looking for the kernels k_ε, $k_{\mathrm{tr}\varepsilon}$ in the form

$$k_\varepsilon(t) = \sum_{k=1}^{m_\varepsilon} w_{\varepsilon,k} e^{-\lambda_{\varepsilon,k} t}, \qquad k_{\mathrm{tr}\varepsilon}(t) = \sum_{k=1}^{m_{\mathrm{tr}\varepsilon}} w_{\mathrm{tr}\varepsilon,k} e^{-\lambda_{\mathrm{tr}\varepsilon,k} t},$$

with unknown weights $w_{\varepsilon,k}$, $w_{\mathrm{tr}\varepsilon,k}$ and exponents $\lambda_{\varepsilon,k}$, $\lambda_{\mathrm{tr}\varepsilon,k}$. For simplicity, we have fixed $m_\varepsilon = m_{\mathrm{tr}\varepsilon}$. Let us denote by $\theta := \{w_{\varepsilon,k}, w_{\mathrm{tr}\varepsilon,k}, \lambda_{\varepsilon,k}, \lambda_{\mathrm{tr}\varepsilon,k}\}$ the vector of all learnable parameters. We assume that the measurements for the tip displacement (more precisely, the mean displacement over the right end face) are available at some discrete time points in the interval $[0, t_{meas}]$ with $t_{meas} = T/2$. We minimize the objective function

$$J(\theta) = \frac{1}{2} \| u_{tip}(\theta) - u_{tip}^{meas} \|_{[0,t_{meas}]}^2,$$

with respect to the parameters θ, using the ℓ_2-norm on the discretized interval $[0, t_{meas}]$.

We use the FEniCS platform [38] for the finite elements implementation of the forward problem and the dolfin-adjoint package [39] for the adjoint problem in combination with PyTorch [40] through the Torch-FEniCS interface [41]. The minimization is performed using the LBFGS method [42] with the strong Wolfe line search. The beam is discretized using $60 \times 10 \times 5$ linear tetrahedral elements (Fig. 1). The time interval is discretized uniformly with the step size $\Delta t = T/100$.

Here, we have been assuming that the measurement data are given for all discrete time points of the time discretization scheme in the interval $[0, t_{meas}]$. Moreover, in our synthetic setting, we assume that for a given θ_*, it holds that

$$u_{tip}^{meas} = u_{tip}(\theta_*) + \eta,$$

where η is an additive white Gaussian noise. We consider in our numerical test cases different noise levels ranging from 2% to 8%.

First, we consider the case when $k_\varepsilon(t) = k_{\mathrm{tr}\varepsilon}(t) = k(t)$. We define the target kernels $k(t) = k^{true}(t)$ as sum-of-exponentials being an accurate (22 modes) approximation of the fractional kernel $t^{\alpha-1}/\Gamma(\alpha)$ with $\alpha = 0.7$. The expansion is obtained by a rational approximation of the Laplace spectrum using the AAA-algorithm [43]. Using this kernel, we numerically generate synthetic measurements of the tip displacement on the interval $[0, t_{meas}]$, using only bending type load (6.1). Then, we predict the kernel $k(t) = k^{pred}(t)$ only from these measurements. To do this, we infer the parameters $\theta = \{w_k, \lambda_k\}$ minimizing $J(\theta)$. We use a rational approximation (moderate accuracy with 8 modes) of the fractional kernel $t^{\alpha_0-1}/\Gamma(\alpha_0)$ with $\alpha_0 = 0.5$ as initial guess.

In Fig. 2, the resulting tip displacements are plotted on the complete interval $[0, T]$ for the models calibrated from noisy data with distinct noise levels. We observe that fitting the noisy data on a shorter interval quite accurately results in a solution slightly deviating from the truth if considered on a larger time interval. This can be explained by the fact that the measurement time interval $[0, t_{meas}]$ is not sufficiently large to fit accurately the tail of the kernel. Indeed, in Fig. 3, where the resulting kernels are compared with the target, we can see that the prediction fits the target at mid-range but is less accurate in the tail. Besides, we measured in $L^1([\Delta t, t_{meas}])$-norm the error between the target and the predicted kernels: 2% noise - 0.032207, 4% noise - 0.085839, 6% noise - 0.144768, 8% noise - 0.159670.

