VARIATIONAL APPROACHES FOR PHOTO-ACOUSTIC TOMOGRAPHY

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Science and Engineering

2019

Ashkan Javaherian
School of Mathematics

Contents

Word count 71804

List of Tables

List of Figures

The University of Manchester

Ashkan Javaherian
Doctor of Philosophy
Variational approaches for photo-acoustic tomography
January 31, 2019

Photo-acoustic Tomography (PAT) is a hybrid imaging method which simultaneously takes advantage of a high spatial resolution attributed to ultrasound and a rich contrast provided by optics. In PAT, the tissue is irradiated by electromagnetic waves in the visible or near-infrared ranges. A portion of energy from the emitted pulses is absorbed, and induces local pressures which propagate outwards as acoustic waves, and are measured in time by ultrasound detectors. The objective is a reconstruction of a distribution of optical absorption coefficient from the measured data. This inverse problem involves two steps, the first of which is a reconstruction of the initial pressure distribution from the measured data (PAT), and the second step is a reconstruction of optical absorption coefficient from the initial pressure distribution (quantitative PAT). Variational approaches are one of the robust methods for PAT, and are often solved via an iterative implementation of a pair of forward and adjoint operators. In chapter 1, we present a detailed introduction for PAT. In chapter 2, a line search multi-grid method is developed for improving the speed of image reconstruction using variational approaches. We define the forward operator as a linear system of first order wave equations that can be adapted to heterogeneous media, and account for an acoustic attenuation following a frequency power law. We derive the adjoint of this operator using an *adjoint-then-discretise* method. We numerically approximated the forward and adjoint operators using a k-space pseudo-spectral method. In chapter 3, a continuous adjoint was derived for PAT of the brain. We considered a forward operator that describes the propagation of acoustic waves for linear isotropic heterogeneous and lossy elastic media with an acoustic attenuation following a frequency power law. Using a k-space pseudospectral method for an approximation of the forward and adjoint operators, we analytically show that the derived continuous adjoint matches an associated algebraic adjoint. In chapter 4, we solve a single-stage problem of qunatitative PAT for realistic acoustic media. The optical portion of the forward operator is defined using a diffusion approximation (DA) model, and the acoustic portion is described in the same way as chapter 2. We developed two inexact Newton approaches for direct QPAT. Using these approaches, an associated nonlinear objective function is minimised as a sequence of linearised problems using matrix-free Jacobian-based methods.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

I would like to sincerely thank my supervisor, Dr. Sean Holman, for the many broad moral and scientific supports, patient guidance and discussions, and intuitive ideas for my research.

I would also like to offer my special thanks to my official supervisor, Professor Bill Lionheart.

Special thanks to my examiners, Professor Oliver Dorn and Professor Markus Haltmeier, for putting their valuable time on my thesis and their great advice, which definitely improved the quality of this thesis to a high degree.

I am also very grateful to the Faculty of Science and Engineering at the University of Manchester for offering me a Dean's award for my PhD course. Here, I must deeply thank my supervisors, Dr. Sean Holman and Professor Bill Lionheart, for placing their trust in me, and supporting my application.

My thanks to all PhD students and staff at the school of Mathematics for their great helps, supports, and providing a wonderful environment for my research.

A special thank you to my family, especially to my mother, father and sister, whom I haven't seen during my PhD course, for their distant moral supports and patience.

Chapter 1

Introduction

1.1 Introduction

Photo-acoustic tomography (PAT), which is also called Opto-acoustic tomography, is an imaging modality for visualising anatomy and physiology of soft tissues in humans and small animals [80]. This imaging technique combines useful features of light and sound for imaging, the first of which provides a rich contrast and high versatility for a quantitative detection of light absorbing biomolecules, and the latter retains a high spatial resolution because of a low scattering of acoustic waves in tissue media, compared to optical photons [42]. The most interesting absorbing biomolecules in tissue media are hemoglobin, melanin, lipids, DNA-RNA, water and cytochromes [75].

The phenomenon of conversion of light to acoustic sound waves, namely *Photoacoustic*, was first reported by Bell in 1880 [11]. However, it took more than one century for this physical phenomenon to evolve to an idea for biomedical imaging [48, 40, 30]. In PAT, a transparent tissue sample is irradiated by nano-second duration pulses of electromagnetic waves, in the visible or near-infrared ranges [75]. (The latter is used for depth imaging.) A portion of energy from the emitted optical photons is absorbed or scattered by light absorbing structures (chromophores), and is partially converted to heat through a nonradiative relaxation of the excited volume of tissue [75]. The induced heat is proportional to an optical absorption coefficient and photon density at the irradiated point. The generated heat causes a thermal expansion, which induces a rapid local increase in pressure. The local pressure-rises then propagate outwards

as acoustic waves, and are detected in time by ultrasound detectors located on a detection surface outside the sample. The inverse problem is to reconstruct an image of the optical absorption coefficient, given the measured pressure at the boundary [61].

The physics of the PAT problem involves two steps, the first of which is the generation of absorbed optical energy, and the second is the propagation of the induced local pressure as acoustic waves. Correspondingly, the inverse problem involves two steps: an acoustic step for reconstructing an image of initial pressure distribution (or absorbed optical energy) from a set of time series of pressure data at the boundary, and an optical step for reconstruction of a quantitative image of optical absorption coefficient using the calculated initial pressure distribution [61].

The acoustic portion of PAT can be fit into a class of Thermo-acoustic tomography (TAT) techniques, in which a broad spectrum of electromagnetic waves, i.e., visible, near-infrared or radio-frequency ranges, can be used for an induction of acoustic waves. Therefore, the image reconstruction techniques for the acoustic portion of PAT and TAT are equivalent.

In this chapter, we first introduce the image reconstruction techniques for the acoustic portion of PAT/TAT, and then briefly explain the approaches for an extension of the acoustic problem to quantitative photo-acoustic tomography, which combines the optical and acoustic portions of the problem. Throughout this thesis, following the existing literature in the field, the acoustic portion of the problem is referred to as PAT, and the composite *opto-acoustic* problem is referred to as quantitative PAT (QPAT).

In section 1.2, we summarise back-projection inversion algorithms for PAT that are solved in a time domain using Green's function approaches. In section 1.3, we introduce back-projection algorithms that are solved in a frequency domain. These approaches are often much more efficient than time-domain methods. In section 1.4, we explain approaches for an extension of back-projection algorithms using Green's function techniques to acoustically heterogeneous media. In section 1.5, we introduce the time reversal approach, which is known as the least restrictive approach among non-iterative inversion algorithms for PAT, to the best of our knowledge. In section 1.6, we explain iterative inversion algorithms, which are based on an iterative implementation of an operator modelling the propagation and measurement of photo-acoustic

waves, and an associated back-projection (often adjoint) operator. (A common class of these methods are variational approaches, which are used in our study.) In section 1.7, a numerical computation of an arising forward operator using a *k-space pseudo-spectral* method is explained. Section 1.8 is devoted to the optimisation algorithms we have used for solving the inverse problems of PAT arising in chapters 2 and 3. In section 1.9, we introduce quantitative photo-acoustic tomography, and also explain approaches for modelling the optical portion of this inverse problem. In section 1.10, we outline the following chapters of the thesis. Finally, in section 1.11, we explain the contributions of the candidate and supervisor for this thesis.

1.2 Time-domain back-projection algorithms for homogeneous media

In the context of TAT, the first image reconstruction techniques were based on localising thermoacoustic sources via linear [26] or sector scans [45] of the tissue sample using focused ultrasound detectors. In [76, 26], 2D reconstructions were performed from multiple 1D images obtained from a linear scan of the sample using focused ultrasonic transducers. In [30], a synthetic weighted delay-and-sum algorithm was used for a 3D reconstruction from multiple in-plane detections of acoustic waves. Applying a phased-array technique, the measured acoustic signals are temporally shifted in order to compensate for the different transit times of acoustic waves [30].

The next class of image reconstruction techniques for PAT are based on back-projection algorithms using Green's function approaches. Before explaining the image reconstruction techniques for PAT, we introduce the *heat conduction equation*

$$\rho C_p \frac{\partial}{\partial t} T(\boldsymbol{r}, t) - \nabla \cdot (K \nabla T(\boldsymbol{r}, t)) = H(\boldsymbol{r}, T), \tag{1.1}$$

where K is the thermal diffusion coefficient, ρ is the ambient density, C_p is the specific heat capacity, $T(\boldsymbol{r}, t)$ is the rise in the temperature because of the absorption of electromagnetic wave, and $H(\boldsymbol{r}, t)$ is the heat energy per unit volume and per unit time deposited in the irradiated fluid.

Since the electromagnetic pulse duration is much shorter than the thermal diffusion time, the diffusion term in the *heat conduction equation* is about six orders of

magnitude less than the term corresponding to the time derivative of temperature, and is thus neglected [81]. This assumption is called *thermal confinement*.

We now define the acoustic fields, which are a vector-valued acoustic displacement $u(\boldsymbol{r}, t)$ and a scalar-valued acoustic pressure $p(\boldsymbol{r}, t)$, where $\boldsymbol{r} \in \mathbb{R}^d$ is the spatial position with d the number of spatial dimensions, and t denotes time. For an acoustically homogeneous medium, the expansion equation is described by [78]

$$\nabla \cdot u(\boldsymbol{r}, t) = -\frac{p(\boldsymbol{r}, t)}{\rho c^2} + \beta T(\boldsymbol{r}, t), \tag{1.2}$$

where c is the sound speed, and β denotes the isobaric volume expansion coefficient.

In this section, we briefly summarise back-projection formulae using Green's function approaches for solving the PAT problem in acoustically homogeneous media, i.e., c and ρ are constant with respect to $\boldsymbol{r}$. To do this, we start with derivation of an associated forward operator.

The propagation of the generated acoustic waves in an acoustically homogeneous medium obeys a linear inviscid force equation [78]

$$\rho \frac{\partial^2}{\partial t^2} u(\boldsymbol{r}, t) = -\nabla p(\boldsymbol{r}, t). \tag{1.3}$$

Using equations (1.1), (1.2) and (1.3), the propagation of acoustic waves induced by an optical absorbed energy $H(\boldsymbol{r}, t)$ follows [78]

$$\nabla^2 p(\boldsymbol{r}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(\boldsymbol{r}, t) = -\frac{\beta}{C_p} \frac{\partial}{\partial t} H(\boldsymbol{r}, t). \tag{1.4}$$

A general solution of (1.4) can be obtained in terms of Green's functions using [17]

$$\nabla^2 G - \frac{1}{c^2} \frac{\partial^2 G}{\partial t^2} = -\delta(\boldsymbol{r} - \boldsymbol{r}')\delta(t - t'), \tag{1.5}$$

where G is the free-space Green's function. The Green's function G in dimension $d = 3$ is

$$G\left(\boldsymbol{r}, t; \boldsymbol{r}', t'\right) = \frac{\delta\left(|\boldsymbol{r} - \boldsymbol{r}'| - c(t - t')\right)}{4\pi |\boldsymbol{r} - \boldsymbol{r}'|}. \tag{1.6}$$

Using (1.6), the solution of (1.4) will be in the form [78]

$$p(\boldsymbol{r}, t) = \frac{\beta}{4\pi C_p} \iiint \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} \frac{\partial}{\partial t'} H(\boldsymbol{r}', t') \, d\boldsymbol{r}' \Big|_{t' = t - (|\boldsymbol{r} - \boldsymbol{r}'|/c)}. \tag{1.7}$$

Following [78], the heating function is now expressed as a product of a function of absorbed energy in $\boldsymbol{r}$ and an illumination function in t. This yields

$$H(\boldsymbol{r}, t) = H(\boldsymbol{r})\tilde{I}(t), \tag{1.8}$$

where $H(\boldsymbol{r})$ is the heat deposited in the fluid per unit volume, and $\tilde{I}(t)$ is a temporal illumination function. (For brevity, we used the same notations for $H(\boldsymbol{r}, t)$ and $H(\boldsymbol{r})$.) Plugging (1.8) into (1.7), for a given $H(\boldsymbol{r})$, $p(\boldsymbol{r}, t)$ is calculated as

$$p(\boldsymbol{r}, t) = \frac{\beta}{4\pi C_p} \iiint \frac{H(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \frac{\partial}{\partial t'} \tilde{I}(t') \, d\boldsymbol{r}' \Big|_{t'=t-(|\boldsymbol{r}-\boldsymbol{r}'|/c)}. \tag{1.9}$$

Under the condition of *stress confinement*, for which the duration of an absorption of electromagnetic energy is much shorter than the time required by the sound to travel across the heated volume, $I(t)$ is assumed a Dirac delta function, i.e., $I(t) = \delta(t)$ [78]. Using this assumption, the pressure on a spherical detection surface defined by $r = r_0$ will be in the form

$$p(\boldsymbol{r}_0, t) = \frac{\beta}{4\pi C_p} \iiint H(\boldsymbol{r}) \frac{\delta'\left(t - \frac{|\boldsymbol{r}_0 - \boldsymbol{r}|}{c}\right)}{|\boldsymbol{r}_0 - \boldsymbol{r}|} \, d\boldsymbol{r}. \tag{1.10}$$

An inverse problem associated with (1.10) is the calculation of the absorbed optical energy $H(\boldsymbol{r})$, given $p(\boldsymbol{r}_0, t)$ for all $\boldsymbol{r}_0$ in some set. This amounts to the acoustic portion of the inverse problem of QPAT, and is referred to as PAT in this thesis.

The relation between initial pressure and heat. Under an isochoric condition, i.e., the time for absorption of the electromagnetic wave is much shorter than the time for a change in density of the heated volume, the initial pressure distribution $p_0(\boldsymbol{r})$ follows [17]

$$p(\boldsymbol{r}, t = 0) = p_0(\boldsymbol{r}) = \Gamma(\boldsymbol{r})H(\boldsymbol{r}), \tag{1.11}$$

where Γ is Grüneisen parameter, and is defined by

$$\Gamma = \frac{\beta c^2}{C_p}. \tag{1.12}$$

1.2.1 Back-projection algorithm using spherical harmonics

In [78], a back-projection algorithm was developed for PAT using the spherical harmonic function basis in a frequency domain [78]. Before explaining this approach, we

introduce the temporal Fourier transform (FT) of $p(\boldsymbol{r}_0, t)$ in the form

$$\tilde{p}(\boldsymbol{r}_0, w) = \int_{-\infty}^{+\infty} p(\boldsymbol{r}_0, t) e^{iwt} dt, \tag{1.13}$$

where w denotes the temporal frequency, and i is the imaginary number. Also, an inverse of the temporal FT is defined by

$$p(\boldsymbol{r}_0, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{p}(\boldsymbol{r}_0, w) e^{-iwt} dw. \tag{1.14}$$

Using (1.13), taking an FT from both sides of (1.10), together with using the dispersion relation for electromagnetic waves in vacuum ($k = w/c$), yields [78]

$$\tilde{p}(\boldsymbol{r}_0, w) = -iw \frac{\beta}{C_p} \iiint H(\boldsymbol{r}) \frac{e^{ik|\boldsymbol{r}_0 - \boldsymbol{r}|}}{4\pi |\boldsymbol{r}_0 - \boldsymbol{r}|} d\boldsymbol{r}, \tag{1.15}$$

where k denotes the spatial frequency (wavenumber) [78].

Now, we briefly introduce the principle of spherical harmonics, which will be used later for derivation of the back-projection algorithm. A complete orthogonal integral of spherical harmonic functions $Y_l^m(\theta_0, \varphi_0)$ is defined by

$$\iint_{\Omega_0} Y_l^m(\theta_0, \varphi_0) Y_k^{n*}(\theta_0, \varphi_0) d\Omega_0 = \delta_{l,k}\ \delta_{m,n} \tag{1.16}$$

where $d\Omega_0 = \sin(\theta_0) d\theta_0 d\varphi_0$, and $*$ denotes the complex conjugate [78]. Here, $\theta \in (0, \pi]$ is the polar angle and $\varphi_0 \in (0, 2\pi]$ is the azimuthal angle. Also, the Legendre polynomial is defined by [78]

$$\mathcal{P}_l(\boldsymbol{n} \cdot \boldsymbol{n}_0) = \frac{4\pi}{2l+1} \sum_{m=-l}^{+l} Y_l^m(\theta, \varphi) Y_l^{m*}(\theta_0, \varphi_0), \tag{1.17}$$

where the unit vectors $\boldsymbol{n} = \boldsymbol{r}/r$ and $\boldsymbol{n}_0 = \boldsymbol{r}_0/r_0$ are aligned by the directions θ, φ and θ_0, φ_0, respectively [78]. Also, $r = |\boldsymbol{r}|$ and $r_0 = |\boldsymbol{r}_0|$. Using a spherical harmonic function basis, the following replacement is enforced to (1.15).

$$\frac{e^{ik|\boldsymbol{r}_0 - \boldsymbol{r}|}}{4\pi |\boldsymbol{r}_0 - \boldsymbol{r}|} = \frac{ik}{4\pi} \sum_{l=0}^{+\infty} (2l+1)\, j_l(kr) h_l^{(1)}(kr_0) \mathcal{P}_l(\boldsymbol{n} \cdot \boldsymbol{n}_0),\ (k > 0). \tag{1.18}$$

Here, $j_l(\cdot)$ and $h_l^{(1)}(\cdot)$ are spherical Bessel function and Hankel function of the first kind, respectively [78]. By some calculations given in [78], an exact solution of $H(\boldsymbol{r})$ given $\tilde{p}(\boldsymbol{r}_0, w)$ can be obtained using

$$H(\boldsymbol{r}) = \frac{C_p}{2\pi^2 c\beta} \iint_{\Omega_0} \int_0^{+\infty} \tilde{p}(\boldsymbol{r}_0, w) \cdot \sum_{m=0}^{+\infty} \frac{(2m+1) j_m(kr)}{h_m^{(1)}(kr_0)} P_m(\boldsymbol{n} \cdot \boldsymbol{n}_0)\ dk\ d\Omega_0. \tag{1.19}$$

For further details on derivation of this formula, the reader is referred to [78]. The equation (1.19) gives an exact inverse solution of (1.10), but it is computationally expensive because it involves a summation of series which may be very slowly converging. [78].

1.2.2 Modified (filtered) Back-projection algorithm

Since the low-frequency components of the photo-acoustic (PA) waves do not contribute to the spatial resolution, they are not useful for imaging, and can be removed by a filter. Because of this, the radius of the detection surface r_0 is much larger than the wavelengths of PA waves. Using this, an assumption $|k|r_0 \gg 1$ is used to derive an asymptotic form of the Hankel function as [78]

$$h_m^{(1)}(kr_0) \approx \frac{1}{h_m^{(2)}(kr_0)} \left(\frac{1}{(kr_0)^2} + O(\frac{1}{(kr_0)^4}) \right), \tag{1.20}$$

where $h_l^{(2)}(\cdot)$ is the spherical Hankel function of the second kind. Using (1.19) and (1.20), $H(r)$ can be approximated by the formula [78]

$$H(\boldsymbol{r}) = -\frac{r_0^2 C_p}{2\pi c^3 \beta} \iint_{\Omega_0} \frac{1}{2\pi} \int_{-\infty}^{+\infty} (-iw)\tilde{p}(\boldsymbol{r}_0, w) \cdot \frac{e^{-iw\frac{|\boldsymbol{r}_0 - \boldsymbol{r}|}{c}}}{|\boldsymbol{r}_0 - \boldsymbol{r}|}\, dw \;\; d\Omega_0 \tag{1.21}$$

By taking an inverse FT of $\tilde{p}(\boldsymbol{r}_0, w)$ using (1.14), (1.21) can be solved the time domain. This finally gives

$$H(\boldsymbol{r}) = -\frac{r_0^2 C_p}{2\pi c^4 \beta} \iint_{\Omega_0} \frac{1}{t} \frac{\partial p(\boldsymbol{r}_0, t)}{\partial t}\bigg|_{t=|\boldsymbol{r}_0 - \boldsymbol{r}|/c} d\Omega_0. \tag{1.22}$$

Exact inversion formulae were also derived for a detection surface with planar and cylindrical geometries. As opposed to the back-projection formula in (1.22), the inversion formulae for these geometries are solved in a frequency domain. The reader is referred to [81] for a planar detection surface, and to [84] for a cylindrical detection surface.

1.2.3 Universal back-projection algorithm

In [79], a *universal back-projection* formula, which can be adapted to three detection geometries, i.e, spherical, planar and cylindrical, is derived. Following [79], we apply

an FT and its inverse using a variable $\bar{t} = ct$. By plugging (1.11) and the dispersion relation into (1.15), an FT of $p(\boldsymbol{r}_0, \bar{t})$ obeys

$$\tilde{p}(\boldsymbol{r}_0, k) = -ik \iiint \tilde{G}_k^{(\text{out})}(\boldsymbol{r}', \boldsymbol{r}_0) p_0(\boldsymbol{r}') \; d\boldsymbol{r}', \tag{1.23}$$

where

$$\tilde{G}_k^{(\text{out})}(\boldsymbol{r}', \boldsymbol{r}_0) = \frac{e^{ik|\boldsymbol{r}' - \boldsymbol{r}_0|}}{4\pi|\boldsymbol{r}' - \boldsymbol{r}_0|}. \tag{1.24}$$

Using this, a back-projection formula for a spherical, planar or cylindrical (universal) detection surface S is derived, i.e., [79]

$$p_0^b(r) = \frac{1}{\pi} \int_S \int_{-\infty}^{+\infty} \tilde{p}(\boldsymbol{r}_0, k) \Big[\boldsymbol{n}_0^s \cdot \nabla_{\boldsymbol{r}_0} \tilde{G}_k^{(\text{in})}(\boldsymbol{r}, \boldsymbol{r}_0) \Big] \; dk \;\; dS, \tag{1.25}$$

where $\boldsymbol{n}_0^s$ is the normal of surface S that points to the source, and $\nabla_{\boldsymbol{r}_0}$ is the gradient with respect to variable $\boldsymbol{r}_0$. Also,

$$\tilde{G}_k^{(\text{in})}(\boldsymbol{r}, \boldsymbol{r}_0) = \frac{e^{-ik|\boldsymbol{r} - \boldsymbol{r}_0|}}{4\pi|\boldsymbol{r} - \boldsymbol{r}_0|}. \tag{1.26}$$

Taking an inverse FT of $\tilde{p}(\boldsymbol{r}_0, k)$ gives an inverse formula in the time domain in the form

$$p_0^b(\boldsymbol{r}) = -\frac{2}{\bar{\Omega}_0} \nabla \cdot \int_S \left[\frac{p(\boldsymbol{r}_0, \bar{t})}{\bar{t}} \right]_{\bar{t} = |\boldsymbol{r} - \boldsymbol{r}_0|} \boldsymbol{n}_0^s \; dS_0, \tag{1.27}$$

where $\bar{\Omega}_0$ is a solid angle of a full-view detection surface S_0 with respect to a reconstruction point inside S_0. (For a planar detection geometry, $\Omega_0 = 2\pi$, and for spherical and cylindrical geometries, $\Omega_0 = 4\pi$). A similar approach was taken in [23] for a spherical detection geometry. Equation (1.27) can be simplified to

$$p_0^b(\boldsymbol{r}) = \int_{\boldsymbol{\Omega}_0} b(\boldsymbol{r}_0, \bar{t} = |\boldsymbol{r} - \boldsymbol{r}_0|) d\bar{\Omega}_0 / \bar{\Omega}_0, \tag{1.28}$$

where

$$b(\boldsymbol{r}_0, \bar{t}) = 2p(\boldsymbol{r}_0, \bar{t}) - 2\bar{t} \frac{\partial p(\boldsymbol{r}_0, \bar{t})}{\partial \bar{t}} \tag{1.29}$$

is the back-projection term for the measured data at position $\boldsymbol{r}_0$, and $d\bar{\Omega}_0$ is the solid angle for a detection element dS_0 with respect to a reconstruction point at position $\boldsymbol{r}$, and is defined by

$$d\bar{\Omega}_0 = \frac{dS_0}{|\boldsymbol{r} - \boldsymbol{r}_0|^2} \cdot \Big[\boldsymbol{n}_0^s \cdot \frac{\boldsymbol{r} - \boldsymbol{r}_0}{|\boldsymbol{r} - \boldsymbol{r}_0|} \Big]. \tag{1.30}$$

From (1.30), $d\bar{\Omega}_0 / \bar{\Omega}_0$ can be expressed as a weighting factor that weights the contribution to reconstruction of the source at point $\boldsymbol{r}$ from a detection element dS_0 [79].

1.2.4 Far-field approximation for arbitrary detection surface

In biomedical applications, the detection surface may differ from the regular geometries mentioned above. For instance, it may be required that the detection surface follows the shape of the imaged organ in order to reduce the propagation distance of acoustic waves as much as possible. This may be needed, for exmaple for minimising the acoustic attenuation [13]. Note that the favourable attenuation of high frequency parts of acoustic waves deteriorates the spatial resolution. To meet this requirement, a far-field approximation of (1.22) that does not depend on the actual shape of the detection surface was derived in [13]. This approach uses a Fraunhofer approximation of (1.15) via enforcing a replacement $|\boldsymbol{r}_0 - \boldsymbol{r}| = r_0 - \boldsymbol{r} \cdot \hat{\boldsymbol{r}}_0$ for the exponent, where $\hat{\boldsymbol{r}}_0 = \boldsymbol{r}_0/r_0$, and a replacement $|\boldsymbol{r}_0 - \boldsymbol{r}| = r_0$ for the denominator. Using these replacements, together with (1.11) and (1.12), (1.15) can be rewritten as

$$\tilde{p}(\boldsymbol{r}_0, w) \approx -\frac{ik}{4\pi c}\frac{e^{ikr_0}}{r_0}\int p_0(\boldsymbol{r})e^{-ik\boldsymbol{r}\cdot\hat{\boldsymbol{r}}_0}\, d\boldsymbol{r}. \tag{1.31}$$

Defining $\boldsymbol{k} := k\hat{\boldsymbol{r}}_0$, together with $\hat{p}_0(\boldsymbol{k}) := \int p_0(\boldsymbol{r})e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}d\boldsymbol{r}$, gives

$$\tilde{p}(\boldsymbol{r}_0, w) \approx -\frac{ik}{4\pi c}\frac{e^{ikr_0}}{r_0}\hat{p}_0(\boldsymbol{k}). \tag{1.32}$$

Equation (1.32) gives an operator that links the spatial FT of the absorbed energy $p_0(\boldsymbol{r})$ in direction $\hat{\boldsymbol{r}}_0$ to a one-dimensional temporal FT of the measured data $p(\boldsymbol{r}_0, t)$. Using the Fourier slice theorem for the classical Radon transform, the one-dimensional FT of a function at k of a projection orthogonal to $\hat{\boldsymbol{r}}_0$ is equal to the three dimensional FT of the same function at $k\hat{\boldsymbol{r}}_0$ [13]. Using this theorem, together with applying an inverse of one-dimensional FT to (1.32), results in

$$\begin{aligned} p(\boldsymbol{r}_0, t) &\approx \frac{1}{4\pi r_0 c}\frac{\partial}{\partial t}g\left(\hat{\boldsymbol{r}}_0, r_0 - ct\right) \\ &= \frac{1}{4\pi c}\frac{\partial}{\partial t}\left(\int_{\boldsymbol{r}\cdot\hat{\boldsymbol{r}}_0 = r_0 - ct}\frac{p_0(\boldsymbol{r})}{r_0}\, dS(\boldsymbol{r})\right), \end{aligned} \tag{1.33}$$

where $g\left(\hat{\boldsymbol{r}}_0, r_0 - ct\right)$ is a classical three-dimension Radon Transform of p_0, i.e., an integral of p_0 over a plane with a normal vector $\hat{\boldsymbol{r}}_0$ and distance $r_0 - ct$ from the origin [13]. Using the above and given $p(\boldsymbol{r}_0, t)$, the inverse problem can be solved as follows:

1- Take a temporal FT for a calculation of $\tilde{p}(\boldsymbol{r}_0, w)$.

2- Calculate $\hat{p}_0(\boldsymbol{k})$ from $\tilde{p}(\boldsymbol{r}_0, w)$ using (1.32).

3- Calculate $p_0(\boldsymbol{r})$ by taking an inverse spatial FT of $\hat{p}_0(\boldsymbol{k})$.

These steps can be combined in order to solve the inverse problem directly in a time domain [13]. Using a classical Radon Tranform, a generalised inversion formula can be obtained, i.e.,

$$p_0(\boldsymbol{r}) = \frac{1}{8\pi^2} \int_S g'' \left(\hat{\boldsymbol{r}}_0, \boldsymbol{r} \cdot \boldsymbol{r}_0\right) d\bar{\Omega}_0(\hat{\boldsymbol{r}}_0), \tag{1.34}$$

where $\boldsymbol{n}(\boldsymbol{r}_0)$ is the normal vector at $\boldsymbol{r}_0$ that points outside the surface S, and $d\bar{\Omega}_0(\hat{\boldsymbol{r}}_0) := \boldsymbol{n}(\boldsymbol{r}_0) \cdot \hat{\boldsymbol{r}}_0 dS/r_0^2$ is the solid angle of the surface area element dS with respect to the origin [13].

Using (1.33) and (1.34), an inversion formula can be obtained, i.e.,

$$p_0(\boldsymbol{r}) \approx -\frac{1}{2\pi c} \int_S r_0 \frac{\partial p(\boldsymbol{r}_0, t)}{\partial t}\bigg|_{ct=r_0-\boldsymbol{r}\cdot\hat{\boldsymbol{r}}_0} d\bar{\Omega}_0(\hat{\boldsymbol{r}}_0). \tag{1.35}$$

Equation (1.35) gives an exact solution for any rotationally symmetric p_0 with origin as the centre of rotation. Equation (1.35) for any centre of rotation at position $\boldsymbol{r}$ inside S follows

$$\begin{aligned} p_0(\boldsymbol{r}) &\approx -\frac{1}{2\pi c} \int_S |\boldsymbol{r}_0 - \boldsymbol{r}| \frac{\partial p(\boldsymbol{r}_0, t)}{\partial t}\bigg|_{ct=|\boldsymbol{r}_0-\boldsymbol{r}|} d\boldsymbol{\Omega}_r(\boldsymbol{r}_0) \\ &= -\frac{1}{2\pi} \int_S t \frac{\partial p(\boldsymbol{r}_0, t)}{\partial t}\bigg|_{ct=|\boldsymbol{r}_0-\boldsymbol{r}|} d\bar{\Omega}_r(\boldsymbol{r}_0), \end{aligned} \tag{1.36}$$

where $d\bar{\Omega}_r(\boldsymbol{r}_0)$ is the solid angle element for a detection surface element dS with respect to a centre of rotation at $\boldsymbol{r}$, and is approximated as

$$d\Omega_r(\boldsymbol{r}_0) \approx \frac{dS}{|\boldsymbol{r}_0 - \boldsymbol{r}|^2} \left[\boldsymbol{n}(\boldsymbol{r}_0) \cdot \frac{\boldsymbol{r}_0}{|\boldsymbol{r}_0|}\right]. \tag{1.37}$$

This is derived using an approximation $r \ll r_0$ to (1.30).

1.3 Frequency-domain techniques

Frequency-domain inversion approaches are based on solving the inverse problem in a spatial and temporal frequency domain, and then transforming the solution back to the original spatial and temporal domain. Using this approach for $d = 3$, the acoustic field is expressed as an infinite sum of a product of functions [55]

$$p(\boldsymbol{r}, t) = \sum_n a_n f_{1,n}(r_1 := x)\ f_{2,n}(r_2 := y)\ f_{3,n}(r_3 := z) f_{4,n}(t). \tag{1.38}$$

For the forward problem, the pressure field $p(\boldsymbol{r}, t)$ is known over d spatial Cartesian coordinates at a time instant $t = 0$, whereas for the inverse problem, the pressure field is known in two spatial Cartesian coordinates (detection surface) and over time [55]. Using this approach, the forward problem is solved by a direct calculation of the pressure field at a detection surface and over time using a decomposition of fields in Cartesian coordinates in a spatial frequency domain, and the inverse problem is solved by a direct reconstruction of the field in spatial coordinates at the time origin, given $p(\boldsymbol{r}_0, t)$ [55]. (See [41, 81] for planar detection surface, or [84] for cylindrical detection surface.) Here, we restrict our attention on the planar detection geometry, because this is more favourable than cylindrical geometry because of mathematical simplicity and numerical efficiency [55]. The approach we will explain in the sequel has been proposed in [17]. Using (1.6), (1.10), (1.11) and (1.12), the calculation of $p(\boldsymbol{r}, t)$ from $p_0(\boldsymbol{r})$ can be expressed as

$$p(\boldsymbol{r}, t) = \frac{1}{c^2} \int \; p_0(\boldsymbol{r}') \frac{\partial G}{\partial t} \left(\boldsymbol{r}, t; \boldsymbol{r}', t'\right) d\boldsymbol{r}'. \tag{1.39}$$

The k-space method uses an FT of the free-space Green's function G with respect to both $\boldsymbol{r}$ and t. Using equation (1.5), the Green's function satisfies

$$-k^2 G(w, \boldsymbol{k}) + \frac{w^2}{c^2} G(w, \boldsymbol{k}) = -e^{-i\boldsymbol{k}\cdot\boldsymbol{r}'} e^{iwt'}, \tag{1.40}$$

where $\boldsymbol{k} = (k_x, k_y, k_z)$ is the wavenumber vector, and w is the temporal frequency. Also, a spatio-temporal inverse FT of $G(k, w)$ satisfies

$$G(\boldsymbol{r}, t; \boldsymbol{r}', t') = \frac{1}{(2\pi)^4} \int \int \frac{e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')} e^{-iw(t-t')}}{k^2 - (w/c)^2} dw d\boldsymbol{k}, \tag{1.41}$$

where $k = |\boldsymbol{k}|$. Using (1.41), the Green's function is expressed as a sum of plane waves with direction defined by $\boldsymbol{k}$ and temporal frequency w [17]. Now, we will take two approaches for calculation of $p(\boldsymbol{r}, t)$ from $p_0(\boldsymbol{r})$ using (1.41).

1.3.1 Calculation of the pressure field everywhere at a single time instant

In the first approach, the pressure field at a single time instant t is calculated in one step as follows. An integral of (1.41) with respect to w using Cauchy's residue theorem

for handling the singularity of G, together with setting $t' = 0$, gives for $t > 0$

$$G(\boldsymbol{r}, t; \boldsymbol{r}') = \frac{c}{(2\pi)^3} \int \frac{\sin(ckt)}{k} e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')} d\boldsymbol{k}, \tag{1.42}$$

Plugging (1.42) into (1.39) gives

$$p(\boldsymbol{r}, t) = \frac{1}{(2\pi)^3} \iint p_0(\boldsymbol{r}') \cos(ckt) e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')} d\boldsymbol{k}\ dr'. \tag{1.43}$$

Using the above equation, $p(\boldsymbol{r}, t)$ can be calculated from $p_0(\boldsymbol{k})$ using

$$p(\boldsymbol{r}, t) = \frac{1}{(2\pi)^3} \int p_0(\boldsymbol{k}) \cos(ckt) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} d\boldsymbol{k}, \tag{1.44}$$

where $p_0(\boldsymbol{k})$ is calculated by

$$p_0(\boldsymbol{k}) = \int p_0(\boldsymbol{r}) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} d\boldsymbol{r}. \tag{1.45}$$

Using this approach, the pressure field at time t is directly calculated using the exact time propagator $\cos(ckt)$ without a need for calculation of the field over time [17].

