

A CONTINUOUS ADJOINT FOR PHOTOACOUSTIC TOMOGRAPHY OF THE BRAIN

ASHKAN JAVAHERIAN

SEAN HOLMAN

*School of Mathematics, University of Manchester, M13 9PL, UK
e-mail: sean.holman@manchester.ac.uk*

ABSTRACT

We present an optimisation framework for photo-acoustic tomography of brain based on a system of coupled equations that describe the propagation of sound waves in linear isotropic inhomogeneous and lossy elastic media with the absorption and physical dispersion following a frequency power law using fractional Laplacian operators. The adjoint of the associated continuous forward operator is derived, and a numerical framework for computing this adjoint based on a k-space pseudo-spectral method is presented. We analytically show that the derived continuous adjoint matches the adjoint of an associated discretised forward operator. We include this adjoint in a first-order positivity constrained optimisation algorithm that is regularised by total variation minimisation, and show that the iterates monotonically converge to a minimiser of an objective function, even in the presence of some error in estimating the physical parameters of the medium.

1. INTRODUCTION

Quantitative Photo-acoustic Tomography (QPAT) is a hybrid imaging modality which simultaneously takes advantage of the rich contrast attributed to optical imaging and the high spatial resolution brought up by ultrasound. In this technique, short pulses of near-infrared light are used to irradiate tissue. The energy from these pulses is absorbed as a function of the optical absorption map of the tissue. This generates local increases in pressure which propagate outwards as photo-acoustic (PA) waves, and are then measured by broadband detectors placed at the surface. The inverse problem of QPAT is to reconstruct the spatially varying optical absorption coefficient from the recorded PA signals. This involves two inverse problems, namely acoustic and optical [29]. The acoustic inverse problem is to reconstruct the initial pressure distribution from the recorded PA signals, and the optical inverse problem is to recover optical coefficients from the computed initial pressure map [29]. These two inverse problems can be solved distinctly [16, 2, 17], or alternatively as a direct hybrid problem [9]. In this work we consider only the acoustic portion of the inverse problem which we simply call Photo-acoustic Tomography (PAT).

Considering the acoustic inverse problem, Time reversal (TR) is a comprehensive inversion approach for PAT since it can be used for media with heterogeneous acoustic properties and arbitrary detection geometries [14, 13, 39]. However, this method is based on a continuous domain with idealised conditions such as a closed detection surface or exactly known medium's properties [14],

which do not hold in real cases. Problems such as finite sampling, a limited accessible angle for detection surface, errors in estimation of medium’s properties, or errors in data measurement make the acoustic inverse problem ill-posed [16]. In these cases, model-based iterative methods are often used, e.g., TR-based iterative techniques [28], or optimisation algorithms [16, 2, 17]. The optimisation approaches are often based on computation of the gradient of an objective function in terms of a forward model and a corresponding adjoint model. Using an explicit formulation of the adjoint operator, the convergence of an arising iterative algorithm was established in Hilbert spaces using a Landweber algorithm for media with inhomogeneous sound speed [4, 32], and was later improved using Nesterov’s fast gradient and the CG methods [10].

Because of the dependance of shape, spectrum and amplitude of PA signals on physical properties of tissue media, it will be advantageous if the image reconstruction in PAT is enriched by tissue-realistic models that account for the absorption behaviours evident in tissues [36, 8, 30]. Among model-based iterative approaches for absorbing media, the adjoint was computed by a “discretise-then-adjoint” method in [16], or by an “adjoint-then-discretise” method in [17].

It is well-known that modelling the propagation of sound waves can be considerably expedited compared to Finite difference time-domain (FDTD) methods by using Pseudo-spectral time-domain (PSTD) methods. Applying these techniques, the spatial gradients are computed in frequency domain, while the temporal gradients are computed using finite difference methods, similar to FDTD techniques. The efficiency of PSTD methods is because of a fast computation of the spatial gradients using Fast Fourier Transforms (FFTs), as well as a dramatic relaxation in the mesh requirement and time step [5, 6].

PAT has shown its potential for characterization of the vasculature in small animals or within a few mm of the skin’s surface in humans [43]. Furthermore, PAT has been utilised successfully for transcranial brain imaging in small animals [42, 23]. In these cases, the effect of the skull on the propagation of PA waves is neglected because of the low thickness of the skull (≈ 1 mm), and thus the image reconstruction is done based on scalar acoustic wave equations [42, 23].

To account for aberration of PA signals because of the heterogeneous properties of the skull, a subject-specific imaging model was proposed, where the inhomogeneity of the skull is taken into account using adjunct information about the skull anatomy and composition [15]. This information must be obtained from x-ray computed tomography image data, or some other imaging modalities.

The application of PAT in transcranial brain imaging of humans is very limited since PA signals are aberrated to a high degree by absorption, scattering and compressional-to-shear mode conversion effects due to the high thickness of the skull (4mm-7mm). Similar to the scalar acoustic problem, the iterative methods for transcranial brain imaging can be categorised into time-reversal (convergent Neumann-series) [31, 20], or optimisation algorithms. Recently, an optimisation framework for transcranial PAT was proposed, where the forward problem describes the wave propagation in a linear isotropic, heterogeneous and lossy elastic medium, and a corresponding adjoint model is algebraically derived from the discretised forward operator, i.e., *discretise-then-adjoint* method [25]. This forward and adjoint pair was approximated by a finite-difference time domain (FDTD) method using a fourth-order FD method for computing the spatial gradient of fields. In addition, attenuation was described by a diffusive model [27], which does not account for the dependency of the wavefield attenuation on frequency, and assumes that the shear-to-compressional wave absorption ratio is proportional to the compressional-to-shear velocity ratio. The authors mentioned that this absorption model is accurate on the condition that the induced PA waves are sufficiently band-limited (cf. [25], section 2.1).

In PAT, the compartmentalised distribution of light absorbing molecules composing tissues induces step-like discontinuities in the generated pressure field. As a result, the generated PA waves

are considerably more broadband than ultrasonic waves [39, 17]. Furthermore, the absorption of sound waves in many media such as tissues has been experimentally shown to follow a frequency power law with a non-integer power, which can be described by fractional derivatives [36, 37]. Classical attenuation models used the fractional time derivatives, which are non-local in time, and thus require storing the time history of field variables [24]. It has been established that the fractional time derivatives can be replaced by fractional space derivatives, which are nonlocal in space rather than time, and are thus more memory efficient [36, 37]. This is done using the dispersion relation for lossless wave equation. The cost of this method is that the spatially non-local operators violate causality [22].

In elastic solids, compressional and shear waves propagate at different speeds. As a result, using the dispersion relation for describing fractional space derivatives requires splitting the field variables into compressional and shear parts [37]. This is done using a dyadic wave number tensor in the frequency domain [6, 37]. Additionally, by splitting the fields, the numerical dispersion errors accumulated by the time integrations can also be minimised via applying the k-space correction to the spatial gradients, which allows larger time steps without loss of stability or accuracy in heterogeneous media [34, 6].

Contribution. We consider a forward map in the PAT problem in which a system of coupled first-order equations describes the propagation of PA waves in linear isotropic, heterogeneous and lossy elastic media, where the absorption and physical dispersion follow a frequency-power law. We derive the adjoint of the PAT forward map in this context. This adjoint, referred to here as the *analytic viscoelastic adjoint*, is derived on a continuous domain, and is in the form of a system of partial differential equations. We shall analytically show that a numerical computation of the derived *analytic viscoelastic adjoint* using the k-space pseudo-spectral method matches the algebraic adjoint of the associated forward model. By setting viscosity coefficients to zero in our derived analytic (continuous) adjoint, the general form of the adjoint model for lossless media is derived. This can be used as a basic model, when other existing attenuation models are considered, e.g. [22, 21]. We emphasise that in the absence of attenuation the difference between our adjoint model and that of [25] arises from the different methods of discretisation of the spatial gradient. Our derived analytic adjoint including attenuation effects is numerically validated using an adjoint test, and then the forward and adjoint pair is included in a positivity constrained and total-variation regularised solver based on the Iterative Shrinkage Thresholding algorithm (ISTA) [3] for image reconstruction in 2D and 3D scenarios.

2. ACOUSTIC WAVE PROPAGATION FOR VISCOELASTIC MEDIA

Here, we briefly review the model we use for the propagation of acoustic waves in viscoelastic media. For further details, the reader is referred to [37]. To do this, we start with defining the associated fields, which are the particle displacement vector $u_i(x, t)$, particle velocity vector $v_i(x, t) = du_i(x, t)/dt$, scalar pressure $p(x, t)$ and stress tensor $\sigma_{ij}(x, t)$, where x and t denote the position and time, respectively. We also define the medium's parameters, which are the Lamé elastic parameters $\mu(x)$ and $\lambda(x)$, and are related to the shear and compressional wave speeds, $c_s(x)$ and $c_p(x)$ respectively, by the equations

$$\mu = \rho c_s^2, \quad \lambda = \rho c_p^2 - 2\mu, \quad (1)$$

where $\rho(x)$ denotes the medium's mass density (See [6]). In this section, for brevity the dependence on x and t is neglected, and following [6] and [37], we use the Einstein summation notation. We

also define the strain tensor field as [6]

$$\epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right). \quad (2)$$

Using these, the deformation in an isotropic lossless elastic medium is described by the relation between σ and ϵ in the form

$$\sigma_{ij} = \lambda \delta_{ij} \epsilon_{ll} + 2\mu \epsilon_{ij}. \quad (3)$$

One method of incorporating attenuation is to modify the stress-strain relation (3) as

$$\sigma_{ij} = \lambda \delta_{ij} \epsilon_{ll} + 2\mu \epsilon_{ij} + \chi \delta_{ij} \frac{\partial}{\partial t} \epsilon_{ll} + 2\eta \frac{\partial}{\partial t} \epsilon_{ij}, \quad (4)$$

where χ and η denote the compressional and shear viscosity coefficients. The equation (4) is called Kelvin-Voigt model. Plugging (2) into (4) gives [37]

$$\frac{\partial \sigma_{ij}}{\partial t} = \lambda \delta_{ij} \frac{\partial v_l}{\partial x_l} + \mu \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) + \chi \delta_{ij} \frac{\partial^2 v_l}{\partial x_l \partial t} + \eta \left(\frac{\partial^2 v_i}{\partial x_j \partial t} + \frac{\partial^2 v_j}{\partial x_i \partial t} \right). \quad (5)$$

Let us define the temporal frequency by ω . For low-frequency ranges, the Kelvin-Voigt model describes an acoustic absorption proportional to ω^2 and a constant sound speed (no dispersion), whereas for high frequency ranges, both the absorption and dispersion vary proportional to $\omega^{1/2}$ [37].

As discussed in section 1, experimental studies have shown that attenuation in many materials of interests, including tissue media such as bone, is proportional to ω^y with y a non-integer between 0 and 2 [33]. Because of the broadband nature of PA signals, as well as the high level of the attenuation in the skull, this behaviour cannot be neglected. To account for the non-integer power law dependence, the integer temporal derivatives in equations (4) and (5) can be replaced by fractional time derivatives [12]. For an isotropic medium, this gives the fractional Kelvin-Voigt model in the form [37]

$$\sigma_{ij} = \lambda \delta_{ij} \epsilon_{ll} + 2\mu \epsilon_{ij} + \chi \delta_{ij} \frac{\partial^{y-1}}{\partial t^{y-1}} \epsilon_{ll} + 2\eta \frac{\partial^{y-1}}{\partial t^{y-1}} \epsilon_{ij}, \quad (6)$$

where

$$\eta = -\frac{2\rho c_s^3}{\cos(\pi y/2)} \alpha_{0,s}, \quad \chi = -\frac{2\rho c_p^3}{\cos(\pi y/2)} \alpha_{0,p} - 2\eta, \quad (7)$$

with $\alpha_{0,s}$ and $\alpha_{0,p}$, respectively the attenuation coefficients pertaining to shear and compressional waves in $\text{Np}(\text{rad/s})^{-y} \text{m}^{-1}$ [37].

The temporal fractional derivatives in equation (6) are non-local in time, and thus their numerical computation requires the storage of the time history of fields, which is very computationally expensive. To overcome this problem, the dispersion formula for lossless media, i.e., the relation between ω and spatial frequency k ($\omega \approx ck$) with c the sound speed, is used to replace the fractional time derivatives by fractional space derivatives, which are non-local in space, rather than time [37]. Using this method, the fractional time derivative is written as two fractional Laplacian operators [36, 37].

This method provides a significant computational memory benefit, since at each time step the wavefield at all spatial positions is readily accessible. The system of viscoelastic wave equations enriched by fractional Laplacian operators can be used to describe absorption and physical dispersion behaviours over a wide range of frequencies and absorption values [37]. However, in elastic media since the compressional and shear waves travel at different speeds, separate dispersion relations

must be considered for the compressional and shear parts of the wavefield. This requires that the particle velocity field is split into the compressional and shear components [37]. Throughout this work, superscripts p and s denote the compressional and shear parts of the fields, respectively. v_i^p and v_i^s are calculated in the form

$$\begin{aligned} v_i^p &= q^p(v_i) = F^{-1} \left\{ \hat{k}_i \hat{k}_j F \{ v_j \} \right\} \\ v_i^s &= q^s(v_i) = F^{-1} \left\{ (\delta_{ij} - \hat{k}_i \hat{k}_j) F \{ v_j \} \right\}, \end{aligned} \quad (8)$$

where F represents the Fourier transform operator, and $\hat{k}_i \hat{k}_j$ is the unit dyadic wavenumber tensor with $\hat{k}_i = k_i/k$ the normalised wavenumber in direction i and $k = (\sum_i k_i^2)^{1/2}$ the magnitude of wavenumber. By splitting the particle velocity vector, the stress tensor is updated distinctly for compressional and shear parts in the form

$$\begin{aligned} \frac{\partial \sigma_{ij}^{p,s}}{\partial t} &= \lambda \left(\delta_{ij} \frac{\partial}{\partial x_l} v_l^{p,s} \right) + \mu \left(\frac{\partial}{\partial x_j} v_i^{p,s} + \frac{\partial}{\partial x_i} v_j^{p,s} \right) + \chi \left(\delta_{ij} \frac{\partial}{\partial x_l} \frac{\partial^{y-1}}{\partial t^{y-1}} v_l^{p,s} \right) \\ &+ \eta \left(\frac{\partial}{\partial x_j} \frac{\partial^{y-1}}{\partial t^{y-1}} v_i^{p,s} + \frac{\partial}{\partial x_i} \frac{\partial^{y-1}}{\partial t^{y-1}} v_j^{p,s} \right). \end{aligned} \quad (9)$$

Now, the fractional temporal derivatives in (9) can be replaced by fractional Laplacian operators with different sound speed for the compressional and shear waves. To make this replacement, we introduce the operator

$$L_{c_{p,s}}^y = c_{p,s}^y (-\nabla^2)^{y/2}. \quad (10)$$

Using this notation and the comments above, we will be using the following definition throughout the rest of this work including in (9) [37]

$$\frac{\partial^{y-1}}{\partial t^{y-1}} v_i^{p,s} = \sin(\pi y/2) L_{c_{p,s}}^{y-1} v_i^{p,s} - \cos(\pi y/2) L_{c_{p,s}}^{y-2} \frac{1}{\rho} \frac{\partial}{\partial x_j} \sigma_{ij}^{p,s}. \quad (11)$$

We will make this replacement in our forward model. The continuous forward model for the wave propagation is completed with the conservation of momentum

$$\frac{\partial v_i}{\partial t} = \frac{1}{\rho} \sum_{p,s} \frac{\partial}{\partial x_j} \sigma_{ij}^{p,s}. \quad (12)$$

Note that in the second term of (11), following [37] $\partial v_i^{p,s}/\partial t$ has been replaced by the split terms of (12). This is done to avoid having to compute time differences in the discretised model (cf. [37], section III).