The convergence of the loss function during the optimization process is depicted in Fig. 4. There we can see that for all the noise level cases, the calibration process converges after 8 optimization steps. Besides, the minimum of the loss function grows with the noise level. The evolution of the energy in time is depicted in Fig. 5 for the case of a 2% noise level. We observe that the energy of the calibrated model fits accurately the truth on the measurement interval $[0, t_{meas}]$. Moreover, the calibrated kernel allows predicting the following evolution.

Next, we consider the case when $k_{\mathrm{tr}\varepsilon}^{true}(t)$ is given by a 22 modes rational approximation of $t^{\alpha_1-1}/\Gamma(\alpha_1)$ with $\alpha_1 = 0.9$, and $k_\varepsilon^{true}(t)$ is given by a 22 modes rational approximation of $t^{\alpha_2-1}/\Gamma(\alpha_2)$ with $\alpha_2 = 0.7$. We use a 8 modes rational approximation of $t^{\alpha_0-1}/\Gamma(\alpha_0)$ with $\alpha_0 = 0.5$ as initial guess for both kernels. Using this kernel, we numerically generate synthetic measurements of the tip displacement on the interval $[0, t_{meas}]$, using both loading types, bending (6.1) and
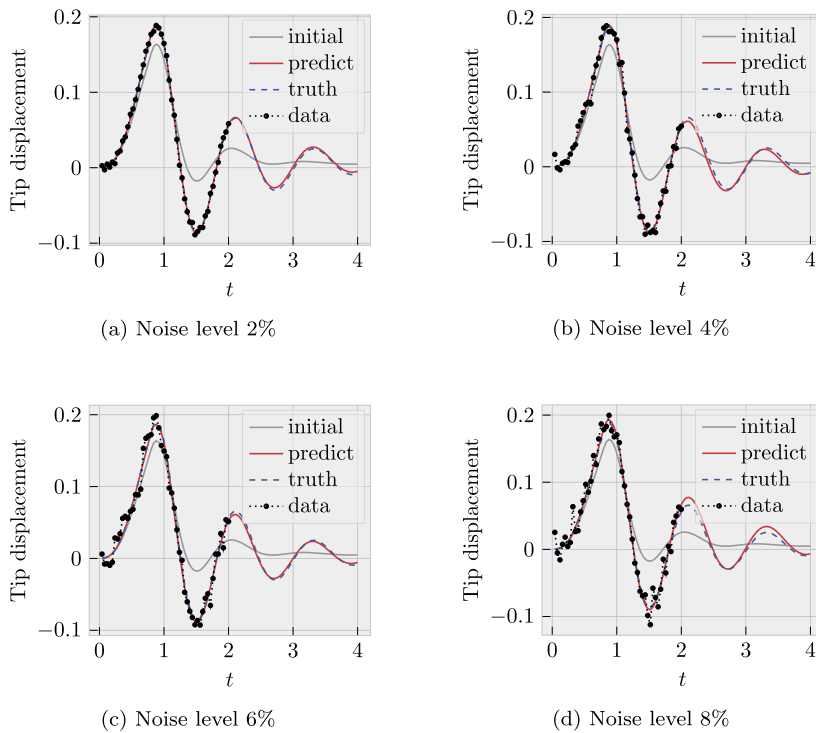
(a) Noise level 2%

(b) Noise level 4%

(c) Noise level 6%

(d) Noise level 8%

Fig. 2. Evolution of the tip displacement comparing the "true" model and the model calibrated from the noisy data with the noise level $2, 4, 6, 8\%$.
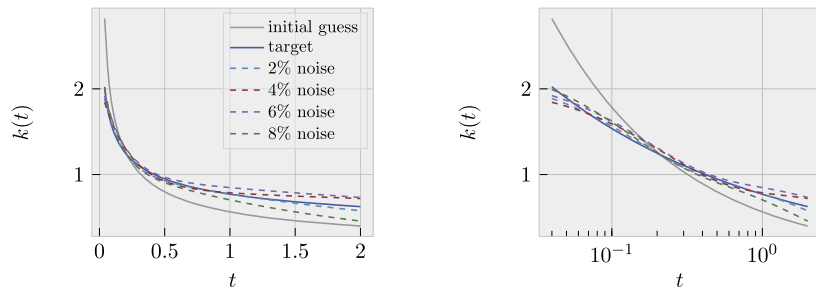