1.3.2 Calculation of time series of pressure field on a chosen line or plane

The second model is useful for cases in which the time series of measurement is required at only a number of points placed on a line or plane in the field, for example a simulation of the signals measured by an array of detectors. To derive this model, we remove singularities in the integrand in (1.41) via taking an integral over the vertical component of k, rather than an integral over w [17]. Because of a decomposition of Cartesian coordinates in a frequency domain, we will use separate integrals for k_x, k_y and k_z. This gives

$$\int \frac{e^{\mathbf{i}\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')} e^{-iwt}}{(k_z - \zeta)(k_z + \zeta)} dk_z, \tag{1.46}$$

where,

$$\zeta = \begin{cases} \operatorname{sgn}(w) \left((w/c)^2 - k_r^2\right)^{1/2}, & |w/c| \geq k_r \\ +i \left(k_r^2 - (w/c)^2\right)^{1/2}, & |w/c| < k_r \end{cases} \tag{1.47}$$

with sgn(·) the signum function. Applying the Cauchy residue theorem for solving (1.46) gives the Green's function for $t > 0$ as

$$G(\boldsymbol{r}, t; \boldsymbol{r}') = \frac{i\pi}{(2\pi)^4} \iiint \left(\frac{1}{\zeta}\right) e^{i\left[k_x(x-x')+k_y(y-y')+\zeta|z-z'|-wt\right]}\, dw\, dk_x\, dk_y. \tag{1.48}$$

Equation (1.48) gives an expression of G as a sum of plane waves in x, y and t, whereas in (1.42), G is expressed as a sum of plane waves in $\boldsymbol{r} = (x, y, z)$. Taking a temporal derivative of G gives

$$\frac{\partial G}{\partial t} = \frac{\pi}{(2\pi)^4} \iiint \left(\frac{w}{\zeta}\right) e^{i\left[k_x(x-x')+k_y(y-y')+\zeta|z-z'|-wt\right]}\, dw\, dk_x\, dk_y. \tag{1.49}$$

Plugging (1.49) into (1.39), $p(\boldsymbol{r}, t)$ is calculated as

$$p(\boldsymbol{r}, t) = \frac{1}{16\pi^3 c^2} \int_{\mathbb{R}^3} \iiint \left(\frac{w}{\zeta}\right) p_0(\boldsymbol{r}') e^{i\left[k_x(x-x')+k_y(y-y')+\zeta|z-z'|-wt\right]} dw\, dk_x dk_y d\boldsymbol{r}'. \tag{1.50}$$

Since (1.50) gives the pressure field on a z plane, we can set $z = 0$.

Case 1 (real ζ): For $|w/c| \geq k_r$,

$$\int_{|w|\geqslant ck_r} \left|\frac{w}{\zeta}\right| e^{i\zeta|z-z'|} e^{-iwt}\, dw = 2\Re\left\{\int_{ck_r}^{+\infty} \left(\frac{w}{\zeta}\right) e^{i\zeta|z-z'|} e^{-iwt}\, dw\right\}, \tag{1.51}$$

where $\Re\{\cdot\}$ denotes the real part. Using $z = 0$, together with plugging (1.51) into (1.50), gives

$$p_{\text{prop}}(x, y, t) = \frac{1}{(2\pi)^3 c^2} \Re\left\{\int_{w=ckr}^{w=+\infty} \iint \left(\frac{w}{\zeta}\right) p_0(k_x, k_y, w) e^{i(k_x x + k_y y - wt)} dk_x dk_y\, dw\right\}, \tag{1.52}$$

where p_{prop} denotes the *propagating* plane waves, i.e., waves that propagate away from the source, and arrive at the acoustic far field. Also, $p_0(k_x, k_y, w)$ is determined by an interpolation from $p_0(k_x, k_y, \zeta)$, which is calculated using

$$p_0(k_x, k_y, \zeta) = \iiint p_0(\boldsymbol{r}') e^{-i(k_x x' + k_y y' - \zeta|z'|)}\, d\boldsymbol{r}' \tag{1.53}$$

The above equation can be modified in the form

$$p_0(k_x, k_y, \zeta) = \iiint p_0(\boldsymbol{r}') e^{-i(k_x x' + k_y y')} \cos(\zeta|z'|)\, d\boldsymbol{r}', \tag{1.54}$$

Using this, the propagating plane waves can be written as a cosine transform in the form

$$p_{\text{prop}}(x, y, t) = \frac{1}{(2\pi)^3 c^2} \int_{w=ck_r}^{w=+\infty} \iint \left(\frac{w}{\zeta}\right) p_0(k_x, k_y, w) e^{i(k_x x + k_y y)} \cos(wt) dk_x dk_y\, dw. \tag{1.55}$$

Case 2 (imaginary ζ): For$|w/c| < k_r$, ζ is imaginary. Using this condition, the integral over w will be in the form

$$\int_{|w|<ck_r} \left|\frac{w}{\zeta}\right| e^{i\zeta|z-z'|} e^{-iwt}\, dw = -2i \int_0^{ck_r} \left(\frac{w}{\zeta}\right) e^{i\zeta|z-z'|} \sin(wt)\, dw. \tag{1.56}$$

The same as in case 1, substituting (1.56) into (1.50), together with setting $z = 0$, results in

$$p_{\text{evan}}(x,y,t) = \frac{-i}{(2\pi)^3 c^2} \int_{w=0}^{w=ck_r} \iint \left(\frac{w}{\zeta}\right) p_0(k_x,k_y,w) e^{i(k_x x + k_y y)} \sin(wt) dk_x dk_y\, dw, \tag{1.57}$$

where p_{evan} denotes the *evanescent* plane waves, i.e., waves that decay exponentially with z, and thus do not contribute to the field beyond a few wavelengths of the source. However, they must be considered for the fields close to the source [17]. Also, using the fact that ζ is purely imaginary for this case,

$$p_0(k_x,k_y,\zeta) = \iiint p_0(r') e^{-i(k_x x' + k_y y')} e^{-|\zeta z'|}\, d\boldsymbol{r}'. \tag{1.58}$$

The inverse problem is to reconstruct $p(x,y,z,t=0)$, given $p = p(x,y,t)$ in an infinite plane $z = 0$ for all times $t > 0$. This can be done by inverting a cosine transform in (1.55). (See [41]).

1.4 Effects of acoustic Heterogeneity

An important assumption for the inversion algorithms introduced above is that the sound speed c is homogeneous, i.e., c is not varying with r. For many biological tissues, this assumption is not exactly true. For example, the sound speed in female breast varies by 15% [82]. The heterogeneity of sound speed may have deleterious effect on the quality of reconstructed images. Because of refraction of rays passing a tissue interface with a sound mismatch, the wavefronts are distorted in both amplitude and phase [82]. Refractions may make a distortion in amplitude of wavefronts because of an interference between multiple rays induced by refractions. Using a ray approach, this implies that the waves from a point source may reach a single detection point from different paths [82]. A distortion in phase is because of variations in the time-of-flight (TOF) of rays that are induced by heterogeneities in the sound speed. For ultrasound

tomography (UT), it was shown that the amplitude distortion dominates the phase distortion [87]. In contrast, for PAT, it was shown that for small-scale heterogeneities (in the range of the wavelength of acoustic waves), the amplitude distortion of a wavefront is negligible when the detection surface is sufficiently far from the heterogeneities. Also, for large-scale heterogeneities, the multi-path interference induced by reflections is negligible [82]. Using a ray approach, it was proven that no two rays from a source inside a convex boundary can intersect each other following refractions at the boundary. Although for biological tissues, the boundary between soft tissues may be quite irregular and concave, the induced amplitude distortion is insignificant [82].

Let us define a point source $\boldsymbol{r}$ and a detection point $\boldsymbol{r}_0$. For the case of homogeneous sound speed, an acoustic ray from point $\boldsymbol{r}$ reaches point $\boldsymbol{r}_0$ along the straight line $\boldsymbol{r}, \boldsymbol{r}_0$. For the case of heterogeneous sound speed, the ray from point $\boldsymbol{r}$ travels along the path $\boldsymbol{r}, \boldsymbol{r}_b, \boldsymbol{r}_0$, where $\boldsymbol{r}_b$ is a refraction point at a tissue interface. The TOF along path $L(\boldsymbol{r}, \boldsymbol{r}_b, \boldsymbol{r}_0)$ is defined by

$$t_f(\boldsymbol{r}, \boldsymbol{r}_b, \boldsymbol{r}_0) = \int_{L(\boldsymbol{r}, \boldsymbol{r}_b, \boldsymbol{r}_0)} d\boldsymbol{r}''/c(\boldsymbol{r}''), \tag{1.59}$$

where $c(\boldsymbol{r}'')$ is the local sound speed at point $\boldsymbol{r}''$ positioned along a path $L(\boldsymbol{r}, \boldsymbol{r}_b, \boldsymbol{r}_0)$. It was proven for a weakly heterogeneous medium that [82]

$$t_f(\boldsymbol{r}, \boldsymbol{r}_0) = \int_{L(\boldsymbol{r}, \boldsymbol{r}_0)} d\boldsymbol{r}''/c(\boldsymbol{r}'') = t_f(\boldsymbol{r}, \boldsymbol{r}_b, \boldsymbol{r}_0) + O(\varepsilon^2) \tag{1.60}$$

for small value $\varepsilon = (c(\boldsymbol{r}'') - c_0)/c_0$, where c_0 is a reference homogeneous sound speed [82]. Using the approximation (1.60), the forward model defined in (1.10) can be extended to acoustically heterogeneous media. This yields [82, 1]

$$p_{het}(\boldsymbol{r}_0, t) = \frac{\beta}{4\pi C_p} \int H(r) \frac{\partial}{\partial t} \frac{\delta\left(t - t_f(\boldsymbol{r}, \boldsymbol{r}_0)\right)}{|\boldsymbol{r}_0 - \boldsymbol{r}|} \, d\boldsymbol{r}, \tag{1.61}$$

where the subscript p_{het} denotes a heterogeneous medium. Equation (1.61) can be written in an equivalent form [2, 1]

$$g_{het}(\boldsymbol{r}_0, \bar{t}) = \int H(\boldsymbol{r}) \delta(\bar{t} - c t_f\left(\boldsymbol{r}, \boldsymbol{r}_0\right)) \frac{c t_f(\boldsymbol{r}, \boldsymbol{r}_0)}{|\boldsymbol{r}_0 - \boldsymbol{r}|} d\boldsymbol{r}, \tag{1.62}$$

where $g_{het}(\boldsymbol{r}_0, \bar{t})$ is a generalised Radon transform (GRT) of $H(\boldsymbol{r})$, and can be expressed as an integration over non-spherical isochronous surfaces [2, 1, 47]. The relation between p_{het} and g_{het} is described by[1]

$$g_{het}(\boldsymbol{r}_0, \bar{t} = c_0 t) = \frac{4\pi C_p}{\beta} t \int_0^t p_{het}(\boldsymbol{r}_0, t') dt'. \tag{1.63}$$

By using a heuristic layer-stripping procedure, it was shown that g_{het} possesses a two-fold redundancy using the so-called *half time* functions

$$g_{het}^{(1)}(\boldsymbol{r}_0,t)=\begin{cases} g_{het}(\boldsymbol{r}_0,t), & 0\leqslant t\leqslant t_f(0,\boldsymbol{r}_0)\\ 0, & \text{otherwise}\end{cases}$$
$$g_{het}^{(2)}(\boldsymbol{r}_0,t)=\begin{cases} g_{het}(\boldsymbol{r}_0,t), & t_f(0,\boldsymbol{r}_0)\leqslant t\leqslant \infty\\ 0, & \text{otherwise},\end{cases} \tag{1.64}$$

where $t_f(0,\boldsymbol{r}_0)$ is the TOF from the centre of rotation to the detection point $\boldsymbol{r}_0$ [86]. It was shown that $H(\boldsymbol{r})$ can be determined by either $g_{het}^{(1)}$ or $g_{het}^{(2)}$ [1]. Let us define $H_{het}^1(\boldsymbol{r})$ and $H_{het}^2(\boldsymbol{r})$ as images obtained from *half-time* data $g_{het}^{(1)}$ and $g_{het}^{(2)}$, respectively. In the absence of finite sampling effects or measurement errors, $H_{het}^1(\boldsymbol{r})=H_{het}^1(\boldsymbol{r})$ [86].

Now, we assume that the sound speed is piecewise constant, and the geometry of interfaces is known. (This may be obtained from the PAT measurement data.) Based on this, a finite-dimensional parameterisation is applied on $c(\boldsymbol{r})$ using the vector $C=\left[c_1,c_2,...,c_n\right]$ with components the sound speed associated with each subregion of the imaging medium. Given C, one obtains the two-fold redundant data $g_C^{(1)}$ and $g_C^{(2)}$. For a full-view data, i.e., specified for all points $\boldsymbol{r}_0$ on a closed measurement surface S, and in the absence of data inconsistencies, these two sets of data must give the *half-time* reconstructions $H_C^1(\boldsymbol{r})=H_C^2(\boldsymbol{r})$ because of redundancy in data. Using this, the objective is to calculate C that satisfies this equality [86]. In practical cases, because of noise in data and finite sampling effects, we will obtain $H_C^1(\boldsymbol{r})\approx H_C^2(\boldsymbol{r})$. Thus, C can be obtained by solving

$$\underset{C}{\operatorname{argmin}}\, F(C):=\|H_C^1(\boldsymbol{r})-H_C^2(\boldsymbol{r})\big]\|. \tag{1.65}$$

The solution $H(\boldsymbol{r})$ can be obtained by solving an inversion of equation (1.61) for a given C_* that has been obtained using (1.65). This may lead to an alternative reconstruction of H and C [86].

1.4.1 Geometrical acoustic approximation

In the GRT model defined by (1.62), a first-order geometrical acoustic approximation of travel times $t_f(\boldsymbol{r}_0,\boldsymbol{r})$ is used. This approximation was made using an assumption

that acoustic rays travel along linear paths [82, 86]. In [47], an approach was proposed for derivation of higher-order perturbations to travel times. Using this approch, the effects of ray bending are also included in calculation of TOFs. In the sequel, we briefly describe this approach. Let us define a heterogeneous sound speed map as a perturbation to a homogeneous background c_0 with a magnitude of perturbation ϵ. This gives

$$c(\boldsymbol{r}) = c_0 + \epsilon c_1(\boldsymbol{r}). \tag{1.66}$$

For an acoustically homogeneous medium with sound speed c_0, applying a Fourier transform on (1.5) gives an unperturbed Green's function $G_0(\boldsymbol{r}, \boldsymbol{r}', w)$ that satisfies a Helmotz equation in the form

$$\left(\nabla^2 + \frac{w^2}{c_0^2}\right) G_0(\boldsymbol{r}, \boldsymbol{r}', w) = -\delta(\boldsymbol{r} - \boldsymbol{r}'). \tag{1.67}$$

In a same way, for an acoustically heterogeneous medium with sound speed $c(\boldsymbol{r})$, a generalised Green's function $G(\boldsymbol{r}, \boldsymbol{r}', w)$ satisfies a Helmotz equation in the form [47]

$$\left(\nabla^2 + \frac{w^2}{c^2(\boldsymbol{r})}\right) G(\boldsymbol{r}, \boldsymbol{r}', w) = -\delta(\boldsymbol{r} - \boldsymbol{r}'). \tag{1.68}$$

A geometrical approximation is valid when the size of heterogeneities is much larger than the wave-length, i.e., the sound speed does not change considerably over one wavelength [47]. Using this assumption, the Green's function for a heterogeneous medium can be expressed as [47]

$$G(\boldsymbol{r}, \boldsymbol{r}', w) = \left(g_0 + \frac{g_1}{ik} + \frac{g_2}{(ik)^2} + ...\right) e^{iw\tau(\boldsymbol{r}, \boldsymbol{r}')}, \tag{1.69}$$

where k is the wavenumber, and τ is called *eikonal* function. A zeroth-order approximation of (1.69) yields

$$G(\boldsymbol{r}, \boldsymbol{r}', w) = g_0(\boldsymbol{r}, \boldsymbol{r}') e^{iw\tau(\boldsymbol{r}, \boldsymbol{r}')}. \tag{1.70}$$

Using the above equation, together with a temporal FT of (1.61), gives

$$p(\boldsymbol{r}, w) \approx -\frac{iw\beta}{C_p} \int H(\boldsymbol{r}') g_0(\boldsymbol{r}, \boldsymbol{r}') e^{\mathbf{i}wt_f(\boldsymbol{r}, \boldsymbol{r}')} d\boldsymbol{r}'. \tag{1.71}$$

Using (1.68), (1.69) and (1.70) gives in the limit $\lambda \to 0$ [47]

$$\begin{aligned} O(k^2): &\quad \left[\nabla_{\boldsymbol{r}} \tau(\boldsymbol{r}, \boldsymbol{r}')\right]^2 = c^{-2}(\boldsymbol{r}) \\ O(k^1): &\quad 2\nabla_{\boldsymbol{r}} \tau \cdot \nabla_{\boldsymbol{r}} g_0 + g_0 \nabla_{\boldsymbol{r}}^2 \tau = 0 \\ O(k^0): &\quad \nabla_{\boldsymbol{r}}^2 g_0(\boldsymbol{r}, \boldsymbol{r}') = -\delta(\boldsymbol{r} - \boldsymbol{r}'), \end{aligned} \tag{1.72}$$

where λ is the wavelength, and $\nabla_{\boldsymbol{r}}$ denotes the gradient with respect to $\boldsymbol{r}$. The first line of the above equation is called *eikonal* equation, and gives a general form of (1.59) in the form [47]

$$\tau(\boldsymbol{r}, \boldsymbol{r}') = \int_{L_c(\boldsymbol{r},\boldsymbol{r}')} d(\boldsymbol{r}'')/c(\boldsymbol{r}''), \tag{1.73}$$

where $L_c(\boldsymbol{r}, \boldsymbol{r}')$ denotes an arc length along a curved path between $\boldsymbol{r}$ and $\boldsymbol{r}'$. (Note that this path is not necessarily straight line.) Based on the Fermat's principle, the ray trajectory follows a path that minimises the acoustic path length. For an acoustically homogeneous medium, this is a straight line. For a medium with a low variation in the sound speed, it was shown that with a first order approximation of perturbation, this trajectory can be chosen to be the path along a reference ray that satisfies (1.73). Additionally, the second line of (1.72) yields [47]

$$\begin{aligned} g_0(\boldsymbol{r}, \boldsymbol{r}') &= \frac{1}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|} e^{-\frac{1}{2}\int_{L_c(\boldsymbol{r},\boldsymbol{r}')} c(\boldsymbol{r}'')(\nabla^2\tau)\ d\boldsymbol{r}''} \\ &= \frac{1}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|} \left(\frac{c(\boldsymbol{r})}{c(\boldsymbol{r}')}\right)^{1/2} \end{aligned} \tag{1.74}$$

1.4.2 Higher-order geometrical acoustic approximation

Although the acoustic rays bend towards the region that has a lower sound speed, such effects are not taken into account using a first-order perturbation to travel times using (1.73). Motivated by this, in [47], higher-order perturbations to time travels were included in the generalised Radon Transform equation. To do this, the travel times are expressed as

$$\tau(\boldsymbol{r}, \boldsymbol{r}') = \tau_0(\boldsymbol{r}, \boldsymbol{r}') + \epsilon\tau_1(\boldsymbol{r}, \boldsymbol{r}') + \epsilon^2\tau_2(\boldsymbol{r}, \boldsymbol{r}') + ... \tag{1.75}$$

It was shown that [47]

$$\tau_0(\boldsymbol{r}, \boldsymbol{r}') = \int_{L_0} d\boldsymbol{r}''/c_0, \tag{1.76}$$

$$\tau_1(\boldsymbol{r}, \boldsymbol{r}') = -\int_{L_0} \frac{c_1(\boldsymbol{r}'')}{c_0^2} d\boldsymbol{r}'' \tag{1.77}$$

and

$$\tau_2(\boldsymbol{r}, \boldsymbol{r}') = \int_{L_0} \frac{c_0}{2}\left[\frac{c_1^2(\boldsymbol{r}'')}{c_0^4} - |\nabla\tau_1(\boldsymbol{r}, \boldsymbol{r}')|^2\right] d\boldsymbol{r}'', \tag{1.78}$$

where L_0 is the reference ray associated with the reference eikonal $\tau_0(\boldsymbol{r}, \boldsymbol{r}')$.

1.5 Time reversal

In sections 1.2.4 and 1.4, we briefly explained the generalised Green's function methods that can address an irregularity of geometry of a detection surface or heterogeneity of sound speed, respectively. The approaches that use Green's functions are often based on some specific assumptions, for example the region of support for the acoustic source is inside the detection surface, or the effects of acoustic attenuation or variations in ambient density on the shape, spectrum and amplitude of propagating acoustic waves are neglected. Also, these approaches often cannot benefit from advances in modelling acoustic wave propagation.

Motivated by this, a universal approach was proposed for PAT. This inversion technique, which is called *Time reversal* (TR), is based on retransmitting the time series of boundary data in a time reversed order. The TR method was first suggested by [23] and [83], and was then used in many studies of PAT, e.g., [13]. Here, we explain the TR approach for solving an inverse problem associated with a second-order photoacoustic wave equation in the form

$$\begin{aligned} \left(\nabla^2 - \frac{1}{c^2(\boldsymbol{r})}\frac{\partial^2 p}{\partial t^2}\right) p(\boldsymbol{r},t) = 0, \quad (\boldsymbol{r},t) \in \mathbb{R}^d \times \big[0,\infty\big) \\ p(\boldsymbol{r},0) = p_0(\boldsymbol{r}) = \Gamma H(\boldsymbol{r}), \quad \frac{\partial p}{\partial t}(\boldsymbol{r},0) = 0, \end{aligned} \tag{1.79}$$

where d denotes the dimension of the domain. The TR approach uses an assumption that the source function $p_0(\boldsymbol{r})$ and the sound speed $c(\boldsymbol{r})$ are sufficiently smooth. Also, p_0 is compactly supported. Let us denote the domain bounded by the detection surface S by $\Omega \subset \mathbb{R}^d$. One advantage of the TR inversion approach is that p_0 is not required to be supported inside Ω, as opposed to the inversion approaches based on Green's function. The time reversal (TR) approach also uses the following assumptions:

1. The imaging region is bounded by a closed detection surface S.
2. The detection surface S is sufficiently dense and smooth.
3. The pressure field decays inside Ω when $t \to \infty$.

Based on the Huygen's principle, for an odd d and homogeneous c, and for any initial source p_0 with a bounded support and any bounded domain Ω, there is a finite time instant T for which the wave inside Ω vanishes for $t \geqslant T$ [32, 31]. For a non-trapping sound speed, the decay is sufficiently rapid, e.g., exponential for odd

dimensions [32]. Note that the Huygen's principle does not hold for an even d, or when c is trapping.

Given a time-varying $p(\boldsymbol{r}_0, t)$ recorded on an arbitrary detection surface S for time $t \in [0, T]$, the inverse problem is the reconstruction of the initial pressure distribution $p_0(\boldsymbol{r})$ inside Ω. This amounts to solving an initial boundary value problem (IBVP) by applying $p(\boldsymbol{r}_0, t)$ as a Dirichlet boundary condition in a time-reversed order. This yields the field

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) p_{\mathrm{tr}}(\boldsymbol{r}, t) = 0, \quad (\boldsymbol{r}, t) \in \Omega \times \big[0, T\big] \tag{1.80}$$

with initial values

$$p_{\mathrm{tr}}(\boldsymbol{r}, 0) = 0, \ \frac{\partial p_{\mathrm{tr}}}{\partial t}(\boldsymbol{r}, 0) = 0, \quad \boldsymbol{r} \in \Omega, \tag{1.81}$$

and Dirichlet boundary condition

$$p_{\mathrm{tr}}(\boldsymbol{r}_0, t) \leftarrow p(\boldsymbol{r}_0, T - t), \quad (\boldsymbol{r}_0, t) \in S \times \big[0, T\big]. \tag{1.82}$$

Although the TR approach is based on the Huygen's principle from a theoretical point of view, this approach was also investigated numerically for cases in which the Huygen's principle does not hold, e.g., a 2D medium with variable sound speed [32, 31]. The main numerical results that have been reported in these studies are as follows.

1- *Trapping variable sound speed.* The quality of reconstructions significantly depends on the cut-off time T. Also, for a circular detection geometry S, the interfaces of source $p_0(\boldsymbol{r})$ that are normal to the radius of S cannot be stably reconstructed. This is because wavefronts from a radial interface travel perpendicular to the radial direction, and can be trapped. This implies that the singularity associated with a radial interface does not show up as a singularity in the measured data, and thus cannot be fully reconstructed. This effect appears as blurriness of the radial interfaces in the reconstructed image.

2- *Erroneous sound speed.* The PAT problem uses an assumption that the sound speed is known. To meet this requirement, the sound speed is either assumed constant, or may be obtained from adjunct imaging data such as ultrasound computed tomography [44]. It was shown that using an erroneous sound speed deteriorates accuracy for both amplitude and location of interfaces, the first of which is very important for quantitative photo-acoustic tomography, i.e., the reconstruction of an image of optical coefficients from measured pressure data [61].

1.6 Model-based approaches

As discussed above, the TR approach is based on the assumption that the boundary data is densely sampled on a detection surface that encloses the imaging region. This condition does not hold in many biomedical applications of PAT because of a low degree of freedom for positioning detectors. Additionally, the accuracy of the TR approach is contingent on a sufficiently accurate estimation of the sound speed, which is not practical for many biomedical applications, unless the information from adjunct imaging modalities such as ultrasound or x-ray computed tomography is used [33, 44]. In general, practical difficulties such as a finite sampling, a limited-view or few-view detection surface, errors in estimation of medium's properties, or errors in data measurement make the acoustic inverse problem ill-posed [34]. For these cases, model-based iterative algorithms are used, in which the forward operator and an associated back-projection operator (often the adjoint operator) are iteratively applied in order to find p_0 that minimises an error function, which is the discrepancy between the measured data and a predicted data.

To do this, we use (1.79) in order to define an acoustic forward operator on a continuous domain in the form

$$\begin{aligned} &\Lambda : C_0^\infty \to \mathbb{R}^{N_s N_t}, \\ &\Lambda[p_0] = \mathcal{M}p(\boldsymbol{r}, t), \quad (\boldsymbol{r}, t) \in \mathbb{R}^d \times \left[0, T\right] \\ &g(\boldsymbol{r}_s, t_s) = \Lambda[p_0] \end{aligned} \tag{1.83}$$

where $\mathcal{M}$ is an operator that at each time instant maps the pressure field to detector positions $\boldsymbol{r}_s$ and measurement time instants t_s. Since the operator Λ is linear, its discretisation can be written in a matrix form

$$\hat{P} = \mathbb{H}P_0 + \eta, \tag{1.84}$$

where $\hat{P} \in \mathbb{R}^{N_s N_t}$ is a stack of time series of measured data with N_s the number of detectors and N_t the number of measurement time instants, and η denotes the noise in data. Also $P_0 \in R^{N_n}$ is a discretised initial pressure distribution with N_n the number of discretisation points. Additionally, $\mathbb{H} \in \mathbb{R}^{N_s N_t \times N_n}$ is a linear operator that represents a discretisation of the forward operator Λ. Note that throughout this thesis, the Grüneisen parameter Γ is assumed to be 1, and thus the initial pressure distribution is equal to the absorbed optical energy.

Having defined a discrete formulation of the forward operator Λ, denoted by $\mathbb{H}$, the inverse problem is the calculation of P_0 from $\hat{P}$ using (1.84). Using model-based inversion approaches, this inverse problem is solved using an alternative implementation of a forward and back-projection (or adjoint) pair. In general, the model-based inversion approaches for solving (1.84) can be categorised into convergent Neumann-series based algorithms or optimisation approaches.

1.6.1 Convergent Neumann-series based approach

In [59], an explicit formula is derived based on a convergent Neumann-series approach for cases in which data is measured (available) on the whole boundary. (The numerical results for cases in which the data is available only on a part of the boundary have also been reported [52].) The Neumann-series method uses a forward operator, and a back-projection operator that is a modified variant of the time reversal operator (equations (1.80), (1.81) and (1.82)), for which the pressure field is initialised by a harmonic extension of boundary data $g(\boldsymbol{r}, T)$.

Considering (1.83), the act of the left pseudo-inverse of Λ, denoted by Λ^+, on $g \in \mathbb{R}^{N_s \times N_t}$ is defined by [59, 52]

$$\Lambda^+ g := p_0(\cdot), \text{ in } \Omega. \tag{1.85}$$

If we assume that p_0 is in an energy space, i.e., $p_0 \in H_D(\Omega)$, the operator Λ^+ will amount to a applying the TR approach using (1.80), (1.81) and (1.82), for which the pressure field is initialised by a function ϕ that minimises the energy norm $\|\cdot\|_{H_D(\Omega)}$. Using the Dirichlet principle, ϕ is a solution to an elliptic boundary value problem of the form

$$\nabla\phi(\boldsymbol{r}) = 0 \tag{1.86}$$

with boundary condition $\phi|_S = g(\cdot, T)$, which satisfies compatibility conditions of first order, i.e., $g(\cdot, T) \in S \times \{T\}$ is sufficiently smooth [59, 52]. Using this, we replace the initialisation in (1.81) by

$$p_{\text{tr}}(\boldsymbol{r}, 0) = \phi(\boldsymbol{r}), \ \frac{\partial p_{\text{tr}}}{\partial t}(\boldsymbol{r}, 0) = 0, \quad \boldsymbol{r} \in \Omega. \tag{1.87}$$

For an odd N, a constant c and a sufficiently large T, we will have $\Lambda^+\Lambda = I$, i.e., Λ^+ is an actual inverse of Λ [52]. In practice, Λ^+ satisfies

$$\Lambda^+\Lambda = I - K, \tag{1.88}$$

where K is an error operator, and satisfies

$$\|Kp_0\|_{H_D(\Omega)} \leqslant \|p_0\|_{H_D(\Omega)}, \forall p_0 \in H_D(\Omega), \tag{1.89}$$

for any smooth c (trapping or non-trapping), and for any time $T > 0$. For a sufficiently large T, the inequality is strict. Using this, the inverse problem of PAT can be solved using

$$f = \sum_{m=0}^{\infty} K^m \Lambda^+ g, \; g := \Lambda p_0. \tag{1.90}$$

For further details on the theory of the convergent Neumann-series algorithm, specifically conditions for stability and convergence, the reader is referred to [59] and [52].

1.6.2 Variational approaches

The variational approaches are based on an iterative minimisation of an error functional f, which is often defined as an $\mathcal{L}^2$ norm of the discrepancy between the measured data and a predicted (modelled) data. This leads to a minimisation problem of the form

$$\underset{x}{\operatorname{argmin}} \left\{ f(X) := \|\mathbb{H}X - \hat{P}\|_2 \right\}, \tag{1.91}$$

where X is a discretised variant of a sought after initial pressure distribution p_0.

Using Green's functions methods, a large number of model-based approaches have been proposed. These approaches are often based on storing a discretisation of the Poisson's integral (or its variant) in a matrix form, and using algebraic reconstruction techniques for solving the arising minimisation problem. Using these approaches, an iterative computation of the forward and adjoint operators is avoided at the expense of storing a matrix representing the whole forward operator [19, 56]. A storage of the forward matrix is practical for these cases, because the matrix that models a forward operator using a Poisson's integral is often sparse. This is due to the fact the acoustic

signal measured at a specific point and a time instant depends only on a small portion of the image [55]. Additionally, a representation of the forward operator as a sparse matrix enables the application of fast inversion algorithms that utilise sparsity for achieving a high numerical efficiency [56].

The ultimate goal of PAT is quantitative photo-acoustic tomography (QPAT), a quantitative reconstruction of the optical absorption coefficient from a set of time series of pressure data. We will show in section 1.9 that the inverse problem of QPAT is highly non-linear and ill-posed [61, 25, 24]. This indicates that any errors in modelling the acoustic forward operator rapidly grow, and may dominate the signal data. This motivates improving the accuracy of the acoustic forward operator as much as possible using numerical schemes that can be adapted to local acoustic properties of the medium, i.e., variable sound speed and ambient density, and also account for attenuation of acoustic waves. For this class of forward models, an integration of the associated acoustic fields over time is required. This makes an explicit storage of the forward operator impractical (See [34].) To avoid an explicit storage of a discretised forward operator, we solve the minimisation problem (1.91) via an alternative implementation of a pair of forward and adjoint operators in an iterative manner [34, 4, 3, 37]. (Note that this also applies to convergent Neumann series approach.)

The arising iterative algorithm is computationally expensive [34, 37], but it can be handled, for example by a fast implementation of reconstruction algorithms using the advances on data casting and parallelisation using GPUs [65, 20, 73], or FPGA [85]. Some of advantages of using variational methods for PAT are as follows.

1- Variational algorithms can be adapted to different schemes for modelling the forward problem, including techniques based on Green's function methods [19, 56, 50, 12, 74] or those accounting for tissue realistic acoustic properties of the medium such as heterogeneity of acoustic properties and attenuation [34, 37].

2- In a real-world setting, because of a finite size of detectors, an assumption that the detection of the pressure field is point-wise ($\boldsymbol{r}_0 \in S$) fails. Therefore, it will be advantageous if the geometrical effects of the detectors' finite aperture are included in the reconstruction [6, 72, 53, 55]. In addition, the detection process also depends on the acoustic and electric impedance mismatch between the tissue and detectors [49, 54]. One advantage of using optimisation algorithms is that these effects can be included in

the forward model, and accounted for in the inversion process, e.g. [53]. Additionally, using variational approaches, it will be possible to compensate for modelling errors, e.g., measurement errors, for the reconstruction [64].

3- For ill-posed problems of PAT, either stability or both uniqueness and stability may fail. (See for example [52].) This implies that a large number of different sources may produce boundary data sets that are close to the measured data [55]. Using variational methods, it will be possible to either impose constraints [22, 34, 37], or enforce regularisation [19, 50, 12, 74, 34, 37] on the initial pressure distribution in order to stabilise the inverse problem.

It turns out that the shape, spectrum and amplitude of propagating acoustic waves are dependent on the acoustic properties of biological tissues [18]. This motivates using models for acoustic wave propagation that account for acoustic properties of the medium, e.g., variations in the sound speed and ambient density, or acoustic attenuation. To do this, here we define a forward operator for our PAT problem using a linear system of three-coupled first-order wave equations. For a lossless acoustic medium, this system of equations, which is equivalent to (1.79), is defined by [60, 69]

$$\begin{aligned} \frac{\partial}{\partial t}u(\boldsymbol{r},t) &= -\frac{1}{\rho_0(\boldsymbol{r})}\nabla p(\boldsymbol{r},t) \\ \frac{\partial}{\partial t}p(\boldsymbol{r},t) &= -\rho_0(\boldsymbol{r})\nabla\cdot u(\boldsymbol{r},t) \\ p(\boldsymbol{r},t) &= c^2(\boldsymbol{r})\rho(\boldsymbol{r},t), \end{aligned} \tag{1.92}$$

with initial conditions

$$p(\boldsymbol{r},0) = p_0(\boldsymbol{r}), \quad u(\boldsymbol{r},0) = 0, \tag{1.93}$$

where u is the vector-valued particle velocity field, and ρ is the acoustic density field. Additionally, $\rho_0(\boldsymbol{r})$ denotes the ambient density of medium. Here, the first equation is the *linearised equation of motion (conservation of momentum)*, the second equation is the *linearised equation of continuity (conservation of mass)*, and the third equation represents the adiabatic equation of state [69].