Equations (9) and (12), together give a system of coupled partial differential equations which describe the propagation of acoustic waves in linear isotropic, heterogeneous and lossy elastic media with an attenuation following the frequency power law.

3. CONTINUOUS FORWARD AND ADJOINT OPERATORS

Let $\Omega \subset \mathbb{R}^d$ be a d -dimensional open, bounded set containing the initial pressure. For describing the measurement done by detectors, we define the operator \mathcal{G} , which maps the compressional part of the stress tensor field to the pressure field in the form

$$\mathcal{G}\sigma^p(x, t) = -\frac{1}{d} \delta_{ij} \sigma_{ij}^p(x, t) = p(x, t). \quad (13)$$

We also introduce the measurement operator $\mathcal{M} : \mathcal{L}^2(\mathbb{R}^d) \rightarrow \mathbb{R}^{N_s N_t}$ which maps the pressure field p to the measurements detected at each of N_t instants in time and N_s detectors. Each detector has a limited access to the pressure field, i.e., it measures the pressure field over a small, finite volume of space, and the measurement takes a finite time T . In the next definition we give the forward map for our inverse problem.

Definition 1. *We define the PAT forward operator using the viscoelastic model in the form*

$$\begin{aligned} \Lambda : C_0^\infty(\Omega) &\rightarrow \mathbb{R}^{N_s N_t} \\ \Lambda[p_0](x, t) &= \mathcal{M}\mathcal{G}\sigma^p(x, t), \end{aligned} \quad (14)$$

where $\sigma_{ij}^{p,s}$ and v_i satisfy (9) and (12) with initial conditions

$$\sigma_{ij}^p(x, 0) = -\delta_{ij}p_0(x), \quad \sigma_{ij}^s(x, 0) = 0, \quad v_i(x, 0) = 0. \quad (15)$$

The inverse problem is to reconstruct an approximation of p_0 given $\Lambda[p_0]$. The forward operator Λ is well-defined under fairly basic conditions such as $c_p > 0$, $c_s > 0$, $\rho > 0$, $y \in (0, 2) \setminus \{1\}$, $\alpha_{0,s} > 0$, $\alpha_{0,p} > 0$, and all of the fields c_p , c_s , ρ , $\alpha_{0,s}$, and $\alpha_{0,p}$ infinitely differentiable (it is likely that weaker hypotheses such as $c_s \geq 0$, or some of the fields non-smooth can be used as well). Intuitively the forward map is well-defined as it models a physical process, and for more detail on this modelling we refer to [37]. Mathematically the given model for the forward map can be proven to be well-defined by considering the system of equations formed by (9) and (12) as a linear evolution equation for v , σ^p , and σ^s in the form

$$\frac{\partial}{\partial t} \begin{pmatrix} v \\ \sigma^p \\ \sigma^s \end{pmatrix} = \mathcal{A} \begin{pmatrix} v \\ \sigma^p \\ \sigma^s \end{pmatrix}$$

for a linear operator \mathcal{A} initially defined on $C_0^\infty(\mathbb{R}^n)$ functions. Considering \mathcal{A} as an unbounded linear operator on the appropriate \mathcal{L}^2 space with domain that depends on y , we can apply the Hille-Phillips Theorem [11] to show that \mathcal{A} generates a quasi-contractive semi-group, which is what describes the evolution of the fields v , σ^p , and σ^s . The main step in the application of this theorem is to prove that for $\gamma > 0$ sufficiently large

$$\gamma - \mathcal{A}$$

maps from the domain of \mathcal{A} onto the \mathcal{L}^2 space. It is in this step that the requirements on the positivity of the parameters arise, as certain bilinear forms must be shown to be coercive. Finally, the measurement operator \mathcal{M} must be chosen to be an appropriate linear functional on $\mathcal{L}^2(\Omega)$, although this may also be relaxed to allow sampling individual point values if p_0 is assumed to be smooth.

In the next Lemma, we will calculate the adjoint of Λ with respect to the \mathcal{L}^2 inner product, i.e.,

$$\Lambda^* : \mathbb{R}^{N_s N_t} \rightarrow \mathcal{L}^2(\Omega). \quad (16)$$

For this, we need the adjoint of $L_{c_p, s}^y$, i.e.,

$$L_{c_p, s}^{y*} = (-\nabla^2)^{y/2} c_{p, s}^y, \quad (17)$$

which is formal adjoint since $-\nabla^2$ is self-adjoint. We assume that y is constant over the entire medium in the same way as [37]. We also need the time reversal operator \mathcal{R} defined by

$$\mathcal{R}[p](x, t) = p(x, T - t).$$

Lemma 1. *The adjoint map Λ^* can be calculated from $\Lambda^*[\hat{P}](x, t) = p_0^*(r)$, where $p_0^* = -\delta_{ij}\sigma_{ij}^{p,s*}(x, T)$, and $\sigma_{ij}^{p,s*}$ and $v_i^{p,s*}$ satisfy the coupled equations*

$$\begin{aligned} \rho \frac{\partial v_i^*}{\partial t} &= \sum_{p,s} q^{p,s} \left[\left(\frac{\partial}{\partial x_i} (\lambda \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\mu \sigma_{ij}^{p,s*}) \right) \right. \\ &\quad \left. + \sin(\pi y/2) L_{c_{p,s}}^{(y-1)*} \left(\frac{\partial}{\partial x_i} (\chi \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\eta \sigma_{ij}^{p,s*}) \right) \right] \end{aligned} \quad (18)$$

$$\begin{aligned} v_i^{p,s*} &= v_i^* - \frac{1}{\rho} \cos(\pi y/2) L_{c_{p,s}}^{(y-2)*} \left(\frac{\partial}{\partial x_i} (\chi \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\eta \sigma_{ij}^{p,s*}) \right) \\ \frac{\partial \sigma_{ij}^{p*}}{\partial t} &= \frac{1}{2} \left(\frac{\partial v_i^*}{\partial x_j} + \frac{\partial v_j^*}{\partial x_i} \right) + (\mathcal{R}\mathcal{G}^* \mathcal{M}^* \hat{P}) \\ \frac{\partial \sigma_{ij}^{s*}}{\partial t} &= \frac{1}{2} \left(\frac{\partial v_i^*}{\partial x_j} + \frac{\partial v_j^*}{\partial x_i} \right) \end{aligned} \quad (19)$$

with initial conditions

$$\sigma_{ij}^{p,s*}(x, 0) = 0, \quad v_i^*(x, 0) = 0. \quad (20)$$

Remark 1. Solutions for the set of equations (18) and (19) can be shown to exist by a similar strategy as for the equations describing the evolution of the forward fields.

Proof. We will show that when $\sigma_{ij}^{p,s}$ and v_i satisfy equations (9), (12) and (15), and also $\sigma_{ij}^{p,s*}$ and v_i^* satisfy (18), (19) and (20), then the forward map Λ and adjoint Λ^* must satisfy

$$\langle \Lambda[p_0], \hat{P} \rangle_{\mathbb{R}^{N_s N_t}} = \langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)} \quad (21)$$

for any $p_0 \in C_0^\infty(\Omega)$ and $\hat{P} \in \mathbb{R}^{N_s N_t}$. Because it will make calculations easier, we first deal with the adjoint fields in a time reversed order (i.e. we make the change of variable $t \mapsto T - t$) and with v_i^* replaced by $-v_i^*$ so that the initial conditions (20) are actually final conditions $\sigma_{ij}^*(x, T) = 0$ and $v_i^*(x, T) = 0$. Accordingly, we have the following relation between v_i and the adjoint field v_i^*

$$\int_0^T \int_{\mathbb{R}^d} \rho \left(\frac{\partial v_i}{\partial t} \overline{v_i^*} + v_i \overline{\frac{\partial v_i^*}{\partial t}} \right) dx dt = 0. \quad (22)$$

Plugging (12) into the first integrand in the above equation gives

$$\int_0^T \int_{\mathbb{R}^d} \sum_{p,s} \frac{\partial \sigma_{ij}^{p,s}}{\partial x_j} \overline{v_i^*} + \rho v_i \overline{\frac{\partial v_i^*}{\partial t}} dx dt = 0. \quad (23)$$

Integrating-by-parts, and using that $\sigma_{ij}^{p,s} \rightarrow 0$ at infinity, we end up with

$$\int_0^T \int_{\mathbb{R}^d} \sum_{p,s} -\sigma_{ij}^{p,s} \frac{1}{2} \overline{\left(\frac{\partial v_i^*}{\partial x_j} + \frac{\partial v_j^*}{\partial x_i} \right)} + \rho v_i \overline{\frac{\partial v_i^*}{\partial t}} dx dt = 0. \quad (24)$$

In the above equation, we also used the symmetry of the stress tensor $\sigma_{ij}^{p,s} = \sigma_{ji}^{p,s}$. Now we apply the same procedure to the stress tensor. Using the final conditions $\sigma_{ij}^{p,s*}(x, T) = 0$ yields

$$\int_0^T \int_{\mathbb{R}^d} \frac{\partial \sigma_{ij}^p}{\partial t} \overline{\sigma_{ij}^{p*}} + \sigma_{ij}^p \overline{\frac{\partial \sigma_{ij}^{p*}}{\partial t}} dx dt = \int_{\mathbb{R}^3} p_0 \overline{\sigma_{ii}^{p*}(x, 0)} dx = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}. \quad (25)$$

and

$$\int_0^T \int_{\mathbb{R}^d} \frac{\partial \sigma_{ij}^s}{\partial t} \overline{\sigma_{ij}^{s*}} + \sigma_{ij}^s \overline{\frac{\partial \sigma_{ij}^{s*}}{\partial t}} dx dt = 0. \quad (26)$$

Now, plugging (9) into the first integrands in the left-hand sides of (25) and (26) and then adding these two equations results in

$$\begin{aligned} & \int_0^T \int_{\mathbb{R}^d} \sum_{p,s} \left[\left(\lambda \delta_{ij} \frac{\partial v_l^{p,s}}{\partial x_l} + \mu \left(\frac{\partial v_i^{p,s}}{\partial x_j} + \frac{\partial v_j^{p,s}}{\partial x_i} \right) + \chi \left(\delta_{ij} \frac{\partial}{\partial x_l} \frac{\partial v_l^{p,s}}{\partial t^{y-1}} \right) \right) \right. \\ & \left. + \eta \left(\frac{\partial}{\partial x_j} \frac{\partial v_i^{p,s}}{\partial t^{y-1}} + \frac{\partial}{\partial x_i} \frac{\partial v_j^{p,s}}{\partial t^{y-1}} \right) \right] \overline{\sigma_{ij}^{p,s*}} + \sigma_{ij}^{p,s} \overline{\frac{\partial \sigma_{ij}^{p,s*}}{\partial t}} dx dt = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}. \end{aligned} \quad (27)$$

Taking integration-by-parts to the first term in the bracket in the above equation, together with the fact that σ_{ij}^* is symmetric, gives

$$\begin{aligned} & \int_0^T \int_{\mathbb{R}^d} \sum_{p,s} \left[- \left(\frac{\partial}{\partial x_i} \left(\lambda \sigma_{ll}^{p,s*} \right) + 2 \frac{\partial}{\partial x_j} \left(\mu \sigma_{ij}^{p,s*} \right) + \frac{\partial}{\partial x_i} \left(\chi \sigma_{ll}^{p,s*} \right) \right) \frac{\partial v_i^{p,s}}{\partial t^{y-1}} \right. \\ & \left. + 2 \frac{\partial}{\partial x_j} \left(\eta \sigma_{ij}^{p,s*} \right) \frac{\partial v_i^{p,s}}{\partial t^{y-1}} + \sigma_{ij}^{p,s} \overline{\frac{\partial \sigma_{ij}^{p,s*}}{\partial t}} \right] dx dt = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}. \end{aligned} \quad (28)$$

Now, adding equations (24) and (28) yields

$$\begin{aligned} & \int_0^T \int_{\mathbb{R}^d} \rho v_i \overline{\frac{\partial v_i^*}{\partial t}} - \sum_{p,s} \left[\frac{\partial}{\partial x_i} \left(\lambda \sigma_{ll}^{p,s*} \right) + 2 \frac{\partial}{\partial x_j} \left(\mu \sigma_{ij}^{p,s*} \right) \right. \\ & \left. + \left(\frac{\partial}{\partial x_i} \left(\chi \sigma_{ll}^{p,s*} \right) + 2 \frac{\partial}{\partial x_j} \left(\eta \sigma_{ij}^{p,s*} \right) \right) \frac{\partial v_i^{p,s}}{\partial t^{y-1}} \right] v_i^{p,s} \\ & + \sum_{p,s} \left[\frac{\partial \sigma_{ij}^{p,s*}}{\partial t} - \frac{1}{2} \left(\frac{\partial v_i^*}{\partial x_j} + \frac{\partial v_j^*}{\partial x_i} \right) \right] \sigma_{ij}^{p,s} dx dt = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}. \end{aligned} \quad (29)$$