Fig. 3. Comparison of the resulting kernels calibrated using different noise level (right: log-scale).
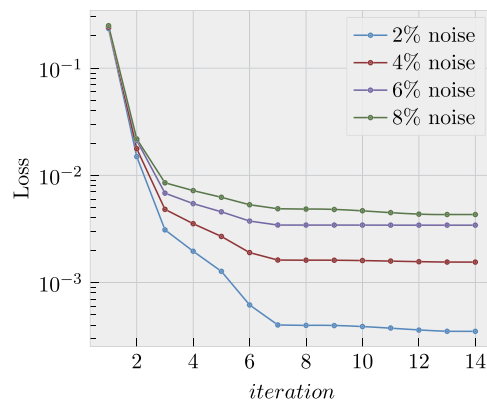

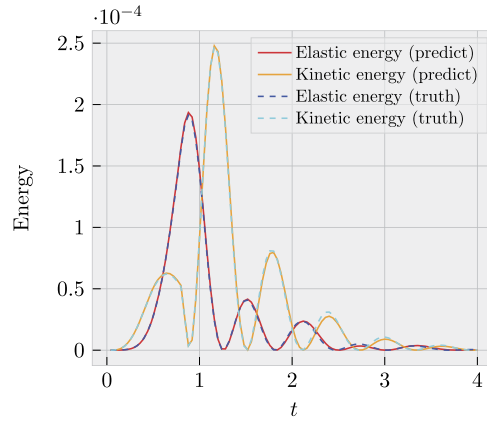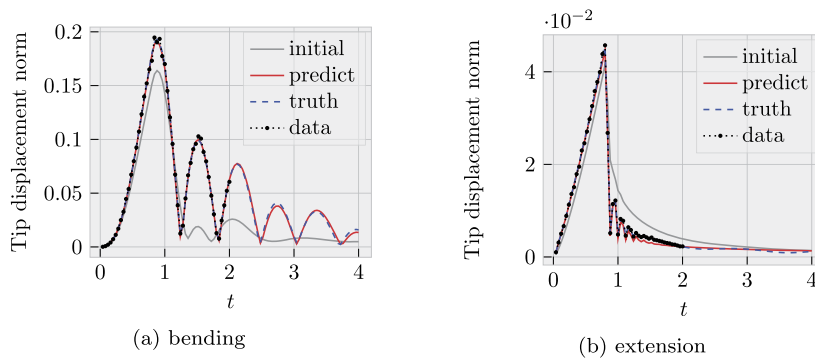**Fig. 4.** Convergence of the loss function during the optimization.

Fig. 5. Evolution of the kinetic and elastic energies for the "true" and calibrated models (noise level 2%).

(a) bending

(b) extension

Fig. 6. Evolution of the tip displacement comparing the "true" model and the model calibrated from the noisy data with the noise level 2%, using two loading types: (a) - bending, (b) - extension.

extension (6.2), and adding a noise of level 2%. Then, we predict the kernel $k_{\mathrm{tr}\,\varepsilon}^{pred}(t)$ and $k_{\varepsilon}^{pred}(t)$ from these measurements with the number of modes $m_{\mathrm{tr}\,\varepsilon} = m_{\varepsilon} = 8$. To do this, we infer the parameters $\theta = \{w_{\mathrm{tr}\,\varepsilon,k}, \lambda_{\mathrm{tr}\,\varepsilon,k}, w_{\varepsilon,k}, \lambda_{\varepsilon,k}\}$ minimizing the loss function $J(\theta)$ defined as

$$J(\theta) = \frac{1}{2} \frac{\|u_{tip}^{bend}(\theta) - u_{tip}^{bend,meas}\|^2_{[0,t_{meas}]}}{\|u_{tip}^{bend,meas}\|^2_{[0,t_{meas}]}} + \omega \cdot \frac{1}{2} \frac{\|u_{tip}^{ext}(\theta) - u_{tip}^{ext,meas}\|^2_{[0,t_{meas}]}}{\|u_{tip}^{ext,meas}\|^2_{[0,t_{meas}]}},$$

where the superscripts *bend* and *ext* correspond to the bending and extension excitation types, respectively; the norm is again the ℓ_2-norm on the discretized interval $[0, t_{meas}]$. Above, the weight ω is introduced to balance the contributions due to the bending and the extension terms. In our experiment, we fix $\omega = 10$ in order to take into account the tip displacement measurements under extension after the time t_{meas}, which have a scale of an order of magnitude smaller than the global scale $\|u_{tip}^{ext,meas}\|_{[0,t_{meas}]}$ (see Fig. 6b). In Fig. 6, the norm of the resulting tip displacements is plotted on the complete interval $[0, T]$ for the model calibrated from the noisy data. Again, we can see that the calibrated model can predict the evolution after the time t_{meas}. In Fig. 7, where the resulting kernels are compared with the target, we can observe that providing measurements for different loading types results in a good approximation of the integral kernels.