To account for an absorption and dispersion following a frequency power law, the third line in (1.92) is replaced by an alternative variant of equation of state including two fractional Laplacian operators in the form [69]

$$p(\boldsymbol{r},t) = c^2(\boldsymbol{r})\left[1 - \tau(\boldsymbol{r})\frac{\partial}{\partial t}(-\nabla^2)^{y/2-1} - \eta(\boldsymbol{r})(-\nabla^2)^{(y-1)/2}\right]\rho(\boldsymbol{r},t), \tag{1.94}$$

where τ and η are absorption and dispersion proportionality coefficients, and y is a power law exponent [66, 69]. We will give further details on this in chapters 2 and 4.

1.7 Numerical schemes for solving the forward operator

As discussed in the previous section, model-based approaches are based on an iterative implementation of the forward and adjoint operators. This requires an accurate, yet efficient, modelling of the forward operator (and an associated adjoint). (See for example [34].) The finite element (FE) method is a popular method for solving partial differential equations (PDEs) because of a high accuracy and an adaptability to irregular shapes for source and heterogeneities. For an FE mesh with N_n nodes, the solution is represented as a linear combination of N_n basis functions. Using this, the solution of the forward operator amounts to a calculation of coefficients in terms of the basis functions via solving an arising sparse linear equation often using Krylov methods. In [39], an FE scheme is proposed for solving a Helmholtz variant of photo-acoustic wave equation (cf. (1.67) and (1.68)), which is derived by applying a temporal Fourier transform to (1.79). However, this numerical scheme is very expensive for high-resolution 3D PAT, as 10 nodes per wave-length are required for an accurate approximation of the acoustic field.

1.7.1 Finite-difference time-domain methods

Finite difference time-domain (FDTD) methods, which are are based on an approximation of derivatives (in both space and time) using finite differences, are less flexible than FE methods for handling the shape of heterogeneities, but are more popular for PAT because of simplicity of their implementation. In [36], an FDTD method was used for solving a system of first-order photo-acoustic wave equations that include the second-order nonlinear terms and classical dissipative effects. Among the recent applications, an FDTD method has been used for an approximation of a system of elastic wave propagation that was used for an implementation of a pair of forward and adjoint operators for PAT of the brain [46].

Similar to FE methods, an accurate approximation of the acoustic wave propagation using FDTD methods requires 10 grid points per wave-length [18]. Therefore, an approximation of high frequency parts of photo-acoustic waves requires a very dense spatial grid [18].

Using FD schemes, the spatial gradient of a field is approximated by fitting a polynomial to a number of points, and calculating the derivative of the polynomial [18]. For example, using a first-order variant of FD schemes, the gradient is calculated by fitting a straight line between two grid points [18]. Therefore, the FD schemes use the local properties of the field for computing the gradient, and the accuracy of estimation of the gradient is dependent on the order of the fitted polynomial.

1.7.2 K-space pseudo-spectral methods

Motivated by this fact, Pseudo-spectral (PS) techniques have been proposed, for which the gradient at each point along each Cartesian coordinate is calculated by fitting a Fourier series to all grid points on a line passing that point along that Cartesian coordinate [18]. Using PS methods has two main advantages:

1- The gradient associated with each Cartesian coordinate can be calculated globally (for all points) using a Fast Fourier Transform (FFT) and its inverse.

2- Using FFT, a field is approximated as a finite sum of periodic functions, which can be described by only two nodes per wave-length [60, 18]. This dramatically reduces the mesh requirement for modelling the high-frequency parts of the acoustic field.

Using Fourier series for interpolating functions imposes a periodicity to the field. This implies that the value of the field and its derivative at both ends of each row of the grid are the same. Because of this, the waves leaving one side of the grid reenters from the opposite side. This phenomenon is called *wave wrapping* [60].

To explain our used numerical scheme, we start with discretisation of medium's parameters and acoustic fields on a $d \in \{2, 3\}$ dimensional uniform rectilinear grid. The position of grid points is denoted by $\boldsymbol{r}_\zeta$ with $\zeta = (\zeta_1, ..., \zeta_d) \in \{1, ..., N_1\} \times ... \times \{1, ..., N_d\}$ and $N_n = \prod_{i=1}^{d} N_i$ the total number of grid points. Discretised variants of sound speed c and ambient density ρ_0 are denote by $\bar{c}$ and $\bar{\rho}_0$, respectively. We denote a separation of grid in space along coordinate i by Δr_i and a separation in time by Δt. Accordingly, we define a discretisation of the pressure field p at position ζ and

time step n by $p_{(\zeta,n)}$. We will also use $v_{(i,\zeta,n)}$ and $\rho_{(i,\zeta,n)}$ as discretised variants of the vector-valued particle velocity v and acoustic density ρ with i signifying a dependence on Cartesian coordinate i. (Note that ρ is a vector because of the way the PML is implemented on the staggered grid [60].) For brevity, we use the same notations as the continuous formulae for a discretisation of the spatial frequency (wavenumber) along the coordinate i, and a discretisation of the magnitude of the wavenumber. These satisfy

$$k = \left(\sum_{i=1}^{d} k_i^2 \right)^{1/2}. \tag{1.95}$$

In a same way, the vector-valued wavenumber is denoted by $\boldsymbol{k}$. We also hope i, which denotes a Cartesian coordinate, is not confused with the imaginary number $\mathbf{i}$.

Using PSTD methods, a second-order photo-acoustic wave equation defined by (1.79) is transformed into a spatial frequency domain in the form [60]

$$\frac{\partial^2 p(\boldsymbol{k},t)}{\partial t^2} = -(c_0 k)^2 p(\boldsymbol{k},t). \tag{1.96}$$

Here, we assumed a homogeneous medium, and c_0 indicates an associated constant reference sound speed [60, 18]. (ρ_0 is not necessarily constant.) Also, $p(\boldsymbol{k},t)$ is the d dimensional spatial FT of the acoustic pressure field $p(\boldsymbol{k},t)$. The same as FDTD schemes, using PSTD methods, the temporal derivatives are approximated by finite differences. Accordingly, applying a second-order finite-difference scheme for an approximation of (1.96) yields [60, 18]

$$\frac{\bar{p}_{(\zeta,n+1)} - 2\bar{p}_{(\zeta,n)} + \bar{p}_{(\zeta,n-1)}}{(\Delta t)^2} = -\bar{c}_0^2 \mathbb{F}^{-1}\left[\bar{k}^2 \mathbb{F}\left[\bar{p}_{(\zeta,n)}\right] \right], \tag{1.97}$$

where $\bar{c}_0$ is a discretisation c_0, and $\mathbb{F}$ and $\mathbb{F}^{-1}$ denote the discrete Fast Fourier Transform (FFT) and its inverse, respectively.

For a lossless acoustic medium, the relation between the temporal frequency w and spatial frequency k (wavenumber) satisfies $w = ck$ [60, 18]. This is called the *analytic dispersion relation.* Although this relation is exact for an analytic model, it turns out that using numerical schemes for propagation of acoustic waves leads to an error in approximation of the temporal frequency w. Using the relation $c = w/k$, this also introduces an error in the sound speed, which is called *numerical dispersion* error. By an approximation of the pressure field using a numerical integration of (1.97)

over time, the numerical dispersion errors are accumulated in time, and deteriorate an approximation of acoustic waves. The numerical dispersion errors can be mitigated using higher-order temporal integration schemes [77].

It has been shown that for an acoustically homogeneous medium, the temporal integrations can be computed without any dispersion errors using [43, 60, 18]

$$\big((\Delta t)^2 \operatorname{sinc}^2(\bar{c}_0 \Delta t \bar{k}/2)\big)^{-1} \left(\bar{p}_{(\boldsymbol{k},n+1)} - 2\bar{p}_{\boldsymbol{k},n)} + \bar{p}_{(\boldsymbol{k},n-1)} \right) = -(\bar{c}_0 k)^2 \bar{p}_{(\boldsymbol{k},n)}, \tag{1.98}$$

where $\operatorname{sinc}(X) = \sin(X)/X$. Using (1.98), the temporal iterations are done in a wavenumber domain [43]. Alternatively, a discrete spatial inverse FT of (1.98) gives an equivalent temporal iteration in real space in the form [60]

$$\frac{\bar{p}_{(\zeta,n+1)} - 2\bar{p}_{(\zeta,n)} + \bar{p}_{(\zeta,n-1)}}{(\Delta t)^2} = -\bar{c}_0^2 \mathbb{F}^{-1} \Big[k^2 \operatorname{sinc}^2(\bar{c}_0 \Delta t k/2) \mathbb{F}\big[\bar{p}_{(\zeta,n)}\big] \Big]. \tag{1.99}$$

The equation (1.99) can be expressed as a modified variant of (1.97), where a Laplacian operator (in a wavenumber domain) included in the right-hand-side of (1.97) has been replaced by

$$\bar{\nabla}^2 \bar{p}_{(\zeta,n)} = -\mathbb{F}^{-1} \Big[k^2 \operatorname{sinc}^2(\bar{c}_0 \Delta t k/2) \mathbb{F}\big[\bar{p}_{(\zeta,n)}\big] \Big], \tag{1.100}$$

where the operator $\bar{\nabla}^2(\cdot)$ is called *second-order k-space* operator, and $\operatorname{sinc}^2(\bar{c}\Delta t k/2)$ is a k-space correction to the spatial gradient [60].

The k-space operator can be extended to a system of three-coupled first-order wave equations [60, 18]. To do this, the second-order k-space operator is factored into parts corresponding to each Cartesian coordinate using (1.95).

It turns out that by staggering the grid in space and time, it will be possible to improve accuracy and stability for discretisation of odd-order spatial and temporal derivatives [60, 18]. Using a staggered spatial grid, a factorisation of (1.100) into coordinates gives [60, 18]

$$\bar{\nabla}^2 \bar{p}_{(\zeta,n)} = \sum_{i=1}^{d} \frac{\partial}{\partial r_i^+} \frac{\partial}{\partial r_i^-} \bar{p}_{(\zeta,n)}, \tag{1.101}$$

where

$$\frac{\partial \big[\,\cdot\,\big]}{\partial r_i^{\pm}} = \mathbb{F}^{-1} \Big[\mathbf{i} k_i e^{\pm k_i \Delta r_i/2} \operatorname{sinc}(\bar{c}_0 \Delta t k/2) \mathbb{F}\big[\,\cdot\,\big] \Big]. \tag{1.102}$$

Here, $\bar{c}_0$ is a discretisation of a reference (nominal) sound speed for a weakly heterogeneous medium [60, 18]. Following [60], here we briefly explain the advantages of using a staggered spatial grid.

1- Using a fixed grid for approximation of a first-order spatial gradient, the coefficient $\mathbf{i}k$ (corresponding to a non-staggered grid) would have a discontinuity because of a jump from minimum spatial frequency $k_{\min} = -\pi/\Delta r$ to maximum $k_{\max} = +\pi/\Delta r$. This discontinuity is the source of ringing (Gibbs) artefact [60]. The coefficients of the form $\mathbf{i}ke^{\pm k_i \Delta r_i/2}$ (corresponding to a staggered grid) removes this discontinuity, and reduces artifacts when the wavefield is spatially undersampled [60].

2- An approximation of an odd-order spatial derivative using a staggered spatial grid significantly improves the spatial frequency response for high frequency ranges. For example, a spatial FT of an approximation of a first-order derivative of a field on a staggered spatial grid is closer to an associated spatial FT of that field on a continuous domain, compared to using a fixed grid [60]. Particularly, an approximation of the first-order spatial derivative using (1.102) provides an ideal frequency response up to the spatial Nyquist frequency $\pi/\Delta r$ [60].

3- Using PSTD techniques, the spatial derivatives are computed as a finite sum of sinusoidal functions. Sampling of these sinusoidal functions at the center of time steps provides a better approximation than sampling at the ends of time steps (See [60], figure 2(b)). Using a staggered temporal grid for an approximation of (1.92), the temporal iterations are interleaved, i.e., the time derivative of the particle velocity (resp. acoustic density) field is approximated as a function of the spatial gradient of the acoustic density (resp. particle velocity) field [60, 18]. Note that using a staggered temporal grid, dispersion errors may still occur for large time steps. These errors can be minimised using a k-space correction to the spatial gradients, as discussed above [60].

In general, using a k-space pseudo-spectral method for solving a system of photoacoustic wave equations defined by (1.92) provides exact temporal iterations (without any dispersion error) for acoustically homogeneous media. Also, these approaches are highly accurate for weakly heterogeneous media, and are unconditionally stable for $c(r) \leqslant c_0$ [60, 18].

The k-space methods will also be able to improve the accuracy of temporal integrations for media having sharp or high-contrast discontinuities in acoustic properties (sound speed and density distribution), if an appropriate smoothing operator is applied on the acoustic properties. (For further information, see the manual on the *k-Wave* website [65].)

1.7.3 Stability criteria for k-space pseudo-spectral methods

Using a k-space method for a numerical implementation of (1.92), the algorithm is stable under the condition

$$|\sin(\bar{c}_0 k \Delta t/2)| \leqslant \frac{\bar{c}_0}{\bar{c}}, \tag{1.103}$$

where $\bar{c}_0$ is the reference sound speed. For an acoustically homogeneous medium, by choosing $\bar{c}_0 = \bar{c}$, the k-space method is unconditionally stable since the sinusoid function on the left-hand side of (1.103) is always equal to 1.

In the k-wave manual [65], a PSTD method with and without a k-space correction was applied on a homogeneous medium, and an analysis on the performance of the k-space correction in reducing the numerical dispersion error was given. It has been shown that the performance of the k-space method significantly depends on the choice of the reference sound speed. Considering (1.100), the k-space method will match a PSTD variant, if we choose $\bar{c}_0 = 0$. Also, for a choice of $0 < \bar{c}_0 < \bar{c}$ with $\bar{c}$ the true sound speed for a homogeneous medium, the k-space method provides a numerical dispersion (phase speed) error smaller than a PSTD variant, until the dispersion error becomes zero for $\bar{c}_0 = \bar{c}$. For a choice of $\bar{c}_0 > \bar{c}$, the dispersion error associated with the k-space scheme exponentially grows, and becomes larger than the error due to the PSTD scheme. (See figure 2.4 in the *k-Wave* manual [65].)

Considering this analysis, although by choosing $\bar{c}_0 > \bar{c}$, the stability condition (1.103) is strictly satisfied, the main advantage of the k-space schemes for minimising dispersion errors will be lost. As discussed in the above paragraph, a choice of $\bar{c}_0 < \bar{c}$ reduces the dispersion error, but it imposes a condition on the stability, i.e., Δt must be sufficiently small so that

$$\Delta t \leqslant \frac{2}{\bar{c}_0 k_{\max}} \sin^{-1}\left(\frac{\bar{c}_0}{\bar{c}}\right), \tag{1.104}$$

where $k_{\max}$ denotes the maximum wavenumber.

Based on this, for an acoustically heterogeneous medium, which amounts to the replacement of $\bar{c}$ by $\bar{c}_\zeta$, the choice of $\bar{c}_0$ for the k-space scheme makes a trade-off between the stability and accuracy of the temporal iterations. For example, by choosing $\bar{c}_0 = \max(\bar{c}_\zeta)$, the algorithm will be unconditionally stable, but this requires choosing a small Δt in order to avoid corruption of the solution by the phase speed errors accumulated in time. Conversely, a choice of $\bar{c}_0 = \min(\bar{c}_\zeta)$ reduces the phase speed errors, but it requires a small time step that satisfies a stability condition

$$\Delta t \leqslant \frac{2}{\bar{c}_0 k_{\max}} \sin^{-1}\left(\frac{\bar{c}_0}{\max(\bar{c}_\zeta)}\right). \tag{1.105}$$

The stability of the k-space scheme may be expressed in terms of the Courant-Fredrichs-Lewy (CFL) number. For a homogeneous medium, the CFL number is defined by the ratio of the spatial distance a wave travels during a temporal iteration Δt to the separation distance Δr, i.e.,

$$CFL = \frac{\bar{c}\Delta t}{\Delta r}. \tag{1.106}$$

A CFL number is used in order to express the stability condition independently from the grid spacing. Therefore, the CFL number can be thought of a time step that only depends on the numerical scheme, and not the spatial and temporal separation of the grid. Using the stability condition in (1.104), the choice of the CFL number must satisfy

$$CFL \leqslant \frac{2}{\pi}\left(\frac{\bar{c}}{\bar{c}_0}\right) \sin^{-1}\left(\frac{\bar{c}_0}{\bar{c}}\right), \tag{1.107}$$

where we have used $k_{\max} = \pi/\Delta r$. Based on the above, given the CFL number and having defined our spatial grid, the time step can be chosen using

$$\Delta t = \frac{CFL \Delta r_{\min}}{\max(\bar{c}_\zeta)}. \tag{1.108}$$

For applying k-space pseudo-spectral methods on weakly heterogeneous media, a choice of Δt using (1.108) with a CFL number of 0.3 provides a good balance between accuracy and computational speed [65]. For further details on the accuracy and stability of the k-space pseudo-spectral methods, the reader is referred to the manual on the *k-Wave* website [65].

1.8 Forward-Backward Splitting (FBS) methods for PAT

Among gradient-based optimisation algorithms for minimisation of non-smooth convex functions, *Iterative Shrinkage Thresholding* Algorithms (ISTA) are very popular for solving the inverse problem of PAT, because they are computationally cheap regarding speed and memory storage [9]. (For application of this class of methods in PAT, see [4, 3, 34, 37].) Considering the minimisation problem (1.91) for an ill-conditioned $\mathbb{H}$, the associated objective function is not sufficiently meaningful, implying that there exist a large number of solutions that minimise the objective function up to a level of accuracy. The ill-posedness of the problem can be mitigated by replacing the original ill-posed problem by a nearby well-posed problem, for which the solution is close to the true solution of the original ill-posed problem. To do this, the solution is constrained by enforcing a non-negativity constraint (using the fact that the initial pressure distribution is non-negative) [22, 34], or a regularisation (penalty) function $\mathcal{R}(X)$ is added to the objective function [34]. Using these, a *constrained* and *regularised* variant of the minimisation problem (1.91) is defined by

$$\underset{X\geqslant 0}{\operatorname{argmin}} \left\{F_C(X) := f(X) + 2\lambda\mathcal{R}(X)\right\}, \tag{1.109}$$

where as in (1.91)

$$f(X) = \|\mathbb{H}X - \hat{P}\|^2. \tag{1.110}$$

Here, $\|\cdot\|$ is an Euclidean norm, and $\lambda > 0$ is a *regularisation parameter*, which makes a balance between fidelity to measurements and fidelity to an assumption about the true solution.

1.8.1 FBS methods

Here, an *unconstrained* form of (1.109) is of interest. This is derived by replacing the non-negativity constraint by an indicator function on a non-negativity set $C = \{X \geqslant 0\}$ that is added to the objective function F_C. Using this, an *unconstrained* variant of the minimisation problem (1.109) will be in the form

$$\underset{X}{\operatorname{argmin}} \left\{F(X) := f(X) + 2\lambda\ g(X)\right\}, \tag{1.111}$$

where $g(X) = 2\lambda\mathcal{R}(X) + \delta_C(X)$ with

$$\delta_C = \begin{cases} 0, & X \succcurlyeq 0 \\ +\infty, & X < 0. \end{cases} \tag{1.112}$$

Here, $f : \mathbb{R}^{N_n} \to \mathbb{R}$ is a continuously differentiable function with a Lipschitz continuous gradient having a Lipschitz constant $L > 0$ that satisfies [9, 10]

$$\|\nabla f(X) - \nabla f(Y)\| \leqslant L\|X - Y\| \tag{1.113}$$

for every $X, Y \in \mathbb{R}^{N_n}$. Additionally, $g : \mathbb{R}^{N_n} \to \mathbb{R}$ is a proper, closed, convex and possibly non-smooth function. We now define the *proximal operator* associated with function g as [10]

$$\text{prox}_g(X, \tau) := \underset{u}{\text{argmin}} \left\{ g(u) + \frac{1}{2\tau}\|u - X\|^2 \right\}. \tag{1.114}$$

The proximal operator finds a point close to the minimiser of g in a neighborhood of a starting point X. From the optimality condition of (1.111) and for any constant $\tau > 0$, an optimal solution X_* satisfies [9]

$$0 \in \tau\nabla f(X_*) + \tau\partial g(X_*), \tag{1.115}$$

where ∂g denotes the sub-gradient of g. Using (1.115), the solution X_* can be derived from [9]

$$\begin{aligned} 0 &\in \tau\nabla f(X_*) - X_* + X_* + \tau\partial g(X_*) \\ (I + \tau\partial g)(X_*) &\in (I - \tau\nabla f)(X_*). \end{aligned} \tag{1.116}$$

This finally gives

$$X_* = (I + \tau\partial g)^{-1}(I - \tau\nabla f)(X_*). \tag{1.117}$$

From (1.117), the solution x_* can be obtained using a fixed point iterative scheme of the form [9]

$$X_k = (I + \tau_k\partial g)^{-1}(I - \tau_k\nabla f)(X_{k-1}), \ (\tau_k > 0) \tag{1.118}$$

with $X_0 \in \mathbb{R}^{N_n}$. The operator $(I - \tau_k\nabla f)(X_{k-1})$ is a *forward gradient-descent* step on f, starting from X_{k-1} and moving along the direction $-\nabla f$ (negative gradient)

with a step-size τ_k [10, 27]. This operator amounts to a *steepest descent* step on f [9]. Additionally, the operator $(I + \tau_k \partial g)^{-1}$ is a *Backward gradient descent* step, and can be solved using [9]

$$(I + \tau \partial g)^{-1}(X) \equiv \text{prox}_g(X, \tau_k), \ \forall X \in \mathbb{R}^{N_n}. \tag{1.119}$$

Using the above, the iteration (1.118) can be fit into a class of *Forward-Backward Splitting* (FBS) methods for minimisation of a convex and possibly non-smooth function subject to constraints on the solution [9, 27]. Using *Forward-Backward Splitting* methods, each iteration involves two steps, the first of which is a steepest descent step on a smooth function f and the second step is a descent step on a possibly non-smooth g [9]. This method is outlined in Algorithm 1.1 [9, 27].

Algorithm 1.1 FBS method

1: Initialise: $X_0 \in \mathbb{R}^{N_n}$
2: **while** Not converged **do**
3: $\quad \hat{X}_k = X_{k-1} - \tau_k \nabla f(X_{k-1})$
4: $\quad X_k = \text{prox}_g(\hat{X}_k, \tau_k)$
5: **end while**

Here, for our PAT problem, the gradient of f is computed by

$$\nabla f(X) = 2\mathbb{H}^T \left(\mathbb{H}X - \hat{P} \right), \tag{1.120}$$

where the operators $\mathbb{H}$ and $\mathbb{H}^T$ are discretised variants of the acoustic forward and adjoint operators. Considering (1.120), the main cost of each iteration in Algorithm 1.1 is a computation of ∇f.

For cases in which the function g is separable, for example when g is an $\mathcal{L}^1$ norm, the proximal operator can be solved separately for each component of $\hat{X}$ as a one-dimensional problem. Specifically, using $g(X) = \|X\|_1$, the proximal operator is simplified into a Shrinkage Thresholding (ST) operator that satisfies [9, 10]

$$\begin{gathered} \mathcal{T} : \mathbb{R}^{N_n} \to \mathbb{R}^{N_n} \\ \mathcal{T}(X, \alpha)_i = (|X_i| - \alpha)_+ \, sgn(X_i), \end{gathered} \tag{1.121}$$

where α is a thresholding parameter, and sgn denotes the signum function. FBS approaches that use a proximal operator in the form (1.121) are called *Iterative Shrinkage Thresholding* Algorithms (ISTA) [9, 10].

Using a separable g, the main cost per iteration is the computation of a gradient using (1.120), and the computational cost of solving the proximal operator is negligible [37]. FBS algorithms (like ISTA) are very popular for large-scale linear inverse problems, and thus have been widely used for the acoustic inverse problem of PAT [4, 3, 34, 37].

The convergence of FBS algorithms is contingent on a certain stability bound on the sequence τ_k. Accordingly, the step-size can be either assumed constant, or selected adaptively using an appropriate backtracking line search [10]. Assuming a constant step-size $\tau_k = \tau$ for all k, it turns out that the convergence of the FBS algorithm is guaranteed using a step-size

$$\tau \in \left(0, \frac{2}{L_{\min}}\right), \tag{1.122}$$

where $L_{\min}$ is the smallest Lipschitz constant of ∇f (cf. (1.113)) [9, 10, 27]. Using (1.120), $L_{\min} = 2\|\mathbb{H}^T\mathbb{H}\| = 2\mathcal{S}_{\max}\left(\mathbb{H}^T\mathbb{H}\right)$, where $\mathcal{S}_{\max}(\cdot)$ denotes the maximum eigenvalue [9, 10]. Therefore, the FBS algorithm will converge, if $\tau \in \left(0, 1/\mathcal{S}_{\max}(\mathbb{H}^T\mathbb{H})\right)$ [9, 10].

Using a constant step-size $\tau = 1/L$, the sequence X_k generated by Algorithm 1.1 satisfies [9, 10]

$$F(X_k) - F(X_*) \leqslant \frac{L\|X_0 - X_*\|^2}{2k} \quad (\forall k \geqslant 1). \tag{1.123}$$

Here, X_* is a minimiser of function F, which is defined by (1.111). Considering (1.123), an FBS algorithm that uses Algorithm 1.1 provides a *sublinear* rate of convergence for function values $F(X_k)$ [9, 10].

1.8.2 Accelerated FBS methods

The global rate of convergence for FBS algorithms can be significantly improved using a specific linear combination of the previous two iterates at each iteration [9, 10]. An outline of this method is given in Algorithm 1.2 [9, 10, 27].

It has been shown that using a constant $\tau = 1/L$, the sequence X_k generated by Algorithm 1.2 satisfies [9, 10]

$$F(X_k) - F(X_*) \leqslant \frac{2L\|X_0 - X_*\|^2}{(k+1)^2} \quad (\forall k \geqslant 1). \tag{1.124}$$

Algorithm 1.2 Accelerated FBS method

1: Initialise: $Y_1 = X_0 \in \mathbb{R}^{N_n}, t_1 = 1$
2: **while** Not converged **do**
3: $\quad \hat{X}_k = Y_k - \tau_k \nabla f(Y_k)$
4: $\quad X_k = \text{prox}_g(\hat{X}_k, \tau_k)$
5: $\quad t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$
6: $\quad Y_{k+1} = X_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(X_k - X_{k-1})$
7: **end while**

This indicates that a worst-case rate of convergence for values $F(X_k)$ provided by an FBS approach that is accelerated using Algorithm 1.2 is *quadratic*. It is also worth mentioning that the convergence of an FBS algorithm with a fixed step-size requires $\tau \in (0, 2/L_{\min})$, whereas an accelerated FBS algorithm will converge, if $\tau \in (0, 1/L_{\min})$ [16].

1.8.3 Total variation (TV) regularisation

Having defined a general form of the FBS algorithm and its accelerated variant, here we will introduce our choice for function g and an associated proximal operator. Considering (1.111), here we are interested in total variation (TV) regularisation approaches, which remove noise and produce sharp interfaces at the cost of removing small-scale details in the reconstructed image [70]. This class of approaches are based on a minimisation of an integral of the gradient of the sought parameter distribution. In the context of image denoising, a TV regularisation approach was proposed by Chambolle based on a dual approach [14, 15]. Using this, a TV regularised variant of an FBS algorithm (and an accelerated variant) was derived, in which the dual approach of Chambolle was used for solving an associated proximal operator subject to a box constraint $X_l \leqslant X \leqslant X_u$ with X_l and X_u the lower and upper bounds [9]. Using the dual approach of Chambolle, the proximal operator involves a Shrinkage Thresholding (ST) step [9]. Because of this, an FBS approach that uses the dual approach of Chambolle can be fit into the class of ISTA approaches [10]. (See [4, 3, 34, 35] for applications in PAT). Using the dual approach of Chambolle, the proximal operator included in the FBS algorithm amounts to solving a denoising problem for an image that has been already calculated by a forward gradient step (cf. Algorithms 1.1 and 1.2).

Given a noisy image $f \in \mathbb{R}^d$, the denoising problem is to calculate $u \in \Omega$ ($\Omega \in \mathbb{R}^d$)

from

$$f = u + \eta, \tag{1.125}$$

where η is the noise in the image. One way for solving the denoising problem (1.125) is using the so-called *Rudin-Osher-Fatemi* (ROF) model [57]. The ROF minimisation problem is defined as

$$u = \underset{u \in BV(\Omega)}{\arg\min} \quad |u|_{BV} = \int_{\Omega} |\nabla u(\boldsymbol{r})| d\boldsymbol{r}, \tag{1.126}$$

subject to the constraints

$$\int_{\Omega} u(\boldsymbol{r}) d\boldsymbol{r} = \int_{\Omega} f(\boldsymbol{r}) d\boldsymbol{r}, \quad \int_{\Omega} |u(\boldsymbol{r}) - f(\boldsymbol{r})|^2 d\boldsymbol{r} = \sigma^2. \tag{1.127}$$

Here, $\boldsymbol{r}$ denotes the spatial space, ∇u is the gradient of u, and BV denotes the space of images with a bounded variation. Also,

$$\mathcal{J}(u) = \int_{\Omega} |\nabla u(\boldsymbol{r})| d\boldsymbol{r} \tag{1.128}$$

is the total variation functional. Note that the constraints in (1.127) indicate that η is independently and identically distributed as a Gaussian random variable with zero mean and standard deviation σ [57]. An unconstrained variant of the ROF minimisation problem gives

$$u = \underset{u \in BV(\Omega)}{\arg\min} \left\{ 2\mathcal{J}(u) + \nu \|f - u\|^2_{\mathcal{L}^2} \right\}, \tag{1.129}$$

where the first term is the TV functional, and the second term is a data fidelity norm associated with the constraints in (1.127) with ν a Lagrangian multiplier. Note that the coefficient 2 is multiplied by the first term in order to make (1.129) consistent with (1.111).

Also note that a comparison between (1.111) and a non-negativity constrained variant of (1.129) implies that $\nu \equiv 1/\lambda$. This problem has a minimiser in the space of functions with bounded variations $u \in BV(\Omega)$, if and only if $u \in \mathcal{L}^1(\Omega)$, and also

$$\mathcal{J}(u) = \sup_{\vec{\xi} \in C^1_c(\Omega;\mathbb{R}^d),\ |\vec{\xi}(\boldsymbol{r})| \leqslant 1} \left\{ \int_{\Omega} u(\boldsymbol{r}) \nabla \cdot \vec{\xi}(\boldsymbol{r}) d\boldsymbol{r} \right\} < \infty, \quad \forall \boldsymbol{r} \in \Omega, \tag{1.130}$$

where $\nabla \cdot \vec{\xi}$ is the divergence of ξ with $|\vec{\xi}| = \left(\sum_{i=1}^{d} \xi_i^2 \right)^{1/2}$ [70, 14].

Using (1.126) and (1.127), (or alternatively (1.129)), an Euler-Lagrange equation can be derived, i.e., [57]

$$u = f + \frac{1}{\nu}\nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right). \tag{1.131}$$

The non-differentiability of the right-hand-side of (1.131) can be handled by enforcing an approximation $|\nabla u| \approx \sqrt{|\nabla u|^2 + \beta}$, where $\beta > 0$ is a smoothing parameter [71]. Using (1.131), together with assuming u as a function of time t (as well as space), the minimisation problem can be solved by [57]

$$\frac{\partial u}{\partial t} = -\nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta}}\right) + \nu(u - f) \ (t > 0, r \in \Omega) \tag{1.132}$$

using an initialisation $u|_{t=0} = u^0$ and a homogeneous Neumann boundary condition $\partial u/\partial \hat{n} = 0$ with $\hat{n}$ a vector normal to the surface $\partial\Omega$. A spatial discretisation of (1.132) gives a forward Euler time marching scheme [57, 71]. Using an appropriate step size for the time integration, this scheme amounts to using a steepest descent algorithm, and thus suffers from a slow asymptotic convergence rate [57, 71]. Using higher-order methods such as nonlinear conjugate gradient or Newton's method provides a better rate of convergence for cases in which the objective function depends smoothly on u using a sufficiently large β. However, the efficiency and accuracy are deteriorated using a small β [71].

An alternative way is to linearise a fixed point iteration derived from the optimality condition of (1.129) using the last iterate u_k. This gives

$$\left(1 + \frac{1}{\nu}L(u_k)\right) u_{k+1} = f, \quad k = 0, 1, ..., \tag{1.133}$$

where

$$L(u)v = -\nabla \cdot \left(\frac{1}{\sqrt{|\nabla u|^2 + \beta}}\nabla v\right). \tag{1.134}$$

Now, we define a discretisation of our TV minimisation problem. Let us define a discretised image $\bar{u} \in \mathbb{R}^{N_n}$ with $N_n = N_1 \times ... \times N_d$. Also let $\zeta_i \in \{1, ..., N_i\}$ denote the position of sampled points along Cartesian coordinates $i \in \{1, ..., d\}$ as in section 1.7. Using this, we define a discretised gradient operator $D : \mathbb{R}^{N_1 \times ... \times N_d} \rightarrow \mathbb{R}^{N_1 - 1 \times .... N_i .... \times N_d} \times \mathbb{R}^{N_1 \times .... N_i - 1 .... \times N_d} \times \mathbb{R}^{N_1 \times .... N_i .... \times N_d - 1}$ in the form

$$D\bar{u} = \Big((D\bar{u})^1, ..., (D\bar{u})^i, ..., (D\bar{u})^d\Big), \tag{1.135}$$

where, $(D_i\bar{u})$ is defined by

$$(D\bar{u})^i_{(\zeta_1,...,\zeta_i,...,\zeta_d)} = \bar{u}_{(\zeta_1,...,\zeta_i,...,\zeta_d)} - \bar{u}_{(\zeta_1,...,\zeta_i+1,...,\zeta_d)}, \quad 1 \leqslant \zeta_i < N_i \tag{1.136}$$

Using (1.135) and (1.136), a discretised TV function is defined by [14]

$$J(\bar{u}) = \sum_{\zeta_1,,...,\zeta_d} \left[\sum_{i=1}^{d} \left((D\bar{u})^i_{(\zeta_1,...,\zeta_d)} \right)^2 \right]^{1/2}. \tag{1.137}$$

We also define the transpose of the gradient $D^T : \mathbb{R}^{N_1-1\times....N_i....\times N_d} \times \mathbb{R}^{N_1\times....N_i-1....\times N_d} \times \mathbb{R}^{N_1\times....N_i....\times N_d-1} \rightarrow \mathbb{R}^{N_1\times...\times N_d}$, which is defined by [14, 34]

$$D^T X = \sum_{i=1}^{d} (D^T X)^i, \tag{1.138}$$

where

$$(D^T X)^i_{(\zeta_1,...,\zeta_i,...,\zeta_d)} = \begin{cases} X^i_{(\zeta_1,...,\zeta_i,...,\zeta_d)} - X^i_{(\zeta_1,...,\zeta_i-1,...,\zeta_d)}, & 1 < \zeta_i < N_i \\ X^i_{(\zeta_1,...,\zeta_i,...,\zeta_d)}, & \zeta_i = 1 \\ -X^i_{(\zeta_1,...,\zeta_i-1,...,\zeta_d)}, & \zeta_i = N_i. \end{cases} \tag{1.139}$$

To include a non-negativity constraint, we also define an operator for an orthogonal projection to the set $C = \{X > 0\}$ using [14, 9, 34]

$$(\mathcal{P}_C \bar{u})_{(\zeta_1,...,\zeta_d)} = \max\left\{0, \bar{u}_{(\zeta_1,...,\zeta_d)}\right\} \tag{1.140}$$

Using the dual approach of Chambolle [14], we also define an operator $\mathcal{P}_p : \mathbb{R}^{N_1-1\times....N_i....\times N_d} \times \mathbb{R}^{N_1\times....N_i-1....\times N_d} \times \mathbb{R}^{N_1\times....N_i....\times N_d-1} \rightarrow \mathbb{R}^{N_1-1\times....N_i....\times N_d} \times \mathbb{R}^{N_1\times....N_i-1....\times N_d} \times \mathbb{R}^{N_1\times....N_i....\times N_d-1}$ in the form

$$\mathcal{P}_p X = \Big((\mathcal{P}_{p,1} X^1), ..., (\mathcal{P}_{p,i} X^i), ..., (\mathcal{P}_{p,d} X^d) \Big), \tag{1.141}$$

where

$$(\mathcal{P}_{p,i} X)_{(\zeta_1,...,\zeta_d)} = \frac{X^i_{(\zeta_1,...,\zeta_d)}}{\max\left\{1, \left[\sum_{i=1}^{d} (X^i_{(\zeta_1,...,\zeta_d)})^2\right]^{1/2}\right\}}. \tag{1.142}$$

Using the operators defined above and using a fixed $\tau = \tau_k$ $(\forall k)$, we outline our proximal operator using the dual approach of Chambolle in Algorithm 1.3 [9, 34] . Here, the superscript k denotes the (inner) iteration number associated with our proximal operator with $*$ the final iteration.