Now, plugging the fractional Laplacian operators defined in equation (11) into the second line in equation (29), together with (8), yields

$$\begin{aligned}
& \int_0^T \int_{\mathbb{R}^d} \left(\rho \frac{\partial v_i^*}{\partial t} - \sum_{p,s} \left[\frac{\partial}{\partial x_i} (\lambda \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\mu \sigma_{ij}^{p,s*}) \right. \right. \\
& \quad \left. \left. + \left(\frac{\partial}{\partial x_i} (\chi \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\eta \sigma_{ij}^{p,s*}) \right) \sin(\pi y/2) L_{c_{p,s}}^{y-1} q^{p,s} \right] v_i \right. \\
& \quad \left. + \sum_{p,s} \left(\frac{\partial}{\partial x_i} (\chi \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\eta \sigma_{ij}^{p,s*}) \right) \cos(\pi y/2) L_{c_{p,s}}^{y-2} \frac{1}{\rho} \frac{\partial}{\partial x_j} \sigma_{ij}^{p,s} \right. \\
& \quad \left. + \sum_{p,s} \left[\frac{\partial \sigma_{ij}^{p,s*}}{\partial t} - \frac{1}{2} \left(\frac{\partial v_i^*}{\partial x_j} + \frac{\partial v_j^*}{\partial x_i} \right) \right] \sigma_{ij}^{p,s} \right) dx dt = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}.
\end{aligned} \tag{30}$$

In the above equation, we also used the linearity of operator $q^{p,s}$ with respect to v_i . By taking integration-by-parts to the third line and using the symmetry of $\sigma_{ij}^{p,s}$ in the same way as the first integrand in (23), we can see from (19) that if the integral in the first two lines of (30) is equal to zero, then (21) holds and the proof is complete (recall again that relative to (18), we have reversed the time, and changed v_i^* to $-v_i^*$). So we now focus on the first two lines of (30) which we will denote \mathcal{I} . Considering that $q^{p,s}$ is self-adjoint yields

$$\begin{aligned}
\mathcal{I} = & \int_0^T \int_{\mathbb{R}^d} \left[\rho \frac{\partial v_i^*}{\partial t} - \sum_{p,s} q^{p,s} \left[\left(\frac{\partial}{\partial x_i} (\lambda \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\mu \sigma_{ij}^{p,s*}) \right) \right. \right. \\
& \quad \left. \left. + \sin(\pi y/2) L_{c_{p,s}}^{(y-1)*} \left(\frac{\partial}{\partial x_i} (\chi \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\eta \sigma_{ij}^{p,s*}) \right) \right] \right] v_i dx dt.
\end{aligned} \tag{31}$$

Since v_i^* and $\sigma_{ij}^{p,s*}$ satisfy (18) we can now see that in fact $\mathcal{I} = 0$, and so the proof is complete. \square

Setting $\chi, \eta = 0$ and ignoring the splitting operator results in the general form of the adjoint for lossless media, which can be adapted to other attenuation models (see for example [22, 21]).

4. NUMERICAL COMPUTATION

Having found an analytically exact method of computing the adjoint operator Λ^* in the previous section, we now consider in more detail the discretisation and computation of the forward operator Λ and adjoint operator Λ^* . We denote the position of a given grid point in Cartesian coordinates by x_ζ where $\zeta = (\zeta_1, \dots, \zeta_d) \in \{1, \dots, N_1\} \times \dots \times \{1, \dots, N_d\}$ with $N = \prod_{i=1}^d N_i$ the total number of grid points along d dimensions. The grid spacing along the i th direction will be denoted by Δx_i . Also let n denote the iteration corresponding to time $t_n = n\Delta t$ with $n \in \{-1, \dots, N_t - 1\}$.

We discretise the spatial derivatives by a pseudo-spectral method. The k-space correction is also applied to the spatial derivatives in order to minimise the numerical dispersion errors due to the time integration. We approximate the fields on a uniform rectilinear grid staggered in space and time. It turns out that these staggered configurations increase accuracy and stability for approximation of odd-order spatial and temporal derivatives [34, 6]. Applying a k-space pseudo-spectral method

on a staggered grid, the spatial gradient in direction i will be in the form

$$\frac{\partial_{p,s}\{\cdot\}}{\partial x_i^\pm} = F^{-1} \left\{ \mathbf{i}k_i \operatorname{sinc}(C_0^{p,s}k\Delta t/2) e^{\pm \mathbf{i}k_i \Delta x_i/2} F\{\cdot\} \right\}, \quad (32)$$

where, as opposed to (8), F and F^{-1} denote the discrete Fourier transform and its inverse. Additionally, $\operatorname{sinc}(C_0^{p,s}k\Delta t/2)$ is the k-space operator which enforces a k-space correction to the gradient, where $C_0^{p,s}$ is the reference sound speed associated with the compressional and shear parts of the fields. The reader is referred to [34, 6] for further details on the k-space pseudo-spectral method.

To simulate wave propagation in an infinite domain using a computational grid with limited size, it is necessary that the outward travelling waves that reach the edge of the domain are absorbed by perfectly matched layers (PMLs) [6, 25]. Using PMLs, the general evolution equation $\frac{\partial R(x,t)}{\partial t} = \beta(x,t)$ is transformed into the form [34]

$$\frac{\partial R(x,t)}{\partial t} + \alpha_a R(x,t) = \beta(x,t), \quad (33)$$

where α_a is the attenuation coefficient associated with the PML, which is tapered within the PML thickness at each side of the grid (cf. [6], Eq.(42)). This yields

$$\frac{\partial (e^{\alpha_a t} R(x,t))}{\partial t} = e^{\alpha_a t} \beta(x,t). \quad (34)$$

Using a staggered temporal grid, this is approximated as

$$\frac{e^{\alpha_a(t+\Delta t)} R(x_\zeta, t + \Delta t) - e^{\alpha_a t} R(x_\zeta, t)}{\Delta t} = e^{\alpha_a(t+\Delta t/2)} \beta(x_\zeta, t + \Delta t/2). \quad (35)$$

This gives the update

$$R(x_\zeta, t + \Delta t) = e^{-\alpha_a \Delta t/2} \left[e^{-\alpha_a \Delta t/2} R(x_\zeta, t) + \Delta t \beta(x_\zeta, t + \Delta t/2) \right]. \quad (36)$$

Using direction-dependent PMLs, the field variables are split into directions along the Cartesian coordinates $m \in \{1, \dots, d\}$ [6]. In the sequel, the directions associated with PMLs are written to the left of the fields. We define the diagonal PML attenuation matrices $A_m \in \mathbb{R}^{N \times N}$ by

$$A_m = \operatorname{diag}(e^{-m\alpha_a \Delta t/2}). \quad (37)$$

Note that $-m\alpha_a$ depends on the grid point here.

For the staggered temporal grid we consider also the time points $t_{n+1/2} = n\Delta t + \Delta t/2$. To accommodate the staggered spatial grid we introduce the operators \mathbb{T}_i which shift the point x by $\Delta x_i/2$ in the i th coordinate, i.e., x_i changes to $x_i + \Delta x_i/2$. We will also use the same notation for the corresponding operator acting on functions defined by

$$\mathbb{T}_i f(x) = f(\mathbb{T}_i x). \quad (38)$$

The discretised particle velocity vector field is denoted by ${}_m v_{(i;\zeta;n)} \in \mathbb{R}_m^d \times \mathbb{R}_i^d \times \mathbb{R}_\zeta^N \times \mathbb{R}_n^{N_t+1}$ and is approximated on a staggered spatial grid as

$${}_m v_{(i;\zeta;n)} \approx \mathbb{T}_i {}_m v_i(x_\zeta, t_n). \quad (39)$$

The p and s parts of the discretised stress tensor field are denoted by ${}_m \sigma_{(ij;\zeta;n)}^{p,s} \in \mathbb{R}_m^d \times \mathbb{R}_i^d \times \mathbb{R}_j^d \times \mathbb{R}_\zeta^N \times \mathbb{R}_n^{N_t+1}$ and are approximated on a staggered grid as

$${}_m \sigma_{(ij;\zeta;n)}^{p,s} \approx \begin{cases} {}_m \sigma_{ij}^{p,s}(x_\zeta, t_n) & \text{if } i = j \\ \mathbb{T}_i \mathbb{T}_j {}_m \sigma_{ij}^{p,s}(x_\zeta, t_n) & \text{if } i \neq j. \end{cases} \quad (40)$$

Because of using a staggered grid, the unit dyadic tensor in (8) will be in the form [6]

$$(\hat{k}_i \hat{k}_j)_{\text{staggered}} = (\hat{k}_i \hat{k}_j)_{\text{nonstaggered}} \times \xi_{ij}, \quad (41)$$

where

$$\xi_{ij} = e^{+\mathbf{i}(k_i \Delta x_i - k_j \Delta x_j)/2} \quad (42)$$

is the shifting operator with \mathbf{i} standing for the imaginary number [6].

We also define

$$\tau_{\text{dis}}^{p,s} = C_{p,s}^{y-1} \sin(\pi y/2) \quad (43)$$

$$\tau_{\text{abs}}^{p,s} = C_{p,s}^{y-2} \cos(\pi y/2) \quad (44)$$

with $C_{p,s} \in \mathbb{R}^N$ the discretised form of $c_{p,s}$. Using a staggered grid, we define the medium's parameters as diagonal matrices of size $N \times N$ in the form

$$\begin{aligned} \bar{\rho}_i &= \text{diag}(\mathbb{T}_i \rho) \\ \bar{\lambda} &= \text{diag}(\lambda) \\ \bar{\mu}_{ij} &= \begin{cases} \text{diag}(\mu) & \text{if } i = j \\ \text{diag}(\mathbb{T}_i \mathbb{T}_j \mu) & \text{if } i \neq j. \end{cases} \\ \bar{\chi} &= \text{diag}(\chi) \\ \bar{\eta}_{ij} &= \begin{cases} \text{diag}(\eta) & \text{if } i = j \\ \text{diag}(\mathbb{T}_i \mathbb{T}_j \eta) & \text{if } i \neq j. \end{cases} \\ \bar{\tau}_{i,\text{dis}} &= \text{diag}(\mathbb{T}_i \tau_{\text{dis}}) \\ \bar{\tau}_{i,\text{abs}} &= \text{diag}(\mathbb{T}_i \tau_{\text{abs}}) \end{aligned} \quad (45)$$

where ρ, λ, μ, χ , and η on the right hand sides in (45) are the medium parameters evaluated at the N grid points. In the formulas that follow for the discretised model, these matrices are always understood to act on discretised fields in the index ζ corresponding to the spatial grid.

We also introduce the $N \times N$ matrices discretising the relevant fractional Laplacian operators as

$$\begin{aligned} \bar{Y}_{\text{dis}} &= F^{-1} \left\{ k^{y-1} F\{\cdot\} \right\} \\ \bar{Y}_{\text{abs}} &= F^{-1} \left\{ k^{y-2} F\{\cdot\} \right\}. \end{aligned} \quad (46)$$

Finally, we define the following function which we will use to simplify some of the formulas

$$h(i, j) = \begin{cases} +1 & \text{if } i = j \\ -1 & \text{if } i \neq j. \end{cases} \quad (47)$$

4.1. Forward model. In the sequel, the approximation of the system of viscoelastic wave equations defined by equations (9), (12) and (15) based on the details given above will be outlined. It is worth mentioning that a code is available in the open-source *k-Wave* toolbox for describing wave propagation in heterogeneous lossy elastic media based on equations (5), (12) and (15) using the k-space pseudo-spectral method [35, 38]. As discussed in section 2, the results of this code give an attenuation that is not evident in tissue media [37]. Therefore, we modified the pre-existing *k-Wave* toolbox code so that it includes two fractional Laplacian operators in order to account for absorption and physical dispersion following the frequency power law. This code is outlined as follows.

While in the continuous model we assume the initial pressure is instantaneous, in the discretised model we introduce the initial pressure at $t = 0$ to the forward model as an additive source split over the time interval $t \in [-\Delta t/2, +\Delta t/2]$. For this, $p(t = 0) = 1$ is approximated as $P(n = [-1/2 + 1/2]) = \frac{1}{\Delta t} [0.5 \ 0.5]$ (cf. [2], Appendix B). Considering this, together with (15) and dividing the source by PML directions, gives a source in the form

$$m^s(ij;\zeta;n+1/2) = \begin{cases} -\frac{\delta_{ij}}{2d\Delta t} \mathbb{S}P_0 & n = -1, 0 \\ 0 & \text{otherwise,} \end{cases} \quad (48)$$

where P_0 denotes the discretised form of p_0 , and \mathbb{S} is a symmetric smoothing operator that is used for mitigating unexpected oscillations in propagation of the initial pressure P_0 (For further details, the reader is referred to [2], Appendix B).

Start at iterate $n = -1$ with initial conditions $m\sigma_{(ij;\zeta;n=-1)}^{p,s} = 0$ and $mv_{(i;\zeta;n=-3/2)} = 0$, and terminate at iterate $n = N_t - 2$.

1. Update the particle velocity field:

$$mv_{(i;\zeta;n+\frac{1}{2})} = A_m \left[A_m mv_{(i;\zeta;n-\frac{1}{2})} + \frac{\Delta t}{\bar{\rho}_i} \sum_{p,s} \frac{\partial_{p,s}}{\partial x_m^{h(i,m)}} \sigma_{(im;\zeta;n)}^{p,s} \right] \quad (49)$$

$$v_{(i;\zeta;n+\frac{1}{2})} = \sum_{m=1}^d mv_{(i;\zeta;n+\frac{1}{2})}.$$

2. Split the particle velocity field into compressional and shear parts:

$$v_{(i;\zeta;n+\frac{1}{2})}^{p,s} = Q^{p,s} v_{(i;\zeta;n+\frac{1}{2})}. \quad (50)$$

Here, $Q^{p,s}$ denotes the discretised form of functions $q^{p,s}$ defined in (8).