CRediT authorship contribution statement

Barbara Kaltenbacher: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Ustim Khristenko:** Software, Validation, Writing – original draft. **Vanja Nikolić:** Formal analysis, Writing – original draft, Writing – review & editing. **Mabel Lizzy Rajendran:** Formal analysis. **Barbara Wohlmuth:** Conceptualization, Methodology, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Barbara Kaltenbacher reports financial support was provided by Austrian Science Fund FWF under the
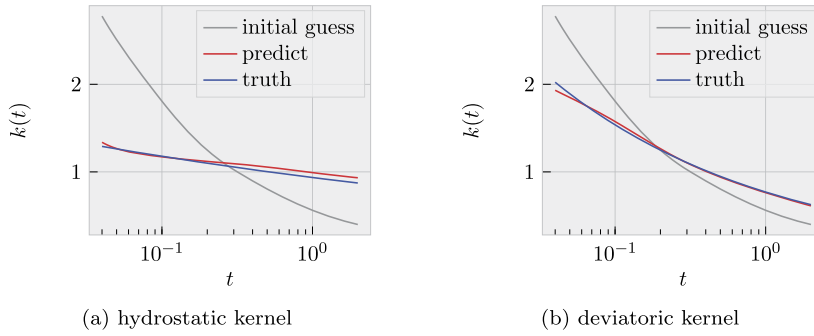
(a) hydrostatic kernel   (b) deviatoric kernel

Fig. 7. Comparison of the resulting kernels $k_{\mathrm{tr}\,\varepsilon}(t)$ (left) and $k_\varepsilon(t)$ (right) in the log-scale.

Acknowledgement

Appendix A. Regularity

We briefly discuss here the higher regularity of the displacement **u** in a setting with only Dirichlet boundary conditions, i.e., $\Gamma_N = \emptyset$ and $\partial\Omega$ is sufficiently smooth. More precisely, we present the formal derivation of a higher-order estimate in terms of **z** and note that additional bounds on **u** can be obtained by distinguishing the smooth and singular cases for k_σ, as in Propositions 3.1 and 3.2.

Abbreviating $\boldsymbol{w} = \mathbb{C}\varepsilon(\mathbf{z})$, $\boldsymbol{v} = \mathbb{D} * \varepsilon(\mathbf{u}_t)$ in (2.12) gives

$$\rho\mathbf{z}_{tt} - \mathrm{div}\,[\boldsymbol{w} + \boldsymbol{v}] = \mathbf{g} \tag{A.1}$$

where we recall that $\mathbb{D}$ is defined in (2.4) and should satisfy assumptions (2.7) and (2.8). Testing (A.1) with $-\mathrm{div}(\boldsymbol{w}_t + \boldsymbol{v}_t)$ after integration over $\Omega \times (0, t)$ yields

$$\frac{1}{2}\|\varrho^{1/2}\mathbb{C}^{1/2}\varepsilon(\mathbf{z}_t(t))\|^2_{L^2(\Omega)^d} + \frac{1}{2}\|\mathrm{div}(\boldsymbol{w} + \boldsymbol{v})(t)\|^2_{L^2(\Omega)^d}$$

$$+ \int_0^t \int_\Omega ((k_\sigma * \varepsilon(\mathbf{u}_t))_{tt} + \varepsilon(\mathbf{u}_{tt}))\,\mathbb{D} * \varepsilon(\mathbf{u}_{tt})\,\mathrm{d}x\mathrm{d}t$$

$$= -\int_0^t \int_\Omega \mathbf{g}_t(s)\mathrm{div}(\boldsymbol{w} + \boldsymbol{v})(s)\,\mathrm{d}x\mathrm{d}s + \int_\Omega \mathbf{g}(t)\mathrm{div}(w + v)(t)\,\mathrm{d}x$$