Algorithm 1.3 Proximal operator $\text{prox}_g\left(\hat{X}, \tau^k\right)$

1: Input: $\hat{X} \in \mathbb{R}^{N_1-1\times....N_i....\times N_d} \times \mathbb{R}^{N_1\times....N_i-1....\times N_d} \times \mathbb{R}^{N_1\times....N_i....\times N_d-1}, \lambda' = \lambda\tau, d \in \{2,3\}$
2: Initialise: $P^0 = 0^{N_1-1\times....N_i....\times N_d} \times 0^{N_1\times....N_i-1....\times N_d} \times 0^{N_1\times....N_i....\times N_d-1}$
3: **while** Not converged **do**
4: $\quad P^k = \mathcal{P}_p\Big[P^{k-1} + \frac{1}{2d\lambda'} D\left(\mathcal{P}_C\Big[\hat{X} - 0.5\lambda' D^T P^{k-1}\Big]\right)\Big]$
5: **end while**
6: $X^* = \mathcal{P}_C\left(\hat{X} - \lambda D^T P^*\right)$

Algorithm 1.4 Accelerated proximal operator $\text{prox}_g\left(\hat{X}, \tau^k\right)$

1: Input: $\hat{X} \in \mathbb{R}^{N_1-1\times....N_i....\times N_d} \times \mathbb{R}^{N_1\times....N_i-1....\times N_d} \times \mathbb{R}^{N_1\times....N_i....\times N_d-1}, \lambda' = \lambda\tau, d \in \{2,3\}, k_{\max}, t^1 = 1$
2: Initialise: $Q^1 = P^0 = 0^{N_1-1\times....N_i....\times N_d} \times 0^{N_1\times....N_i-1....\times N_d} \times 0^{N_1\times....N_i....\times N_d-1}$
3: **while** Not converged **do**
4: $\quad P^k = \mathcal{P}_p\Big[Q^k + \frac{1}{2d\lambda'} D\left(\mathcal{P}_C\Big[\hat{X} - 0.5\lambda' D^T Q^k\Big]\right)\Big]$
5: $\quad t^{k+1} = \frac{1+\sqrt{1+4t^{k^2}}}{2}$
6: $\quad Q^{k+1} = P^k + \left(\frac{t^k-1}{t^{k+1}}\right)(P^k - P^{k-1})$
7: **end while**
8: $X^* = \mathcal{P}_C\big[\hat{X} - \lambda D^T P^*\big]$

In a same way as Algorithm 1.2, an accelerated version for the proximal operator defined in Algorithm 1.3 can be derived. This is outlined in Algorithm 1.4 [9, 34]. We terminate Algorithms 1.3 and 1.4 when

$$k > m \ \cap \ \|X^k - X^{k-m}\| < \varepsilon_g, \tag{1.143}$$

where ε_g is a terminating threshold, and m is a user-defined number of iterations. We will define the stopping criteria for Algorithms 1.1 and 1.2 in chapters 2 and 3.

1.9 Optical problem

In the previous section, we explained the acoustic portion of photo-acoustic tomography, which is a calculation of an initial pressure distribution (or absorbed optical energy) from time series of measured pressure data at the boundary. We also discussed iterative model-based approaches for solving ill-posed problems that arise for example from a limited or few-view data set or an erroneous estimation of acoustic properties [34, 37].

The main goal of photo-acoustic tomography is a quantitative reconstruction of a distribution of optical absorption coefficients from time series of boundary data [28]. This is referred to as *Quantitative* photo-acoustic tomography (QPAT). This involves two inverse problems, the former of which is a calculation of an image of absorbed optical energy from measured pressure data, and the latter is a calculation of the optical coefficients from a distribution of absorbed energy [24, 25, 28]. The optical portion of the QPAT problem is highly nonlinear and ill-posed in the sense of Hadamard [24, 28]. It was shown that this problem is non-unique for a single optical source [7] or frequency [8]. These two inverse problems can be either solved separately [25, 29, 61, 62], or combined as a direct composite problem [24, 28, 51]. The latter approach has recently received much attention, because the optical coefficients, as opposed to the absorbed optical energy, are independent from the used optical source or frequency, and therefore, by solving a direct variant of the inverse problem, it will be possible to improve stability of the inversion process using multi-source [7, 25, 24, 51] or multi-frequency settings [8]. Another advantage of using a direct composite form of QPAT is that both the optical and acoustic portions of the problem can simultaneously use information about noise in data [51] and an assumption about the true optical coefficients [24]. In contrast, by solving the optical and acoustic inverse problems distinctly, the acoustic portion of the inverse problem cannot benefit from an assumption about the optical coefficients, and also the optical portion of the inverse problem fails to utilise information about noise in boundary data [24, 28, 51].

Using both approaches for the QPAT problem, the inverse problem is often solved by model-based algorithms in order to deal with the nonlinearity and ill-condition of an associated optical forward operator. The same as an acoustic inverse problem, model-based approaches are based on modelling the physics of an optical problem.

1.9.1 Radiative Transfer Equation (RTE)

A very accurate model for simulating the propagation of optical photons is the Radiative Transfer Equation (RTE). This model is a simplified version of a *Transport* equation using an assumption that the particles undergo elastic collisions, i.e., the energy of the particles does not change in collisions [63]. To explain the RTE model, we define an optical domain $\Omega_o \subset \Omega$, which is isotropic, i.e., the probability of the

scattering between two directions depends only on the relative angle between two directions. Also, we assume that the support of the optical medium Ω_o is inside the acoustic medium Ω.

Using RTE, the propagation of optical photons in a medium $\Omega_o \subset \Omega$ is defined by [63]

$$\begin{aligned} &\frac{1}{c_o(\boldsymbol{r})}\frac{\partial\phi(\boldsymbol{r},\hat{s})}{\partial t} + \hat{s}\cdot\nabla\phi(\boldsymbol{r},\hat{s}) + (\mu_s(\boldsymbol{r}) + \mu_a(\boldsymbol{r}))\,\phi(\boldsymbol{r},\hat{s}) = \\ &\mu_s(\boldsymbol{r})\int_{S^{d-1}}\theta(\hat{s}\cdot\hat{s}')\phi(\boldsymbol{r},\hat{s}')d\hat{s}' + q(\boldsymbol{r},\hat{s}), \end{aligned} \tag{1.144}$$

where c_o denotes the speed of light in tissue, μ_a and μ_s are the absorption and scattering coefficients, $\phi(\boldsymbol{r},\hat{s})$ is the radiance, and $q(\boldsymbol{r},\hat{s})$ is an internal source (inside Ω_o). Also $\theta(\hat{s}\cdot\hat{s}')$ is the scattering phase function, and describes the probability that a photon with a direction $\hat{s}'$ has a direction $\hat{s}$ after a scattering event. It satisfies [63]

$$\int_{S^{d-1}}\theta(\hat{s}\cdot\hat{s}')d\hat{s} = \int_{S^{d-1}}\theta(\hat{s}\cdot\hat{s}')d\hat{s}' = 1. \tag{1.145}$$

For an isotropic optical medium, the phase function can be described by a Henyey-Greenstein scattering function in the form

$$\theta(\hat{s}\cdot\hat{s}') = \begin{cases} \frac{1}{2\pi}\frac{1-g_o^2}{1+g_o^2-2g_o\cos\gamma}, & \text{if } d=2 \\ \\ \frac{1}{4\pi}\frac{1-g_o^2}{(1+g_o^2-2g_o\cos\gamma)^{3/2}}, & \text{if } d=3. \end{cases} \tag{1.146}$$

Here, γ is the angle between directions $\hat{s}$ and $\hat{s}'$, and $-1 < g_o < 1$ is the scattering anisotropy factor, and describes the shape of the probability density [63].

Using a boundary source $\phi_0(\boldsymbol{r},\hat{s})$ ($\boldsymbol{r} \in \partial\Omega_o$) which is nonzero at location of optical sources, the boundary condition is defined by

$$\phi(\boldsymbol{r},\hat{s}) = \phi_0(\boldsymbol{r},\hat{s}), \quad \boldsymbol{r} \in \partial\Omega_o,\ \hat{s}\cdot\hat{n} < 0, \tag{1.147}$$

where $\hat{n}$ is an outward unit normal.

1.9.2 Diffusion Approximation (DA) to RTE

Although RTE is a very accurate model for simulating the propagation of light, it requires a computation of the radiance in space and phase, and is thus computationally very expensive. As a result, RTE is not practical for 3D QPAT [58]. A more efficient

model for propagation of light in tissue media is provided by a diffusion approximation (DA) to RTE. This can be derived by a P_1 approximation to an expansion of the radiance, source term and the phase function into series of spherical harmonics [63]. The P_1 approximation is based on an assumption that the angular distribution of the mentioned functions is uniform for a highly scattering medium. To introduce the PDEs describing the DA model, we start with a definition of the photon density as

$$\phi(\boldsymbol{r}) = \int_{S^{d-1}} \phi(\boldsymbol{r}, \hat{s})\ d\hat{s}. \tag{1.148}$$

We also define the photon current as

$$J_o(\boldsymbol{r}) = \int_{S^{d-1}} \hat{s}\phi(\boldsymbol{r}, \hat{s})\ d\hat{s}. \tag{1.149}$$

Using a P_1 approximation, the radiance is defined by [63]

$$\phi(\boldsymbol{r}, \hat{s}) \approx \frac{1}{|S^{d-1}|}\phi(\boldsymbol{r}) + \frac{d}{|S^{d-1}|}\hat{s} \cdot J_o(\boldsymbol{r}). \tag{1.150}$$

The same approximation for the source term gives

$$q(\boldsymbol{r}, \hat{s}) \approx \frac{1}{|S^{d-1}|}q_0(\boldsymbol{r}) + \frac{d}{|S^{d-1}|}\hat{s} \cdot q_1(\boldsymbol{r}). \tag{1.151}$$

Here, q_0 and q_1 are the isotropic and dipole components of an internal source, and are respectively defined by [63]

$$\begin{aligned} q_0(\boldsymbol{r}) &= \int_{S^{d-1}} q(\boldsymbol{r}, \hat{s})\ d\hat{s} \\ q_1(\boldsymbol{r}) &= \int_{S^{d-1}} \hat{s}q(\boldsymbol{r}, \hat{s})\ d\hat{s}. \end{aligned} \tag{1.152}$$

In the same way, a P_1 approximation of the scattering phase function yields

$$\theta(\hat{s} \cdot \hat{s}') \approx \frac{1}{|S^{d-1}|}g_0 + \frac{d}{|S^{d-1}|}(\hat{s} \cdot \hat{s}')g_1, \tag{1.153}$$

where,

$$\begin{aligned} g_0(\boldsymbol{r}) &= \int_{S^{d-1}} \theta(\hat{s} \cdot \hat{s}')\ d\hat{s} = 1 \\ g_1(\boldsymbol{r}) &= \int_{S^{d-1}} (\hat{s} \cdot \hat{s}')\theta(\hat{s} \cdot \hat{s}')\ d\hat{s}. \end{aligned} \tag{1.154}$$

Using a Henyey-Greenstein scattering function (cf.(1.146)) yields $g_1 = g_o$. Plugging (1.153) into the first term in the right-hand-side of (1.144) gives [63]

$$\mu_s(\boldsymbol{r}) \int_{S^{d-1}} \theta(\hat{s} \cdot \hat{s}')\phi(\boldsymbol{r}, \hat{s}')d\hat{s}' = \mu_s(\boldsymbol{r})\frac{1}{|S^{d-1}|}\phi(\boldsymbol{r}) + \mu_s(\boldsymbol{r})\frac{d}{|S^{d-1}|}g_1\hat{s} \cdot J_o(\boldsymbol{r}). \tag{1.155}$$

In the same way, plugging the approximations in (1.150) and (1.151) into the left-hand-side of (1.144) yields

$$\frac{1}{|s^{d-1}|}\left(\frac{1}{c_o(\boldsymbol{r})}\frac{\partial}{\partial t}+\hat{s}\cdot\nabla+\mu_a(\boldsymbol{r})\right)\phi(\boldsymbol{r})+\frac{d}{|S^{d-1}|}\left(\frac{1}{c_o(\boldsymbol{r})}\frac{\partial}{\partial t}+\hat{s}\cdot\nabla+\mu_a(\boldsymbol{r})+\mu_s'(\boldsymbol{r})\right)\hat{s}\cdot J_o(\boldsymbol{r})$$
$$=\frac{1}{|s^{d-1}|}q_0(\boldsymbol{r})+\frac{d}{|s^{d-1}|}\hat{s}\cdot q_1(\boldsymbol{r}), \tag{1.156}$$

where, μ_s' is the *reduced scattering coefficient*, and satisfies

$$\mu_s'=\mu_s(1-g_1). \tag{1.157}$$

An integration of both sides of (1.156) over angular directions S^{d-1} gives

$$\left(\frac{1}{c_o(\boldsymbol{r})}\frac{\partial}{\partial t}+\mu_a\right)\phi(\boldsymbol{r})+\nabla\cdot J_o(\boldsymbol{r})=q_0(\boldsymbol{r}) \tag{1.158}$$

Also, multiplying both sides of (1.156) with $\hat{s}$ and an integration over angular directions S^{d-1}, gives

$$\left(\frac{1}{c_o(\boldsymbol{r})}\frac{\partial}{\partial t}+\mu_a+\mu_s'\right)J_o(\boldsymbol{r})+\frac{1}{d}\nabla\phi(\boldsymbol{r})=q_1(\boldsymbol{r}). \tag{1.159}$$

For derivation of a DA model, we assume that the light source is isotropic ($q_1=0$), and also $\frac{1}{c_o}\frac{\partial J_o(\boldsymbol{r})}{\partial t}=0$. For the latter assumption, one uses the fact that $\mu_a \ll \mu_{s'}$. Applying these assumptions, (1.159) gives

$$J_o(\boldsymbol{r})=-\kappa(\boldsymbol{r})\nabla\phi(\boldsymbol{r}), \tag{1.160}$$

where κ denotes the diffusion coefficient, and is in the form

$$\kappa(\boldsymbol{r})=\frac{1}{d\left(\mu_a(\boldsymbol{r})+\mu_s'(\boldsymbol{r})\right)^{-1}}. \tag{1.161}$$

Equation (1.160) is called *Fick's law* [63]. Now, by plugging (1.160) into (1.158), the DA form of RTE is derived. This is in the form

$$-\nabla\cdot\kappa(\boldsymbol{r})\nabla\phi(\boldsymbol{r})+\mu_a\phi(\boldsymbol{r})+\frac{1}{c_o}\frac{\partial\phi(\boldsymbol{r})}{\partial t}=q_0(\boldsymbol{r}). \tag{1.162}$$

Using the P_1 approximation, the total inward-directed photon flux at a point $\boldsymbol{r}\in\partial\Omega_0$ is defined by

$$\Gamma_-(\boldsymbol{r})=\int_{\hat{s}\cdot\hat{n}<0}(\hat{s}\cdot\hat{n})\,\phi(\boldsymbol{r},\hat{s})d\hat{s}=\boldsymbol{\gamma}_d\phi(r)-\frac{1}{2}\hat{n}\cdot J_o(\boldsymbol{r}),\ (\boldsymbol{r}\in\partial\Omega_o) \tag{1.163}$$

where $\boldsymbol{\gamma}_2 = 1/\pi$ and $\boldsymbol{\gamma}_3 = 1/4$. In the same way, the total outward-directed photon flux is defined by

$$\Gamma_+(\boldsymbol{r}) = \int_{\hat{s}\cdot\hat{n}>0} (\hat{s}\cdot\hat{n})\,\phi(\boldsymbol{r},\hat{s})d\hat{s} = \boldsymbol{\gamma}_d\phi(\boldsymbol{r}) + \frac{1}{2}\hat{n}\cdot J_o(\boldsymbol{r}),\ (\boldsymbol{r}\in\partial\Omega_o) \tag{1.164}$$

In an absence of optical sources at the boundary, we have

$$\Gamma_-(\boldsymbol{r}) = 0\ (\boldsymbol{r}\in\Omega_o). \tag{1.165}$$

Using this assumption, together with plugging the Fick's law into (1.163), gives a boundary condition in the form

$$\phi(\boldsymbol{r}) + \frac{1}{2\boldsymbol{\gamma}_d}\kappa\frac{\partial\phi(\boldsymbol{r})}{\partial\hat{n}} = 0, \quad \boldsymbol{r}\in\partial\Omega_o. \tag{1.166}$$

Additionally, reflections may occur at the boundary because of different refractive indices between the tissue and the surrounding medium. To account for these effects, (1.165) is replaced by

$$\Gamma_-(\boldsymbol{r}) = R_o\Gamma_+(\boldsymbol{r})\ (\boldsymbol{r}\in\Omega_o), \tag{1.167}$$

where R_o is the reflection coefficient at the boundary $\partial\Omega_o$. Plugging (1.163) and (1.164) into (1.167) gives

$$\phi(\boldsymbol{r}) + \frac{1}{2\boldsymbol{\gamma}_d}\kappa(\boldsymbol{r})\boldsymbol{\mathcal{A}}\frac{\partial\phi(\boldsymbol{r})}{\partial\hat{n}} = 0, \quad \boldsymbol{r}\in\partial\Omega_o. \tag{1.168}$$

where $\boldsymbol{\mathcal{A}}$ is the coefficient associated with the mismatched refractive indices, and satisfies

$$\boldsymbol{\mathcal{A}} = \frac{1+R_o}{1-R_o}. \tag{1.169}$$

Using DA, an optical (external) source is also enforced to the boundary condition in order to simulate an optical illumination. For optical tomography or QPAT, this can be defined either as a *collimated*, or *diffuse* source. Using a diffusive source model, which is more common for QPAT [63], the boundary condition is modified as

$$\phi(\boldsymbol{r}) + \frac{1}{2\boldsymbol{\gamma}_d}\kappa(\boldsymbol{r})\boldsymbol{\mathcal{A}}\frac{\partial\phi(\boldsymbol{r})}{\partial\hat{n}} = \frac{I_s}{\gamma_d}, \quad \boldsymbol{r}\in\partial\Omega_o, \tag{1.170}$$

where I_s denotes an inward-directed diffusive boundary current. Note that I_s is nonzero at the position of optical sources. Equation (1.170), which is called a *Robin*

boundary condition, has been widely used as a boundary condition for simulating the light illumination on tissue media [63].

Using a steady state variant of the DA model (1.162) using $\partial\phi(\boldsymbol{r})/\partial t = 0$, together with an assumption of no internal source $q_0(\boldsymbol{r}) = 0$, yields

$$-\nabla \cdot \kappa(\boldsymbol{r})\nabla\phi(\boldsymbol{r}) + \mu_a(\boldsymbol{r})\phi(\boldsymbol{r}) = 0, \ \boldsymbol{r} \in \Omega_o \tag{1.171}$$

Equation (1.171), together with the boundary condition (1.170), gives a system of coupled PDEs for modelling illumination and propagation of optical photons for QPAT [61, 62].

The absorbed optical energy, which is equal to the initial pressure distribution under an assumption $\Gamma(\boldsymbol{r}) = 1$, is defined by

$$H(\boldsymbol{r}) = \mu_a(\boldsymbol{r})\phi(\boldsymbol{r}). \tag{1.172}$$

The most common numerical approach for approximation of this system of equations is *Finite element method* (FEM) [63], and is also used in our study. An FEM approach is applied on a variational form of these equations. For further details on a numerical implementation of (1.171) and boundary condition (1.170), the reader is referred to chapter 4.

Considering (1.172), together with the fact that ϕ also depends on μ_a, the dependence of H on μ_a is highly non-linear [24, 25]. Because of this, together with the fact that tissue media are highly scattering, the inverse problem of solving the optical coefficients μ_a and μ_s' from H is a highly-nonlinear and ill-posed problem [61, 62]. We emphasise that the parameter of interest for QPAT is only the absorption coefficient μ_a, but the forward operator is also dependent on μ_s', which is not available in practice. Therefore, a simultaneous reconstruction of μ_a and μ_s' is often indicated for QPAT [61, 51]. An alternative inverse problem has also been considered in many studies, for which a simultaneous reconstruction of μ_a and κ is of interest [24, 25, 29]. In this thesis, we used the latter approach for solving the inverse problem of QPAT (cf. chapter 4).

1.10 Outline of thesis

This thesis has been written in a journal (alternative) format. Here, the main reasons for choosing this format are briefly explained. At the time of a decision on this format, the candidate and supervisor felt that the obtained analytic and numerical results might be of interest to researchers working in the related fields, and a presentation of the obtained results for larger related communities might be more appropriate. Therefore, the candidate preferred to put more time on publishing the obtained results in some peer-reviewed journals, and to write a thesis as a collection of the papers.

Chapters 2 and 3 of this thesis are identical copies of the corresponding published papers in peer-reviewed journals. In addition, chapter 4 is an identical copy of a manuscript accepted in a peer-reviewed journal. A paragraph has been included at the beginning of chapters 2, 3 and 4 for referring to the corresponding published or accepted papers. In addition, we provided a section, entitled *Supplementary materials*, at the end of these chapters. These sections provide a link between the papers, and also include some information that have not been embedded in the papers because of space constraints. In chapter 5, we will present a summary of this thesis, and give some discussions and suggestions for future works. Note that the contents of chapters 1 and 5, together with sections entitled *Supplementary materials* (cf. at the end of chapters 2-4), have not been published or submitted for publication elsewhere, and also we will not publish them elsewhere after the submission of this thesis. In the sequel, we give a brief introduction for the following chapters.

1.10.1 Chapter 2: A multi-grid iterative method for photoacoustic tomography

To cope with a high computation cost due to the iterative nature of variational approaches for PAT, we develop a line-search multi-grid (MG) version of first-order Forward-Backward Splitting (FBS) methods in order to improve the speed of image reconstruction using this class of inversion approaches [37]. The forward operator we consider for our PAT problem is defined using a system of three-coupled first-order wave equations defined in (1.92) with the third equation replaced by an equation of state accounting for an absorption and dispersion following a frequency power law, as

defined in (1.94). Also, the initial values are set using (1.93).

As discussed in section 1.6, variational approaches are popular for PAT, because they are tolerant in dealing with practical difficulties, e.g., errors in estimation of medium's properties, errors in data collection, and a limited-view or a few-view detection surface [34]. Variational approaches significantly improve the image reconstruction for PAT, but they are computationally very expensive because of a need for an iterative implementation of the forward operator and an associated adjoint operator [4]. As explained in section 1.8, among variational approaches for the acoustic inverse problem of PAT (for cases in which we are interested in solving the optical and acoustic inverse problems distinctly), FBS methods perform well regarding accuracy and speed [34, 4, 3].

The forward gradient step in an FBS algorithm uses a steepest descent search direction, which is a negative gradient of the smooth part of the objective function (data fidelity norm) as a function of a forward operator and a corresponding adjoint (cf. Algorithms 1.1 and 1.2). In [34], by deriving a matrix form of a numerical implementation of our considered forward operator using a k-space pseudo-spectral approach, an algebraic adjoint was proposed using a *discretise-then-adjoint* method. In chapter 2, we will also derive an adjoint of this forward operator on a continuous domain using an *adjoint-then-discretise* method. We shall show that in an absence of absorption and dispersion effects, our derived continuous adjoint matches a continuous adjoint that has been previously derived in [4] for lossless media. A continuous adjoint can be expressed as a generalised variant of a corresponding disretised adjoint in the sense that it is independent from numerical schemes for solving the forward and adjoint pair [4], and thus any modifications in the forward modelling can be straightforwardly incorporated into the optimisation framework.

We numerically approximate our forward and associated adjoint operators using a k-space pseudo-spectral method, as explained in section 1.7. To do this, we use an open-source toolbox, which is freely available on the *k-Wave* website [65, 68].

For an application of FBS methods on PAT, the main computational cost of each iteration is because of the search direction (negative gradient) in the forward step, and the cost of the proximal operator is negligible [37]. Motivated by this fact, instead of a computation of the search direction at a target (fine) level for all iterations, we

compute the search direction alternatively by deciding between a negative gradient at a target level or alternatively an approximate error correction at a coarse level [37]. From a practical point of view, our proposed MG algorithm performs well for cases in which the coarse level has a spacing of grid points not larger than the spatial distance of ultrasound detectors. By applying our forward operator on a lossy and heterogeneous acoustic medium, together with a derivation of an associated adjoint on a continuous domain, we will show that an application of our MG algorithm on two levels improves the speed of reconstruction, compared to a fixed grid [37]. We use a TV regularised variant of ISTA and FISTA, as explained in section 1.8. For each level, the forward operator is applied on a smoothed version of initial pressure distribution, and the acoustic properties (sound speed and density distributions) are also smoothed in order to mitigate errors arising from finite sampling (See the manual on the *k-Wave* website [65]).

1.10.2 Chapter 3: A continuous adjoint for photo-acoustic tomography of the brain

An algebraic adjoint operator has been recently derived for variational approaches in transcranial PAT using an isotropic heterogeneous elastic and lossy medium, for which the acoustic absorption is described by a diffusive model which neglects the dependency of absorption on temporal frequency [46]. This adjoint operator is algebraically derived from a matrix form of a numerical computation of the forward operator using a Finite difference time-domain (FDTD) method [46]. Motivated by this study [46], we derive a continuous adjoint for transcranial PAT using an isotropic heterogeneous elastic and lossy medium, for which the absorption and physical dispersion follow a frequency power law using two fractional Laplacian operators [67], similar to (1.94). For a numerical implementation of the forward operator, we modified an open-source code that is available on the *k-Wave* toolbox [65] so that it includes two fractional Laplacian operators in order to account for the frequency-dependent absorption and dispersion effects (See [67]). We analytically show that by a numerical computation of the forward and adjoint pair using a k-space pseudo-spectral method, our continuous adjoint matches an associated discretised adjoint [38]. From a numerical point

of view, we validate our derived continuous adjoint using an inner product test [38]. Using these forward and adjoint operators, we applied a variational approach using a TV regularised variant of ISTA in order to iteratively solve our PAT problem for brain imaging (cf. section 1.8).

1.10.3 Chapter 4: Direct quantitative photoacoustic tomography for realistic acoustic media

A reconstruction of a distribution of the optical absorption coefficient from time series of pressure data, which is referred to as quantitative photo-acoustic tomography, involves two inverse problems, namely acoustic and optical [28]. As explained at the beginning of section 1.9, for a traditional variant of QPAT, these two inverse problems are solved separately [61, 62], but it has been shown that solving the QPAT problem directly using a composite *opto-acoustic* forward operator and an associated *acousto-optic* adjoint operator improves the stability of the reconstruction [28, 24]. The reasons have been explained in section 1.9 (See also section 1 in chapter 4.) To the best of our knowledge, for existing studies on the direct composite problem of QPAT, the acoustic portion of the forward operator is solved using methods based on Green's function, and an associated adjoint is derived using exact inversion formulae. (See [28, 24, 51].)

As mentioned before, these approaches are often based on assuming an acoustically homogeneous and lossless medium, which does not hold for tissue media. Additionally, the optical inverse problem of QPAT is highly nonlinear and ill-posed. Therefore, if the acoustic heterogeneity or attenuation are not accounted for in modelling the acoustic portion of the problem, the arising errors may be magnified by the small singular values of the optical forward operator, and deteriorate the convergence of the inversion algorithm.

Motivated by this fact, we solve the direct problem of QPAT using a composite *opto-acoustic* forward operator, in which the acoustic portion solves a linear system of PDEs using (1.92), which can be adapted to variations in acoustic properties, and uses an equation of state defined by (1.94), which has two fractional Laplacian operators in order to account for an acoustic absorption and dispersion following a frequency power law [66]. The same as previous chapters, we use a k-space pseudo-spectral method for a

numerical implementation of the acoustic forward operator using an open-source code available on the *k-wave* website [65]. (Note that the acoustic portion of the forward operator is the same as in chapter 2.) We also use the DA model explained in section 1.9.2 for solving the optical portion of the composite forward operator.

We use a *discretise-then-adjoint* method for deriving an associated *acousto-optic* adjoint operator, which is an algebraic adjoint of the Fréchet derivative of the composite forward operator. The acoustic portion of our algebraic adjoint operator is novel in the sense that we include the effects of PMLs in calculation of the adjoint, as opposed to an algebraic adjoint derived in [34] for the same acoustic forward operator. Using a finite element method (FEM) for a numerical implementation of the optical portion of the forward operator, we analytically show that an adjoint of the Fréchet derivative of the optical forward operator using an *adjoint-then-discretise* approach with a discretisation using an $\mathcal{L}^2$ projection of the nodal values matches an algebraic adjoint of Fréchet derivative of an associated discretised optical forward operator.

We use an inexact Newton method [21] for solving the direct problem of QPAT. To do this, a nonlinear objective function is iteratively linearised, and an arising linear system of equations, which amounts to a left multiplication of the inverse of Hessian matrix by a vector of negative gradient, is solved using a preconditioned conjugate gradient (PCG) method with a total variation matrix as a preconditioner [29]. For solving each linearised subproblem, the Jacobian matrix and its transpose are computed implicitly in order to handle memory issues. The regularisation is adjusted implicitly by enforcing an early stopping criterion to each linearised subproblem. (See [5].) Our numerical results show that the developed inexact Newton algorithm performs better than non-linear gradient-based methods that use a Quasi-Newton search direction. The nonlinear gradient-based methods that use a Quasi-Newton search direction are popular for QPAT, e.g. [25, 24]. Specifically, using a TV regularisation, we develop two inexact Newton methods using Lagged diffusivity (LD) and Primal Dual Interior Point Method (PD-IPM), and compare them with a gradient-based method that uses a limited-memory BFGS approach for computing a search direction and an alternating direction method of multipliers (ADMM) for applying a TV regularisation. (See [25, 24] for an application in QPAT and direct variant of QPAT, respectively.)

1.10.4 Chapter 5: Conclusion, discussion and future works

In addition to conclusions and discussions given at the end of chapters 2-4, we will present a summary of this thesis in chapter 5. This also includes some discussions about the obtained numerical results. Furthermore, using the numerical results, we will give some ideas and suggestions for future works.

Notations for the discretised forward operators

Here, we introduce the notations we use for the discretised forward operators in this thesis. In chapter 2, the discretised forward operator has been denoted by H, and in chapter 3, we will use $\bar{\mathbb{H}}$ for the discretised forward operator. In chapter 4, in which the forward operator is a composite opto-acoustic map, we will use $\mathbb{H}_o$ and $\mathbb{H}_a$ for the optical and acoustic portions of a discretisation of our composite forward operator $\mathbb{H}$. Throughout this thesis, we use the superscript $*$ for denoting the adjoint of an operator for both continuous and discretised domains.

1.11 Contributions of the Supervisor/Candidate

The published papers that will be presented in the following chapters are the results of collaborations between the PhD candidate and his supervisor. Here, the contributions of the supervisor and candidate are clarified.

For chapter 2, an adjoint was calculated using an *adjoint-then-discretise* approach (continuous adjoint) by the supervisor. The rest of this work including the numerical experiments was done by the candidate, and he also used the advice and ideas provided by the supervisor during this work.

For chapter 3, an adjoint using an *adjoint-then-discretise* approach was calculated by the supervisor for lossless media. The derived continuous adjoint was extended to lossy media by the candidate. This involved an incorporation of acoustic absorption and dispersion following a frequency power law in the forward operator and a corresponding continuous adjoint operator. The candidate also derived the discretised adjoint, and established a relation between the continuous and discretised adjoint. The

rest of this work including the numerical experiments were done by the candidate. The supervisor also applied some amendments to the paper before submission to a corresponding journal. These amendments were applied in order to make the manuscript concise and suitable for pulication in a mathematical journal. The supervisor also provided useful suggestions and ideas during the work.

For chapter 4, the Fréchet derivative of the optical forward operator and its adjoint on a continuous domain have been derived by the supervsior. He also gave some advice to the candidate for an analytic derivation of a Primal-Dual Interior-Point-Method (PD-IPM) algorithm for a total variation regularisation adapted to a nonlinear optimisation problem. The candidate managed the rest of this work using advice and guidance provided by the supervisor.

Bibliography

[1] M. A. Anastasio, J. Zhang, and X. Pan. Image reconstruction in thermoacoustic tomography with compensation for acoustic heterogeneities. *Proc.SPIE*, 5750:5750 – 5750 – 7, 2005.

[2] M. A. Anastasio, J. Zhang, X. Pan, Y. Zou, G. Ku, and L. Wang. Half-time image reconstruction in thermoacoustic tomography. *IEEE Trans. Med. Imag.*, 24(2):199 – 210, 2005.

[3] 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. pages 1–19, 2016. arXiv:1605.00133v1.

[4] 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.

[5] S. R. Arridge, M. M. Betcke, and L. Harhanen. Iterated preconditioned lsqr method for inverse problems on unstructured grids. *Inverse Problems*, 30(7):075009, 2014.

[6] D. Baek, J. A. Jensen, and M. Willatzen. Modeling transducer impulse responses

for predicting calibrated pressure pulses with the ultrasound simulation program field ii. *The Journal of the Acoustical Society of America*, 127(5):2825–2835, 2010.

[7] G. Bal and K. Ren. Multi-source quantitative photoacoustic tomography in a diffusive regime. *Inverse Problems*, 27(7):075003, 2011.

[8] G. Bal and K. Ren. On multi-spectral quantitative photoacoustic tomography in diffusive regime. *Inverse Problems*, 28(2):025010, 2012.

[9] 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.

[10] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[11] A. Bell. Upon the production and reproduction of sound by light. *Am. J. Sci.*, 20:305–324, 1880.

[12] A. Buehler, A. Rosenthal, T. Jetzfellner, A. Dima, D. Razansky, and V. Ntziachristos. Model-based optoacoustic inversions with incomplete projection data. *Medical Physics*, 38(3):1694–1704.

[13] P. Burgholzer, G. J. Matt, M. Haltmeier, and G. Paltauf. Exact and approximative imaging methods for photoacoustic tomography using an arbitrary detection surface. *Phys. Rev. E*, 75:046706, Apr 2007.

[14] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, Jan 2004.

[15] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, Apr 1997.

[16] P. Combettes and V. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

[17] B. T. Cox and P. C. Beard. Fast calculation of pulsed photoacoustic fields in fluids using k-space methods. *The Journal of the Acoustical Society of America*, 117(6):3616–3627, 2005.

[18] 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.

[19] X. L. Dean-Ben, A. Buehler, V. Ntziachristos, and D. Razansky. Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography. *IEEE Transactions on Medical Imaging*, 31(10):1922–1928, Oct 2012.