3. Update the stress tensor field:

$$m\sigma_{(ij;\zeta;n+1)}^{p,s} = A_m \left[A_m m\sigma_{(ij;\zeta;n)}^{p,s} + \Delta t \left[\bar{\lambda} \delta_{ij} \frac{\partial_{p,s}}{\partial x_m} v_{(m;\zeta;n+\frac{1}{2})}^{p,s} \right. \right. \\ \left. \left. + \bar{\mu}_{ij} \left(\delta_{mj} \frac{\partial_{p,s}}{\partial x_j^{-h(i,j)}} v_{(i;\zeta;n+\frac{1}{2})}^{p,s} + \delta_{mi} \frac{\partial_{p,s}}{\partial x_i^{-h(i,j)}} v_{(j;\zeta;n+\frac{1}{2})}^{p,s} \right) \right. \right. \\ \left. \left. + \bar{\chi} \delta_{ij} \frac{\partial_{p,s}}{\partial x_m} \left(\bar{\tau}_{m,\text{dis}}^{p,s} \bar{Y}_{\text{dis}} v_{(m;\zeta;n+\frac{1}{2})}^{p,s} - \bar{\tau}_{m,\text{abs}}^{p,s} \bar{Y}_{\text{abs}} \partial_t v_{(m;\zeta;n+1/2)}^{p,s} \right) \right. \right. \\ \left. \left. + \bar{\eta}_{ij} \delta_{mj} \frac{\partial_{p,s}}{\partial x_j^{-h(i,j)}} \left(\bar{\tau}_{i,\text{dis}}^{p,s} \bar{Y}_{\text{dis}} v_{(i;\zeta;n+\frac{1}{2})}^{p,s} - \bar{\tau}_{i,\text{abs}}^{p,s} \bar{Y}_{\text{abs}} \partial_t v_{(i;\zeta;n+1/2)}^{p,s} \right) \right. \right. \\ \left. \left. + \bar{\eta}_{ij} \delta_{mi} \frac{\partial_{p,s}}{\partial x_i^{-h(i,j)}} \left(\bar{\tau}_{j,\text{dis}}^{p,s} \bar{Y}_{\text{dis}} v_{(j;\zeta;n+\frac{1}{2})}^{p,s} - \bar{\tau}_{j,\text{abs}}^{p,s} \bar{Y}_{\text{abs}} \partial_t v_{(j;\zeta;n+1/2)}^{p,s} \right) \right] \right]. \quad (51)$$

where

$$\partial_t v_{(i;\zeta;n+1/2)}^{p,s} = \sum_{m=1}^d A_m \frac{1}{\bar{\rho}_i} \frac{\partial_{p,s}}{\partial x_m^{h(i,m)}} \sigma_{(im;\zeta;n)}^{p,s} \quad (52)$$

4. *Add source:*

$$\begin{aligned} m\sigma_{(ij;\zeta;n+1)}^p &\leftarrow m\sigma_{(ij;\zeta;n+1)}^p + \Delta t \ m^s_{(ij;\zeta;n+1/2)} \\ \sigma_{(ij;\zeta;n+1)}^{p,s} &= \sum_{m=1}^d m\sigma_{(ij;\zeta;n+1)}^{p,s}. \end{aligned} \quad (53)$$

5. *Compute the pressure field and map it to detected data at ultrasound detectors:*

We use \mathbb{G} and \mathbb{M} for denoting the discretised variants of \mathcal{G} and \mathcal{M} . Correspondingly, at each iterate the pressure field is computed by

$$p_{(\zeta;n+1)} = \mathbb{G}\sigma_{(ij;\zeta;n+1)}^p = -\frac{1}{d} \sum_{l,m=1}^d m\sigma_{(ll;\zeta;n+1)}^p, \quad (54)$$

and is then interpolated to ultrasound detectors using trilinear interpolation [2, 25], i.e.,

$$\hat{P}_{n+1} = \mathbb{M}p_{(\zeta;n+1)}, \quad (55)$$

where $\mathbb{M} \in \mathbb{R}^{N_s \times N}$ is a map from the pressure at grid points to the pressure measured by the detectors, and $\hat{P}_n \in \mathbb{R}^{N_s}$ is the vector of measured pressure data at iteration $n = 0, \dots, N_t - 1$.

4.2. Analytic adjoint model. The continuous adjoint model defined by equations (18), (19) and (20) are solved numerically as follows. For brevity, we ignore the superscript $*$ for denoting the adjoint fields in the discretised case. Before defining the time stepping procedure for the adjoint, we first define the additive source. To account for splitting of P_0 over the first two temporal iterations in the forward model (cf. equation (48)), we define the order reversed adjoint measured data in the form [2]

$$\hat{P}_{n+1/2}^{\text{adj}} = \frac{1}{2\Delta t} \begin{cases} \hat{P}_{N_t-1}, & n = -1 \\ \hat{P}_{N_t-n-1} + \hat{P}_{N_t-n-2}, & n = 0, \dots, N_t - 2. \\ \hat{P}_0, & n = N_t - 1 \end{cases} \quad (56)$$

The adjoint measured data is mapped from ultrasound detector positions to an additive source that is defined at grid points using

$$m^s_{(ij;\zeta;n+1/2)} = \mathbb{G}^T \mathbb{M}^T \hat{P}_{n+1/2}^{\text{adj}}, \quad (57)$$

where

$$\mathbb{G}^T = -1_m \otimes \frac{\delta_{ij}}{d}. \quad (58)$$

Start at iterate $n = -1$ with initial conditions $m\sigma_{(ij;\zeta;n=-1)} = 0$ and $m^v_{(i;\zeta;n=-3/2)} = 0$, and terminate at iterate $n = N_t - 2$.

1. *Update the particle velocity field:*

$$\begin{aligned} m^v_{(i;\zeta;n+\frac{1}{2})} &= A_m \left[A_m \ m^v_{(i;\zeta;n-\frac{1}{2})} \right. \\ &+ \frac{\Delta t}{\bar{\rho}_i} \left[\sum_{p,s} \sum_{j=1}^d Q^{p,s} \left[\frac{\partial_{p,s}}{\partial x_i^+} (\bar{\lambda}_{ij} \sigma_{(jj;\zeta;n)}^{p,s}) + 2 \frac{\partial_{p,s}}{\partial x_j^{h(i,j)}} (\bar{\mu}_{ij} \sigma_{(ij;\zeta;n)}^{p,s}) \right. \right. \\ &\left. \left. + \bar{Y}_{\text{dis}} \bar{\tau}_{i,\text{dis}}^{p,s} \left(\frac{\partial_{p,s}}{\partial x_i^+} (\bar{\chi}_{ij} \sigma_{(jj;\zeta;n)}^{p,s}) + 2 \frac{\partial_{p,s}}{\partial x_j^{h(i,j)}} (\bar{\eta}_{ij} \sigma_{(ij;\zeta;n)}^{p,s}) \right) \right] \right] \end{aligned} \quad (59)$$

2. Add the absorption term to the particle velocity field:

$$m \mathbf{v}_{(i;\zeta;n+\frac{1}{2})}^{p,s} = m v_{(i;\zeta;n+\frac{1}{2})} - A_m \frac{1}{\bar{\rho}_i} \sum_{j=1}^d \bar{Y}_{\text{abs}} \bar{\tau}_{i,\text{abs}}^{p,s} \left(\frac{\partial_{p,s}}{\partial x_i^+} (\bar{\chi}_i \sigma_{(jj;\zeta;n)}^{p,s}) + 2 \frac{\partial_{p,s}}{\partial x_j^{h(i,j)}} (\bar{\eta}_{ij} \sigma_{(ij;\zeta;n)}^{p,s}) \right) \quad (60)$$

Note that we are not using the summation convention in these formulas.

3. Update the stress tensor field:

$$m \sigma_{(ij;\zeta;n+1)}^{p,s} = A_m \left[A_m m \sigma_{(ij;\zeta;n)}^{p,s} + \frac{\Delta t}{2} \left(\frac{\partial_{p,s}}{\partial x_j^{-h(i,j)}} j \mathbf{v}_{(i;\zeta;n+\frac{1}{2})}^{p,s} + \frac{\partial_{p,s}}{\partial x_i^{-h(i,j)}} i \mathbf{v}_{(j;\zeta;n+\frac{1}{2})}^{p,s} \right) \right] \quad (61)$$

4. Add source:

$$m \sigma_{(ij;\zeta;n+1)}^p \leftarrow m \sigma_{(ij;\zeta;n+1)}^p + \Delta t m s_{(ij;\zeta;n+1/2)} \quad (62)$$

5. Compute the pressure field at final iterate and apply smoothing:

$$p_{(\zeta;n=N_t-1)} = -\mathbb{S} \left(\frac{1}{d} \sum_{l,m=1}^d m \sigma_{(ll;\zeta;n=N_t-1)}^p \right). \quad (63)$$

5. ADJOINT FOR DISCRETISED VISCOELASTIC FORWARD MODEL

In this section, we will calculate the adjoint of the viscoelastic forward model Λ defined by (9), (12) and initial conditions in (15) based on the *discretise-then-adjoint* method. To do this, we consider the discretised equations (49) and (51) in a matrix form. Accordingly, let the particle velocity vector $\bar{v}_{n-1/2} \in \mathbb{R}^{Nd^2}$ at each time step be made up of the components $m v_{(i;\zeta;n-1/2)} \in \mathbb{R}^N$ ($i, m \in \{1, \dots, d\}$). Let us also define the stress tensor as vector $\bar{\sigma}_n \in \mathbb{R}^{12N}$ ($d = 2$) or $\bar{\sigma}_n \in \mathbb{R}^{30N}$ ($d = 3$) composed of the components $m \sigma_{(ij;\zeta;n)}^{p,s} \in \mathbb{R}^N$ ($i, j, m \in \{1, \dots, d\}$). Note that for the latter, we used the symmetry of σ_{ij} , together with the fact that $m \sigma_{ij} = 0$ if $m \notin \{i, j\}$ [6] to reduce the number of degrees of freedom.

We also define $X_n \in \mathbb{R}^{39N}$ (3D case) as a stack of the particle velocity and stress fields at time step n in the form $X_n = [(\bar{v}_{n-1/2})^T (\bar{\sigma}_n)^T]^T$. Let also $\mathcal{S} \in \mathbb{R}^{39N N_t \times N}$ give a map from the discretised initial pressure $P_0 \in \mathbb{R}^N$ to an additive source (cf. (48)), which we will write as

$$S = \mathcal{S} P_0 \in \mathbb{R}^{39N N_t}. \quad (64)$$

We will also write $S_{n+1/2} = \mathcal{S}_n P_0$ for the source at time step n . In particular $\mathcal{S}_n = 0$ except when $n = -1$ or 0 . The time sequence of fields at steps ($n \in \{-1, \dots, N_t - 2\}$) is then given by

$$X_{n+1} = T X_n + S_{n+1/2}, \quad (65)$$

where $T \in \mathbb{R}^{39N \times 39N}$ implements (49) and (51), and $X_{-1} = 0$ (cf. section 4.1). Here for brevity the operators \mathcal{S}_n are given using (48) multiplied by Δt , and thus multiplication by Δt is neglected in the second term of (65). We will look in more detail at the matrix T later in section 5.1. Finally, we introduce a measurement matrix $\mathcal{M} \in \mathbb{R}^{N_s \times 39N}$ that at each time step maps the field X_n to the measured data at the sensors (i.e. implements formulas (54) and (55)). Note that \mathcal{M} for the discretised formulae is defined not the same as for the continuous formulae. We first consider the map from the source S to the measurements.

Definition 2. The map $\mathbb{H} : \mathbb{R}^{39N_t} \rightarrow \mathbb{R}^{N_s N_t}$ is defined by

$$\hat{P} = \mathbb{H}S, \quad \hat{P}_n = \mathcal{M}X_n \quad (n \in \{0, \dots, N_t - 1\}), \quad \hat{P} = \left[\hat{P}_{n+1} \right]_{n=-1}^{N_t-2}, \quad (66)$$

where X_n is updated by (65) with initial condition $X_{-1} = 0$, and $\hat{P} \in \mathbb{R}^{N_s N_t}$ is the time series stack of measured data at iterates $n \in \{0, \dots, N_t - 1\}$.

In the next lemma, we show how to compute the adjoint of \mathbb{H} . Note that in fact this lemma applies more generally for the adjoint of any discretised problem taking the form described here.

Lemma 2. The adjoint \mathbb{H}^* of \mathbb{H} defined in definition 2 is given by

$$\left[X_n^* \right]_{n=-1}^{N_t-2} = \mathbb{H}^* \hat{P} \in \mathbb{R}^{39N_t}, \quad X_{N_t-1}^* = 0, \quad X_{n-1}^* = T^* X_n^* + \mathcal{M}^* \hat{P}_n \quad (n \in \{0, \dots, N_t - 1\}). \quad (67)$$

Proof. Let us assume that X_n satisfies (65) with initial condition $X_{-1} = 0$, and X_n^* satisfies the last two equations in (67). Then using the conditions $X_{-1} = 0$ and $X_{N_t-1}^* = 0$ we have

$$\sum_{n=-1}^{N_t-2} (X_{n+1} - X_n) \cdot \overline{X_n^*} = \sum_{n=0}^{N_t-1} X_n \cdot \overline{(X_{n-1}^* - X_n^*)}.$$

Then applying (65) on the left and (67) on the right we have

$$\sum_{n=-1}^{N_t-2} (TX_n - X_n + S_{n+1/2}) \cdot \overline{X_n^*} = \sum_{n=0}^{N_t-1} X_n \cdot \overline{(T^* X_n^* - X_n^* + \mathcal{M}^* \hat{P}_n)}.$$

Rearranging this slightly gives

$$\sum_{n=-1}^{N_t-2} (TX_n - X_n) \cdot \overline{X_n^*} + \sum_{n=-1}^{N_t-2} S_{n+1/2} \cdot \overline{X_n^*} = \sum_{n=0}^{N_t-1} (TX_n - X_n) \cdot \overline{X_n^*} + \sum_{n=0}^{N_t-1} (\mathcal{M}X_n) \cdot \overline{\hat{P}_n}.$$

Applying again the conditions $X_{-1} = 0$ and $X_{N_t-1}^* = 0$ we see that

$$\sum_{n=-1}^{N_t-2} S_{n+1/2} \cdot \overline{X_n^*} = \sum_{n=0}^{N_t-1} (\mathcal{M}X_n) \cdot \overline{\hat{P}_n}$$

which is equivalent to $\langle S, \mathbb{H}^* \hat{P} \rangle_{\mathbb{R}^{39N_t}} = \langle \mathbb{H}S, \hat{P} \rangle_{\mathbb{R}^{N_s N_t}}$, and so completes the proof. \square

The forward map actually defined in section 4.1 is

$$\mathbb{H} \mathcal{S},$$

and so the adjoint is

$$\mathcal{S}^* \mathbb{H}^*.$$

If we incorporate a time reversal, which amounts to changing $X_n^* \mapsto X_{N_t-2-n}^*$ in Lemma 2, as well as including \mathcal{S}^* , we obtain the following corollary which gives the full method of calculating the adjoint in our case incorporating time reversal. The sum in (68) is actually just two terms which can also be used to explain (56) if we commute the summing operation with the computation of X^* .