$$\leq 2\|\mathbf{g}_t\|^2_{L^1(0,t;L^2(\Omega)^d)} + \frac{1}{8}\|\mathrm{div}(\boldsymbol{w} + \boldsymbol{v})\|^2_{L^\infty(0,t;L^2(\Omega)^d)} + 2\|\mathbf{g}(t)\|^2_{L^2(\Omega)^d}$$

$$+ \frac{1}{8}\|\mathrm{div}(\boldsymbol{w} + \boldsymbol{v})(t)\|^2_{L^2(\Omega)^d},$$

where we have used $\boldsymbol{w}(0) = \boldsymbol{v}(0) = 0$. Note that the second and fourth term on the right-hand side can be absorbed into the second term on the left-hand side. Due to the assumptions on $\mathbb{D}$, we have

$$\int_0^t \int_\Omega ((k_\sigma * \varepsilon(\mathbf{u}_t))_{tt} + \varepsilon(\mathbf{u}_{tt}))\,\mathbb{D} * \varepsilon(\mathbf{u}_{tt})\,\mathrm{d}x\mathrm{d}t \geq 0.$$

Altogether, we obtain the estimate

$$\|\varepsilon(\mathbf{z}_t(t))\|^2_{L^2(\Omega)^d} + \|\mathrm{div}(\boldsymbol{w} + \boldsymbol{v})(t)\|^2_{L^2(\Omega)^d} \lesssim \|\mathbf{g}_t\|^2_{L^1(0,t;L^2(\Omega)^d)} + \|\mathbf{g}(t)\|^2_{L^2(\Omega)^d},$$

a.e. in time. The derived bounds allow us to conclude that

$$\varepsilon(\mathbf{u}_t) + (k_\sigma * \varepsilon(\mathbf{u}_t))_t \in L^\infty(0, T; L^2(\Omega)^{d^2})$$

and

$$\mathbb{C}\varepsilon(\mathbf{u}) + k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\mathrm{tr}\,\varepsilon} * \mathbb{I}\,\mathrm{tr}\,\varepsilon(\mathbf{u}_t) \in L^\infty(0, T; L^2(\Omega)^{d^2}),$$

where by using the $L^\infty(H^1)$ estimate on $\mathbf{u}$ that we obtain from both Propositions 3.1 and 3.2 we further have $k_\varepsilon * \mathbb{A}\varepsilon(\mathbf{u}_t) + k_{\mathrm{tr}\,\varepsilon} * \mathbb{I}\,\mathrm{tr}\,\varepsilon(\mathbf{u}_t) \in L^\infty(0, T; L^2(\Omega)^{d^2})$.

Appendix B. Auxiliary inequalities

Analogously to [44, Lemma 1], one can prove the following lower bound.

Lemma B.1. *Given a Hilbert space H and any $v \in H^1(0, T; H)$, $k \in L^p(0, T, H)$ for some $p > 1$, $k \geq 0$ a.e. such that for all $\epsilon > 0$, $k \in W^{1,1}(\epsilon, T, H)$ and $k'|_{[\epsilon, T]} \leq 0$ a.e.*

$$\int_0^T (v(t), k * v_t(t))_H \, dt \geq \frac{1}{2}(k * \|v\|^2)(T) - \frac{1}{2}\int_0^T k(t)\, dt\, \|v(0)\|^2. \tag{B.1}$$

Proof. Consider the sequence $(k_n)_{n\in\mathbb{N}} \subseteq W^{1,1}(0, T; H)$, $k_n \to k$ in $L^q(0, T; H)$ defined by

$$k_n(t) = 1_{[0,1/n)}(t)\|k|_{[1/n,T]}\|_{L^\infty([0,1/n])} + 1_{[1/n,T)}(t)k(t).$$