[20] X. L. Deán-Ben, A. Ozbek, and D. Razansky. Volumetric real-time tracking of peripheral human vasculature with gpu-accelerated three-dimensional optoacoustic tomography. *IEEE Trans. Med. Imag.*, 32(11):2050–2055, 2013.

[21] R. Dembo, S. Eisenstat, and T. Steihaug. Inexact newton methods. *SIAM Journal on Numerical Analysis*, 19(2):400–408, 1982.

[22] L. Ding, X. L. Deán-Ben, C. Lutzweiler, D. Razansky, and V. Ntziachristos. Efficient non-negative constrained model-based inversion in optoacoustic tomography. *Phys. Med. Biol*, 60(17):6733–6750, 2015.

[23] D. Finch and R. Sarah K. Patch. Determining a function from its mean values over a family of spheres. *SIAM Journal on Mathematical Analysis*, 35(5):1213–1240, 2004.

[24] H. Gao, J. Feng, and L. Song. Limited-view multi-source quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065004, 2015.

[25] H. Gao, S. Osher, and H. Zhao. *Quantitative Photoacoustic Tomography*, pages 131–158. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[26] K. Geng and W. L. V. Scanning microwave-induced thermoacoustic tomography: Signal, resolution, and contrast. *Medical Physics*, 28(1):4–10.

[27] T. Goldstein, C. Studer, and R. G. Baraniuk. A field guide to forward-backward splitting with a FASTA implementation. *CoRR*, abs/1411.3406, 2014.

[28] M. Haltmeier, L. Neumann, and S. Rabanser. Single-stage reconstruction algorithm for quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065005, 2015.

[29] A. Hannukainen, N. Hyvönen, H. Majander, and T. Tarvainen. Efficient inclusion of total variation type priors in quantitative photoacoustic tomography. *SIAM Journal on Imaging Sciences*, 9(3):1132–1153, 2016.

[30] C. G. A. Hoelen, F. F. M. de Mul, R. Pongers, and A. Dekker. Three-dimensional photoacoustic imaging of blood vessels in tissue. *Opt. Lett.*, 23(8):648–650, Apr 1998.

[31] Y. Hristova. Time reversal in thermoacoustic tomography—an error estimate. *Inverse Problems*, 25(5):055008, 2009.

[32] 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.

[33] 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.

[34] 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.

[35] C. Huang, K. Wang, R. W. Schoonover, L. V. Wang, and M. A. Anastasio. Joint reconstruction of absorbed optical energy density and sound speed distributions in photoacoustic computed tomography: A numerical investigation. *IEEE Transactions on Computational Imaging*, 02(02):136 – 149, 2016.

[36] D. H. Huang, C. Liao, C. W. Wei, and P. C. Li. Simulations of optoacoustic wave propagation in light-absorbing media using a finite-difference time-domain method. *The Journal of the Acoustical Society of America*, 117(5):2795–2801, 2005.

[37] A. Javaherian and S. Holman. A multi-grid iterative method for photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 36(3):696–706, March 2017.

[38] A. Javaherian and S. Holman. A continuous adjoint for photo-acoustic tomography of the brain. *Inverse Problems*, 34(8):085003, 2018.

[39] H. Jiang, Z. Yuan, and X. Gu. Spatially varying optical and acoustic property reconstruction using finite-element-based photoacoustic tomography. *J. Opt. Soc. Am. A*, 23(4):878–888, Apr 2006.

[40] R. A. Kruger. Photoacoustic ultrasound. *Medical Physics*, 21(1):127–131, 1994.

[41] K. P. Köstli, M. Frenz, H. Bebie, and H. P. Weber. Temporal backward projection of optoacoustic pressure transients using fourier transform methods. *Physics in Medicine & Biology*, 46(7):1863, 2001.

[42] C. Lutzweiler and D. Razansky. Optoacoustic imaging and tomography: Reconstruction approaches and outstanding challenges in image performance and quantification. *Sensors*, 13(6):7345–7384, 2013.

[43] T. D. Mast, L. P. Souriau, D. L. D. Liu, M. Tabei, A. I. Nachman, and R. C. Waag. A k-space method for large-scale models of wave propagation in tissue. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 48(2):341–354, March 2001.

[44] T. P. Matthews and M. A. Anastasio. Joint reconstruction of the initial pressure and speed of sound distributions from combined photoacoustic and ultrasound tomography measurements. *Inverse Problems*, 33(12):124002, 2017.

[45] G. K. Minghua Xu and L. V. Wang. Microwave-induced thermoacoustic tomography using multi-sector scanning. *Medical Physics*, 28(9):1958–1963.

[46] K. Mitsuhashi, J. Poudel, T. P. Matthews, A. Garcia-Uribe, 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 Journal on Imaging Sciences*, 10(4):2022–2048, 2017.

[47] D. Modgil, M. A. Anastasio, and P. J. L. Rivière. Image reconstruction in photoacoustic tomography with variable speed of sound using a higher-order geometrical acoustics approximation. *Journal of Biomedical Optics*, 15:15 – 15 – 9, 2010.

[48] A. A. Oraevsky, S. L. Jacques, and F. K. Tittel. Determination of tissue optical properties by piezoelectric detection of laser-induced stress waves. In *Proc. SPIE 1882 Laser-Tissue Interaction IV*, volume 1882, pages 86–101, 1993.

[49] F. P. Paul C. Beard and T. N. Mills. Transduction mechanisms of the fabry-perot polymer film sensing concept for wideband ultrasound detection. *IEEE Trans Ultrason Ferroelectr Freq Control*, 46(6):1575– 1582, 1999.

[50] J. Provost and F. Lesage. The application of compressed sensing for photo-acoustic tomography. *IEEE Transactions on Medical Imaging*, 28(4):585–594, April 2009.

[51] A. Pulkkinen, B. T. Cox, S. R. Arridge, H. Goh, J. P. Kaipio, and T. Tarvainen. Direct estimation of optical parameters from photoacoustic time series in quantitative photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 35(11):2497–2508, Nov 2016.

[52] 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.

[53] A. Rosenthal, V. Ntziachristos, and D. Razansky. Model-based optoacoustic inversion with arbitrary-shape detectors. *Medical Physics*, 38(7):4285–4295, 2011.

[54] A. Rosenthal, V. Ntziachristos, and D. Razansky. Optoacoustic methods for frequency calibration of ultrasonic sensors. *IEEE Trans. Med. Imag.*, 58(2):316 – 326, 2011.

[55] A. Rosenthal, V. Ntziachristos, and D. Razansky. Acoustic inversion in optoacoustic tomography: A review. *Current Medical Imaging Reviews*, 9(4):318–336, 2013.

[56] A. Rosenthal, D. Razansky, and V. Ntziachristos. Fast semi-analytical model-based acoustic inversion for quantitative optoacoustic tomography. *IEEE Trans. Med. Imag.*, 29(6):1275 – 1285, 2010.

[57] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259 – 268, 1992.

[58] T. Saratoon, T. Tarvainen, B. T. Cox, and S. R. Arridge. A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation. *Inverse Problems*, 29(7):075006, 2013.

[59] P. Stefanov and G. Uhlmann. Thermoacoustic tomography with variable sound speed. *Inverse Problems*, 25(7):075011, 2009.

[60] 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.

[61] T. Tarvainen, B. T. Cox, J. P. Kaipio, and S. R. Arridge. Reconstructing absorption and scattering distributions in quantitative photoacoustic tomography. *Inverse Problems*, 28(8):084009, 2012.

[62] T. Tarvainen, A. Pulkkinen, B. T. Cox, J. P. Kaipio, and S. R. Arridge. Bayesian image reconstruction in quantitative photoacoustic tomography. *IEEE Trans. Med. Imag.*, 32(12):2287–2298, 2013.

[63] T. Tarvainen, M. Vauhkonen, V. Kolehmainen, and J. P. Kaipio. Finite element model for the coupled radiative transfer equation and diffusion approximation. *International Journal for Numerical Methods in Engineering*, 65(3):383–405.

[64] J. Tick, A. Pulkkinen, and T. Tarvainen. Image reconstruction with uncertainty quantification in photoacoustic tomography. *The Journal of the Acoustical Society of America*, 139(4):1951–1961, 2016.

[65] 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.

[66] 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.

[67] 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.

[68] B. E. Treeby, J. Jaros, A. P. Rendell, and B. T. Cox. Modeling nonlinear ultrasound propagation in heterogeneous media with power law absorption using a k-space pseudospectral method. *The Journal of the Acoustical Society of America*, 131(6):4324–4336, 2012.

[69] 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.

[70] L. A. Vese and S. J. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19(1):553–572, Dec 2003.

[71] C. Vogel and M. Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.

[72] K. Wang, S. A. Ermilov, R. Su, H. P. Brecht, A. A. Oraevsky, and M. A. Anastasio. An imaging model incorporating ultrasonic transducer properties for three-dimensional optoacoustic tomography. *IEEE Trans. Med. Imag.*, 30(2):203–214, 2011.

[73] K. Wang, C. Huang, Y. J. Kao, C. Y. Chou, A. A. Oraevsky, and M. A. Anastasio. Accelerating image reconstruction in three-dimensional optoacoustic tomography on graphics processing units. *Medical Physics*, 40(2):023301.

[74] K. Wang, R. Su, A. A. Oraevsky, and M. A. Anastasio. Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography. *Phys. Med. Biol*, 57(17):5399–5423, 2012.

[75] L. V. Wang and J. Yao. A practical guide to photoacoustic tomography in the life sciences. *Nature Methods*, 13:627–638, 2016.

[76] L. V. Wang, X. Zhao, H. Sun, and G. Ku. Microwave-induced acoustic imaging of biological tissues. *Review of Scientific Instruments*, 70(9):3744–3748, 1999.

[77] G. Wojcik, B. Fomberg, R. Waag, L. Carcione, J. Mould, L. Nikodym, and T. Driscoll. Pseudospectral methods for large-scale bioacoustic models. In *1997 IEEE Ultrasonics Symposium Proceedings. An International Symposium (Cat. No.97CH36118)*, volume 2, pages 1501–1506 vol.2, Oct 1997.

[78] M. Xu and L. V. Wang. Time-domain reconstruction for thermoacoustic tomography in a spherical geometry. *IEEE Trans. Med. Imag.*, 21(7):814–822, 2002.

[79] M. Xu and L. V. Wang. Universal back-projection algorithm for photoacoustic computed tomography. *Phys. Rev. E*, 71(1):016706, 2005.

[80] M. Xu and L. V. Wang. Photoacoustic imaging in biomedicine. *Review of Scientific Instruments*, 77(4):041101, 2006.

[81] Y. Xu, D. Feng, and L. V. Wang. Exact frequency-domain reconstruction for thermoacoustic tomography. i. planar geometry. *IEEE Transactions on Medical Imaging*, 21(7):823–828, July 2002.

[82] Y. Xu and L. V. Wang. Effects of acoustic heterogeneity in breast thermoacoustic tomography. *IEEE Trans. Med. Imag.*, 50(9):1134 – 1146, 2003.

[83] Y. Xu and L. V. Wang. Time reversal and its application to tomography with diffracting sources. *Phys. Rev. Lett.*, 92:033902, Jan 2004.

[84] Y. Xu, M. Xu, and L. V. Wang. Exact frequency-domain reconstruction for thermoacoustic tomography. ii. cylindrical geometry. *IEEE Transactions on Medical Imaging*, 21(7):829–833, July 2002.

[85] F. Ye, Z. Ji, W. Ding, C. Lou, S. Yang, and D. Xing. Ultrashort microwave-pumped real-time thermoacoustic breast tumor imaging system. *IEEE Transactions on Medical Imaging*, 35(3):839–844, March 2016.

[86] J. Zhang and M. A. Anastasio. Reconstruction of speed-of-sound and electromagnetic absorption distributions in photoacoustic tomography. *Proc.SPIE*, 6086:6086 – 6086 – 7, 2006.

[87] Q. Zhu and B. D. Steinberg. Large-transducer measurements of wavefront distortion in the female breast. *Ultrasonic Imaging*, 14(3):276 – 299, 1992.

Chapter 2

A multi-grid iterative method for photoacoustic tomography

The content of this chapter has been published in: A. Javaherian and S. Holman. A multi-grid iterative method for photoacoustic tomography. IEEE Transactions on Medical Imaging, 36(3):696-706, March 2017.

Abstract

Inspired by the recent advances on minimizing nonsmooth or bound-constrained convex functions on models using varying degrees of fidelity, we propose a line search multi-grid (MG) method for full-wave iterative image reconstruction in photoacoustic tomography (PAT) in heterogeneous media. To compute the search direction at each iteration, we decide between the gradient at the target level, or alternatively an approximate error correction at a coarser level, relying on some predefined criteria. To incorporate absorption and dispersion, we derive the analytical adjoint directly from the first-order acoustic wave system. The effectiveness of the proposed method is tested on a total-variation penalized Iterative Shrinkage Thresholding algorithm (ISTA) and its accelerated variant (FISTA), which have been used in many studies of image reconstruction in PAT. The results show the great potential of the proposed method in improving speed of iterative image reconstruction.

2.1 Introduction

Photoacoustic Tomography (PAT) is a hybrid imaging technique, which combines the advantage of rich contrast attributed to optical imaging and high spatial resolution brought up by ultrasound. Typically near-infrared pulses of light are used to irradiate tissue, which are then absorbed preferentially as a function of the optical absorption of the tissue. The absorbed energy produces local increases in pressure, which move outwards because of the elasticity of soft tissues, and are then sampled temporally by surface detectors [40].

To estimate the optical absorption distribution of the irradiated tissue from the recorded surface data, one faces two distinct inverse problems, namely acoustic [38] and optical [35]. To solve the acoustic inverse problem for media with relatively homogeneous acoustic properties and simple detection surfaces, numerous methods based on filtered back-projection [24, 44, 45] or eigenfunction expansion techniques [46] have been proposed.

From a practical point of view, real-time 3D reconstruction has been provided for photo-acoustic [48] and thermo-acoustic tomography [47] via 2D reconstruction of slices of the sample, and composing the reconstructed slices into a volume image. Techniques to reduce the effects of out-of-plane acoustic signals were successfully introduced in [48], and in order to achieve the speeds required for real-time imaging reconstruction was done by an onboard FPGA in [47]. In these cases, a dense and simple (circular) detection geometry, as well as relatively homogeneous acoustic properties allow a one-step image reconstruction based on a filtered backprojection algorithm [49].

TR is a more versatile inversion approach for PAT since it is practical for media with heterogeneous acoustic properties and arbitrary detection geometries [38, 19, 18]. TR and the other inversion approaches mentioned above are inherently based on continuous models, and thus require the detection surface to be very dense and enclose the object [19, 18]. This is problematic for 3D PAT especially in medical applications. The dependance of shape, spectrum and amplitude of propagating acoustic waves on the characteristic properties of tissue media impels enriching the image reconstruction of PAT by simulation of tissue-realistic acoustic propagation [37]. To achieve this aim here, the forward problem was solved by a first order acoustic system of three coupled

equations which includes two fractional Laplacian operators in order to account separately for absorption and dispersion according to a frequency power law [37]. The main advantage of this acoustic wave propagation model is that it can be efficiently implemented by the k-space pseudospectral method [10, 37] in which the spatial gradient of field parameters is globally computed in frequency domain [34, 37].

To mitigate the effects of data incompleteness and noise, iterative methods are often used, e.g., TR-based iterative algorithms [31, 32] or optimization techniques [42, 20]. Among a great number of optimization approaches, a total variation penalized variant of FISTA [6, 5] has been very popular for iterative PAT [42, 20]. The key element of these optimization approaches is the computation of the gradient of an objective function in terms of the forward model and its adjoint. For heterogeneous media, the adjoint was computed by a "discretize-then-adjoint" method in [20]. Recently, an adjoint was derived for PAT, based on an "adjoint-then-discretize" method [3], from the second order acoustic wave equation which does not include absorption and dispersion. Instead, here the adjoint will be derived using the aforementioned system of three coupled acoustic wave equations [37]. In the absence of absorption and dispersion this matches the adjoint in [3].

In medical PAT, the compartmentalised distribution of chromophores composing tissues induces step-like pressure discontinuities within absorbing regions, which make the generated waves highly broadband [38]. To cover such a broad range of frequencies in the reconstruction, very dense grids are needed which make iterative PAT computationally burdensome. To mitigate this problem, numerous methods have been proposed to accelerate wave propagation models[11, 26, 2]. By recent advances on data casting and parallelization using GPUs [12, 41, 36] or FPGA-based hardware implementation of the reconstruction algorithms [47], wave propagation models were accelerated notably. In the present work, we take a different approach, and look at a method to improve the performance of the underlying algorithm.

In general, whenever a finite-dimensional optimization problem arises from an infinite-dimensional continuous problem, it is possible to control the fidelity with which the optimization model captures the underlying continuous problem [30]. In the case that the arising discretized model is very large-scale, multi-grid (MG) schemes, which exploit a hierarchy of discretized models (levels) of varying size, are very popular.

A MG scheme for unconstrained smooth optimization problems was first proposed by Nash [28], in which the information at the coarse level is utilized to compute the search direction at the target level. This method was recently extended to composite convex functions involving a smooth term plus a non-smooth ℓ_1 term [30, 17], relying on the recent theoretical advances on minimization of smoothable functions [7]. Additionally, Nash's method was recently extended to smooth bound-constrained optimization [21].

To mitigate the burdensome computational requirements of iterative PAT, we propose a line search multi-grid method for full-wave iterative image reconstruction in PAT so that at some iterations, a recursive search direction is computed by minimizing the objective function at some coarser levels. Here the proposed MG method is applied to ISTA and FISTA on two levels, but it can be easily extended to other first order methods such as Primal-Dual algorithms, or to more than two levels.

2.2 Background

2.2.1 Forward Problem

Lossless media

Acoustic propagation in lossless heterogeneous media can be described relying on three coupled equations, i.e, equation of motion, equation of continuity, and adiabatic equation of state, respectively in the form [38]

$$\frac{\partial u}{\partial t}(r,t) = -\frac{1}{\rho_0(r)}\nabla p(r,t), \tag{2.1}$$

$$\frac{\partial \rho}{\partial t}(r,t) = -\rho_0(r)\ \nabla \cdot u(r,t), \tag{2.2}$$

$$p(r,t) = c_0(r)^2 \rho(r,t) \tag{2.3}$$

with initial conditions

$$p(r,0) = p_0(r), \quad u(r,0) = 0. \tag{2.4}$$

Here, $p(r,t)$ denotes the acoustic pressure at position $r \in \mathbb{R}^d$ ($d = 2$ or 3), and time $t \in \mathbb{R}^+$. Additionally, $u(r,t)$ denotes the vector-valued acoustic particle velocity, $c_0(r)$ denotes the varying sound speed, and $\rho(r,t)$ and $\rho_0(r)$ represent the acoustic and ambient densities, respectively.

Lossy media

To simulate wave propagation in lossy media, a trade-off is typically needed between agreement with experimental observations [33], meeting causality conditions [27, 23, 22], and efficiency of numerical computations [9, 37]. Over frequencies of generated ultrasound waves in PAT, the absorption in tissue obeys a frequency power law in the form

$$\alpha = \alpha_0 w^y, \tag{2.5}$$

where α_0 is the absorption coefficient in dB MHz^{-y} cm^{-1}, w is the angular frequency in MHz, and y is the power law exponent [38].

Describing the wave attenuation as effects of viscosity and thermal conduction leads to the so-called thermo-viscous attenuation model, which yields a frequency-squared attenuation ($y = 2$). This model does not match the observed frequency dependence of attenuation in tissues [33], and also violates the causality condition [23]. The attenuation model was later described by a superposition of relaxation mechanisms [27]. This model meets the causality condition [23], and can be implemented efficiently by the k-space method [34]. However, for simulating broadband acoustic propagation in PAT, estimation of the distribution of relaxation parameters for each relaxation process is troublesome. To account for the power law dependence on frequency evident in biological tissue ($1 < y < 1.5$), a lossy wave equation based on temporal convolution, the so-called Szabo's model [33], was proposed, which was later rewritten as a time-domain fractional derivative operator, e.g., [25, 8]. Szabos's model has been shown to be noncausal for $y > 1$ [23]. Furthermore, the time-domain convolution or fractional derivative operators inherently require storing the complete pressure field at previous times. A memory-efficient power law absorption model based on fractional Laplacian operators was proposed in [9], and was then modified to incorporate the dispersive sound speed [37]. This model can be easily incorporated into the k-space pseudospectral method without storing the computed pressure field at previous time steps [37], as opposed to classical absorption and dispersion models that involve time-domain fractional operators. This model is very popular in PAT [38, 36, 37], and was thus used in the present study.

Applying this model, the absorption and dispersion effects are incorporated into

the wave propagation by adding two fractional Laplacian operators to the equation of state in the form [37]

$$p(r,t) = c_0(r)^2\Big\{1 - \tau(r)\frac{\partial}{\partial t}(-\nabla^2)^{\frac{y}{2}-1} - \eta(r)(-\nabla^2)^{\frac{y-1}{2}}\Big\}\rho(r,t). \tag{2.6}$$

The absorption and dispersion proportionality coefficients, $\tau(r)$ and $\eta(r)$ respectively, are calculated by

$$\tau(r) = -2\alpha_0 c_0(r)^{y-1}, \quad \eta(r) = 2\alpha_0 c_0(r)^y \tan(\pi y/2). \tag{2.7}$$

A toolbox for modeling of acoustic wavefield propagation based on the k-space pseudo-spectral method is freely available [36], and was used in this study.

Recently it has been shown that this attenuation model encounters some noncausality problems, since the corresponding equation of state (2.6) is nonlocal in space at each time instant, and in addition the Green's function of the resulting wave equations does not have a finite wave front speed [22]. To maintain these causality conditions, the state equation was spatially localized by enforcing a local time shift to the attenuation model. This also leads to a Green's function with a finite wave front speed, which is known to be a strong causality condition for systems of wave equations [23, 22]. However, similar to classical lossy wave equations based on temporal fractional derivatives [25, 8], the numerical implementation of this model is very expensive for 3D PAT.

2.2.2 Inverse Problem

The acoustic inverse problem is to estimate the initial pressure $p_0(r)$ inside a bounded region $\Omega \subset \mathbb{R}^d$ from the measurements $p_m(r_s,t)$ taken at positions r_s within an open set $\Gamma \subset \mathbb{R}^{d-1}$ from time $t = 0$ to T.

Time Reversal (TR)

Employing the time reversal method, $p_m(r_s,t)$ is enforced as a Dirichlet boundary condition in a time-reversed order, yielding

$$p_{\text{tr}}(r_s,t) = p_m(r_s, T-t). \tag{2.8}$$

Here, the time T is assumed to be sufficiently large so that all waves leave the medium, yielding $p_{\text{tr}}(r, 0) = 0$ [38]. This is, however, not exactly held for even dimensions or heterogeneous media[19, 18]. It was shown that to account for the absorption and physical dispersion in TR, the absorption term in (2.6) must be reversed in sign, while the physical dispersion term remains unchanged [38].

Variational methods

The accuracy of inversion approaches, including TR, is limited for sub-sampled or noisy data. In these cases, variational image reconstruction methods provide an effective alternative [3, 20]. Let p_0 denote the sought after initial pressure distribution, and $\hat{p} \in \mathbb{R}^M (M \in \mathbb{N})$ and ε denote the time series of measured data at sensors and the corresponding noise, respectively. Additionally, let $\mathcal{H}$ represent the forward model discussed in 2.2.1. We then have

$$\hat{p} = \mathcal{H}p_0 + \varepsilon. \tag{2.9}$$

Inferring p_0 from $\hat{p}$ amounts to solving a regularized least-square optimization problem in the form

$$p = \underset{p_0 \geqslant 0}{\operatorname{argmin}} \frac{1}{2} \left\| \mathcal{H}p_0 - \hat{p} \right\|^2 + \lambda \mathcal{J}(p_0). \tag{2.10}$$

Here, $\lambda > 0$ is a regularization parameter, and $\mathcal{J}(p_0)$ is a regularization functional that can be used to impose a-priori information about the true solution. Here, regularization functional $\mathcal{J}$ is taken to be total variation (TV) since it is very popular in PAT because of accounting for feature edges in the reconstruction [2, 20]. Solving (2.10) requires the computation of gradient of the objective function as a function of the forward operator $\mathcal{H}$ and its adjoint $\mathcal{H}^*$. Specifically in PAT, the main computational cost of the minimization problem is the implementation of $\mathcal{H}$ and $\mathcal{H}^*$, and the cost of other steps is negligible in comparison. To derive the adjoint for the forward model described in 2.2.1, a "discretize-then-adjoint" method was proposed, where the computational steps of the discretized forward problem are explicitly reversed [20]. The adjoint obtained by this strategy may not correspond exactly with a discretization of the adjoint in the continuous domain. Very recently, a general analytic form of the adjoint in PAT was derived [3], where the time-reversed pressure $p_m(r_s, T-t)$ is added

as a time-dependent mass source term $s(r,t)$ to (2.2). In comparison, in the TR approach this is enforced as an explicit Dirichlet boundary condition [38]. However, the method in [3] gives the adjoint using the second order acoustic equation, which does not include absorption and dispersion. In order to include absorption and dispersion, we will derive the adjoint using (2.1), (2.2) and (2.6).

2.3 Adjoint of the Three-Coupled First Order Wave Propagation Equation

Similar to [3], the continuous forward operator is the map

$$\begin{aligned} &\mathcal{H} : C_0^\infty(\Omega) \to \mathbb{R}^M, \\ &\mathcal{H}[p_0](r,t) = \mathcal{M} w(r,t) p(r,t), \end{aligned} \tag{2.11}$$

where $w(r,t) \in C_0^\infty(\Gamma \times [0,T])$ restricts the pressure $p(r,t)$ to the spatio-temporal field accessible to the sensors, and $\mathcal{M}$ maps the accessible part of the pressure field into the measured data at sensors $\hat{p} \in \mathbb{R}^M$. Like [3], we will assume that $\mathcal{M}^*$ is given. Now, we have $\mathcal{H}^* = \mathcal{P}^* \mathcal{M}^*$, where $\mathcal{P}^* : C_0^\infty(\Gamma \times [0,T]) \to C_0^\infty(\Omega)$ is the adjoint of

$$\begin{aligned} &\mathcal{P} : C_0^\infty(\Omega) \to C_0^\infty(\Gamma \times [0,T]) \\ &\mathcal{P}[p_0](r,t) = w(r,t) p(r,t). \end{aligned} \tag{2.12}$$

Let us first define the time-reversed adjoint fields, p^*, u^*, ρ^* by

$$\frac{\partial u^*}{\partial t}(r,t) = -\frac{1}{\rho_0(r)} \nabla p^*(r,t), \tag{2.13}$$

$$\begin{aligned} \frac{\partial \rho^*}{\partial t}(r,t) = \rho_0(r) \Big(&- \nabla \cdot u^*(r,t) \\ &+ w(r, T-t) h(r, T-t) \Big), \end{aligned} \tag{2.14}$$

$$\begin{aligned} p^*(r,t) = \rho_0(r) \Big\{ 1 - \frac{\partial}{\partial t} (-\nabla^2)^{\frac{y}{2}-1} \tau(r) \\ - (-\nabla^2)^{\frac{y-1}{2}} \eta(r) \Big\} \frac{c_0(r)^2}{\rho_0(r)} \rho^*(r,t) \end{aligned} \tag{2.15}$$

with initial conditions

$$p^*(r,0) = 0, \quad u^*(r,0) = 0. \tag{2.16}$$

By definition of the adjoint, for any $h(r,t) \in C_0^\infty(\Gamma \times [0,T])$

$$\int_0^T \int_{\mathbb{R}^d} \mathcal{P}[p_0](r,t)\ h(r,t)\ \mathrm{d}r\ \mathrm{d}t = \int_{\mathbb{R}^d} p_0(r)\ \mathcal{P}^*[h](r)\ \mathrm{d}r. \tag{2.17}$$

The claim is that

$$\mathcal{P}^*[h](r) = \frac{\rho^*(r,T)}{\rho_0(r)}. \tag{2.18}$$

To prove this we start with (2.2), which yields

$$\begin{aligned} 0 &= \int_0^T \int_{\mathbb{R}^d} \left(\frac{\partial \rho}{\partial t} + \rho_0 \nabla \cdot u\right) \frac{p^*(r,T-t)}{\rho_0} \mathrm{d}r \ \mathrm{d}t. \\ &= \int_0^T \int_{\mathbb{R}^d} \left(\frac{\partial \rho}{\partial t} + \rho_0 \nabla \cdot u\right) \Big\{1 - \frac{\partial}{\partial t}(-\nabla^2)^{\frac{y}{2}-1}\tau \\ &\qquad - (-\nabla^2)^{\frac{y-1}{2}}\eta\Big\} c_0^2 \frac{\rho^*(r,T-t)}{\rho_0} \ \mathrm{d}r \ \mathrm{d}t. \end{aligned} \tag{2.19}$$

where we also used (2.15) for the second equality, and suppressed the dependence on (r,t) in some places for brevity. Now, the Laplacian $-\nabla^2$ is self-adjoint, and so the fractional powers of the Laplacian are as well and we can move them from acting on $c_0^2\rho^*/\rho_0$ to the first term as in integration-by-parts. This gives

$$\begin{aligned} 0 = \int_0^T \int_{\mathbb{R}^d} &\left[\frac{\partial}{\partial t} c_0^2 \Big\{1 - \eta(-\nabla^2)^{\frac{y-1}{2}}\Big\}\rho\right] \frac{\rho^*(r,T-t)}{\rho_0} \\ &- \left[c_0^2 \tau \frac{\partial}{\partial t}(-\nabla^2)^{\frac{y}{2}-1}\rho\right] \frac{\frac{\partial \rho^*}{\partial t}(r,T-t)}{\rho_0} \\ &+ \nabla \cdot u \ p^*(r,T-t) \ \mathrm{d}r \ \mathrm{d}t. \end{aligned} \tag{2.20}$$

Now we apply integration-by-parts to the terms on the first and third lines as well as (2.6), and the initial conditions in (2.16), which together yield

$$\begin{aligned} 0 = &\int_0^T \int_{\mathbb{R}^d} \frac{p}{\rho_0}\frac{\partial \rho^*}{\partial t}(r,T-t) - u \cdot \nabla p^*(r,T-t) \ \mathrm{d}r \ \mathrm{d}t \\ &- \int_{\mathbb{R}^d} \left[c_0^2 \Big\{1 - \eta(-\nabla^2)^{\frac{y-1}{2}}\Big\}\rho(r,0)\right] \frac{\rho^*(r,T)}{\rho_0} \ \mathrm{d}r. \end{aligned} \tag{2.21}$$

Considering that $\frac{\partial \rho}{\partial t}(r,0) = -\rho_0(r)\nabla \cdot u(r,0) = 0$ by (2.2), and in light of (2.6), we have

$$c_0^2 \Big\{1 - \eta(-\nabla^2)^{\frac{y-1}{2}}\Big\}\rho(r,0) = p_0(r). \tag{2.22}$$

Putting this into the previous formula finally gives

$$\begin{aligned} &\int_{\mathbb{R}^d} p_0(r) \frac{\rho^*(r,T)}{\rho_0(r)} \ \mathrm{d}r = \\ &\int_0^T \int_{\mathbb{R}^d} \frac{p}{\rho_0}\frac{\partial \rho^*}{\partial t}(r,T-t) - u \cdot \nabla p^*(r,T-t) \ \mathrm{d}r \ \mathrm{d}t. \end{aligned} \tag{2.23}$$

Putting (2.23) aside for a moment, we next use (2.1) which gives

$$0 = \int_0^T \int_{\mathbb{R}^d} \left(\rho_0 \frac{\partial u}{\partial t} + \nabla p\right) \cdot u^*(r,T-t) \ \mathrm{d}r \ \mathrm{d}t. \tag{2.24}$$

Applying integration-by-parts to both terms, and enforcing the initial conditions (2.4) and (2.16) gives

$$0 = \int_0^T \int_{\mathbb{R}^d} \rho_0 \; u \cdot \frac{\partial u^*}{\partial t}(r, T-t) - p \; \nabla \cdot u^*(r, T-t) \; \mathrm{d}r \; \mathrm{d}t. \tag{2.25}$$

Subtracting (2.25) from (2.23) and gathering on the right-hand-side the terms involving p and u respectively in lines two and three of the next formula, we have

$$\begin{aligned} &\int_{\mathbb{R}^d} p_0(r) \frac{\rho^*(r,T)}{\rho_0(r)} \; \mathrm{d}r = \\ &\int_0^T \int_{\mathbb{R}^d} \left(\frac{1}{\rho_0} \frac{\partial \rho^*}{\partial t}(r, T-t) + \nabla \cdot u^*(r, T-t) \right) p \; \mathrm{d}r \; \mathrm{d}t \\ &- \int_0^T \int_{\mathbb{R}^d} \left(\rho_0 \frac{\partial u^*}{\partial t}(r, T-t) + \nabla p^*(r, T-t) \right) \cdot u \; \mathrm{d}r \; \mathrm{d}t. \end{aligned} \tag{2.26}$$

Now, by (2.14) the integrand in the second line above is equal to $p(r,t)w(r,t)h(r,t)$, and by (2.13) the third line is equal to zero. Therefore

$$\begin{aligned} \int_{\mathbb{R}^d} p_0(r) \frac{\rho^*(r,T)}{\rho_0(r)} \; \mathrm{d}r &= \int_0^T \int_{\mathbb{R}^d} w(r,t) p(r,t) h(r,t) \; \mathrm{d}r \; \mathrm{d}t \\ &= \int_0^T \int_{\mathbb{R}^d} \mathcal{P}[p_0](r,t) \; h(r,t) \; \mathrm{d}r \; \mathrm{d}t. \end{aligned} \tag{2.27}$$

Finally, using (2.17), we see now that $\mathcal{P}^*[h](r) = \frac{\rho^*(r,T)}{\rho_0(r)}$, and thus the claim (2.18) about the adjoint is proven. Taking $\tau(r) = 0$ and $\eta(r) = 0$ makes this adjoint the same as that proposed in [3] for lossless media.

2.4 First-order Optimization Methods for PAT

The numerical implementation of the derived forward and adjoint operators requires discretization of the models. Accordingly, the discretized variant of the sought after initial pressure p_0 is denoted by $x \in \mathbb{R}^N (N \in \mathbb{N})$ with N the number of grid points, and the discretized forward model linking x to data $\hat{p}$ is denoted by $H \in \mathbb{R}^{M \times N}$. Problem (2.10) is in a class of non-smooth constrained convex minimization problems of the form

$$\underset{x}{\operatorname{argmin}} \left\{ F(x) := f(x) + g(x) \right\}. \tag{2.28}$$

Here, $f(x) = \frac{1}{2} \left\| Hx - \hat{p} \right\|^2$ is a continuously differentiable function with Lipschitz continuous gradient having smallest Lipschitz constant $L_f = \sigma_{\max}(H^*H)$, where $\sigma_{\max}(.)$

stands for the largest singular value. The gradient of f is computed by

$$\nabla f(x) = H^* \left(Hx - \hat{p} \right). \tag{2.29}$$

Additionally, we take $g(x) = \lambda \mathcal{J}(x) + \delta_C(x)$ where δ_C is the indicator function for the set of constraints $C = \{x \geqslant 0\}$.

Applying the so-called forward-backward splitting method to a fixed point iterative scheme arising from the optimality conditions of problem (2.28) gives the two-step Iterative Shrinkage Thresholding Algorithm (ISTA) shown in Algorithm 2.1.