Corollary 1. $\mathcal{S}^*\mathbb{H}^*$ can be computed as

$$\mathcal{S}^*\mathbb{H}^*\hat{P} = \sum_{n=-1}^{N_t-2} \mathcal{S}_n^* X_{N_t-2-n}^* \quad (68)$$

where X_n^* is determined by

$$X_{-1}^* = 0, \quad X_{n+1}^* = T^* X_n^* + \mathcal{M}^* \hat{P}_{N_t-2-n} \quad (n \in \{-1, \dots, N_t-2\}). \quad (69)$$

5.1. The Matrices T and T^* . In this section we write the matrices T and T^* explicitly using the forward model presented in section 4.1 to show how multiplication by each of them may be computed. Considering corollary 1, we define the adjoint measured data for the discretised adjoint as $\bar{P}_{n+1/2}^{\text{adj}} = \Delta t \hat{P}_{n+1/2}^{\text{adj}}$. To start we can write (49) and (51) in the condensed forms

$$\bar{v}_{n+\frac{1}{2}} = A_v [A_v \bar{v}_{n-\frac{1}{2}} + \Phi \bar{\sigma}_n], \quad (70)$$

and

$$\bar{\sigma}_{n+1} = A_\sigma [A_\sigma \bar{\sigma}_n + \Psi_{\text{dis}} \bar{v}_{n+\frac{1}{2}} - \Psi_{\text{abs}} \bar{\sigma}_n] + \theta S_{n+1/2} \quad (71)$$

where θ is a sparse matrix that maps $S_{n+1/2}$ to the space of vector $\bar{\sigma}_n$. Also A_v , A_σ , Φ , Ψ_{dis} , and Ψ_{abs} are matrices that will be described in more detail below although for now we note that A_v and A_σ are both diagonal. Based on this we see that T can be written as the following product of matrices in block form

$$T = \begin{pmatrix} I_v & 0 \\ A_\sigma \Psi_{\text{dis}} & A_\sigma (A_\sigma - \Psi_{\text{abs}}) \end{pmatrix} \begin{pmatrix} A_v^2 & A_v \Phi \\ 0 & I_\sigma \end{pmatrix} \quad (72)$$

where I_v and I_σ are the identity matrices. From (72) we have

$$T^* = \begin{pmatrix} A_v^2 & 0 \\ \Phi^* A_v & I_\sigma \end{pmatrix} \begin{pmatrix} I_v & \Psi_{\text{dis}}^* A_\sigma \\ 0 & (A_\sigma - \Psi_{\text{abs}}^*) A_\sigma \end{pmatrix}. \quad (73)$$

Using Corollary 1, the above equation gives the updates for the adjoint problem as

$$\begin{aligned} \bar{v}_{n+1/2} &= A_v^2 (\bar{v}_{n-1/2} + \Psi_{\text{dis}}^* A_\sigma \bar{\sigma}_n) \\ \bar{\sigma}_{n+1} &= A_\sigma^2 \bar{\sigma}_n + \Phi^* (A_v)^{-1} \bar{v}_{n+1/2} - \Psi_{\text{abs}}^* A_\sigma \bar{\sigma}_n + \theta'^* \mathcal{M}^* \bar{P}_{n+1/2}^{\text{adj}}, \end{aligned} \quad (74)$$

where θ' is a sparse matrix mapping the space of vector $\bar{\sigma}_n$ to the space of vector X_n . Defining $\tilde{\sigma} = A_\sigma \bar{\sigma}$ and $\tilde{v} = A_v \bar{v}$ gives

$$\begin{aligned} \tilde{v}_{n+1/2} &= A_v (A_v \tilde{v}_{n-1/2} + \Psi_{\text{dis}}^* \tilde{\sigma}_n) \\ \tilde{\sigma}_{n+1} &= A_\sigma (A_\sigma \tilde{\sigma}_n + \Phi^* \tilde{v}_{n+1/2} - \Psi_{\text{abs}}^* \tilde{\sigma}_n) + \theta'^* \mathcal{M}^* \bar{P}_{n+1/2}^{\text{adj}}. \end{aligned} \quad (75)$$

Now let us consider the matrices A_v , A_σ , Φ , Ψ_{dis} , and Ψ_{abs} . First we note that A_v and A_σ can be found from (37). Though the others can be read off from (49) and (51), we will write them down explicitly here in order to show how we can explicitly calculate multiplication by their adjoints as required in (73).

Toward this goal, let us define the k-space discretised gradient operator

$$\nabla_{(m,i)}^{\pm p,s} = \frac{\partial_{p,s}}{\partial x_i^{\pm h(m,i)}}, \quad (76)$$

which is defined by (32). Note that the superscript p, s accounts for $C_0^{p,s}$ used in the k-space method (see (32)). Here we are considering $\nabla_{(m,i)}^{\pm p,s}$ to be an $N \times N$ matrix, and based on (32) and the unitarity of the discrete Fourier transform we see that

$$\left(\nabla_{(m,i)}^{\pm p,s}\right)^* = -\nabla_{(m,i)}^{\mp p,s}. \quad (77)$$

Using a pseudo-spectral method without k-space correction, the gradient operator is the same for compressional and shear parts of the fields, and thus this superscript would not be required in that case. We will also need the symmetrisation operator \mathbb{S} acting in the ij indices defined by

$$\mathbb{S}[a]_{ij} = \frac{a_{ij} + a_{ji}}{2}. \quad (78)$$

We apply \mathbb{S} to objects having more indices, but specify that it always acts on the pair ij . In (70), the action of matrix $\Phi \in \mathbb{R}^{9N \times 30N}$ on $\bar{\sigma}$ can be written as

$$m(\Phi\bar{\sigma})_i = \sum_{p,s} m(\Phi'\bar{\sigma})_i^{p,s}, \quad (79)$$

where from (49) and the second line in (53),

$$m(\Phi'\bar{\sigma})_i^{p,s} = \frac{\Delta t}{\bar{\rho}_i} \left(\nabla_{(i,m)}^+ p,s\right) \sum_{m'=1}^d m'\bar{\sigma}_{im'}^{p,s}. \quad (80)$$

Here and in what follows we do not include the spatial index ζ explicitly, but understand that for every value of the other indices (i and m here) we have a vector of dimension N , and that the discretised gradient and multiplication by $1/\bar{\rho}_i$ are implemented as operators acting on this spatial index. Thus, from (75) and using (77),

$$m(\Phi^*\bar{v})_{ij}^{p,s} = -\Delta t \mathbb{S} \left[\left(\nabla_{(i,j)}^- p,s\right) \frac{1}{\bar{\rho}_i} j\bar{v}_i \right]. \quad (81)$$

The symmetrisation \mathbb{S} must be added since this should map into the space of symmetric tensors.

Next, from (51), together with the second line in (49), the operator $\Psi_{\text{dis}} \in \mathbb{R}^{30N \times 9N}$ acts on \bar{v} by

$$\begin{aligned} m(\Psi_{\text{dis}}\bar{v})_{ij}^{p,s} &= \Delta t \left[\delta_{ij} \left[\bar{\lambda} \left(\nabla_{(m,m)}^- p,s\right) + \bar{\chi} \left(\nabla_{(m,m)}^- p,s\right) \bar{\tau}_{m,\text{dis}}^{p,s} \bar{Y}_{\text{dis}} \right] Q^{p,s} \sum_{m'=1}^d m'\bar{v}_m \right. \\ &\quad \left. + 2 \mathbb{S} \left[\left[\bar{\mu}_{ij} \delta_{mj} \left(\nabla_{(i,j)}^- p,s\right) + \bar{\eta}_{ij} \delta_{mj} \left(\nabla_{(i,j)}^- p,s\right) \bar{\tau}_{i,\text{dis}}^{p,s} \bar{Y}_{\text{dis}} \right] Q^{p,s} \sum_{m'=1}^d m'\bar{v}_i \right] \right] \end{aligned} \quad (82)$$

From this we can find the formula for the action of the adjoint in (75) as

$$\begin{aligned} m(\Psi_{\text{dis}}^*\bar{\sigma})_i &= -\Delta t \sum_{p,s} \sum_{j=1}^d Q^{p,s} \left[\left[\left(\nabla_{(i,i)}^+ p,s\right) \bar{\lambda} + \bar{Y}_{\text{dis}} \bar{\tau}_{i,\text{dis}}^{p,s} \left(\nabla_{(i,i)}^+ p,s\right) \bar{\chi} \right] i\bar{\sigma}_{jj}^{p,s} \right. \\ &\quad \left. + 2 \left[\left(\nabla_{(i,j)}^+ p,s\right) \bar{\mu}_{ij} + \bar{Y}_{\text{dis}} \bar{\tau}_{i,\text{dis}}^{p,s} \left(\nabla_{(i,j)}^+ p,s\right) \bar{\eta}_{ij} \right] j\bar{\sigma}_{ij}^{p,s} \right]. \end{aligned} \quad (83)$$

Additionally, from (51), (52) and (80), in (71) the operator $\Psi_{\text{abs}} \in \mathbb{R}^{30N \times 30N}$ acts on $\bar{\sigma}$ by

$$m(\Psi_{\text{abs}}\bar{\sigma})_{ij}^{p,s} = m \left(\Psi'_{\text{abs}} \left(\frac{1}{\Delta t} \sum_{m'=1}^d A_{m'm'}(\Phi'\bar{\sigma}) \right)_{ij} \right)^{p,s}, \quad (84)$$

which is actually the action of Ψ'_{abs} on $\partial_t \bar{v}$ using (51) in the form

$$\begin{aligned} m(\Psi'_{\text{abs}}\partial_t \bar{v})_{ij}^{p,s} &= \Delta t \left[\delta_{ij} \bar{\chi} \left(\nabla_{(m,m)}^- \right)^{p,s} \bar{\tau}_{m,\text{abs}}^{p,s} \bar{Y}_{\text{abs}} \partial_t \bar{v}_m^{p,s} \right. \\ &\quad \left. + 2 \mathbb{S} \bar{\eta}_{ij} \delta_{mj} \left(\nabla_{(i,j)}^- \right)^{p,s} \bar{\tau}_{i,\text{abs}}^{p,s} \bar{Y}_{\text{abs}} \partial_t \bar{v}_i^{p,s} \right] \end{aligned} \quad (85)$$

Using (75), the action of the adjoint is then given by

$$m(\Psi_{\text{abs}}^* \tilde{\sigma})_{ij}^{p,s} = m \left(\frac{1}{\Delta t} \Phi'^* A_m \Psi'_{\text{abs}} \tilde{\sigma} \right)_{ij}^{p,s}, \quad (86)$$

where

$$m(\Psi_{\text{abs}}'^* \tilde{\sigma})_i^{p,s} = -\Delta t \sum_{j=1}^d \bar{Y}_{\text{abs}} \bar{\tau}_{i,\text{abs}}^{p,s} \left[\left(\nabla_{(i,i)}^+ \right)^{p,s} \bar{\chi} \tilde{\sigma}_{jj}^{p,s} + 2 \left(\nabla_{(i,j)}^+ \right)^{p,s} \bar{\eta}_{ij} \tilde{\sigma}_{ij}^{p,s} \right]. \quad (87)$$

Finally, plugging (86) into the second line in (75), together with using (79), gives

$$\begin{aligned} \tilde{v}_{n+1/2} &= A_v (A_v \tilde{v}_{n-1/2} + \Psi_{\text{dis}}^* \tilde{\sigma}_n) \\ \tilde{\mathbf{v}}_{n+1/2} &= \mathbb{1}_{p,s} \otimes \tilde{v}_{n+1/2} - \frac{1}{\Delta t} A_v \left(\Psi_{\text{abs}}'^* \tilde{\sigma}_n \right)^{p,s} \\ \tilde{\sigma}_{n+1} &= A_\sigma \left(A_\sigma \tilde{\sigma}_n + \Phi'^* \tilde{\mathbf{v}}_{n+1/2} \right) + \theta'^* \mathcal{M}^* \bar{P}_{n+1/2}^{\text{adj}}, \end{aligned} \quad (88)$$

where, $\mathbb{1}_{p,s} \otimes$ represents the adjoint of $\sum_{p,s}$ included in Φ , and thus as opposed to $\tilde{v}_{n+1/2} \in \mathbb{R}^{Nd^2}$, $\tilde{\mathbf{v}}_{n+1/2} \in \mathbb{R}^{2Nd^2}$ is composed of the compressional and shear components. Equation (88) gives the same formulae as in section 4.2. This indicates that using a k-space pseudo-spectral method the numerical computation of the continuous forward operator Λ and adjoint Λ^* matches the discretised forward operator $\bar{\mathbb{H}} = \mathbb{H}\mathcal{S}$ and the corresponding algebraic adjoint $\bar{\mathbb{H}}^* = \mathcal{S}^* \mathbb{H}^*$, respectively.

6. FIRST-ORDER OPTIMISATION METHODS FOR PAT

We incorporate the forward and adjoint pair in an inverse solver based on the Iterative Shrinkage Thresholding Algorithm (ISTA), which is popular in PAT, e.g. [2, 17]. A fast variant of this algorithm has also been used in PAT [1, 16, 17]. Let the discretised variant of the sought after initial pressure (P_0) be denoted by $P \in \mathbb{R}^N$. The inverse problem of inferring P_0 from \hat{P} can be fit into a general class of non-smooth constrained convex minimisation problems of the form

$$\underset{P}{\text{argmin}} \{F(P) := f(P) + g(P)\}, \quad (89)$$

where $f(P) = \frac{1}{2} \left\| \bar{\mathbb{H}}P - \hat{P} \right\|^2$ is a continuously differentiable function with Lipschitz continuous gradient having smallest Lipschitz constant $L_f = \mathcal{S}_{i_{\text{max}}}(\bar{\mathbb{H}}^* \bar{\mathbb{H}})$ with $\mathcal{S}_{i_{\text{max}}}(\cdot)$ the largest singular value. The gradient of f is computed by

$$\nabla f(P) = \bar{\mathbb{H}}^* \left(\bar{\mathbb{H}}P - \hat{P} \right). \quad (90)$$

Using a total variation (TV) regularised variant of ISTA, we take $g(P) = \lambda_r \mathcal{J}(P) + \delta_C(P)$, where $\mathcal{J}(P)$ represents a TV penalty functional, λ_r denotes the regularisation parameter, and δ_C is an indicator function for the set of constraints $C = \{P \geq 0\}$ [7, 3].