For each $n \in \mathbb{N}$, we have

$$(v(t), (k_n * v_t)(t))_H - \frac{1}{2}(k_n * \|v\|_t^2)(t) = \int_0^t (v(t) - v(s), k_n(t-s)v_t(s))_H \, ds$$

$$= \int_0^t \left(\int_s^t v_t(r)\, dr, k_n(t-s)v_t(s)\right)_H ds = \int_0^t \int_0^r (v_t(r), k_n(t-s)v_t(s))_H \, ds dr$$

$$= \int_0^t \frac{1}{k_n(t-r)}k_n(t-r)\left(v_t(r), \int_0^r k_n(t-s)v_t(s)\, ds\right)_H dr$$

$$= \frac{1}{2}\int_0^t \frac{1}{k_n(t-r)}\frac{d}{dr}\left\|\int_0^r k_n(t-s)v_t(s)\, ds\right\|^2 dr$$

$$= -\frac{1}{2}\int_0^t \frac{k_n'(x-r)}{k_n^2(x-r)}\left\|\int_0^r k_n(t-s)v_t(s)\, ds\right\|^2 dr + \frac{1}{2}\left[\frac{1}{k_n(t-r)}\left\|\int_0^r k_n(t-s)v_t(s)\, ds\right\|^2\right]_0^t$$

$$\geq 0.$$

After using (2.13) and integrating $\int_0^T \cdot\, dt$, this yields (B.1) with k_n in place of k. Then the assertion for k follows from the fact that $v \in W^{1,1}(0, T; H)$, where $W^{1,1}(0, T; H)$ continuously embeds into $L^{2p^*}(0, T; H)$ as well as the estimates

$$\left|\int_0^T (v(t), ((k - k_n) * v_t)(t))_H \, dt\right| \leq \|v\|_{L^p(0,T;H)}\|(k - k_n) * v_t\|_{L^{p^*}(0,T;H)}$$

$$\leq \|v\|_{L^p(0,T;H)}\|k - k_n\|_{L^{p^*}(0,T)}\|v_t\|_{L^1(0,T;H)}$$

and

$$|((k - k_n) * \|v\|^2)(T)| \leq \|k - k_n\|_{L^p(0,T)}\|v\|^2_{L^{2p^*}(0,T;H)},$$

where we have used Hölder's and Young's convolution inequalities. $\quad\square$

Similarly to [45, Lemma 2.3] (see also [46, Theorem 1]), one obtains the following coercivity estimate.

Lemma B.2. *For some Hilbert space H and any* $v \in H^{-\delta/2}(0, T; H)$ *and* $k \in L^1(0, T)$ *with* $\Re(\mathcal{F}k)(\omega) \geq \gamma (1 + \omega^2)^{-\delta/2}$ *for* $\omega \in \mathbb{R}$,

$$\int_0^T (v(s), k * v(s))_H \, ds \geq \gamma \|v\|^2_{H^{-\delta/2}(0,T;X)}.$$

Proof. We first assume $v \in L^2(0, T; H)$ and approximate $k * v$ by $k_\epsilon * v$ with $k_\epsilon(t) = k_\epsilon(t)e^{-\epsilon t}$ for $\epsilon > 0$ and extend $v \in L^2(0, T)$ to all of $\mathbb{R}$ by zero. Plancherel's Theorem and the Convolution Theorem then yield

$$\int_0^t (v(s), (k_\epsilon * v)(s))_H \, ds = \int_{\mathbb{R}} (\mathcal{F}v(\omega), \overline{\mathcal{F}[k_\epsilon * v]}(\omega))_H \, d\omega$$

$$= \int_{\mathbb{R}} \|\mathcal{F}v(\omega)\|^2 \overline{\mathcal{F}k_\epsilon}(\omega) \, d\omega.$$

Here both the left-hand side is real valued because of real-valuedness of k_ϵ and v and therefore also the expression at the end of this chain of identities has to be real valued. Therefore,

$$\int_0^t (v(s), (k_\epsilon * v)(s))_H \, ds = \int_{\mathbb{R}} \|\mathcal{F}v(\omega)\|^2 \Re(\mathcal{F}k_\epsilon)(\omega) \, d\omega.$$

Letting ϵ tend to zero and recalling that $\|v\|^2_{H^{-\delta/2}(\mathbb{R})} = \int_{\mathbb{R}} |\mathcal{F}v(\omega)|^2 (1 + \omega^2)^{-\delta/2} \, d\omega$, now the assertion follows for general $v \in H^{-\delta/2}(0, T; H)$ by density arguments. $\square$

References

[1] R. Sinkus, K. Siegmann, T. Xydeas, M. Tanter, C. Claussen, M. Fink, MR elastography of breast lesions: understanding the solid/liquid duality can improve the specificity of contrast-enhanced MR mammography, Magn. Reson. Med. 58 (6) (2007) 1135–1144.