Algorithm 2.1 ISTA

1: **Iteration 0:** x_0
2: **Iteration k $\geqslant$ 1:**
3: $z_k = x_{k-1} - \alpha_k \nabla f(x_{k-1})$
4: $x_k = \text{prox}_{\alpha_k}(g)(z_k)$
5: **Output:** x_*.

Here, line 3 is a forward gradient descent step [13], and is in a class of line search techniques which utilize a steepest descent search direction $-\nabla f(y)$ and step size α_k [6]. Applying ISTA, the convergence of the iterates x_k to a minimizer x_* of problem (2.28) is proven if $\alpha_k \in (0, 2/L_f)$ [5]. To determine L_f, the largest singular value of H^*H is computed iteratively by the power method [3]. Since L_f is independent to the unknown x, it can be stored and used for all experiments done in a fixed setting [2]. Otherwise, α_k can be computed adaptively by backtracking line search techniques, although this is inefficient for large-scale PAT problems. Additionally,

$$\text{prox}_{\alpha_k}(g)(z_k) := \underset{x}{\text{argmin}} \left\{ g(x) + \frac{1}{2\alpha_k} \|x - z_k\|^2 \right\}. \tag{2.30}$$

is a backward gradient step, and is called the proximal map [13]. Similar to [20], here the proximal map associated with the TV functional was computed based on a dual approach given in [5].

The computational cost of performing the forward and adjoint solvers necessary to compute ∇f on a grid of size $N_x \cdot N_y \cdot N_z$ in N_t time steps is $\mathcal{O}(N_t \cdot N_x \cdot N_y \cdot N_z \log(N_x \cdot N_y \cdot N_z))$, whereas the computational cost of solving the proximal map by the dual approach given in [5] is $\mathcal{O}(N_x \cdot N_y \cdot N_z)$. As a result, the major cost of each iteration is the forward gradient step, while the cost of the proximal map in (2.30) is almost negligible in comparison.

An acceleration to ISTA is given by FISTA [6], which provides a global convergence rate of $\mathcal{O}(1/\sqrt{\epsilon})$, compared to $\mathcal{O}(1/\epsilon)$ for ISTA, where ϵ denotes the desired accuracy. FISTA is outlined in Algorithm 2.2 [6].

Algorithm 2.2 FISTA

1: **Iteration 0:** $y_1 = x_0, t_k = 1 \quad (\theta_k = 0)$
2: **Iteration k $\geqslant$ 1:**
3: $x_k = P_L(y_k)$
4: $t_{k+1} = \frac{1+\sqrt{1+4{t_k}^2}}{2}$
5: $\theta_k = \frac{t_k - 1}{t_{k+1}}$
6: $y_{k+1} = x_k + \theta_k (x_k - x_{k-1})$
7: **Output:** x_*

Here, operator $P_L(\cdot)$ represents lines 3 and 4 in Algorithm 2.1. Note that replacing line 5 by $\theta_k = 0$ in Algorithm 2.2. gives $y_{k+1} = x_k$, and reduces the algorithm to ISTA. In the next section, the multi-grid algorithm will be described for a general algorithm like FISTA but with line 5 possibly replaced. Thus an extension to ISTA or other first-order optimization methods is straightforward. The convergence of sequence x_k provided by FISTA is proven when $\alpha_k \in (0, 1/L_f)$ [6]. For applications of ISTA in iterative PAT, see [3], and for FISTA, see [20, 42].

2.5 Line search Multi-grid Optimization Method

To improve the speed of Algorithms 2.1 and 2.2, a multi-grid (MG) line search strategy is adopted based on Nash's well-known method [28] so that at each iteration the algorithm decides between two possibilities: a direct search direction computed at the target level, or alternatively a recursive search direction generated from some steps taken at coarser levels.

Considering the computational cost of the forward and adjoint operators given in section 2.4, the cost of performing them on a coarse grid with a size $(N_x/2) \cdot (N_y/2) \cdot (N_z/2)$ in $N_t/2$ time steps is less than 1/16 the cost on the fine grid, as the time step is changed proportionally to the spatial distance of grid points. Accordingly, for 2D PAT, the computational cost of coarse forward and adjoint models is less than 1/8 the cost on the fine model.

2.5.1 First-order Coherence of Levels for Smooth Unconstrained Optimization: An Extension to FISTA

We denote the level that supports the fine resolution, referred to here as the "target level", by subscript h and the next coarse level by $h-1$. The transfer of information from level h to $h-1$ is done by restriction operator I_h^{h-1}. Conversely, prolongation operator I_{h-1}^h is used to transfer information from level $h-1$ to h.

To guarantee convergence on multiple levels, the first order optimality conditions of the levels must match. To attain this, Nash [28] suggests adding a linear term to the objective function at the next coarse level. We extend Nash's method to FISTA so that to compute a recursive search direction, starting from iteration $y_{h,k}$ at the target level, we use as the objective function at the next coarse level $h-1$

$$\phi_{h-1}\left(x_{h-1}\right) = F_{h-1}\left(x_{h-1}\right) + \langle v_{h-1}, x_{h-1}\rangle, \tag{2.31}$$

where v_{h-1} stands for

$$v_{h-1} = I_h^{h-1}\nabla F_h\left(y_{h,k}\right) - \nabla F_{h-1}\left(x_{h-1,0}\right), \tag{2.32}$$

with $x_{h-1,0} = I_h^{h-1} y_{h,k}$ the initial point at the next coarse level $h-1$. Note that $y_{h-1,1} = x_{h-1,0}$ according to the initialization in Algorithm 2.2. In this way, the gradient of the objective functions at the point of transfer between the two levels matches so that

$$\nabla\phi_{h-1}\left(x_{h-1,0}\right) = I_h^{h-1}\nabla F_h\left(y_{h,k}\right). \tag{2.33}$$

This property is called "first-order coherence" [28, 43, 30, 17].

2.5.2 Extension to Non-smooth Unconstrained Optimization

The approach given above is not applicable to non-smooth objective functions since the computation of ∇F is not possible. From a theoretical point of view, an approach for minimizing non-smooth functions via treating the problem as a sequence of smooth problems has been considered. A global convergence rate of $\mathcal{O}(1/\epsilon)$ was first established for functions with so-called "explicit max-structure" [29], and was then extended to the so-called "smoothable" functions [7].

Recently, relying on the mentioned works, Nash's multi-grid method was extended to unconstrained composite functions involving a smooth term plus a nonsmooth ℓ_1 term. The convergence rate of ISTA on a multi-grid setting was established by [30]. Recently, an MG method with an optimal rate of convergence ($\mathcal{O}(1/\sqrt{\epsilon})$) was proposed [17]. This has been inspired by a modified variant of Nesterov's acceleration technique, where the problem is treated as a linear combination of primal gradient and mirror descent steps [1]. The global convergence rate established by this MG method is optimal, but the bound on the worst case convergence rate is greater than that of the standard "gradient and mirror descent" algorithm in [1]. Note that in practice the sequence x_k provided by the "gradient and mirror descent" algorithm matches that of FISTA on a fixed grid [1, 17]. However, as opposed to the MG variant of FISTA proposed here, we observed that the MG algorithm proposed in [17] is not efficient in PAT.

In order to use MG with FISTA, at $y_{h,k}$, we smooth the TV penalty function in the form

$$\mathcal{J}_\rho(y_{h,k}) = \sum_{n_1,n_2,n_3} \sqrt{|(\nabla y_{h,k})_{n_1,n_2,n_3}|^2 + \rho^2} - \rho, \tag{2.34}$$

where ρ is the smoothing parameter. The gradient of F is now computed as [39]

$$\nabla F_\rho(y_{h,k}) = H^* \left(H y_{h,k} - \hat{p}\right) - \lambda \, \nabla \cdot \left(\frac{\nabla y_{h,k}}{\sqrt{|\nabla y_{h,k}|^2 + \rho^2}} \right). \tag{2.35}$$

The implementation of Nash's method via computing v_{h-1} and minimizing ϕ_{h-1} by Eqs. (2.32) and (2.31) is now straightforward in the unconstrained case.

2.5.3 Extension to Constrained Convex Optimization

The coherence formula does not account for the bound constraint that is enforced in the PAT problem. In general, very few studies exist in the literature to extend Nash's method to bound-constrained optimization, e.g., [15]. In [14], a method to deal with bound constraints for MG optimization problems was proposed based on truncation of the set of indices at which the constraints are active. In this method, the active nodes on the fine level are fixed for the next coarse-grid correction. This truncation scheme is very conservative, and thus makes the MG algorithm inefficient.

Recently an MG approach for smooth constrained optimization problems has been developed via restriction of bound constraints, rather than the truncation of active

set [21]. The restriction of constraints is done so that a feasible point remains feasible after the coarse correction step. In our specific case, this MG approach is applied to two levels, and a nonnegativity constraint is enforced globally to all nodes at the fine level. Let $\mathcal{I}_{h,i}$ denote the union of indices at level h that locate at the same position as, or neighbor to, index i at level $h-1$. The restriction of constraints gives the lower bound constraints φ_{h-1} at level $h-1$ in the form

$$(\varphi_{h-1})_i = (I_h^{h-1} x_{h,k})_i - \min\{(x_h)_j | j \in \mathcal{I}_{h,i}\} \tag{2.36}$$

This constraint is enforced to all iterates $x_{h-1,k}$ at the coarse level. Note that in FISTA at the target level, the constraint is enforced to $x_{h,k}$, whereas the transfer between levels is done at $y_{h,k}$.

2.5.4 Decision on Recursive Search Direction

At the beginning of each iteration k the algorithm decides whether to compute a recursive search direction on the coarse level. This depends on the first-order optimality condition at the current iterate $y_{h,k}$ at the two levels, as well as the distance between the current iterate and the point $\tilde{y}_h$ at which the last recursive search direction was performed [43, 30, 17]. In particular, a recursive search direction is used at $y_{h,k}$ if

$$\begin{aligned} &\left(\left\|I_h^{h-1} \nabla F_{\rho_h}(y_{h,k})\right\| > \kappa \left\|\nabla F_{\rho_h}(y_{h,k})\right\|\right) \ \cap \\ &\left(\|y_{h,k} - \tilde{y}_h\| > \vartheta \|\tilde{y}_h\| \ \cup \ K_r = 0 \ \cup \ K_d > q_d\right), \end{aligned} \tag{2.37}$$

where $\kappa \in \left(0, \min(1, \min \left\|I_h^{h-1}\right\|)\right)$, $\vartheta \in (0,1)$, and $q_d \in \mathbb{N}$ are some predefined parameters, K_d is the number of consecutive iterations with direct search direction, and K_r is the number of all iterations already performed with recursive search direction [30, 17]. The first condition implies that a recursive search direction is not efficient if the first-order optimality condition is almost satisfied at the starting point of the coarse error correction, as this makes the minimization of the objective function at the coarse level ineffective. Furthermore, the second condition implies that a recursive search direction is not efficient if the current point $y_{h,k}$ is very close to the point $\tilde{y}_h$ since it gives a result that is similar to what was obtained on the last recursive search direction [43]. This condition is ignored if the algorithm has already performed no step with a recursive search direction ($K_r = 0$), or many consecutive steps with a direct search direction at the fine level, say greater than q_d.

2.5.5 Outline of the MG algorithm

In Algorithm 2.3, an MG variant of FISTA on two grids is outlined. Here, $*$ denotes

Algorithm 2.3 FISTA in MG framework

Iteration 0: $y_{h,1} = x_{h,0},\ \theta_k = 0,\ K_d = 0,\ K_r = 0$
Iteration k $\geqslant$ 1:
if $k > 1 \quad \cap \quad$ (2.37) holds **then**
    Recursive search direction: $K_d = 0,\ K_r = K_r + 1$
    $x_{h-1,0} = I_h^{h-1} y_{h,k}$
    $v_{h-1} = I_h^{h-1} \nabla F_\rho(y_{h,k}) - \nabla F_\rho(x_{h-1,0})$
    compute φ_{h-1} by (2.36)
    compute ϕ_{h-1} by (2.31)
    $x_{h-1,*} = \text{FISTA}(h-1, \phi_{h-1}, x_{h-1,0}, \varphi_{h-1})$
    $x_{h,k} = y_{h,k} + I_{h-1}^h (x_{h-1,*} - x_{h-1,0})$
else
    Direct search direction: $K_d = K_d + 1$
    $x_{h,k} = P_L(y_{h,k})$
end if
update $\theta_{h,k}$
$y_{h,k+1} = x_{h,k} + \theta_{h,k} (x_{h,k} - x_{h,k-1})$
Output: $x_{h,*}$

the last iteration at each level. Since ϕ_{h-1} is smooth, at level $h-1$ the proximal map is reduced to a projection on the feasible set defined by φ_{h-1}. At each iteration with a recursive search direction, the termination of the algorithm at the coarse level is done whenever

$$\frac{\phi_{h-1,k} - \phi_{h-1,k+1}}{\max(\phi_{h,k}, \phi_{h,k+1})} < \varepsilon_c \quad \cup \quad * > q_c, \tag{2.38}$$

where q_c denotes the maximum permitted number of iterations at the coarse level, and is applied to guarantee the efficiency of coarse error corrections. Similarly, at the fine level the algorithm was terminated at iteration k if

$$\frac{F_{h,k} - F_{h,k+1}}{\max(F_{h,k}, F_{h,k+1})} < \varepsilon_d. \tag{2.39}$$

2.6 Numerical Results

Numerical studies were performed to investigate the effectiveness of the proposed multi-grid strategy on performance of ISTA and FISTA for iterative image reconstruction in PAT. To numerically solve the three-coupled first order acoustic wave

equations, which were described in section 2.2.1, the *K-Wave* MATLAB toolbox was used [36]. Additionally, to compute the gradient defined in (2.29) at each iteration at each level, the adjoint operator H^* was computed based on the "adjoint-then-discretize" method (cf. section 2.3). The processor that was employed in this work is an Intel(R) Core(TM) i5-4570 CPU @ 3.20 GHz with a RAM of 8.00 GB and a 64-bit operating system (Windows 7, Microsoft).

2.6.1 2D PAT Simulation

A square grid with a size of $2.36 \times 2.36\,\mathrm{cm}^2$ was created, which is made up of 472×472 grid points evenly spaced with a separation distance of 5×10^{-2}mm in both x and y dimensions, supporting frequencies of up to 13.23MHz. To measure the propagated wavefield, 200 point-wise pressure detectors were equidistantly placed along the left half of a circle having a radius of 11mm so that π radians of the circle were covered by the detectors. A PML having a thickness of 20 grid points and a maximum attenuation coefficient of 2 nepers per grid point was added to each side of the simulated grid in order to reduce spurious reflections at the boundaries [36].

Medium's properties: Figures 2.1(a) and 2.1(b) show sound speed and density maps that were used for reconstruction, respectively. The sound speed and density for the inhomogeneity (vasculature) were set to those of blood, i.e., 1575 ms^{-1} and 1055 kgm^{-3}, respectively, and the red color represents skin with a sound speed of 1730 ms^{-1} and a density of 1150 kgm^{-3}. The background inside the detection surface represents fat tissue with a sound speed of 1450 ms^{-1} and a density of 950 kgm^{-3}, and a sound speed of 1500 ms^{-1} and a density of 1000kgm^{-3} were considered for region outside the detection surface to represent water. These maps were inspired by acoustic properties in tissues given in [4]. Note that acoustic properties in realistic tissues are often smoother than the simulated maps, and do not have sharp interfaces. However, these sharp maps were provided in order to make a challenge for coarse error correction in the MG method.

Furthermore, the absorption coefficient was set to 0.75 $\mathrm{dB\,MHz}^{-y}\,\mathrm{cm}^{-1}$ for the whole medium, except the area that represents water, where it was set to 2×10^{-3} $\mathrm{dB\,MHz}^{-y}$ cm
The attenuation power law exponent was set to 1.5 for the entire medium.

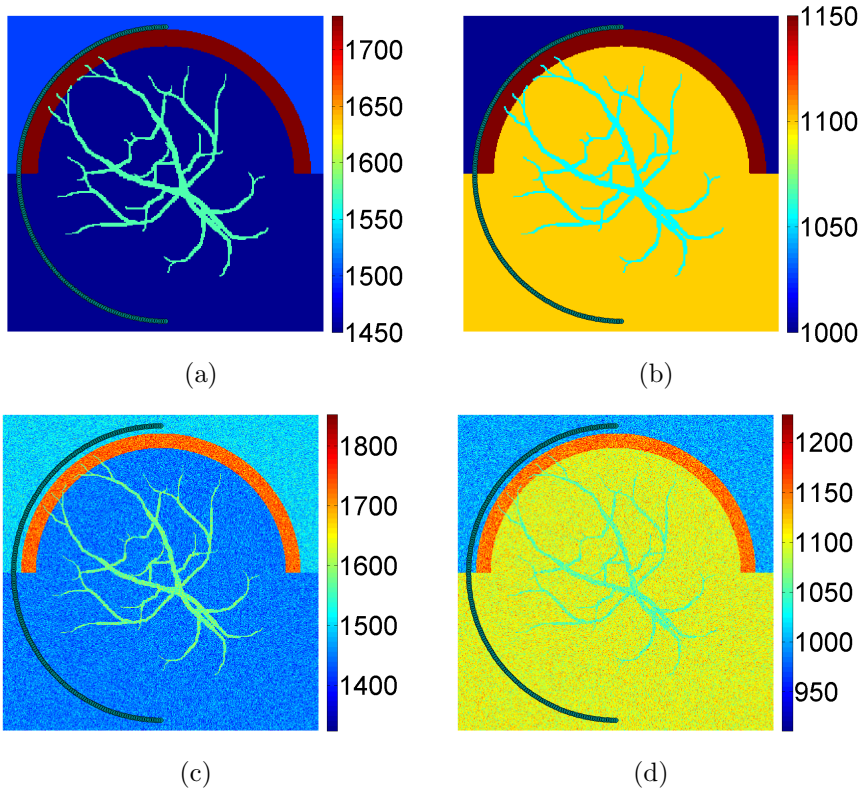


Figure 2.1: Medium's properties for image reconstruction (a) sound speed (b) density, and data generation (c) sound speed (d) density.

Since the exact maps are not readily available for reconstruction, data were generated from a more realistic phantom by contaminating the maps with a 35dB AWGN, as well as shifting the "water-skin" and "skin-soft tissue" interfaces towards the centre of the detection surface by 2% of radius of the circle. Figures 2.1(c) and 2.1(d) show distributions of sound speed and density that were used for data generation, respectively. To mitigate errors arising from aliasing, for all forward and adjoint models, the acoustic properties were smoothed by the k-wave toolbox [36].

The phantom was created so that it simulates the pressure distribution of vessels with a maximal amplitude of 2. Figure 2.2(a) displays the simulated phantom. To avoid spurious oscillations in the computed pressure field, high frequencies of the initial pressure distribution for each forward implementation were filtered by a self-adjoint smoothing operator. This operator was then included in the adjoint (see [3]).

The computed time-dependent pressure field arriving at the sensors was then sampled evenly in 2655 time steps. The interpolation of pressure field to sensors was performed by the well-known linear method. The generated data was then contaminated with a 30 dB AWGN. In order to avoid inverse crime, the reconstruction was

applied to a grid made up of 328×328 grid points, which supports a maximal frequency of 10.07 MHz. The PML at each side of the grid was proportionally reduced to 16 grid points. This grid will further be used as the "target grid" for our proposed multi-grid algorithm.

Iterative methods: The iterative reconstruction was performed by TV-regularized ISTA, i.e., Algorithm 2.1. The step size α_k was chosen to be $2/L_f$ [5] computed by the "power iteration" method, similar to [3, 2]. The regularization parameter was heuristically set to $\lambda = 1 \times 10^{-2}$.

The MG variant of ISTA, Algorithm 2.3 with FISTA replaced by ISTA, was then employed to reconstruct images on two grids having sizes 328×328 and 164×164. The algorithm was implemented by $\kappa = 1/4$, $\vartheta = 10^{-1}$, $q_d = 3$, $q_c = 8$, $\varepsilon_d = 10^{-3}$ and $\varepsilon_c = 10^{-2}$. The coarse model supports a maximal frequency half the fine grid, i.e. 5.038 MHz. For iterations at which a recursive search direction was computed, the TV function was smoothed by $\rho = 1 \times 10^{-2}$ as in (2.34). At the coarse level, the number of grid points associated with the PML was halved so that the thickness of the PML was the same as the target level. The sequences x^k computed by the algorithms were measured by the following parameters.

Relative Error (RE): This is defined at iteration k as

$$RE(x^k) = \frac{\|p_{\text{sol}} - p_{\text{exact}}\|}{\|p_{\text{exact}}\|} \times 100, \tag{2.40}$$

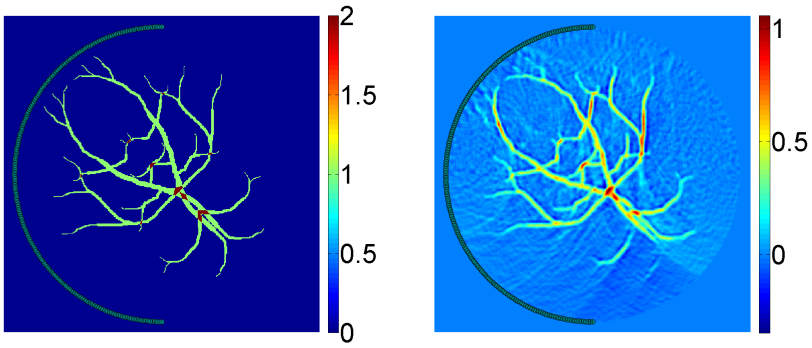
where p_{sol} stands for the sequence x^k interpolated back to the forward grid, and p_{exact} denotes the simulated phantom.

Norm of Residual (RES): This is defined at iteration k as

$$\text{RES}(x^k) = \left\| Ax^k - \hat{p} \right\|. \tag{2.41}$$

Objective function (F): This is defined at iteration k as a discretized variant of (2.10) on the inverse grid. It should be noted that the efficiency of any optimization algorithm, including the MG algorithm we are examining, should be evaluated using the objective function.

Figure 2.2(b) shows the image reconstructed by TR, which has an RE of 68.95%. Figure 2.3(a) shows RE of the images reconstructed by ISTA. To make a fair comparison between the competing algorithms, the image parameters were plotted versus CPU



(a) (b)

Figure 2.2: 2D phantom. (a) Initial pressure distribution (b) image reconstructed by TR.

time although the iterations are also shown as black dots. For the MG algorithm, the iterations at which a recursive search direction was used are designated by hexagrams. As seen in this figure, the fixed-grid algorithm was terminated after 4.87×10^3s (38 iterations), and finally reconstructed an image having an RE of 49.14%, whereas the MG algorithm produced an image having an RE of 48.68% at 1.20×10^3s (6 iterations). The MG algorithm was finally terminated after 2.07×10^3s (11 iterations), and provided an image having an RE of 47.81%. Figure 2.3(b) shows RES in the same way as RE.

Subsequently, FISTA and its MG variant were implemented on the same grids as and with the same parameters as ISTA and MG ISTA, respectively, except that the step size was chosen to be $1/L_f$ [6, 5]. Figures 2.3(c) and 2.3(d) display RE and RES of sequences provided by FISTA, respectively. As shown in figure 2.3(c), FISTA reconstructed a final image having an RE of 50.44% after 3.09×10^3s on a fixed grid (24 iterations), while the MG variant of FISTA provided an image having an RE of 49.31% at 8.09×10^2s (4 iterations). The MG algorithm finally reached an RE of 47.87% at 1.46×10^3s (8 iterations).

The objective function values corresponding to sequences computed by ISTA and FISTA are shown in figure 2.4(a). This figure has been shown in a larger view around the optimum in figure 2.4(b). As seen in these figures, FISTA converged more slowly than ISTA at early iterations, but it was faster than ISTA around the optimum on both fixed grid and two-level grids. ISTA has finally reached an objective function having a value of 2.15×10^2 at 4.87×10^3s on a fixed grid, whereas it has reached an

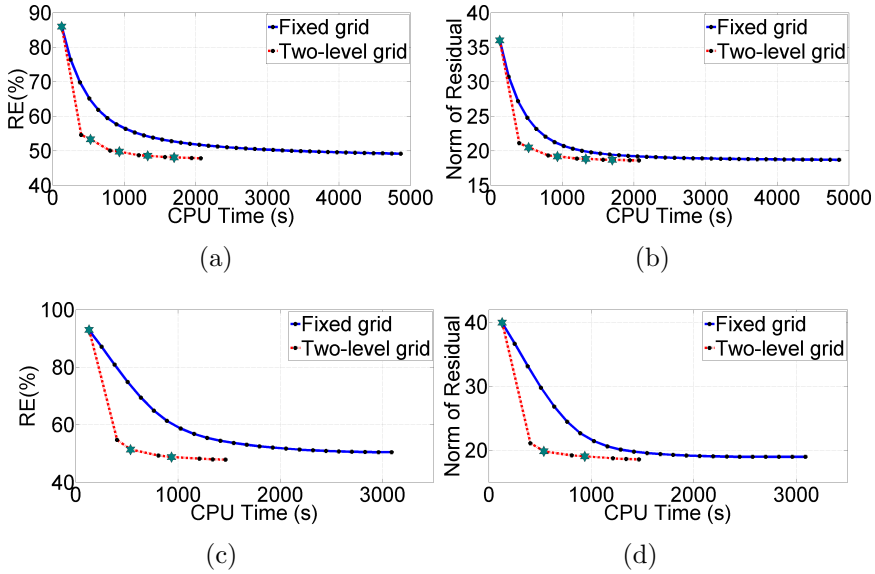


Figure 2.3: Evaluation of 2D images reconstructed on fixed grid (blue) and two-level grid (red). ISTA: (a) RE (b) RES, and FISTA: (c) RE (d) RES.

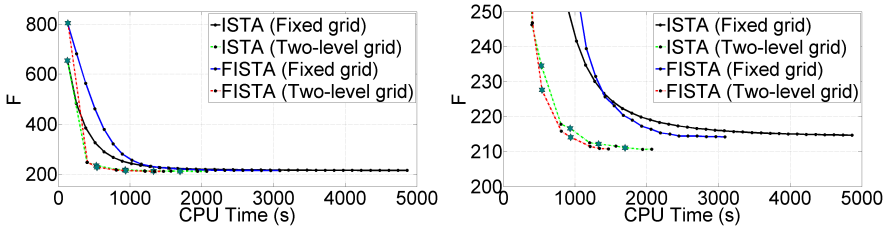


Figure 2.4: Convergence of ISTA on a fixed grid (black), ISTA on a two-level grid (green), FISTA on a fixed grid (blue) and FISTA on a two-level grid (red) for the 2D phantom. (a) objective function (b) larger view.

F of 2.12×10^2 at 1.20×10^3s on a two-level grid. This indicates that the MG variant of ISTA was four times faster than the fixed-grid ISTA. The MG variant of ISTA has finally reached an F of 2.10×10^2 at 2.07×10^3s. Applying FISTA on a fixed grid, the algorithm was terminated at 3.09×10^3s with an F of 2.14×10^2, while FISTA on a two-level grid has reached almost the same value at 9.37×10^2s (5 iterations). This implies that FISTA on a two-level grid was almost three times faster than on a fixed grid. MG FISTA was finally terminated at 1.46×10^3s with an F of 2.10×10^2.

Some of the reconstructed images have been shown in figure 2.5. The color scale of each figure was set independently. It is clear that the MG variant of FISTA reconstructed more accurate images in less times than the fixed-grid algorithm.

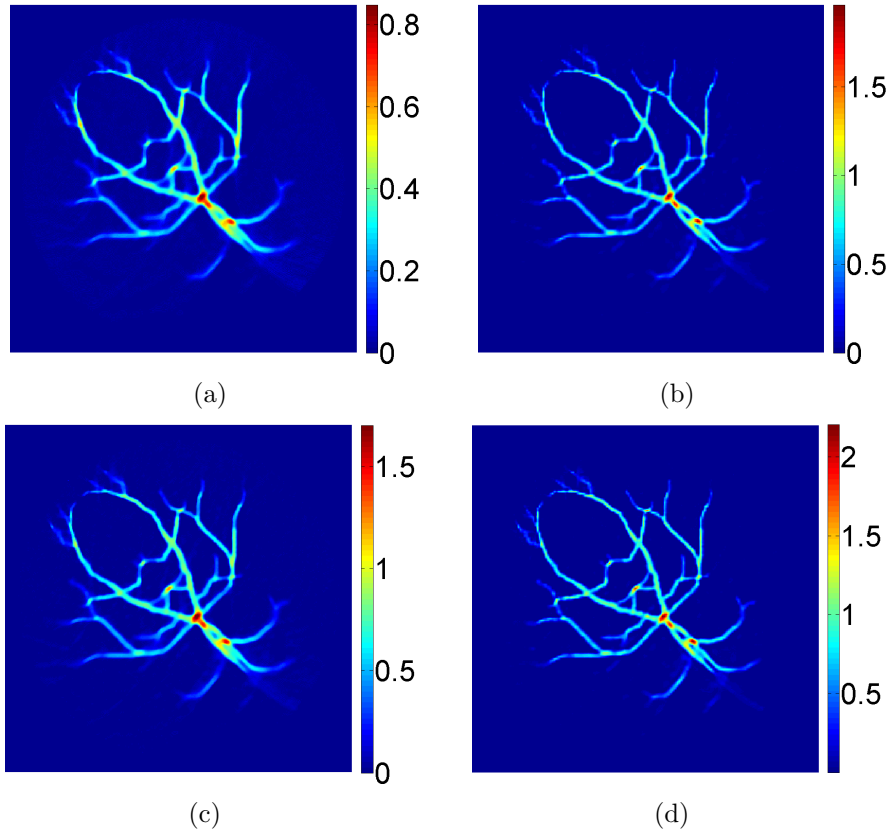


Figure 2.5: 2D images reconstructed by FISTA. (a) Iteration 4 on a fixed grid (5.09×10^2s) (b) Iteration 2 on a two-level grid (4.05×10^2s) (c) Iteration 8 on a fixed grid (1.03×10^3s) (d) Iteration 4 on a two-level grid (8.09×10^2s).

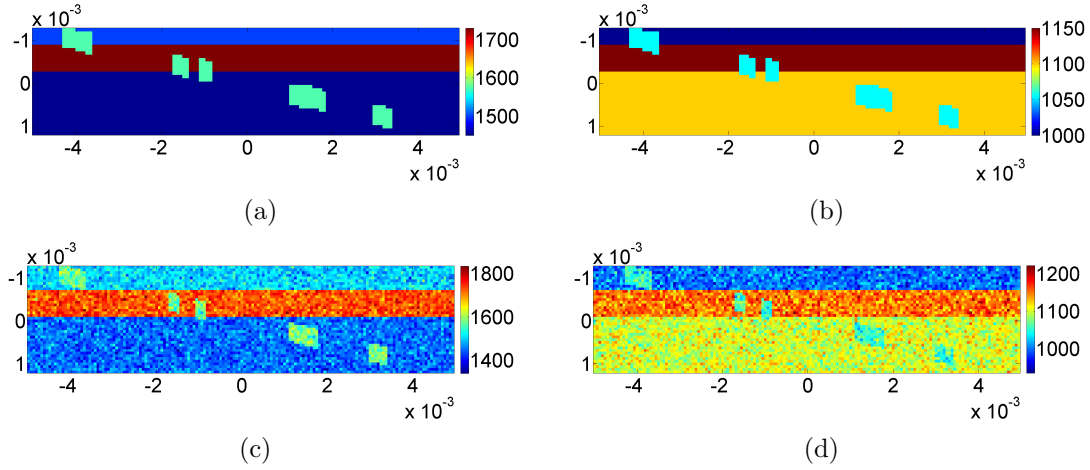


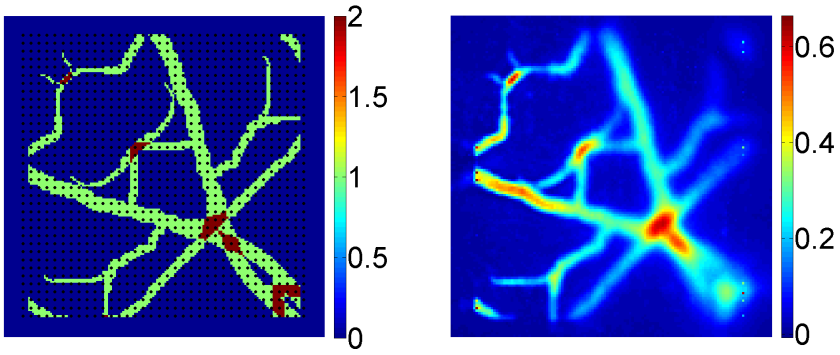
Figure 2.6: Medium's properties for image reconstruction (a) sound speed (b) density, and data generation (c) sound speed (d) density.

2.6.2 3D PAT Simulation

A 3D grid with a size of $1 \times 1 \times 0.25\,\mathrm{cm}^3$ was created, made up of $160 \times 160 \times 40$ grid points with a spatial spacing of 6.25×10^{-2} mm, supporting a maximal frequency of 10.36 MHz in all axes. To measure the propagated wavefield, 36×36 point-wise pressure detectors were placed on the top surface of the grid (see [2]). A PML was added to the grid in the same way as in section 2.6.1. The sound speed and density maps were simulated inhomogeneous, as shown in figures 2.6(a) and 2.6(b), respectively. From the top to bottom, the layers represent properties of water, skin and soft tissue with sound speed and density values given in section 2.6.1 [4]. The sound speed and density of vasculature were set the same as the 2D phantom as well.

To avoid inverse crime, for data generation these maps have been contaminated with a 35dB AWGN, and the "water-skin" and "skin-soft tissue" interfaces were shifted to the bottom by 3 grid points (18.75×10^{-2}mm), as shown in figures 2.6(c) and 2.6(d). To mitigate aliasing artifacts, for all forward and adjoint models, medium's properties were smoothed by the k-wave toolbox [36]. The absorption coefficient and power law exponent of tissues were set the same as those in section 2.6.1.

The phantom that was already created for the 2D scenario was now placed obliquely in the plane $z = y/4$. Figure 2.7(a) displays the simulated phantom from a top view, and the sensors are shown by black dots. The computed pressure field was evenly sampled in 914 time steps, and was linearly interpolated to the sensors. A 30dB AWGN was then incorporated to the generated data.



(a) (b)

Figure 2.7: 3D phantom. (a) Initial pressure distribution (b) image reconstructed by TR.

The reconstruction was applied to a grid made up of $128 \times 128 \times 32$ grid points, supporting a maximal frequency of 9.28 MHz. Figure 2.7(b) displays the image reconstructed by TR. This image has an RE of 87.98%. In our study, 3D visualizations were done by Maximum Intensity Projection (MIP) method (see [20, 2]).

Iterative methods: The iterative reconstruction was performed by ISTA and FISTA. The step sizes were chosen by power iteration method similarly to the algorithms in section 2.6.1. The regularization parameter was heuristically chosen to be $\lambda = 1 \times 10^{-2}$. The MG algorithms were then implemented to reconstruct images on two levels having sizes $128 \times 128 \times 32$ and $64 \times 64 \times 16$. These algorithms were implemented by the same parameters as in 2D scenario, except that κ was set to 1/8, i.e., size of the coarse grid relative to the fine grid, and also the smoothing parameter ρ was set to 3×10^{-2}.

Figures 2.8(a) and 2.8(b) respectively, show RE and RES of the images reconstructed by ISTA versus CPU time in the same way as in section 2.6.1. Figures 2.8(c) and 2.8(d) display RE and RES of sequences provided by FISTA, respectively. As seen in these figures, both ISTA and FISTA exhibited a better performance on a two-level grid than on a fixed grid.