Applying the so-called forward-backward splitting method to a fixed point iterative scheme arising from the optimality conditions of problem (89) gives two steps at each iteration k of the optimisation algorithm. The first step uses a steepest descent search direction $-\nabla f(P^{k-1})$ and step size Γ^k in the form

$$Y^k = P^{k-1} - \Gamma^k \nabla f(P^{k-1}), \quad (91)$$

and is called the forward gradient descent step [7, 3]. Applying ISTA, the iterates P^k converge to a minimiser P^* of problem (89) if $\Gamma^k \in (0, 2/L_f)$ [3]. Here, L_f , the largest singular value of $\bar{\mathbb{H}}^* \bar{\mathbb{H}}$, is computed iteratively by the power method following [2, 1, 17]. Since L_f is independent of the unknown P_0 , it can be stored and used for all experiments done in a fixed setting [1].

The second step is a proximal map in the form

$$\text{prox}_{\Gamma^k}(g)(Y^k) := \underset{P}{\text{argmin}} \left\{ g(P) + \frac{1}{2\Gamma^k} \|P - Y^k\|^2 \right\}, \quad (92)$$

and is called backward gradient step [7]. Following [16, 2], here the proximal map is computed based on Chambolle's dual approach (See [3]). In our study, we will terminate ISTA if the following criterion is satisfied:

$$k > 1 \cap 1 - \frac{F^k}{F^{k-1}} < \epsilon. \quad (93)$$

Here, ϵ is a stopping tolerance, and is chosen close to zero.

7. NUMERICAL RESULTS

The numerical implementation of a system of coupled first-order equations that describe the propagation of PA waves in linear isotropic elastic and lossy media based on an absorption following the classical Kelvin-Voigt model (4) is available on the *k-Wave* website [35, 38]. This code is based on a pseudo-spectral time-domain method [6]. For a numerical implementation of the forward problem, we modified this code so that it includes the absorption and physical dispersion following the frequency power law, using the splitting technique, as discussed in section 4.1 [37]. We also developed a code for implementation of our continuous adjoint, as discussed in section 4.2. We then showed that this adjoint matches the algebraic adjoint of the associated discretised forward operator. To validate the adjoint model, (21) was first used to check if the inner-product test is satisfied for any initial pressure P_0 and data \hat{P} . We then performed reconstructions from simulated data in both 2D and 3D settings as described below.

7.1. 2D phantom.

7.1.1. *Computational grid.* we used a computational grid with a size of 14×14 cm² to simulate the size of the superior surface of the skull.

Data generation: To simulate the propagation of wavefields, the computational grid was made up of 472×472 grid points equidistantly spaced with a separation distance of 2.9661×10^{-2} cm along both Cartesian coordinates. This computational grid was enclosed by a PML having a thickness of 20 grid points and a maximum attenuation coefficient of 2 nepers per grid point so that a good trade-off between mitigating spurious wave wrapping at the boundaries and reflection of waves at the edge of the PML was made [34]. The propagated pressure field was measured in time by 200 detectors that were evenly placed aligned by the top half of periphery of a circle having a radius

of $r = 6.8\text{cm}$ so that π radians were covered by the detectors. The skull was simulated with semi-circular interfaces with distances of $0.85r$ and $0.95r$ to the center of the semi-circle so that it has an even thickness of 6.8mm . This has provided an even radial distance of 3.4mm between the outer edge of the skull and the detectors.

Image reconstruction: To avoid an inverse crime for discretisation [18], the image reconstruction was done on a grid made up of 328×328 grid points which are placed evenly with a separation distance of 4.2683×10^{-2} cm along both Cartesian coordinates. The thickness of the PML was reduced to 16 grid points.

7.1.2. Physical parameters. The maps corresponding to the medium's mass density ρ , compressional wave propagation speed c_p and shear wave propagation speed c_s were shown in figures 1(a), 1(b) and 1(c), respectively. The colour scales are shown to the right of each map, where the blue colour represents the physical parameters of soft tissue with $c_p = 1500 \text{ ms}^{-1}$, $c_s = 0 \text{ ms}^{-1}$ and $\rho = 1000 \text{ kgm}^{-3}$, and the red colour represents the skull with $c_p = 3000 \text{ ms}^{-1}$, $c_s = 1500 \text{ ms}^{-1}$ and $\rho = 1850 \text{ kgm}^{-3}$. These parameters were chosen following [25]. The absorption coefficients were set to $\alpha_{0,p} = 10 \text{ dBMHz}^{-y}\text{cm}^{-1}$ and $\alpha_{0,s} = 20 \text{ dBMHz}^{-y}\text{cm}^{-1}$ in the skull, and $\alpha_{0,p} = 0.75 \text{ dBMHz}^{-y}\text{cm}^{-1}$ and $\alpha_{0,s} = 0.5 \text{ dBMHz}^{-y}\text{cm}^{-1}$ in the soft tissue. Note that we assumed absorption coefficients associated with the skull greater than the experimental values obtained in [41] (cf. Table 1 in [37]). Following [37], the exponent factor was assumed constant across the entire medium, and was set to $y = 1.4$.

7.1.3. Validation of adjoint. The inner-product test is a useful checking method, which estimates the accuracy of implementation of the adjoint operator. Using the setting described above, we numerically measured the accuracy of the computed adjoint model using the inner-product formula in (21). To do this, we used a randomly selected vector for \hat{P} , together with an initial pressure distribution P_0 in the form of a circular disk with a radius of $0.8r$, where the values at each point of the disk are chosen randomly. The relative difference between the left-hand and right-hand sides of (21) was averaged between 10 attempts. This gives values 7.43×10^{-5} and 8.71×10^{-6} for the grids used for image reconstruction and data generation, respectively. Our observations showed us that with an increase in density of the grid, the inner-product test is satisfied with a higher order of accuracy.

7.1.4. Simulation setting. To evaluate the performance of the forward and adjoint pair for image reconstruction, we considered two scenarios as follows.

Scenario 1: In general, the inverse problem in PAT is based on the assumption that the physical parameters of the medium are known. In our first experiment, we used the maps in figures 1(a), 1(b) and 1(c) as physical parameters for both data generation and image reconstruction. This implies that we have an exact knowledge of the physical parameters. Since this assumption does not hold in practical cases, this is considered as an inverse crime [18]. Using these maps, the grid used for data generation supports a maximal frequency up to 2.5286 MHz for propagation of compressional waves through the entire medium and shear waves within the skull.

Scenario 2: In the second experiment, we avoided an inverse crime in estimating medium's parameters by using different maps for data generation and image reconstruction. Correspondingly, for generating data we contaminated the maps in figures 1(a), 1(b) and 1(c) with a 30dB Additive White Gaussian Noise (AWGN). The contaminated maps are displayed in figures 1(d), 1(e) and 1(f). For image reconstruction, we assumed the contaminated maps are not readily available, and thus we used the clean maps. Using the noise contaminated maps for data generation, the associated grid supports maximal frequency up to 2.2047 MHz for propagation of shear waves within the skull

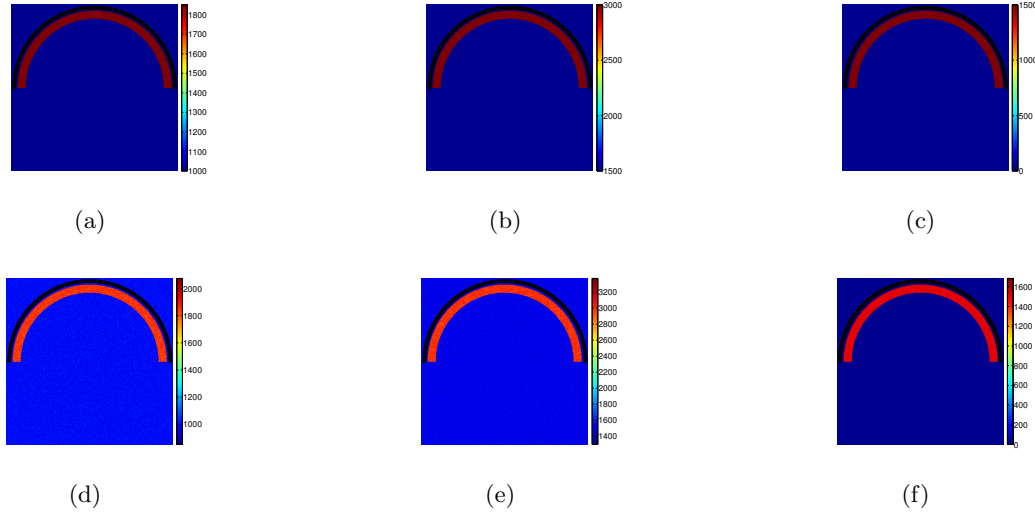


FIGURE 1. 2D phantom. Exact physical maps: (a) ρ (b) c_p (b) c_s , and noise-contaminated physical maps: (d) ρ (e) c_p (f) c_s .

and 2.1889 MHz for compressional waves through the entire medium. In these figures, the location of ultrasound detectors has been shown by the green semi-circle.

The grid used for image reconstruction supports a maximal frequency of 1.7571 MHz for compressional waves through the entire medium and shear waves within the skull. We created the initial pressure map with a maximal amplitude of 2, as shown in figure 2(a). For both scenarios, a CFL of 0.3 was sufficient to guarantee the stability of the forward and adjoint models. Accordingly, the computed pressure wavefield was recorded in 4451 time steps, and interpolated to the detectors using linear interpolation [35]. The generated data (for both scenarios) were then contaminated with a 30dB AWGN.

7.1.5. Image reconstruction. We first reconstructed an image corresponding to each scenario using the time reversal method. This was performed using *k-Wave* toolbox [35, 38]. According to [39], a filtering of the absorption and dispersion terms in the spatial frequency domain may be required to ensure the stability of TR. Here, to make a fair comparison between TR and ISTA, we applied TR optimistically on a non-absorbing medium with $\alpha_{0,p,s} = 0$. The images reconstructed by TR for scenarios 1 and 2 are displayed in figures 2(b) and 2(d), respectively. The computed forward and adjoint operators were then incorporated into the inverse solver discussed in section 6. The regularisation parameter was empirically set to $\lambda_r = 1 \times 10^{-2}$. A step size of $\Gamma_k = 1.8/L_f$ was chosen and used for all iterations k . Here, L_f was computed by the power iteration method [2, 1, 17]. The iterates of power iteration algorithm converged to L_f after around 15 iterations. Using ISTA, the iterates are initialised by zero, and the algorithm was terminated using the stopping tolerance $\epsilon = 1 \times 10^{-4}$. Figures 2(c) and 2(e) show an image of the final iterate computed by ISTA for scenarios 1 and 2, respectively.

The computed sequence of iterates was measured by two parameters:

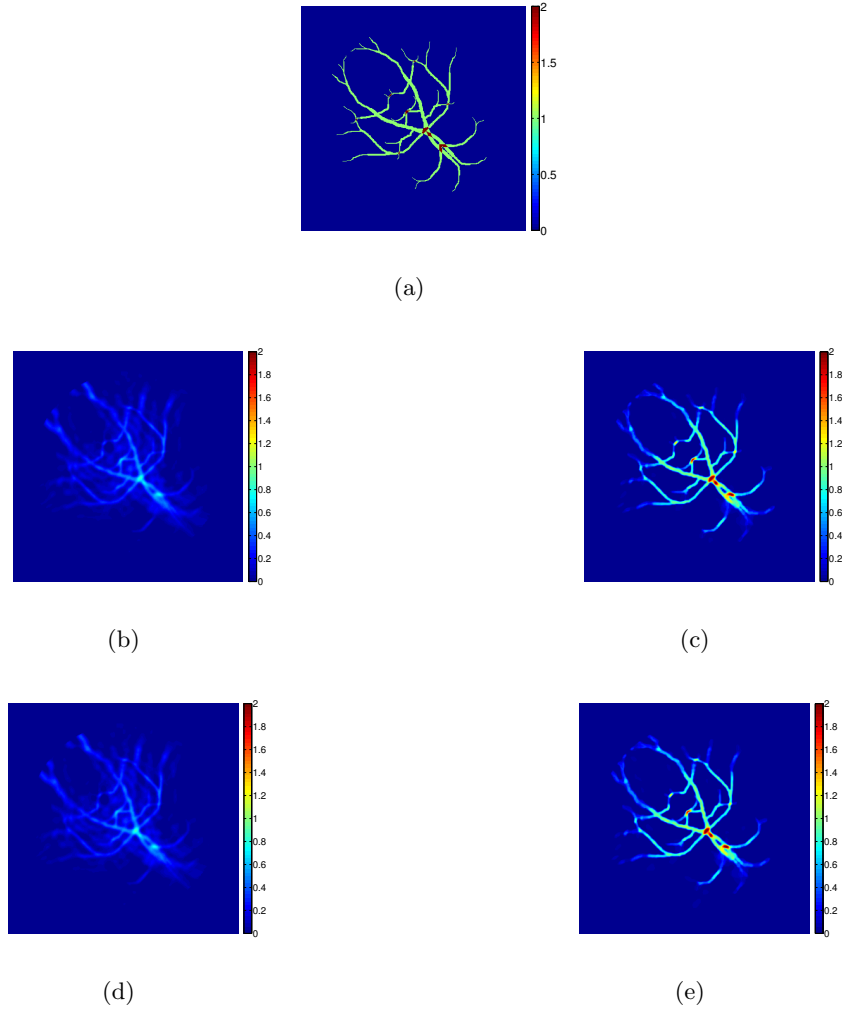


FIGURE 2. 2D phantom. (a) initial pressure map, and reconstructed images using exact physical parameters (inverse crime): (b) TR ($\alpha_{0,p,s} = 0$) (c) ISTA, and erroneous physical parameters: (d) TR ($\alpha_{0,p,s} = 0$) (e) ISTA.

(1) Relative Error (RE):

$$RE(P^k) = \frac{\|P^k - \tilde{P}_{\text{phantom}}\|_2}{\|\tilde{P}_{\text{phantom}}\|_2} \times 100, \quad (94)$$

where P^k denotes the update at iteration k , and $\tilde{P}_{\text{phantom}}$ denotes the initial pressure distribution in the phantom interpolated to the grid used for image reconstruction.

(2) Objective function ($F(P^k)$): (cf. section 6, (89)).