[2] S. Holm, S.P. Näsholm, Comparison of fractional wave equations for power law attenuation in ultrasound and elastography, Ultrasound Med. Biol. 40 (4) (2014) 695–703.

[3] D. Klatt, U. Hamhaber, P. Asbach, J. Braun, I. Sack, Noninvasive assessment of the rheological behavior of human organs using multifrequency mr elastography: a study of brain and liver viscoelasticity, Phys. Med. Biol. 52 (24) (2007) 7281.

[4] S. Buterin, The inverse problem of recovering the Volterra convolution operator from the incomplete spectrum of its rank-one perturbation, Inverse Probl. 22 (6) (2006) 2223.

[5] D.K. Durdiev, Z.D. Totieva, The problem of determining the one-dimensional kernel of viscoelasticity equation with a source of explosive type, J. Inverse Ill-Posed Probl. 28 (1) (2020) 43–52.

[6] J. Janno, L. Von Wolfersdorf, Inverse problems for identification of memory kernels in viscoelasticity, Math. Methods Appl. Sci. 20 (4) (1997) 291–314.

[7] M. Slodička, L. Šeliga, Identification of memory kernels in hyperbolic problems, J. Comput. Appl. Math. 311 (C) (2017) 618–629, https://doi.org/10.1016/j.cam.2016.08.015.

[8] L. Von Wolfersdorf, B. Hofmann, A specific inverse problem for the Volterra convolution equation, Appl. Anal. 87 (1) (2008) 59–81.

[9] Y. Hatano, J. Nakagawa, S. Wang, M. Yamamoto, Determination of order in fractional diffusion equation, J. Math. Ind. 5 (A) (2013) 51.

[10] B. Jin, Y. Kian, Recovery of the order of derivation for fractional diffusion equations in an unknown medium, SIAM J. Appl. Math. (2021).

[11] Z. Li, M. Yamamoto, Uniqueness for inverse problems of determining orders of multi-term time-fractional derivatives of diffusion equation, Appl. Anal. 94 (3) (2015) 570–579, https://doi.org/10.1080/00036811.2014.926335.

[12] G. Li, D. Zhang, X. Jia, M. Yamamoto, Simultaneous inversion for the space-dependent diffusion coefficient and the fractional order in the time-fractional diffusion equation, Inverse Probl. 29 (6) (2013) 065014, https://doi.org/10.1088/0266-5611/29/6/065014.

[13] W. Rundell, Z. Zhang, Fractional diffusion: recovering the distributed fractional derivative from overposed data, Inverse Probl. 33 (3) (2017) 035008, https://doi.org/10.1088/1361-6420/aa573e.

[14] B. Kaltenbacher, W. Rundell, On an inverse problem of nonlinear imaging with fractional damping, Math. Comput. (2021), https://doi.org/10.1090/mcom/3683.

[15] G.C. Gorain, Stabilization for the vibrations modeled by the 'standard linear model' of viscoelasticity, Proc. Math. Sci. 120 (4) (2010) 495–506.

[16] M. Pellicer, J. Sola-Morales, Optimal scalar products in the Moore–Gibson–Thompson equation, Evol. Equ. Control Theory 8 (1) (2019) 203.

[17] M. Conti, V. Pata, M. Pellicer, R. Quintanilla, On the analyticity of the MGT-viscoelastic plate with heat conduction, J. Differ. Equ. 269 (10) (2020) 7862–7880.

[18] M. Enelund, B.L. Josefson, Time-domain finite element analysis of viscoelastic structures with fractional derivatives constitutive relations, AIAA J. 35 (10) (1997) 1630–1637.

[19] F. Saedpanah, Well-posedness of an integro-differential equation with positive type kernels modeling fractional order viscoelasticity, Eur. J. Mech. A, Solids 44 (2014) 201–211.

[20] L. Oparnica, E. Süli, Well-posedness of the fractional Zener wave equation for heterogeneous viscoelastic materials, Fract. Calc. Appl. Anal. 23 (1) (2020) 126–166.

[21] G. Alotta, O. Barrera, A.C. Cocks, M. Di Paola, On the behavior of a three-dimensional fractional viscoelastic constitutive model, Meccanica 52 (9) (2017) 2127–2142.

[22] A. Schmidt, L. Gaul, Finite element formulation of viscoelastic constitutive equations using fractional time derivatives, Nonlinear Dyn. 29 (1) (2002) 37–55.