Fig. 2.9 shows F values versus CPU time for all the used algorithms. As shown in this figure, the F value obtained by MG ISTA after 8.74×10^3s (9 iterations) was less than the optimal value obtained by fixed-grid ISTA at 2.05×10^4s (23 iterations). This indicates that MG ISTA was 2.35 times faster than ISTA on a fixed grid.

Applying FISTA, the F value computed by the MG algorithm after 4.71×10^3s (5 iterations) was less than the optimal value computed by the fixed-grid algorithm

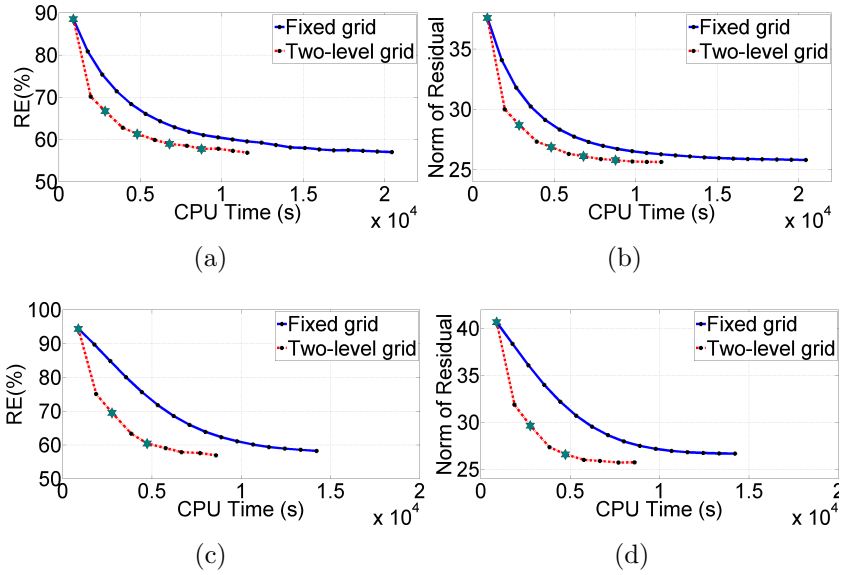


Figure 2.8: Evaluation of 3D images reconstructed on fixed grid (blue) and two-level grid (red). ISTA: (a) RE (b) RES, and FISTA: (c) RE (d) RES.

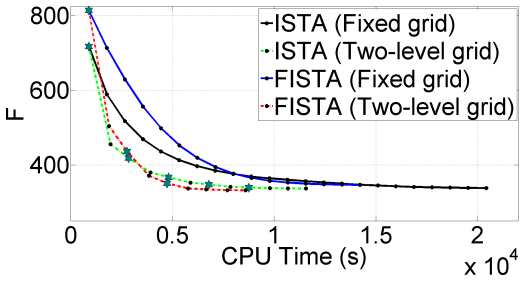


Figure 2.9: Objective function values computed by ISTA on a fixed grid (black), ISTA on a two-level grid (green), FISTA on a fixed grid (blue), and FISTA on a two-level grid (red) for the 3D phantom.

at 1.42×10^4s (16 iterations). This indicates that MG FISTA was almost three times faster than FISTA on a fixed grid.

Furthermore, a comparison between ISTA and FISTA on a fixed grid indicates that the convergence of ISTA was faster than FISTA at early iterations because of using a step size greater than FISTA, but FISTA has finally provided almost the same optimal F as ISTA in a less time. Applying the MG algorithms, MG ISTA was faster than MG FISTA at early iterations, but MG FISTA converged better than MG ISTA around the optimal point, and reached a lower F than MG ISTA in less time. MG ISTA finally reached a value of 3.36×10^2 at 1.15×10^4s (12 iterations), while MG FISTA was terminated at 8.59×10^3s (9 iterations) with an F of 3.32×10^3.

Figures 2.10(a) and 2.10(c) show images reconstructed by FISTA on the fixed grid at iterations 6 and 16 (stopping point), respectively. Figures 2.10(b) and 2.10(d)

display images reconstructed by MG FISTA at iterations 4 and 9 (stopping point). The MIP visualization provided a different scaling for the reconstructed images, compared to the simulated phantom (p_{exact}). This different scaling as well as small scale noise do not affect the evaluation of a human observer [2]. Thus the reconstructed images were rescaled and thresholded before visualization according to [2]. This makes the colorbars equal to that of the simulated phantom, and thus simplifies comparison between the images with respect to the simulated phantom. Accordingly, the visualized image $\bar{x}$ is computed in the form

$$\bar{x} = \text{thres}\left(2\frac{x}{\|x\|_\infty}, 0.1\right) \tag{2.42}$$

where,

$$\text{thres}(v, a) = \begin{cases} v, \text{ if } v \geqslant a \\ 0, \text{ else} \end{cases} . \tag{2.43}$$

Here, a is a thresholding parameter, and the factor 2 in (2.42) accounts for the maximum amplitude of p_{exact}.

2.7 Conclusion

We proposed a line search MG optimization approach for PAT. Applied on two grids, our numerical results show that this strategy has improved the speed of ISTA 4 and 2.35 times, respectively in 2D and 3D scenarios. A better convergence than MG ISTA was reached by MG FISTA, which was 3 times faster than fixed-grid FISTA in both 2D and 3D cases.

We derived the adjoint from a first order acoustic system of equations that includes the absorption and dispersion. Our method for deriving the adjoint is in contrast to the method used in [3], where the adjoint has been derived based on second order acoustic wave equation, and thus does not include absorption and dispersion.

Further studies are needed to extend the proposed MG method to other popular acoustic systems of equations, or other models for describing absorption and dispersion, e.g., [22]. The forward implementation of these acoustic systems is often more expensive than the forward model used in our study, and thus solving the corresponding iterative algorithms in a multi-grid setting can be very useful. Additionally, a

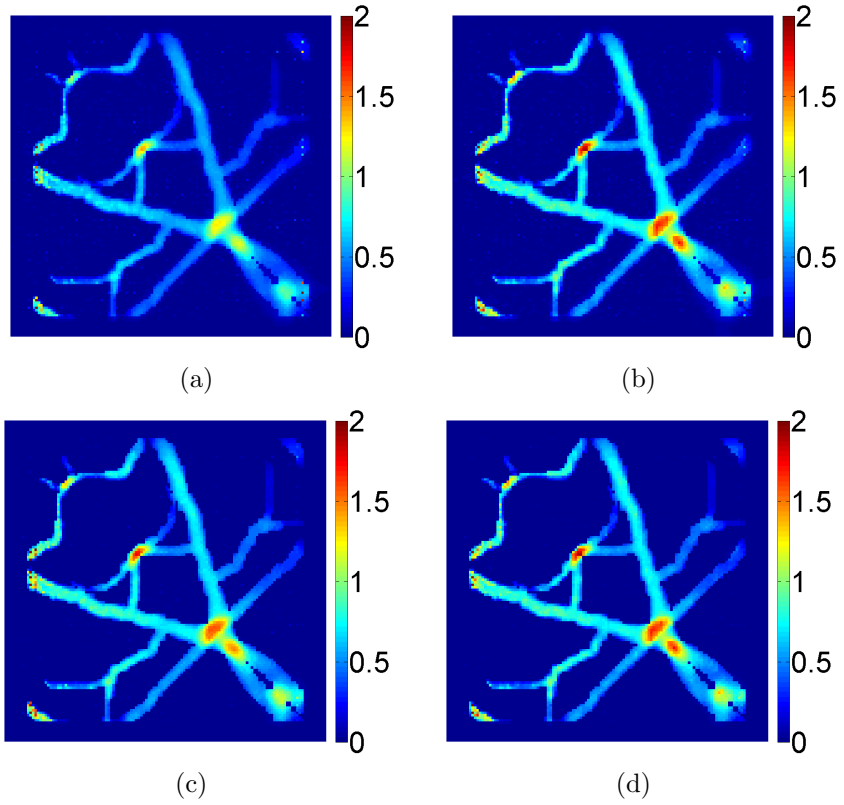


Figure 2.10: Visualization of 3D images reconstructed by FISTA. (a) Iteration 6 on a fixed grid (5.34×10^3s) (b) Iteration 4 on a two-level grid (3.84×10^3s) (c) Iteration 16 on a fixed grid (1.42×10^4s) (d) Iteration 9 on a two-level grid (8.59×10^3s).

direct method for quantitative PAT has recently received much attention, where the forward model is treated as a coupled acoustic and optical model, and thus directly links optical properties of medium to time series of acoustic boundary measurements [16]. Quantitative PAT requires the joint reconstruction of optical absorption and scattering coefficients. This makes the corresponding iterative reconstruction very expensive. It is hoped that an extension of the proposed MG method to the direct quantitative PAT can be useful for improving the reconstruction speed.

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).

2.A Supplementary Materials

In this chapter, we solved the inverse problem of PAT using a class of variational approaches that are based on forward-backward splitting (FBS) methods. In this thesis, we use FBS methods, when we consider only the acoustic portion of the inverse problem (PAT). FBS methods are popular for application of variational approaches for PAT (See [3] for ISTA, and [20, 2] for its accelerated variant (FISTA).) Further details on FBS methods have been given in section 1.8. This approach involves a forward gradient step, which enforces a descent direction on the smooth part of the objective function, and a backward gradient step that enforces a descent direction on the non-smooth part of the objective function, following a projection onto a non-negativity constraint. The backward gradient step is often solved using a proximal operator [6, 5]. In this thesis, we used a total variation (TV) functional for the non-smooth term of the objective function, and the details for an associated proximal operator have been given in section 1.8.3.

2.A.1 Choosing a fixed step size for FBS methods

As explained in sections 2.4 and 1.8, using FBS methods, the step size α can be chosen as a constant value for all iterations, or can be computed adaptively at each iteration using backtracking line search techniques. Using backtracking line search methods, for enforcing a descent direction on the objective function, many implementations of the forward operator H and the adjoint operator H^* may be required for each iteration. This considerably increases the computational cost for an iterative image reconstruction for PAT, and is impractical for large-scale problems. To avoid this, following similar works in [3, 2], we used a fixed step size for all iterations. Choosing a fixed step size for minimisation of a continuously differentiable function f with a Lipschitz continuous gradient (cf. (2.28)) requires a knowledge of the smallest Lipschitz constant L_f for the gradient, that is the largest singular value of H^*H. Using this, the step size must satisfy $\alpha_k \in (0, a/L_f)$ for all k, where $a = 2$ for ISTA and $a = 1$ for FISTA [6, 5]. For PAT, an access to an explicit form of H and H^* is not practical. Therefore, following [3, 2], we used a power method for computation of L_f before applying Algorithms 2.1 and 2.2.

Algorithm 2.4 Power method

1: Choose $q^0 \in \mathbb{R}^N$
2: **while** $k = 1 \ \cup \ \|q^k - q^{k-1}\|_2 > \epsilon$ **do**
3: $\quad p^k = H^*Hq^{k-1}$
4: $\quad q^k = \frac{p^k}{\|p^k\|_2}$
5: $\quad \sigma^k = (q^k)^* H^*Hq^k$
6: **end while**
7: **Output:** σ_*.

Here, q^0 represents a randomly chosen initial pressure distribution inside a volume encompassed by the detectors, and is updated after each iteration of the power method. Also, the superscript k denotes the iteration for the power method, and σ_* denotes the solution of the power method, which is very close to $\sigma_{\max}$, the largest singular value of H^*H. (We hope that the subscript $*$ that denotes an optimal solution is not confused with the superscript $*$ that denotes the adjoint of a matrix or vector.) Additionally, ϵ is a user-adjusted parameter for terminating the algorithm. Note that σ_* is independent of the sought after initial pressure distribution, and can be stored in a look-up table, and reused for all experiments with a fixed setting (cf. section 2.4)

[2].

2.A.2 Methods for derivation of the adjoint operator (continuous and discretised adjoint)

The approaches for derivation of the adjoint operator H^* can be categorised into *adjoint-then-discretise* or *discretise-then-adjoint* methods. In this thesis, an adjoint that is derived using the first (resp. second) approach is referred to as continuous (resp. algebraic (discretised)) adjoint.

The class of *adjoint-then-discretise* methods has been used in many studies of PAT, e.g., [3, 2]. A continuous adjoint can be considered as a generalised variant of a corresponding algebraic adjoint. The reason is that firstly a continuous adjoint, as opposed to an algebraic adjoint, is independent of the numerical approaches for a discretisation of H, and secondly if any modifications are applied on the forward operator H, the applied modifications can be straightforwardly included in a continuous adjoint operator. For example, if we use an alternative approach to equation (2.6) for modelling the acoustic attenuation, or if we use a different approach for a numerical approximation of the forward operator defined in section 2.2.1, the applied modification can be directly applied on the continuous adjoint, using a same aproach as used for the forward operator.

In contrast, using an algebraic (discretised) adjoint, applying any changes on the forward operator may require a recalculation of the adjoint operator. However, considering the forward operator that is defined using (2.1), (2.2) and (2.6), an algebraic adjoint satisfies an adjoint formula more accurately than using a continuous adjoint, although both approaches produce almost the same solutions for the iterative PAT, when we consider only the acoustic portion of the inverse problem, the same as chapter 2.

Since both approaches are poular for PAT, we used both methods in this thesis. In chapter 2, a continuous adjoint, which is a generalised variant of a corresponding algebraic adjoint, may better demonstrate the performance of the proposed multi-grid method.

It is worth mentioning that the forward operator that is described in section 2.2.1

will also be used in chapter 4 as the acoustic portion of a composite opto-acoustic forward operator for quantitative PAT using a single-stage approach [16]. Since the optical portion of the mentioned composite forward operator is highly nonlinear and ill-conditioned, an small error in calculation of the acoustic adjoint may exponentially grow during the inversion process, as opposed to our observation for the pure acoustic portion of PAT. Because of this, in chapter 4, we will use a *discretise-then-adjoint* approach for derivation of the adjoint.

2.A.3 Methods for modelling the forward operator for fluid and solid media

One limitation for the forward operator described above is that it is accurate solely for fluid media, because it does not account for generation and propagation of shear waves. Indeed, for fluid media, these effects can be neglected, and therefore the forward model used in chapter 2 is accurate sufficiently for an iterative PAT in soft tissues, for example the breast. PAT has recently shown its potential for imaging of the brain. For this case, the inverse problem of PAT is a reconstruction of an initial pressure distribution inside the skull from a set of time series of pressure data that is measured outside the skull. To solve this inverse problem using variational (iterative) approaches, the forward operator and its adjoint must account for generation and propagation of the shear waves within the skull. To account for these effects, in chapter 3, we define a forward operator using a system of coupled equations that describe the propagation of acoustic waves in linear isotropic heterogeneous and lossy elastic media. Additionally, an acoustic absorption and dispersion obeying a frequency power law will be modelled using fractional Laplacian operators, the same as the forward operator in chapter 2. We will solve the inverse problem of PAT using this forward operator. Note that the same as in chapter 2, we will consider only the acoustic portion of the inverse problem of photo-acoustic tomography.

Bibliography

[1] Z. Allen-Zhu and L. Orecchia. A novel, simple interpretation of nesterov's accelerated method as a combination of gradient and mirror descent. *CoRR*,

abs/1407.1537, 2014.

[2] 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. pages 1–19, 2016. arXiv:1605.00133v1.

[3] 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.

[4] H. Azhari. *Appendix A: Typical Acoustic Properties of Tissues (in Basics of Biomedical Ultrasound for Engineers)*, pages 313–314. John Wiley Sons, Inc., 2010.

[5] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.*, 18(11):2419–2434, 2009.

[6] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci.*, 2(1):183–202, 2009.

[7] A. Beck and M. Teboulle. Smoothing and first order methods: A unified framework. *SIAM J. Optim.*, 22(2):557–580, 2012.

[8] W. Chen, , and S. Holm. Modified szabo′s wave equation models for lossy media obeying frequency power law. *J. Acoust. Soc. Am.*, 114(5):2570–2574, 2003.

[9] W. Chen and S. Holm. Fractional laplacian time-space models for linear and nonlinear lossy media exhibiting arbitrary frequency power-law dependency. *J. Acoust. Soc. Am.*, 115(4):1424–1430, 2004.

[10] 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.

[11] X. L. Deán-Ben, V. Ntziachristos, and D. Razansky. Acceleration of optoacoustic model-based reconstruction using angular image discretization. *IEEE Trans. Med. Imag.*, 31(5):1154–1162, 2012.

[12] X. L. Deán-Ben, A. Ozbek, and D. Razansky. Volumetric real-time tracking of peripheral human vasculature with gpu-accelerated three-dimensional optoacoustic tomography. *IEEE Trans. Med. Imag.*, 32(11):2050–2055, 2013.

[13] T. Goldstein, C. Studer, and R. G. Baraniuk. A field guide to forward-backward splitting with a FASTA implementation. *CoRR*, abs/1411.3406, 2014.

[14] C. Graser and R. Kornhuber. Multigrid methods for obstacle problems. *J. Comput. Math.*, 27(1):1–44, 2009.

[15] S. Gratton, A. Sartenaer, and P. L. Toint. Recursive trust-region methods for multiscale nonlinear optimization. *SIAM J Optim.*, 19(1):414–444, 2008.

[16] M. Haltmeier, L. Neumann, and S. Rabanser. Single-stage reconstruction algorithm for quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065005, 2015.

[17] V. Hovhannisyan, P. Parpas, and S. Zafeiriou. Magma: Multi-level accelerated gradient mirror descent algorithm for large scale convex composite minimization. 2016.

[18] Y. Hristova. Time reversal in thermoacoustic tomography - an error estimate. *Inverse Problems*, 25(5):055008, 2009.

[19] 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.

[20] 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.

[21] M. Kocvara and S. Mohammad. A first-order multigrid method for bound-constrained convex optimization. *Optim. Method Softw.*, 31(3):622–644, 2016.

[22] R. Kowar. Causality analysis of the fractional state and wave equations based on fractional laplacians. *Wave Motion*, 58:244 – 258, 2015.

[23] R. Kowar and O. Scherzer. *Attenuation Models in Photoacoustics*, pages 85–130. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[24] R. A. Kruger, D. R. Reinecke, and G. A. Kruger. Thermoacoustic computed tomography−technical considerations. *Medical Physics*, 26(9):1832–1837, 1999.

[25] M. Liebler, S. Ginter, T. Dreyer, and R. E. Riedlinger. Full wave modeling of therapeutic ultrasound: Efficient time-domain implementation of the frequency power-law attenuation. *J. Acoust. Soc. Am.*, 116(5):2742–2750, 2004.

[26] C. Lutzweiler, X. L. Deán-Ben, and D. Razansky. Expediting model-based optoacoustic reconstructions with tomographic symmetries. *Medical Physics*, 41(1):013302, 2014.

[27] A. I. Nachman, J. F. Smith, and R. C. Waag. An equation for acoustic propagation in inhomogeneous media with relaxation losses. *J. Acoust. Soc. Am.*, 88(3):1584–1595, 1990.

[28] S. G. Nash. A multigrid approach to discretized optimization problems. *Optim. Method Softw.*, 14:99–116, 2000.

[29] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

[30] P. Parpas, D. V. N. Luong, D. Rueckert, and B. Rustem. A multilevel proximal algorithm for large scale composite convex optimization. 2016.

[31] 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.

[32] P. Stefanov and G. Uhlmann. Thermoacoustic tomography arising in brain imaging. *Inverse Problems*, 27(4), 2011.

[33] T. L. Szabo. Time domain wave equations for lossy media obeying a frequency power law. *J. Acoust. Soc. Am.*, 96(1):491–500, 1994.

[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] T. Tarvainen, A. Pulkkinen, B. T. Cox, J. P. Kaipio, and S. R. Arridge. Bayesian image reconstruction in quantitative photoacoustic tomography. *IEEE Trans. Med. Imag.*, 32(12):2287–2298, 2013.

[36] B. E. Treeby and B. T. Cox. k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields. *J. Biomed. Opt.*, 15(2):021314–021314–12, 2010.

[37] 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.

[38] 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.

[39] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17(1):227–238, 1996.

[40] K. Wang and M. A. Anastasio. *Photoacoustic and Thermoacoustic Tomography: Image Formation Principles*, pages 781–815. Springer New York, New York, NY, 2011.

[41] K. Wang, C. Huang, Y. J. Kao, C. Y. Chou, A. A. Oraevsky, and M. A. Anastasio. Accelerating image reconstruction in three-dimensional optoacoustic tomography on graphics processing units. *Medical Physics*, 40(1):023301, 2013.

[42] K. Wang, R. Su, A. A. Oraevsky, and M. A. Anastasio. Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography. *Phys. Med. Biol*, 57(17):5399–5423, 2012.

[43] Z. Wen and D. Goldfarb. A line search multigrid method for large-scale nonlinear optimization. *SIAM J Optim.*, 20(3):1478–1503, 2010.

[44] M. Xu and L. V. Wang. Time-domain reconstruction for thermoacoustic tomography in a spherical geometry. *IEEE Trans. Med. Imag.*, 21(7):814–822, 2002.

[45] M. Xu and L. V. Wang. Universal back-projection algorithm for photoacoustic computed tomography. *Phys. Rev. E*, 71(1):016706, 2005.

[46] Y. Xu, D. Feng, and L. V. Wang. Exact frequency-domain reconstruction for thermoacoustic tomography. i. planar geometry. *IEEE Trans. Med. Imag.*, 21(7):823–828, 2002.

[47] F. Ye, Z. Ji, W. Ding, C. Lou, S. Yang, and D. Xing. Ultrashort microwave-pumped real-time thermoacoustic breast tumor imaging system. *IEEE Trans. Med. Imag.*, 35(3):839–844, 2016.

[48] F. Ye, S. Yang, and D. Xing. Three-dimensional photoacoustic imaging system in line confocal mode for breast cancer detection. *Applied Physics Letters*, 97(21), 2010.

[49] L. Zeng, D. Xing, H. Gu, D. Yang, S. Yang, and L. Xiang. High antinoise photoacoustic tomography based on a modified filtered backprojection algorithm with combination wavelet. *Medical Physics*, 34(2):556–563, 2007.

Chapter 3

A continuous adjoint for photoacoustic tomography of the brain

The content of this chapter has been published in: A. Javaherian and S. Holman. A continuous adjoint for photo-acoustic tomography of the brain. Inverse Problems, 34(8):085003, 2018.

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.

3.1 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 ($\approx$ 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.

3.2 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

$$\mu = \rho c_s^2, \quad \lambda = \rho c_p^2 - 2\mu, \tag{3.1}$$

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]

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right). \tag{3.2}$$

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

$$\sigma_{ij} = \lambda\delta_{ij}\epsilon_{ll} + 2\mu\epsilon_{ij}. \tag{3.3}$$

One method of incorporating attenuation is to modify the stress-strain relation (3.3) as

$$\sigma_{ij} = \lambda\delta_{ij}\epsilon_{ll} + 2\mu\epsilon_{ij} + \chi\delta_{ij}\frac{\partial}{\partial t}\epsilon_{ll} + 2\eta\frac{\partial}{\partial t}\epsilon_{ij}, \tag{3.4}$$

where χ and η denote the compressional and shear viscosity coefficients. The equation (3.4) is called Kelvin-Voigt model. Plugging (3.2) into (3.4) gives [37]

$$\frac{\partial\sigma_{ij}}{\partial t} = \lambda\delta_{ij}\frac{\partial v_l}{\partial x_l} + \mu\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) + \chi\delta_{ij}\frac{\partial^2 v_l}{\partial x_l\partial t} + \eta\left(\frac{\partial^2 v_i}{\partial x_j\partial t} + \frac{\partial^2 v_j}{\partial x_i\partial t}\right). \tag{3.5}$$

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 3.1, 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 (3.4) and (3.5) can be replaced by fractional time derivatives [12]. For an isotropic medium, this gives the fractional Kelvin-Voigt model in the form [37]

$$\sigma_{ij} = \lambda\delta_{ij}\epsilon_{ll} + 2\mu\epsilon_{ij} + \chi\delta_{ij}\frac{\partial^{y-1}}{\partial t^{y-1}}\epsilon_{ll} + 2\eta\frac{\partial^{y-1}}{\partial t^{y-1}}\epsilon_{ij}, \tag{3.6}$$

where

$$\eta = -\frac{2\rho c_s^3}{\cos(\pi y/2)}\alpha_{0,s}, \qquad \chi = -\frac{2\rho c_p^3}{\cos(\pi y/2)}\alpha_{0,p} - 2\eta, \tag{3.7}$$

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

The temporal fractional derivatives in equation (3.6) 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

$$\begin{aligned} v_i^p &= q^p(v_i) = F^{-1}\Big\{\hat{k}_i\hat{k}_j F\big\{v_j\big\}\Big\} \\ v_i^s &= q^s(v_i) = F^{-1}\Big\{(\delta_{ij} - \hat{k}_i\hat{k}_j) F\big\{v_j\big\}\Big\}, \end{aligned} \tag{3.8}$$

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 $k = \left(\sum_i k_i^2\right)^{1/2}$ the magnitude of wavenumber. By splitting the particle velocity vector, the stress tensor is updated distinctly for compressional and shear parts in the form

$$\begin{aligned} \frac{\partial \sigma_{ij}^{p,s}}{\partial t} = \lambda \left(\delta_{ij}\frac{\partial}{\partial x_l} v_l^{p,s}\right) + \mu \left(\frac{\partial}{\partial x_j} v_i^{p,s} + \frac{\partial}{\partial x_i} v_j^{p,s}\right) + \chi \left(\delta_{ij}\frac{\partial}{\partial x_l}\frac{\partial_{p,s}^{y-1}}{\partial t^{y-1}} v_l^{p,s}\right) \\ + \eta \left(\frac{\partial}{\partial x_j}\frac{\partial_{p,s}^{y-1}}{\partial t^{y-1}} v_i^{p,s} + \frac{\partial}{\partial x_i}\frac{\partial_{p,s}^{y-1}}{\partial t^{y-1}} v_j^{p,s}\right). \end{aligned} \tag{3.9}$$

Now, the fractional temporal derivatives in (3.9) 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

$$L^y_{c_{p,s}} = c^y_{p,s}(-\nabla^2)^{y/2}. \tag{3.10}$$

Using this notation and the comments above, we will be using the following definition throughout the rest of this work including in (3.9) [37]

$$\frac{\partial^{y-1}_{p,s}}{\partial t^{y-1}} v_i^{p,s} = \sin(\pi y/2) L^{y-1}_{c_{p,s}} v_i^{p,s} - \cos(\pi y/2) L^{y-2}_{c_{p,s}} \frac{1}{\rho} \frac{\partial}{\partial x_j} \sigma_{ij}^{p,s}. \tag{3.11}$$

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

$$\frac{\partial v_i}{\partial t} = \frac{1}{\rho} \sum_{p,s} \frac{\partial}{\partial x_j} \sigma_{ij}^{p,s}. \tag{3.12}$$

Note that in the second term of (3.11), following [37] $\partial v_i^{p,s}/\partial t$ has been replaced by the split terms of (3.12). This is done to avoid having to compute time differences in the discretised model (cf. [37], section III).

Equations (3.9) and (3.12), 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.

3.3 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

$$\mathcal{G}\sigma^p(x,t) = -\frac{1}{d}\delta_{ij}\sigma^p_{ij}(x,t) = p(x,t). \tag{3.13}$$

We also introduce the measurement operator $\mathcal{M} : \mathcal{L}^2(\mathbb{R}^d) \to \mathbb{R}^{N_s}$, which at each time instant t_n maps the pressure field p to the measurements detected by all 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 3.1. We define the PAT forward operator using the viscoelastic model in the form

$$\begin{aligned} \Lambda &: C_0^\infty(\Omega) \to \mathbb{R}^{N_s N_t} \\ (\Lambda[p_0])_{t_n} &= \mathcal{M}\mathcal{G}\sigma^p(t_n), \end{aligned} \tag{3.14}$$

where t_n is a discretised representation of time t. Also, $\sigma_{ij}^{p,s}$ and v_i satisfy (3.9) and (3.12) with initial conditions

$$\sigma_{ij}^p(x,0) = -\delta_{ij}p_0(x), \quad \sigma_{ij}^s(x,0) = 0, \quad v_i(x,0) = 0. \tag{3.15}$$

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 (3.9) and (3.12) 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.,

$$\Lambda^* : \mathbb{R}^{N_s N_t} \to \mathcal{L}^2(\Omega). \tag{3.16}$$

For this, we need the adjoint of $L^y_{c_{p,s}}$, i.e.,

$$L^{y*}_{c_{p,s}} = (-\nabla^2)^{y/2} c^y_{p,s}, \tag{3.17}$$

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

$$\begin{aligned} \rho \frac{\partial {v_i}^*}{\partial t} = \sum_{p,s} q^{p,s} \Bigg[\Bigg(\frac{\partial}{\partial x_i}\Big(\lambda \sigma_{ll}^{p,s*}\Big) + 2\frac{\partial}{\partial x_j}\Big(\mu \sigma_{ij}^{p,s*}\Big)\Bigg) \\ + \sin(\pi y/2) L^{(y-1)*}_{c_{p,s}} \Bigg(\frac{\partial}{\partial x_i}\Big(\chi \sigma_{ll}^{p,s*}\Big) + 2\frac{\partial}{\partial x_j}\Big(\eta \sigma_{ij}^{p,s*}\Big)\Bigg)\Bigg] \end{aligned} \tag{3.18}$$

$$\begin{aligned} \mathbf{v}_i^{p,s*} &= v_i^* - \frac{1}{\rho}\cos(\pi y/2) L^{(y-2)*}_{c_{p,s}} \Bigg(\frac{\partial}{\partial x_i}\Big(\chi \sigma_{ll}^{p,s*}\Big) + 2\frac{\partial}{\partial x_j}\Big(\eta \sigma_{ij}^{p,s*}\Big)\Bigg) \\ \frac{\partial {\sigma_{ij}^p}^*}{\partial t} &= \frac{1}{2}\left(\frac{\partial \mathbf{v}_i^{p*}}{\partial x_j} + \frac{\partial \mathbf{v}_j^{p*}}{\partial x_i}\right) + (\mathcal{R}\mathcal{G}^*\mathcal{M}^*\hat{P}) \\ \frac{\partial {\sigma_{ij}^s}^*}{\partial t} &= \frac{1}{2}\left(\frac{\partial \mathbf{v}_i^{s*}}{\partial x_j} + \frac{\partial \mathbf{v}_j^{s*}}{\partial x_i}\right) \end{aligned} \tag{3.19}$$

with initial conditions

$$\sigma_{ij}^{p,s*}(x,0) = 0, \quad v_i^*(x,0) = 0. \tag{3.20}$$

Remark 3.1. Solutions for the set of equations (3.18) and (3.19) 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 (3.9), (3.12) and (3.15), and also σ_{ij}^{p,s^*} and v_i^* satisfy (3.18), (3.19) and (3.20), then the forward map Λ and adjoint Λ^* must satisfy

$$\langle \Lambda[p_0], \hat{P} \rangle_{\mathbb{R}^{N_s N_t}} = \langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)} \tag{3.21}$$

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 (3.20) 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^*

$$\int_0^T \int_{\mathbb{R}^d} \rho \left(\frac{\partial v_i}{\partial t} \overline{v_i^*} + v_i \overline{\frac{\partial v_i^*}{\partial t}} \right) \, \mathrm{d}x \, \mathrm{d}t = 0. \tag{3.22}$$

Plugging (3.12) into the first integrand in the above equation gives

$$\int_0^T \int_{\mathbb{R}^d} \sum_{p,s} \frac{\partial \sigma_{ij}^{p,s}}{\partial x_j} \overline{v_i^*} + \rho v_i \overline{\frac{\partial v_i^*}{\partial t}} \, \mathrm{d}x \, \mathrm{d}t = 0. \tag{3.23}$$

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

$$\int_0^T \int_{\mathbb{R}^d} \sum_{p,s} -\sigma_{ij}^{p,s} \frac{1}{2} \overline{\left(\frac{\partial v_i^*}{\partial x_j} + \frac{\partial v_j^*}{\partial x_i} \right)} + \rho v_i \overline{\frac{\partial v_i^*}{\partial t}} \, \mathrm{d}x \, \mathrm{d}t = 0. \tag{3.24}$$

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

$$\int_0^T \int_{\mathbb{R}^d} \frac{\partial \sigma_{ij}^p}{\partial t} \overline{\sigma_{ij}^{p\,*}} + \sigma_{ij}^p \overline{\frac{\partial \sigma_{ij}^{p\,*}}{\partial t}} \, \mathrm{d}x \, \mathrm{d}t = \int_{\mathbb{R}^3} p_0 \, \overline{\sigma_{ii}^{p\,*}(x,0)} \, \mathrm{d}x = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}. \tag{3.25}$$

and

$$\int_0^T \int_{\mathbb{R}^d} \frac{\partial \sigma_{ij}^s}{\partial t} \overline{\sigma_{ij}^{s\,*}} + \sigma_{ij}^s \overline{\frac{\partial \sigma_{ij}^{s\,*}}{\partial t}} \, \mathrm{d}x \, \mathrm{d}t = 0. \tag{3.26}$$

Now, plugging (3.9) into the first integrands in the left-hand sides of (3.25) and (3.26) and then adding these two equations results in

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} \sum_{p,s} \Bigg[\Bigg(\lambda \delta_{ij} \frac{\partial v_l^{p,s}}{\partial x_l} + \mu \Big(\frac{\partial v_i^{p,s}}{\partial x_j} + \frac{\partial v_j^{p,s}}{\partial x_i} \Big) + \chi \Big(\delta_{ij} \frac{\partial}{\partial x_l} \frac{\partial_{p,s}^{y-1}}{\partial t^{y-1}} v_l^{p,s} \Big) \\ + \eta \Big(\frac{\partial}{\partial x_j} \frac{\partial_{p,s}^{y-1}}{\partial t^{y-1}} v_i^{p,s} + \frac{\partial}{\partial x_i} \frac{\partial_{p,s}^{y-1}}{\partial t^{y-1}} v_j^{p,s} \Big) \Bigg) \overline{\sigma_{ij}^{p,s*}} + \sigma_{ij}^{p,s} \overline{\frac{\partial \sigma_{ij}^{p,s*}}{\partial t}} \Bigg] \, \mathrm{d}x \, \mathrm{d}t = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}. \end{aligned} \tag{3.27}$$

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

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} \sum_{p,s} \Bigg[- \Bigg(\overline{\frac{\partial}{\partial x_i} \Big(\lambda \sigma_{ll}^{p,s*} \Big) + 2 \frac{\partial}{\partial x_j} \Big(\mu \sigma_{ij}^{p,s*} \Big) + \frac{\partial}{\partial x_i} \Big(\chi \sigma_{ll}^{p,s*} \Big)} \frac{\partial_{p,s}^{y-1}}{\partial t^{y-1}} \\ + 2 \overline{\frac{\partial}{\partial x_j} \Big(\eta \sigma_{ij}^{p,s*} \Big)} \frac{\partial_{p,s}^{y-1}}{\partial t^{y-1}} \Bigg) v_i^{p,s} + \sigma_{ij}^{p,s} \overline{\frac{\partial \sigma_{ij}^{p,s*}}{\partial t}} \Bigg] \, \mathrm{d}x \, \mathrm{d}t = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}. \end{aligned} \tag{3.28}$$

Now, adding equations (3.24) and (3.28) yields

$$\begin{aligned}\int_0^T \int_{\mathbb{R}^d} \rho v_i \overline{\frac{\partial v_i^*}{\partial t}} - \sum_{p,s} \Bigg[\overline{\frac{\partial}{\partial x_i}\left(\lambda \sigma_{ll}^{p,s*}\right) + 2\frac{\partial}{\partial x_j}\left(\mu \sigma_{ij}^{p,s*}\right)} \\ + \overline{\left(\frac{\partial}{\partial x_i}\left(\chi \sigma_{ll}^{p,s*}\right) + 2\frac{\partial}{\partial x_j}\left(\eta \sigma_{ij}^{p,s*}\right)\right)} \frac{\partial_{p,s}^{y-1}}{\partial t^{y-1}} \Bigg] v_i^{p,s} \\ + \sum_{p,s} \overline{\left[\frac{\partial \sigma_{ij}^{p,s*}}{\partial t} - \frac{1}{2}\left(\frac{\partial v_i^*}{\partial x_j} + \frac{\partial v_j^*}{\partial x_i}\right) \right]} \sigma_{ij}^{p,s} \, \mathrm{d}x \, \mathrm{d}t = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}.\end{aligned} \tag{3.29}$$

Now, plugging the fractional Laplacian operators defined in equation (3.11) into the second line in equation (3.29), together with (3.8), yields

$$\begin{aligned}\int_0^T \int_{\mathbb{R}^d} \Bigg(\rho \overline{\frac{\partial v_i^*}{\partial t}} - \sum_{p,s} \Bigg[\overline{\frac{\partial}{\partial x_i}\left(\lambda \sigma_{ll}^{p,s*}\right) + 2\frac{\partial}{\partial x_j}\left(\mu \sigma_{ij}^{p,s*}\right)} \\ + \overline{\left(\frac{\partial}{\partial x_i}\left(\chi \sigma_{ll}^{p,s*}\right) + 2\frac{\partial}{\partial x_j}\left(\eta \sigma_{ij}^{p,s*}\right)\right)} \sin(\pi y/2) L_{c_{p,s}}^{y-1} \Bigg] q^{p,s} \Bigg) v_i \\ + \sum_{p,s} \overline{\left(\frac{\partial}{\partial x_i}\left(\chi \sigma_{ll}^{p,s*}\right) + 2\frac{\partial}{\partial x_j}\left(\eta \sigma_{ij}^{p,s*}\right)\right)} \cos(\pi y/2) L_{c_{p,s}}^{y-2} \frac{1}{\rho} \frac{\partial}{\partial x_j} \sigma_{ij}^{p,s} \\ + \sum_{p,s} \overline{\left[\frac{\partial \sigma_{ij}^{p,s*}}{\partial t} - \frac{1}{2}\left(\frac{\partial v_i^*}{\partial x_j} + \frac{\partial v_j^*}{\partial x_i}\right) \right]} \sigma_{ij}^{p,s} \, \mathrm{d}x \, \mathrm{d}t = -\langle p_0, \Lambda^*[\hat{P}] \rangle_{\mathcal{L}^2(\mathbb{R}^d)}.\end{aligned} \tag{3.30}$$

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 (3.23), we can see from (3.19) that if the integral in the first two lines of (3.30) is equal to zero, then (3.21) holds and the proof is complete (recall again that relative to (3.18), we have reversed the time, and changed v_i^* to $-v_i^*$). So we now focus on the first two lines of (3.30) which we will denote $\mathcal{I}$. Considering that $q^{p,s}$ is self-adjoint yields

$$\begin{aligned}\mathcal{I} = \int_0^T \int_{\mathbb{R}^d} \overline{\Bigg[\rho \frac{\partial v_i^*}{\partial t} - \sum_{p,s} q^{p,s} \Bigg[\left(\frac{\partial}{\partial x_i}\left(\lambda \sigma_{ll}^{p,s*}\right) + 2\frac{\partial}{\partial x_j}\left(\mu \sigma_{ij}^{p,s*}\right) \right)} \\ \overline{+ \sin(\pi y/2) L_{c_{p,s}}^{(y-1)*} \left(\frac{\partial}{\partial x_i}\left(\chi \sigma_{ll}^{p,s*}\right) + 2\frac{\partial_{p,s}}{\partial x_j}\left(\eta \sigma_{ij}^{p,s*}\right) \right) \Bigg] \Bigg]} v_i \, \mathrm{d}x \, \mathrm{d}t.\end{aligned} \tag{3.31}$$

Since v_i^* and $\sigma_{ij}^{p,s*}$ satisfy (3.18) we can now see that in fact $\mathcal{I} = 0$, and so the proof is complete. □

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]).