7.1.6. *Observations.* Figures 3(a) and 3(b) show RE and objective function values of iterates computed by ISTA versus the iteration number k , respectively. Figure 3(c) shows F from a large view

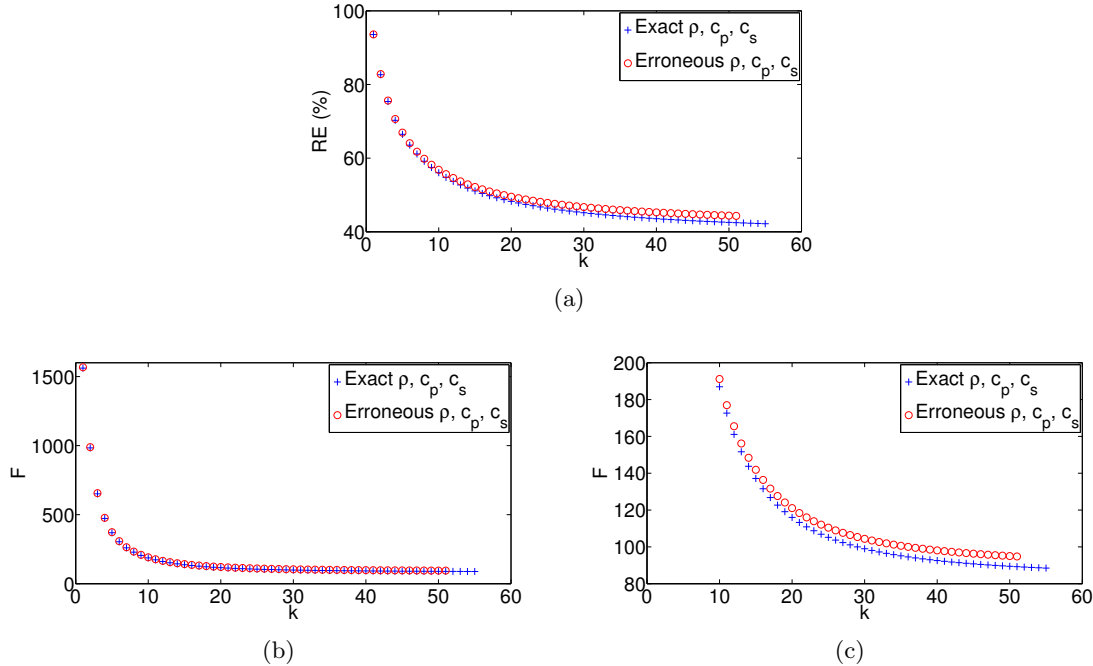


FIGURE 3. 2D phantom. (a) Relative Error (RE) (b) objective function (F) (c) F around the stopping point.

around the stopping point. In these figures, the blue and red plots, respectively correspond to scenarios 1 and 2. Our numerical observations for the two mentioned scenarios are as follows.

Scenario 1: Both RE and F monotonically decreased, and the stopping criterion was satisfied at iteration 55. The RE and F reached values of 42.19% and 87.24 at the final iteration, respectively. The final iteration pertains to the image shown in figure 2(c). From figures 3(b) and 3(c), ISTA has reduced F almost 95%.

Scenario 2: In scenario 2 when we avoided the inverse crime in estimating physical parameters, a monotonic reduction in both RE and F was observed, and the stopping criterion was satisfied at iteration 51. As shown in figures 3(a) and 3(b), RE and F reached values of 44.29% and 93.52 at the final iteration, which corresponds to the image shown in figure 2(e).

These figures indicate that in presence of an error in estimating physical parameters, the inverse solver was sufficiently tolerant to reconstruct almost the same image as using the exact physical maps.

7.2. 3D phantom.

7.2.1. Computational grid. This grid was created as a rectangular cuboid with a size of $14 \times 14 \times 3.5 \text{ cm}^3$ so that it simulates the size of a superior volume of the skull.

Data generation: The grid was made up of $160 \times 160 \times 40$ grid points with a spatial separation of $8.75 \times 10^{-2} \text{ cm}$ along all Cartesian coordinates. Each surface of this grid was enclosed by a PML with 20 grid points, and an attenuation coefficient with a maximum value of 2 nepers per grid point was tapered within the PML [34]. The pressure field was measured by 62×62 point-wise detectors, which were placed equidistantly on the top surface of the grid. The skull was simulated so that its top and bottom surfaces are aligned by the third and tenth horizontal planes of the grid points

from the top surface of the cube. This has provided a thickness of 6.1mm for the skull, as well as a distance of 1.75mm between the top surface of the skull and the detection plane.

Image reconstruction: Here, an inverse crime for discretisation was avoided by using a grid with different size, made up of $128 \times 128 \times 32$ grid points which are positioned with a spatial separation of 1.1mm along all Cartesian coordinates. Proportional to a reduction in size of the computational grid, we reduced the thickness of the PML to 16 grid points. Because of using a coarser computational grid, the thickness of the skull had to be reduced to 5.5mm with the top and bottom edges aligning the third and eighth horizontal planes of the grid points, respectively from the top surface of the grid.

7.2.2. *Physical parameters.* Figures 4(a), 4(b) and 4(c), respectively show the maps associated with the mass density, and the propagation speed of compressional and shear waves. As shown in these figures, the physical parameters of the medium are simulated the same as the 2D phantom.

7.2.3. *Validation of adjoint.* We used an inner-product test based on (21) in order to numerically evaluate the accuracy of the computed adjoint model. To do this, we used a randomly selected vector for \hat{P} , together with a randomly chosen initial pressure distribution P_0 supported in a cuboid region below the skull. This is the region below the 10th (resp. 12th) horizontal plane from the top surface of the grid for image reconstruction (resp. data generation). The mean relative difference between the left-hand and right-hand sides of (21) among 10 attempts was 2.21×10^{-4} and 3.47×10^{-5} for the grids used for image reconstruction and data generation, respectively.

7.2.4. *Simulation setting.* For image reconstruction, two scenarios were considered:

Scenario 1: The maps that are displayed in figures 4(a), 4(b) and 4(c) were used for both data generation and image reconstruction. As discussed above, this is an inverse crime in estimating physical parameters, although because of the shift of soft tissue-skull interfaces between the fine and coarse grids, the inverse crime has been avoided to some degree. Using these maps, the grid used for data generation supports a maximal frequency of 8.7514×10^5 Hz for compressional waves across the entire medium and for shear waves propagated through the skull.

Scenario 2: In addition to the shifting error in physical parameters because of the discretisation, these maps have been contaminated with a 30dB AWGN noise for data generation, whereas the reconstruction is done using the clean maps. This induces an error in estimating physical parameters, as they are not available exactly for image reconstruction. Using the noise contaminated maps, the grid used for data generation supports maximal frequencies up to 7.2018×10^5 Hz and 7.4143×10^5 Hz for compressional waves across the entire grid and shear waves within the skull, respectively.

The grid used for image reconstruction supports a maximal frequency of 6.8571×10^5 Hz for compressional waves through the entire medium and shear waves propagated through the skull. For simulating the initial pressure map, the phantom used for the 2D scenario was placed obliquely inside the grid in a way in which the initial pressure distribution associated with the phantom is compactly supported in the soft tissue. Figure 5(a) shows the simulated phantom from a top view. Here, all 3D images including phantom are displayed from a top view using *maximum intensity projection* technique. Using a *CFL* of 0.3, the simulated pressure wavefield was recorded in 1532 time steps, and interpolated to the detectors using trilinear interpolation [25]. Similar to the 2D phantom, the vector of generated data \hat{P} was contaminated with a 30dB AWGN for both scenarios.

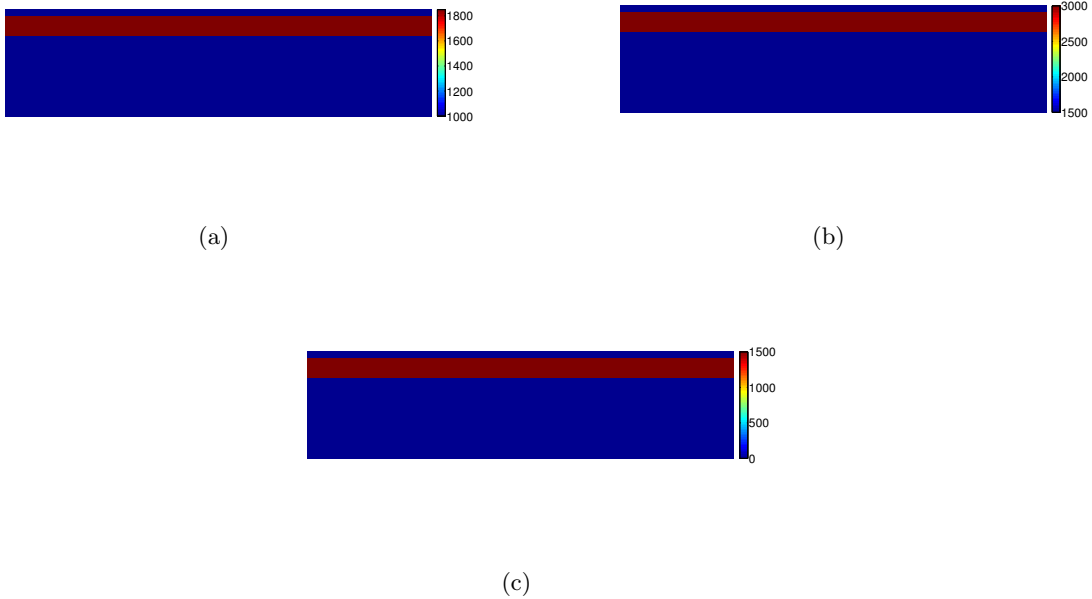


FIGURE 4. 3D phantom. Exact physical maps: (a) ρ (b) c_p (b) c_s .

7.2.5. *Image reconstruction.* We first reconstructed an image for each scenario using TR, which is available on the *k-Wave* website [35, 38]. The images reconstructed by TR for scenarios 1 and 2 are shown in figures 5(b) and 5(d), respectively. Using ISTA, the reconstruction parameters were chosen the same as for the 2D case. We used the power iteration method for computing L_f . Figures 5(c) and 5(e) show an image of the final iterate computed by ISTA for scenarios 1 and 2, respectively. A comparison between these two images indicates that using erroneous physical maps has led to a slight blurriness in the reconstructed image. Note that here the inverse crime has been avoided by exaggeration compared to real cases.

7.2.6. *Observations.* Figure 6(a) shows the RE of the computed iterates versus iteration number. Additionally, figure 6(b) shows the objective function values versus the iteration number around the terminating point. These plots have been displayed using the same colours as for the 2D phantom. From these, our observations for the two discussed scenarios are as follows.

Scenario 1: Both RE and F monotonically decreased until the iteration 54 at which the stopping criterion was satisfied. The final iterate, which is shown in figure 5(c), has an RE of 41.13% and an F of 1.19×10^3 .

Scenario 2: Using the noise contaminated physical maps for data generation, a monotonic reduction for RE and F is observed again, and the terminating criterion was satisfied at iteration 56. The final reconstructed image, which is shown in figure 5(e), has an RE of 48.44% and an F of 1.20×10^3 .

8. DISCUSSION AND CONCLUSION

In this work, we derived the adjoint of the continuous map defined in (14) and (15), which describes the propagation of PA waves in linear isotropic viscoelastic media with the absorption and physical dispersion following a frequency power law. Our derived continuous adjoint is a

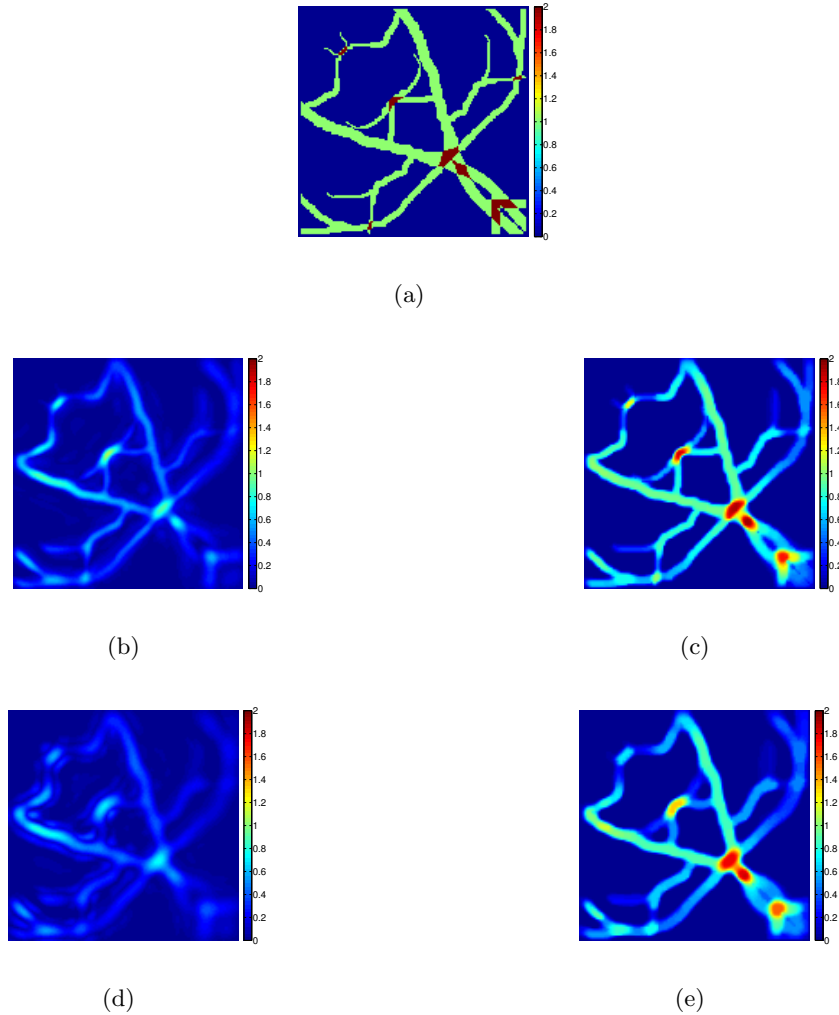


FIGURE 5. 3D phantom. (a) initial pressure map, and reconstructed images using exact physical parameters (inverse crime): (b) TR ($\alpha_{0,p,s} = 0$) (c) ISTA, and erroneous physical parameters: (d) TR ($\alpha_{0,p,s} = 0$) (e) ISTA.

generalised version of the discretised adjoint in [25] in the sense that it can be adapted to any discretisation scheme, and has been extended to include an attenuation evident in tissue media [37]. We analytically showed that a numerical computation of our continuous adjoint using a k-space pseudo-spectral method matches the algebraic adjoint of an associated discretised map defined by (64) and (66).

From a numerical point of view, it was shown that this forward and adjoint pair satisfies the inner-product formula in (21). This pair was then incorporated in a positivity constrained optimisation algorithm based on ISTA that is regularised by the TV denoising approach of Chambolle [3]. We preferred to test the derived forward and adjoint operators on a classical inverse solver (cf. [2, 17] for the application of ISTA in PAT), although this poses some limitations such as a low speed of convergence. The convergence can be improved by using a fast version of ISTA (FISTA) [16, 1, 17].