[23] H.W. Engl, K. Kunisch, A. Neubauer, Convergence rates for Tikhonov regularisation of non-linear ill-posed problems, Inverse Probl. 5 (1989) 523–540, https://doi.org/10.1088/0266-5611/5/4/008.

[24] T.I. Seidman, C.R. Vogel, Well posedness and convergence of some regularisation methods for non-linear ill posed problems, Inverse Probl. 5 (2) (1989) 227, https://doi.org/10.1088/0266-5611/5/2/008.

[25] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, Wiley, New York, 1977.

[26] G. Gripenberg, S.-O. Londen, O. Staffans, Volterra Integral and Functional Equations, Cambridge University Press, 1990.

[27] L.C. Evans, Partial Differential Equations, American Mathematical Society, 2010.

[28] A.A. Kilbas, H.M. Srivastava, J.J. Trujillo, Theory and Applications of Fractional Differential Equations, North-Holland Mathematics Studies, vol. 204, Elsevier Science B.V., Amsterdam, 2006.

[29] I. Podlubny, Fractional Differential Equations, Academic Press, Inc., San Diego, CA, 1999.

[30] W. Feller, An Introduction to Probability Theory and Its Applications, vol. II, 2nd edition, John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[31] R. Gorenflo, M. Yamamoto, Operator-theoretic treatment of linear Abel integral equations of first kind, Jpn. J. Ind. Appl. Math. 16 (1) (1999) 137–161, https://doi.org/10.1007/BF03167528.

[32] G. Leoni, A First Course in Sobolev Spaces, Graduate Studies in Mathematics, vol. 105, American Mathematical Society, Providence, RI, 2009.

[33] M. Burger, S. Osher, Convergence rates of convex variational regularization, Inverse Probl. 20 (5) (2004) 1411–1421, https://doi.org/10.1088/0266-5611/20/5/005.

[34] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, F. Lenzen, Variational Methods in Imaging, Applied Mathematical Sciences, vol. 167, Springer, New York, 2009.

[35] T. Schuster, B. Kaltenbacher, B. Hofmann, K.S. Kazimierski, Regularization Methods in Banach Spaces, Walter de Gruyter, Berlin, 2012.

[36] P. Lam, H. So, C. Chan, Exponential sum approximation for Mittag-Leffler function and its application to fractional Zener wave equation, J. Comput. Phys. 410 (2020) 109389.

[37] J. Thiel, Numerical solution of the time-fractional wave equation in viscoelastic media, BSc thesis, Technical University of Munich..

[38] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. Rognes, G. Wells, Archive of Numerical Software: The FEniCS Project Version 1.5, University Library Heidelberg, 2015.

[39] S.K. Mitusch, S.W. Funke, J.S. Dokken, dolfin-adjoint 2018.1: automated adjoints for FEniCS and Firedrake, J. Open Sour. Softw. 4 (38) (2019) 1292, https://doi.org/10.21105/joss.01292.

[40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: an imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems, vol. 32, Curran Associates, Inc., 2019, pp. 8024–8035.

[41] P. Barkman, J. Moss, Torch-FEniCS, https://github.com/barkm/torch-fenics.

[42] S. Wright, J. Nocedal, Numerical optimization, Springer Sci. 35 (67–68) (1999) 7.

[43] Y. Nakatsukasa, O. Sète, L.N. Trefethen, The AAA algorithm for rational approximation, SIAM J. Sci. Comput. 40 (3) (2018) A1494–A1522.

[44] A. Alikhanov, A priori estimates for solutions of boundary value problems for fractional-order equations, Differ. Equ. 46 (5) (2010) 660–666.

[45] P.P.B. Eggermont, On Galerkin methods for Abel-type integral equations, SIAM J. Numer. Anal. 25 (1988) 1093–1117, https://doi.org/10.1137/0725063.

[46] U. Vögeli, K. Nedaiasl, S.A. Sauter, A fully discrete Galerkin method for Abel-type integral equations, Adv. Comput. Math. 44 (5) (2018) 1601–1626, https://doi.org/10.1007/s10444-018-9598-4.

[47] U. Khristenko, B. Wohlmuth, Solving time-fractional differential equation via rational approximation, IMA J. Numer. Anal., https://doi.org/10.1093/imanum/drac022.