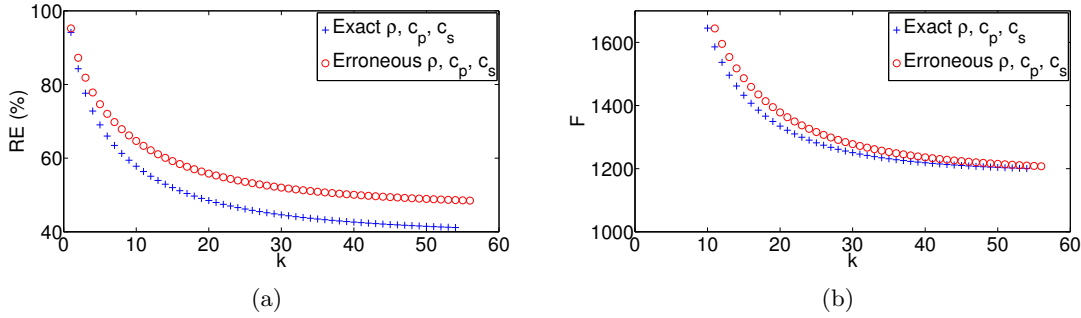


FIGURE 6. 3D phantom. (a) Relative Error (RE) (b) a large view of objective function F around the stopping point.

In addition, an iterative enhancement of solutions based on a *Bregman iteration* [26] may be useful when the collected data is compressively sampled. A *Bregman iteration* algorithm using FISTA has been successfully applied for this case [1].

In both 2D and 3D cases, the iterates monotonically converged to a minimiser of an objective function, and the final reconstructed image was close to the ground truth image. In the presence of an error in estimating physical parameters, the iterates monotonically converge again, but the iterate at the stopping point was slightly less accurate than using the exact physical parameters. This loss of accuracy cannot be detected by eye in the 2D scenario, as shown in figure 2(e). However, for the 3D scenario, figure 5(e) shows that an error in estimating physical parameters has led to a slight blurriness in the reconstructed image, compared to using exact physical maps. Note that in the 3D scenario for the grid used for data generation, in addition to a 30 dB noise added to the physical maps, the skull's thickness is 0.6mm larger than the grid for image reconstruction.

In addition, for the 2D scenario, as shown in figures 2(c) and 2(e), the reconstructed images have some blurriness in regions close to the skull. We believe that this can be attributed to the full internal reflection of wavefronts nearly tangent to the skull, and agrees with theoretical predictions of stability for inversion found in [31] using methods of microlocal analysis. In essence, the blurred region close to the skull is not fully resolved because the wavefronts emanating from that region do not reach the detectors (Note that the geometry of skull in our study is not realistic). To fully understand this a more delicate study on the relation between the theoretical analysis of [31] and PAT of the brain using optimisation algorithms may be needed.

The simplified geometries of the skull we used in our simulations look sufficient to provide an insight on the performance of the derived adjoint, but the geometry and composition of the skull in real cases are much more complicated than our simulations [15]. In addition, in practical cases, to extract all information available from the measured data, the maximal frequency supported by the computational grid must match the maximal frequency that is detectable by detectors [39]. This dramatically increases the computational demands regarding storage space and speed, but it can be handled using GPU accelerated computing [19], or Field-programmable gate array (FPGA) [40]. The 3D detection setting in our study simulated a planar Fabry-Pérot (FP) photoacoustic scanner, which requires several minutes to collect time series of data from PA wavefields [1]. Further studies can be done to apply our adjoint assisted optimisation algorithm on ultra-fast PAT acquisition systems that utilise spatio-temporal sub-sampled data [1].

Using our derived adjoint, an extension of the PAT problem of brain to direct quantitative PAT (QPAT), a direct estimation of the optical parameters inside the skull from the acoustic data collected outside the skull, would be a very interesting topic. The arising *opto-elastic* inverse problem is more challenging than the *opto-acoustic* problem [9] because of the high optical absorption and scattering of the skull and low degrees of freedom for optical illumination. This limits the applicability of multi-source QPAT, which is necessary for uniqueness of the problem when we use a single-frequency optical excitation.

ACKNOWLEDGEMENT

This work was supported in part by a Dean's award from the Faculty of Science and Engineering at the University of Manchester, and the Engineering and Physical Sciences Research Council (EP/M016773/1).

9. APPENDIX

In this appendix, we will show that the adjoint operator can be put into the form of a system of coupled partial differential equations, in the same way as the forward operator, and the update of particle velocity field is actually a sum of the adjoint of absorption and dispersion terms enforced to the stress tensor field.

Continuous adjoint: This is derived by plugging the first formula in (19) into (18)

$$\rho \frac{\partial \mathbf{v}_i^{p,s*}}{\partial t} = \sum_{p,s} q^{p,s} \left(\frac{\partial}{\partial x_i} (\lambda \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\mu \sigma_{ij}^{p,s*}) \right) + \left(\sum_{p,s} q^{p,s} \sin(\pi y/2) L_{c_{p,s}}^{(y-1)*} - \cos(\pi y/2) L_{c_{p,s}}^{(y-2)*} \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial x_i} (\chi \sigma_{ll}^{p,s*}) + 2 \frac{\partial}{\partial x_j} (\eta \sigma_{ij}^{p,s*}) \right) \quad (95)$$

Discretised adjoint: In (88), plugging the second line into the first line yields

$$\tilde{\mathbf{v}}_{n+1/2} = A_v \left(A_v \tilde{\mathbf{v}}_{n-1/2} + \Psi_{\text{dis}}^* \tilde{\sigma}_n - \frac{1}{\Delta t} \left((\Psi_{\text{abs}}'^* \tilde{\sigma}_n)^{p,s} - A_v^2 (\Psi_{\text{abs}}'^* \tilde{\sigma}_{n-1})^{p,s} \right) \right). \quad (96)$$

The numerical computation of (95) is the same as (96), except how the PML acts on the temporal gradient of the stress tensor field. These formulae require an explicit computation of the temporal gradient of the stress tensor using finite difference schemes. To avoid this, we used the formulae (18) and (19) (resp. (88)) for the continuous (resp. discretised) adjoint, which are computed the same, as discussed in sections 4 and 5, respectively.

REFERENCES

- [1] S. Arridge, P. Beard, M. Betcke, B. Cox, N. Huynh, F. Lucka, O. Ogunlade, and E. Zhang. Accelerated high-resolution photoacoustic tomography via compressed sensing. *Physics in Medicine & Biology*, 61(24):8908, 2016.
- [2] S. R. Arridge, M. M. Betcke, B. T. Cox, F. Lucka, and B. E. Treeby. On the adjoint operator in photoacoustic tomography. *Inverse Problems*, 32(11):115012, 2016.
- [3] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, Nov 2009.
- [4] Z. Belhachmi, T. Glatz, and O. Scherzer. A direct method for photoacoustic tomography with inhomogeneous sound speed. *Inverse Problems*, 32(4):045005, 2016.
- [5] B. T. Cox, S. Kara, S. R. Arridge, and P. C. Beard. k-space propagation models for acoustically heterogeneous media: Application to biomedical photoacoustics. *J. Acoust. Soc. Am.*, 121:3453–3464, 2007.
- [6] K. Firouzi, B. T. Cox, B. E. Treeby, and N. Saffari. A first-order k-space model for elastic wave propagation in heterogeneous media. *The Journal of the Acoustical Society of America*, 132(3):1271–1283, 2012.

- [7] T. Goldstein, C. Studer, and R. G. Baraniuk. A field guide to forward-backward splitting with a FASTA implementation. *CoRR*, abs/1411.3406, 2014.
- [8] M. Haltmeier, R. Kowar, and L. V. Nguyen. Iterative methods for photoacoustic tomography in attenuating acoustic media. *Inverse Problems*, 33(11):115009, 2017.
- [9] M. Haltmeier, L. Neumann, and S. Rabanser. Single-stage reconstruction algorithm for quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065005, 2015.
- [10] M. Haltmeier and L. V. Nguyen. Analysis of iterative methods in photoacoustic tomography with variable sound speed. *SIAM Journal on Imaging Sciences*, 10(2):751–781, 2017.
- [11] E. Hille and R. Phillips. *Functional analysis and semigroups*. Amer. Math. Soc., 1974.
- [12] S. Holm and S. P. Näsholm. Comparison of fractional wave equations for power law attenuation in ultrasound and elastography. *Ultrasound in Medicine & Biology*, 40(4):695 – 703, 2014.
- [13] Y. Hristova. Time reversal in thermoacoustic tomography—an error estimate. *Inverse Problems*, 25(5):055008, 2009.
- [14] Y. Hristova, P. Kuchment, and L. Nguyen. Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media. *Inverse Problems*, 24(5):055006, 2008.
- [15] C. Huang, L. Nie, R. W. Schoonover, Z. Guo, C. O. Schirra, M. A. Anastasio, and L. V. Wang. Aberration correction for transcranial photoacoustic tomography of primates employing adjunct image data. *Journal of Biomedical Optics*, 17(6):066016–1–066016–8, 2012.
- [16] C. Huang, K. Wang, L. Nie, L. V. Wang, and M. A. Anastasio. Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media. *IEEE Trans. Med. Imag.*, 32(6):1097–1110, 2013.
- [17] A. Javaherian and S. Holman. A multi-grid iterative method for photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 36(3):696–706, March 2017.
- [18] J. Kaipio and E. Somersalo. Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493 – 504, 2007. Special Issue: Applied Computational Inverse Problems.
- [19] H. Kang, S.-W. Lee, E.-S. Lee, S.-H. Kim, and T. G. Lee. Real-time gpu-accelerated processing and volumetric display for wide-field laser-scanning optical-resolution photoacoustic microscopy. *Biomed. Opt. Express*, 6(12):4650–4660, Dec 2015.
- [20] V. Katsnelson and L. V. Nguyen. On the convergence of the time reversal method for thermoacoustic tomography in elastic media. *Applied Mathematics Letters*, 77:79 – 86, 2018.
- [21] R. Kowar. Causality analysis of the fractional state and wave equations based on fractional laplacians. *Wave Motion*, 58:244 – 258, 2015.
- [22] R. Kowar and O. Scherzer. *Attenuation Models in Photoacoustics*, pages 85–130. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [23] C. Li, A. Aguirre, J. Gamelin, A. Maurudis, Q. Zhu, and L. V. Wang. Real-time photoacoustic tomography of cortical hemodynamics in small animals. *Journal of Biomedical Optics*, 15(1):010509–010509–3, 2010.
- [24] F. Mainardi. *Fractional Calculus and Waves in Linear Viscoelasticity: An Introduction to Mathematical Models*. 01 2010.
- [25] K. Mitsunashi, J. Poudel, T. P. Mathethews, A. Gracia-Urbe, L. V. Wang, and M. A. Anastasio. A forward-adjoint operator pair based on the elastic wave equation for use in transcranial photoacoustic computed tomography. *SIAM J. Imaging Sci.*, 10:2022–2048, 2017.
- [26] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.
- [27] G. Pinton, J. Aubry, E. Bossy, M. Muller, M. Pernot, and M. Tanter. Attenuation, scattering, and absorption of ultrasound in the skull bone. *Medical Physics*, 39(1):299–307.
- [28] J. Qian, P. Stefanov, G. Uhlmann, and H. Zhao. An efficient neumann series-based algorithm for thermoacoustic and photoacoustic tomography with variable sound speed. *SIAM J. Imaging Sci.*, 4(3):850–883, 2011.
- [29] A. Rosenthal, V. Ntziachristos, and D. Razansky. Acoustic inversion in optoacoustic tomography: A review. *Current Medical Imaging Reviews*, 9(4):318–336, 2013.
- [30] O. Scherzer and C. Shi. Reconstruction formulas for photoacoustic imaging in attenuating media. *Inverse Problems*, 34(1):015006, 2018.
- [31] P. Stefanov and G. Uhlmann. Thermoacoustic tomography arising in brain imaging. *Inverse Problems*, 27(4), 2011.

- [32] P. Stefanov and Y. Yang. Multiwave tomography with reflectors: Landweber’s iteration. *Inverse Problems and Imaging*, 11(2):373–401, 2017.
- [33] T. L. Szabo and J. Wu. A model for longitudinal and shear wave propagation in viscoelastic media. *The Journal of the Acoustical Society of America*, 107(5):2437–2446, 2000.
- [34] M. Tabei, T. D. Mast, and R. C. Waag. A k-space method for coupled first-order acoustic propagation equations. *J. Acoust. Soc. Am.*, 111(1):53–63, 2002.
- [35] B. E. Treeby and B. T. Cox. k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields. *Journal of Biomedical Optics*, 15(2):021314–021314–12, 2010.
- [36] B. E. Treeby and B. T. Cox. Modeling power law absorption and dispersion for acoustic propagation using the fractional laplacian. *J. Acoust. Soc. Am.*, 127(5):2741–2748, 2010.
- [37] B. E. Treeby and B. T. Cox. Modeling power law absorption and dispersion in viscoelastic solids using a split-field and the fractional laplacian. *The Journal of the Acoustical Society of America*, 136(4):1499–1510, 2014.
- [38] B. E. Treeby, J. Jaros, D. Rohrbach, and B. T. Cox. Modelling elastic wave propagation using the k-wave matlab toolbox. In *2014 IEEE International Ultrasonics Symposium*, pages 146–149, Sept 2014.
- [39] B. E. Treeby, E. Z. Zhang, and B. T. Cox. Photoacoustic tomography in absorbing acoustic media using time reversal. *Inverse Problems*, 26(11):115003, 2010.
- [40] A. A. U. Alqasemi, H. Li and Q. Zhu. Fpga-based reconfigurable processor for ultrafast interlaced ultrasound and photoacoustic imaging. *IEEE Trans. Ultrason., Ferroelect., Freq. Control.*, 59(7):1344–1353, 2012.
- [41] P. White, G. Clement, and K. Hynynen. Longitudinal and shear mode ultrasound propagation in human skull bone. *Ultrasound in Medicine & Biology*, 32(7):1085 – 1096, 2006.
- [42] Z. Xu, Q. Zhu, and L. V. Wang. In vivo photoacoustic tomography of mouse cerebral edema induced by cold injury. *Journal of Biomedical Optics*, 16(6):066020–066020–4, 2011.
- [43] E. Z. Zhang, J. G. Laufer, R. B. Pedley, and P. C. Beard. In vivo high-resolution 3d photoacoustic imaging of superficial vascular anatomy. *Physics in Medicine & Biology*, 54(4):1035, 2009.