3.4 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 $\zeta = (\zeta_1, \ ... \ , \zeta_d) \in \{1, \ ... \ , N_1\} \times \ ... \ \times \{1, \ ... \ , N_d\}$ with $N = \prod_{i=1}^{d} N_i$ the total number of grid points along d dimensions. The grid spacing along the ith direction will be denoted by Δx_i. Also let n denote the iteration corresponding to time $t_n = n\Delta t$ with $n \in \{-1, ..., N_t - 1\}$.

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

$$\frac{\partial_{p,s}\{\cdot\}}{\partial x_i^{\pm}} = F^{-1}\left\{\mathbf{i}k_i \operatorname{sinc}(C_0^{p,s} k\Delta t/2) e^{\pm \mathbf{i}k_i \Delta x_i/2} F\{\cdot\}\right\}, \tag{3.32}$$

where, as opposed to (3.8), 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]

$$\frac{\partial R(x,t)}{\partial t} + \alpha_a R(x,t) = \beta(x,t), \tag{3.33}$$

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

$$\frac{\partial \left(e^{\alpha_a t} R(x,t)\right)}{\partial t} = e^{\alpha_a t} \beta(x,t). \tag{3.34}$$

Using a staggered temporal grid, this is approximated as

$$\frac{e^{\alpha_a (t+\Delta t)} R(x_\zeta, t+\Delta t) - e^{\alpha_a t} R(x_\zeta, t)}{\Delta t} = e^{\alpha_a (t+\Delta t/2)} \beta(x_\zeta, t+\Delta t/2). \tag{3.35}$$

This gives the update

$$R(x_\zeta, t+\Delta t) = e^{-\alpha_a \Delta t/2} \Big[e^{-\alpha_a \Delta t/2} R(x_\zeta, t) + \Delta t \beta(x_\zeta, t+\Delta t/2) \Big]. \tag{3.36}$$

Using direction-dependent PMLs, the field variables are split into directions along the Cartesian coordinates $m \in \{1, ..., d\}$ [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

$$A_m = \operatorname{diag}(e^{-{}_m\alpha_a \Delta t/2}). \tag{3.37}$$

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 ith 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

$$\mathbb{T}_i f(x) = f(\mathbb{T}_i x). \tag{3.38}$$

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

$${}_m v_{(i;\zeta;n)} \approx \mathbb{T}_i \, {}_m v_i \left(x_\zeta, t_n\right). \tag{3.39}$$

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

$${}_m \sigma^{p,s}_{(ij;\zeta;n)} \approx \begin{cases} {}_m \sigma^{p,s}_{ij} \left(x_\zeta, t_n\right) & \text{if} \quad i = j \\ \mathbb{T}_i \mathbb{T}_j \, {}_m \sigma^{p,s}_{ij} \left(x_\zeta, t_n\right) & \text{if} \quad i \neq j. \end{cases} \tag{3.40}$$

Because of using a staggered grid, the unit dyadic tensor in (3.8) will be in the form [6]

$$(\hat{k}_i\hat{k}_j)_{\text{staggered}} = (\hat{k}_i\hat{k}_j)_{\text{nonstaggered}} \times \xi_{ij}, \tag{3.41}$$

where

$$\xi_{ij} = e^{+\mathbf{i}(k_i\Delta x_i - k_j\Delta x_j)/2} \tag{3.42}$$

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

$$\tau_{\text{dis}}^{p,s} = C_{p,s}^{y-1}\sin(\pi y/2) \tag{3.43}$$

$$\tau_{\text{abs}}^{p,s} = C_{p,s}^{y-2}\cos(\pi y/2) \tag{3.44}$$

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

$$\begin{aligned}
\bar{\rho}_i &= \text{diag}\,(\mathbb{T}_i\rho)\\
\bar{\lambda} &= \text{diag}\,(\lambda)\\
\bar{\mu}_{ij} &= \begin{cases} \text{diag}\,(\mu) & \text{if} \quad i = j\\ \text{diag}\,(\mathbb{T}_i\mathbb{T}_j\mu) & \text{if} \quad i \neq j. \end{cases}\\
\bar{\chi} &= \text{diag}\,(\chi)\\
\bar{\eta}_{ij} &= \begin{cases} \text{diag}\,(\eta) & \text{if} \quad i = j\\ \text{diag}\,(\mathbb{T}_i\mathbb{T}_j\eta) & \text{if} \quad i \neq j. \end{cases}\\
\bar{\tau}_{i,\text{dis}} &= \text{diag}\,(\mathbb{T}_i\tau_{\text{dis}})\\
\bar{\tau}_{i,\text{abs}} &= \text{diag}\,(\mathbb{T}_i\tau_{\text{abs}})
\end{aligned} \tag{3.45}$$

where ρ, λ, μ, χ, and η on the right hand sides in (3.45) 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

$$\begin{aligned}
\bar{Y}_{\text{dis}} &= F^{-1}\Big\{k^{y-1}F\{\cdot\}\Big\}\\
\bar{Y}_{\text{abs}} &= F^{-1}\Big\{k^{y-2}F\{\cdot\}\Big\}.
\end{aligned} \tag{3.46}$$

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

$$h(i,j) = \begin{cases} +1 & \text{if} \quad i = j \\ -1 & \text{if} \quad i \neq j. \end{cases} \tag{3.47}$$

3.4.1 Forward model.

In the sequel, the approximation of the system of viscoelastic wave equations defined by equations (3.9), (3.12) and (3.15) 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 (3.5), (3.12) and (3.15) using the k-space pseudo-spectral method [35, 38]. As discussed in section 3.2, 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 $t \in \big[-\Delta t/2, +\Delta t/2\big]$. For this, $p(t = 0) = 1$ is approximated as $P(n = \big[-1/2 \;\; +1/2\big]) = \frac{1}{\Delta t}\big[0.5 \;\; 0.5\big]$ (cf. [2], Appendix B). Considering this, together with (3.15) and dividing the source by PML directions, gives a source in the form

$${}_m s_{(ij;\zeta;n+1/2)} = \begin{cases} -\frac{\delta_{ij}}{2d\Delta t}\mathbb{S}P_0 & n = -1, 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.48}$$

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* ${}_m\sigma^{p,s}_{(ij;\zeta;n=-1)} = 0$ *and* ${}_m v_{(i;\zeta,n=-3/2)} = 0$, *and terminate at iterate* $n = N_t - 2$.

1. Update the particle velocity field:

$$\begin{aligned} {}_m v_{(i;\zeta;n+\frac{1}{2})} &= A_m \left[A_m \ {}_m v_{(i;\zeta;n-\frac{1}{2})} + \frac{\Delta t}{\bar{\rho}_i} \sum_{p,s} \frac{\partial_{p,s}}{\partial x_m^{h(i,m)}} \sigma^{p,s}_{(im;\zeta;n)} \right] \\ v_{(i;\zeta;n+\frac{1}{2})} &= \sum_{m=1}^{d} {}_m v_{(i;\zeta;n+\frac{1}{2})}. \end{aligned} \tag{3.49}$$

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

$$v^{p,s}_{(i;\zeta;n+\frac{1}{2})} = Q^{p,s} v_{(i;\zeta;n+\frac{1}{2})}. \tag{3.50}$$

Here, $Q^{p,s}$ denotes the discretised form of functions $q^{p,s}$ defined in (3.8).

3. Update the stress tensor field:

$$\begin{aligned} {}_m\sigma^{p,s}_{(ij;\zeta;n+1)} = A_m \Bigg[& A_m \ {}_m\sigma^{p,s}_{(ij;\zeta;n)} + \Delta t \ \Bigg[\bar{\lambda}\delta_{ij} \frac{\partial_{p,s}}{\partial x_m^-} v^{p,s}_{(m;\zeta;n+\frac{1}{2})} \\ & + \bar{\mu}_{ij} \left(\delta_{mj} \frac{\partial_{p,s}}{\partial x_j^{-h(i,j)}} v^{p,s}_{(i;\zeta;n+\frac{1}{2})} + \delta_{mi} \frac{\partial_{p,s}}{\partial x_i^{-h(i,j)}} v^{p,s}_{(j;\zeta;n+\frac{1}{2})} \right) \\ & + \bar{\chi}\delta_{ij} \frac{\partial_{p,s}}{\partial x_m^-} \left(\bar{\tau}^{p,s}_{m,\mathrm{dis}} \ \bar{Y}_{\mathrm{dis}} \ v^{p,s}_{(m;\zeta;n+\frac{1}{2})} - \bar{\tau}^{p,s}_{m,\mathrm{abs}} \ \bar{Y}_{\mathrm{abs}} \ \partial_t v^{p,s}_{(m;\zeta;n+1/2)} \right) \\ & + \bar{\eta}_{ij}\delta_{mj} \frac{\partial_{p,s}}{\partial x_j^{-h(i,j)}} \left(\bar{\tau}^{p,s}_{i,\mathrm{dis}} \ \bar{Y}_{\mathrm{dis}} \ v^{p,s}_{(i;\zeta;n+\frac{1}{2})} - \bar{\tau}^{p,s}_{i,\mathrm{abs}} \ \bar{Y}_{\mathrm{abs}} \ \partial_t v^{p,s}_{(i;\zeta;n+1/2)} \right) \\ & + \bar{\eta}_{ij}\delta_{mi} \frac{\partial_{p,s}}{\partial x_i^{-h(i,j)}} \left(\bar{\tau}^{p,s}_{j,\mathrm{dis}} \ \bar{Y}_{\mathrm{dis}} \ v^{p,s}_{(j;\zeta;n+\frac{1}{2})} - \bar{\tau}^{p,s}_{j,\mathrm{abs}} \ \bar{Y}_{\mathrm{abs}} \ \partial_t v^{p,s}_{(j;\zeta;n+1/2)} \right) \Bigg] \Bigg]. \end{aligned} \tag{3.51}$$

where

$$\partial_t v^{p,s}_{(i;\zeta;n+1/2)} = \sum_{m=1}^{d} A_m \frac{1}{\bar{\rho}_i} \frac{\partial_{p,s}}{\partial x_m^{h(i,m)}} \sigma^{p,s}_{(im;\zeta;n)} \tag{3.52}$$

4. Add source:

$$\begin{aligned} {}_m\sigma^{p}_{(ij;\zeta;n+1)} &\leftarrow {}_m\sigma^{p}_{(ij;\zeta;n+1)} + \Delta t \ {}_m s_{(ij;\zeta;n+1/2)} \\ \sigma^{p,s}_{(ij;\zeta;n+1)} &= \sum_{m=1}^{d} {}_m\sigma^{p,s}_{(ij;\zeta;n+1)}. \end{aligned} \tag{3.53}$$

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

$$p_{(\zeta;n+1)} = \mathbb{G}\sigma^{p}_{(ij;\zeta;n+1)} = -\frac{1}{d} \sum_{l,m=1}^{d} {}_m\sigma^{p}_{(ll;\zeta;n+1)}, \tag{3.54}$$

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

$$\hat{P}_{n+1} = \mathbb{M} p_{(\zeta;n+1)}, \tag{3.55}$$

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 $n = 0, ..., N_t - 1$.

3.4.2 Analytic adjoint model

The continuous adjoint model defined by equations (3.18), (3.19) and (3.20) 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 (3.48)), we define the order reversed adjoint measured data in the form [2]

$$\hat{P}^{\text{adj}}_{n+1/2} = \frac{1}{2\Delta t} \begin{cases} \hat{P}_{N_t-1}, & n = -1 \\ \hat{P}_{N_t-n-1} + \hat{P}_{N_t-n-2}, & n = 0, ..., N_t - 2 \\ \hat{P}_0, & n = N_t - 1 \end{cases}. \tag{3.56}$$

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

$${}_m s_{(ij;\zeta;n+1/2)} = \mathbb{G}^T \mathbb{M}^T \hat{P}^{\text{adj}}_{n+1/2}, \tag{3.57}$$

where

$$\mathbb{G}^T = -1_m \otimes \frac{\delta_{ij}}{d}. \tag{3.58}$$

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:

$$
\begin{aligned}
{}_m v_{(i;\zeta;n+\frac{1}{2})} = A_m \Bigg[& A_m \ {}_m v_{(i;\zeta;n-\frac{1}{2})} \\
& + \frac{\Delta t}{\bar{\rho}_i} \Bigg[\sum_{p,s} \sum_{j=1}^{d} Q^{p,s} \Big[\frac{\partial_{p,s}}{\partial x_i^+} \big(\bar{\lambda} \ {}_i\sigma^{p,s}_{(jj;\zeta;n)} \big) + 2 \frac{\partial_{p,s}}{\partial x_j^{h(i,j)}} \big(\bar{\mu}_{ij} \ {}_j\sigma^{p,s}_{(ij;\zeta;n)} \big) \\
& + \bar{Y}_{\text{dis}} \ \bar{\tau}^{p,s}_{i,\text{dis}} \left(\frac{\partial_{p,s}}{\partial x_i^+} \big(\bar{\chi} \ {}_i\sigma^{p,s}_{(jj;\zeta;n)} \big) + 2 \frac{\partial_{p,s}}{\partial x_j^{h(i,j)}} \big(\bar{\eta}_{ij} \ {}_j\sigma^{p,s}_{(ij;\zeta;n)} \big) \right) \Big] \Bigg] \Bigg]
\end{aligned}
\tag{3.59}
$$

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

$$
\begin{aligned}
{}_m \mathbf{v}^{p,s}_{(i;\zeta;n+\frac{1}{2})} = & {}_m \ v_{(i;\zeta;n+\frac{1}{2})} \\
& - A_m \frac{1}{\bar{\rho}_i} \sum_{j=1}^{d} \bar{Y}_{\text{abs}} \ \bar{\tau}^{p,s}_{i,\text{abs}} \left(\frac{\partial_{p,s}}{\partial x_i^+} \big(\bar{\chi} \ {}_i\sigma^{p,s}_{(jj;\zeta;n)} \big) + 2 \frac{\partial_{p,s}}{\partial x_j^{h(i,j)}} \big(\bar{\eta}_{ij} \ {}_j\sigma^{p,s}_{(ij;\zeta;n)} \big) \right)
\end{aligned}
\tag{3.60}
$$

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

3. Update the stress tensor field:

$$
{}_m\sigma^{p,s}_{(ij;\zeta;n+1)} = A_m \left[A_m \ {}_m\sigma^{p,s}_{(ij;\zeta;n)} + \frac{\Delta t}{2} \Big(\frac{\partial_{p,s}}{\partial x_j^{-h(i,j)}} {}_j\mathbf{v}^{p,s}_{(i;\zeta;n+\frac{1}{2})} + \frac{\partial_{p,s}}{\partial x_i^{-h(i,j)}} {}_i\mathbf{v}^{p,s}_{(j;\zeta;n+\frac{1}{2})} \Big) \right]
\tag{3.61}
$$

4. Add source:

$$
{}_m\sigma^{p}_{(ij;\zeta;n+1)} \leftarrow {}_m\sigma^{p}_{(ij;\zeta;n+1)} + \Delta t \ {}_m s_{(ij;\zeta;n+1/2)}
\tag{3.62}
$$

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

$$
p_{(\zeta;n=N_t-1)} = -\mathbb{S} \left(\frac{1}{d} \sum_{l,m=1}^{d} {}_m\sigma^{p}_{(ll;\zeta;n=N_t-1)} \right).
\tag{3.63}
$$

3.5 Adjoint for discretised viscoelastic forward model

In this section, we will calculate the adjoint of the viscoelastic forward model Λ defined by (3.9), (3.12) and initial conditions in (3.15) based on the *discretise-then-adjoint* method. To do this, we consider the discretised equations (3.49) and (3.51) 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 ${}_m v_{(i;\zeta;n-1/2)} \in \mathbb{R}^N$ $(i, m \in \{1, ..., d\})$. 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 ${}_m\sigma^{p,s}_{(ij;\zeta,n)} \in \mathbb{R}^N$ $(i,j,m \in \{1,...,d\})$. 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 $X_n \in R^{39N}$ (3D case) as a stack of the particle velocity and stress fields at time step n in the form $X_n = \left[\left(\bar{v}_{n-1/2}\right)^T \ \left(\bar{\sigma}_n\right)^T \right]^T$. Let also $\mathcal{S} \in \mathbb{R}^{39NN_t \times N}$ give a map from the discretised initial pressure $P_0 \in \mathbb{R}^N$ to an additive source (cf. (3.48)), which we will write as

$$S = \mathcal{S}P_0 \in \mathbb{R}^{39NN_t}. \tag{3.64}$$

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 $(n \in \{-1,...,N_t-2\})$ is then given by

$$X_{n+1} = TX_n + S_{n+1/2}, \tag{3.65}$$

where $T \in \mathbb{R}^{39N \times 39N}$ implements (3.49) and (3.51), and $X_{-1} = 0$ (cf. section 3.4.1). Here for brevity the operators $\mathcal{S}_n$ are given using (3.48) multiplied by Δt, and thus multiplication by Δt is neglected in the second term of (3.65). We will look in more detail at the matrix T later in section 3.5.1. 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 (3.54) and (3.55)). 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 3.2. The map $\mathbb{H} : \mathbb{R}^{39NN_t} \to \mathbb{R}^{N_s N_t}$ is defined by

$$\hat{P} = \mathbb{H}S, \quad \hat{P}_n = \mathcal{M}X_n \quad (n \in \{0,...,N_t-1\}), \quad \hat{P} = \left[\hat{P}_{n+1}\right]_{n=-1}^{N_t-2}, \tag{3.66}$$

where X_n is updated by (3.65) 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 $n \in \{0,...,N_t-1\}$.

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 3.2. The adjoint $\mathbb{H}^*$ of $\mathbb{H}$ defined in definition 3.2 is given by

$$\left[X_n^*\right]_{n=-1}^{N_t-2} = \mathbb{H}^*\hat{P} \in \mathbb{R}^{39NN_t}, \quad X_{N_t-1}^* = 0, \quad X_{n-1}^* = T^*X_n^* + \mathcal{M}^*\hat{P}_n \quad (n \in \{0, ..., N_t - 1\}). \tag{3.67}$$

Proof. Let us assume that X_n satisfies (3.65) with initial condition $X_{-1} = 0$, and X_n^* satisfies the last two equations in (3.67). 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 (3.65) on the left and (3.67) 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 $\langle S, \mathbb{H}^*\hat{P}\rangle_{\mathbb{R}^{NN_t}} = \langle \mathbb{H}S, \hat{P}\rangle_{\mathbb{R}^{N_s N_t}}$, and so completes the proof.

□

The forward map actually defined in section 3.4.1 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 3.2, 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 (3.68) is actually just two terms which can also be used to explain (3.56) if we commute the summing operation with the computation of X^*.

Corollary 3.1. $\mathcal{S}^*\mathbb{H}^*$ can be computed as

$$\mathcal{S}^*\mathbb{H}^*\hat{P} = \sum_{n=-1}^{N_t-2} \mathcal{S}_n^* X_{N_t-2-n}^* \tag{3.68}$$

where X_n^* is determined by

$$X_{-1}^* = 0, \quad X_{n+1}^* = T^* X_n^* + \mathcal{M}^* \hat{P}_{N_t-2-n} \quad (n \in \{-1,\ \dots\ , N_t - 2\}). \tag{3.69}$$

3.5.1 The Matrices T and T^*

In this section we write the matrices T and T^* explicitly using the forward model presented in section 3.4.1 to show how multiplication by each of them may be computed. Considering corollary 3.1, 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 (3.49) and (3.51) in the condensed forms

$$\bar{v}_{n+\frac{1}{2}} = A_v\big[A_v\ \bar{v}_{n-\frac{1}{2}} + \Phi\bar{\sigma}_n\big], \tag{3.70}$$

and

$$\bar{\sigma}_{n+1} = A_\sigma\big[A_\sigma\bar{\sigma}_n + \Psi_{\text{dis}}\bar{v}_{n+\frac{1}{2}} - \Psi_{\text{abs}}\bar{\sigma}_n\big] + \theta S_{n+1/2} \tag{3.71}$$

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

$$T = \begin{pmatrix} I_v & 0 \\ A_\sigma\Psi_{\text{dis}} & A_\sigma(A_\sigma - \Psi_{\text{abs}}) \end{pmatrix} \begin{pmatrix} A_v^2 & A_v\Phi \\ 0 & I_\sigma \end{pmatrix} \tag{3.72}$$

where I_v and I_σ are the identity matrices. From (3.72) we have

$$T^* = \begin{pmatrix} A_v^2 & 0 \\ \Phi^* A_v & I_\sigma \end{pmatrix} \begin{pmatrix} I_v & \Psi_{\text{dis}}^* A_\sigma \\ 0 & (A_\sigma - \Psi_{\text{abs}}^*)A_\sigma \end{pmatrix}. \tag{3.73}$$

Using Corollary 3.1, the above equation gives the updates for the adjoint problem as

$$\begin{aligned} \bar{v}_{n+1/2} &= A_v^2\left(\bar{v}_{n-1/2} + \Psi_{\text{dis}}^* A_\sigma \bar{\sigma}_n\right) \\ \bar{\sigma}_{n+1} &= A_\sigma^2\bar{\sigma}_n + \Phi^*(A_v)^{-1}\bar{v}_{n+1/2} - \Psi_{\text{abs}}^* A_\sigma\bar{\sigma}_n + \theta'^*\mathcal{M}^*\bar{P}_{n+1/2}^{\text{adj}}, \end{aligned} \tag{3.74}$$

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

$$\begin{aligned} \tilde{v}_{n+1/2} &= A_v \left(A_v \tilde{v}_{n-1/2} + \Psi^*_{\text{dis}} \tilde{\sigma}_n \right) \\ \tilde{\sigma}_{n+1} &= A_\sigma \left(A_\sigma \tilde{\sigma}_n + \Phi^* \tilde{v}_{n+1/2} - \Psi^*_{\text{abs}} \tilde{\sigma}_n \right) + \theta'^* \mathcal{M}^* \bar{P}^{\text{adj}}_{n+1/2}. \end{aligned} \tag{3.75}$$

Now let us consider the matrices $A_v 3$, A_σ, Φ, Ψ_{dis}, and Ψ_{abs}. First we note that A_v and A_σ can be found from (3.37). Though the others can be read off from (3.49) and (3.51), we will write them down explicitly here in order to show how we can explicitly calculate multiplication by their adjoints as required in (3.73).

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

$$\nabla^{\pm}_{(m,i)}{}^{p,s} = \frac{\partial_{p,s}}{\partial x_i^{\pm h(m,i)}}, \tag{3.76}$$

which is defined by (3.32). Note that the superscript p, s accounts for $C_0^{p,s}$ used in the k-space method (see (3.32)). Here we are considering $\nabla^{\pm}_{(m,i)}{}^{p,s}$ to be an $N \times N$ matrix, and based on (3.32) and the unitarity of the discrete Fourier transform we see that

$$\left(\nabla^{\pm}_{(m,i)}{}^{p,s} \right)^* = -\nabla^{\mp}_{(m,i)}{}^{p,s}. \tag{3.77}$$

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

$$\mathbb{S}[a]_{ij} = \frac{a_{ij} + a_{ji}}{2}. \tag{3.78}$$

We apply $\mathbb{S}$ to objects having more indices, but specify that it always acts on the pair ij.

In (3.70), the action of matrix $\Phi \in \mathbb{R}^{9N \times 30N}$ on $\bar{\sigma}$ can be written as

$${}_m(\Phi \bar{\sigma})_i = \sum_{p,s} {}_m(\Phi' \bar{\sigma})_i^{p,s}, \tag{3.79}$$

where from (3.49) and the second line in (3.53),

$${}_m(\Phi' \bar{\sigma})_i^{p,s} = \frac{\Delta t}{\bar{\rho}_i} \left(\nabla^{+}_{(i,m)}{}^{p,s} \right) \sum_{m'=1}^{d} {}_{m'} \bar{\sigma}_{im}^{p,s}. \tag{3.80}$$

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 (3.75) and using (3.77),

$$ {}_m(\Phi^* \tilde{v})^{p,s}_{ij} = -\Delta t \, \mathbb{S} \left[\left(\nabla^-_{(i,j)} {}^{p,s} \right) \frac{1}{\bar{\rho}_i} {}_j \tilde{v}_i \right]. \tag{3.81} $$

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

Next, from (3.51), together with the second line in (3.49), the operator $\Psi_{\text{dis}} \in \mathbb{R}^{30N \times 9N}$ acts on $\bar{v}$ by

$$ \begin{aligned} {}_m(\Psi_{\text{dis}} \bar{v})^{p,s}_{ij} = \Delta t \Bigg[& \delta_{ij} \left[\bar{\lambda} \left(\nabla^-_{(m,m)} {}^{p,s} \right) + \bar{\chi} \left(\nabla^-_{(m,m)} {}^{p,s} \right) \bar{\tau}^{p,s}_{m,\text{dis}} \bar{Y}_{\text{dis}} \right] Q^{p,s} \sum_{m'=1}^{d} {}_{m'} \bar{v}_m \\ & + 2 \, \mathbb{S} \left[\left[\bar{\mu}_{ij} \, \delta_{mj} \, \left(\nabla^-_{(i,j)} {}^{p,s} \right) + \bar{\eta}_{ij} \, \delta_{mj} \left(\nabla^-_{(i,j)} {}^{p,s} \right) \bar{\tau}^{p,s}_{i,\text{dis}} \, \bar{Y}_{\text{dis}} \right] Q^{p,s} \sum_{m'=1}^{d} {}_{m'} \bar{v}_i \right] \Bigg] \end{aligned} \tag{3.82} $$

From this we can find the formula for the action of the adjoint in (3.75) as

$$ \begin{aligned} {}_m(\Psi^*_{\text{dis}} \tilde{\sigma})_i = -\Delta t \sum_{p,s} \sum_{j=1}^{d} Q^{p,s} \Bigg[& \left[\left(\nabla^+_{(i,i)} {}^{p,s} \right) \bar{\lambda} + \bar{Y}_{\text{dis}} \bar{\tau}^{p,s}_{i,\text{dis}} \left(\nabla^+_{(i,i)} {}^{p,s} \right) \bar{\chi} \right] {}_i \tilde{\sigma}^{p,s}_{jj} \\ & + 2 \left[\left(\nabla^+_{(i,j)} {}^{p,s} \right) \bar{\mu}_{ij} + \bar{Y}_{\text{dis}} \bar{\tau}^{p,s}_{i,\text{dis}} \left(\nabla^+_{(i,j)} {}^{p,s} \right) \bar{\eta}_{ij} \right] {}_j \tilde{\sigma}^{p,s}_{ij} \Bigg]. \end{aligned} \tag{3.83} $$

Additionally, from (3.51), (3.52) and (3.80), in (3.71) the operator $\Psi_{\text{abs}} \in \mathbb{R}^{30N \times 30N}$ acts on $\bar{\sigma}$ by

$$ {}_m(\Psi_{\text{abs}} \bar{\sigma})^{p,s}_{ij} = {}_m \left(\Psi'_{\text{abs}} (\frac{1}{\Delta t}) \sum_{m'=1}^{d} A_{m'm'} (\Phi' \bar{\sigma}) \right)^{p,s}_{ij}, \tag{3.84} $$

which is actually the action of Ψ'_{abs} on $\partial_t \bar{v}$ using (3.51) in the form

$$ \begin{aligned} {}_m(\Psi'_{\text{abs}} \partial_t \bar{v})^{p,s}_{ij} = \Delta t \Bigg[& \delta_{ij} \bar{\chi} \left(\nabla^-_{(m,m)} {}^{p,s} \right) \bar{\tau}^{p,s}_{m,\text{abs}} \bar{Y}_{\text{abs}} \partial_t \, \bar{v}^{p,s}_m \\ & + 2 \, \mathbb{S} \, \bar{\eta}_{ij} \, \delta_{mj} \left(\nabla^-_{(i,j)} {}^{p,s} \right) \bar{\tau}^{p,s}_{i,\text{abs}} \, \bar{Y}_{\text{abs}} \partial_t \, \bar{v}^{p,s}_i \Bigg] \end{aligned} \tag{3.85} $$

Using (3.75), the action of the adjoint is then given by

$$ {}_m(\Psi^*_{\text{abs}} \tilde{\sigma})^{p,s}_{ij} = {}_m \, (\frac{1}{\Delta t} \Phi'^* A_m \Psi'^*_{\text{abs}} \tilde{\sigma})^{p,s}_{ij}, \tag{3.86} $$

where

$$ {}_m(\Psi'^{*}_{\text{abs}}\tilde{\sigma})^{p,s}_i = -\Delta t \sum_{j=1}^{d} \bar{Y}_{\text{abs}} \bar{\tau}^{p,s}_{i,\text{abs}} \left[\left(\nabla^{+}_{(i,i)}{}^{p,s} \right) \bar{\chi}_{\ i} \tilde{\sigma}^{p,s}_{jj} + 2 \left(\nabla^{+}_{(i,j)}{}^{p,s} \right) \bar{\eta}_{ij\ j} \tilde{\sigma}^{p,s}_{ij} \right]. \quad (3.87) $$

Finally, plugging (3.86) into the second line in (3.75), together with using (3.79), gives

$$ \begin{aligned} \tilde{v}_{n+1/2} &= A_v \left(A_v \tilde{v}_{n-1/2} + \Psi^{*}_{\text{dis}} \tilde{\sigma}_n \right) \\ \tilde{\mathbf{v}}_{n+1/2} &= 1_{p,s} \otimes \tilde{v}_{n+1/2} - \frac{1}{\Delta t} A_v \left(\Psi'^{*}_{\text{abs}} \tilde{\sigma}_n \right)^{p,s} \\ \tilde{\sigma}_{n+1} &= A_\sigma \left(A_\sigma \tilde{\sigma}_n + \Phi'^{*} \tilde{\mathbf{v}}_{n+1/2} \right) + \theta'^{*} \mathcal{M}^{*} \bar{P}^{\text{adj}}_{n+1/2}, \end{aligned} \quad (3.88) $$

where, $1_{p,s}\otimes$ 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 (3.88) gives the same formulae as in section 3.4.2. 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.

3.6 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

$$ \underset{P}{\operatorname{argmin}} \left\{ F(P) := f(P) + g(P) \right\}, \quad (3.89) $$

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 $L_f = \mathcal{S}_{i\max}(\bar{\mathbb{H}}^*\bar{\mathbb{H}})$ with $\mathcal{S}_{i\max}(.)$ the largest singular value. The gradient of f is computed by

$$ \nabla f(P) = \bar{\mathbb{H}}^* \left(\bar{\mathbb{H}} P - \hat{P} \right). \quad (3.90) $$

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 $C = \{P \geqslant 0\}$ [7, 3].

Applying the so-called forward-backward splitting method to a fixed point iterative scheme arising from the optimality conditions of problem (3.89) 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

$$Y^k = P^{k-1} - \Gamma^k \, \nabla f(P^{k-1}), \tag{3.91}$$

and is called the forward gradient descent step [7, 3]. Applying ISTA, the iterates P^k converge to a minimiser P^* of problem (3.89) 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

$$\text{prox}_{\Gamma^k}(g)(Y^k) := \underset{P}{\text{argmin}} \left\{ g(P) + \frac{1}{2\Gamma^k} \left\| P - Y^k \right\|^2 \right\}, \tag{3.92}$$

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:

$$k > 1 \ \cap \ 1 - \frac{F^k}{F^{k-1}} < \epsilon. \tag{3.93}$$

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

3.7 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 (3.4) 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 3.4.1 [37]. We also developed

a code for implementation of our continuous adjoint, as discussed in section 3.4.2. We then showed that this adjoint matches the algebraic adjoint of the associated discretised forward operator. To validate the adjoint model, (3.21) 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.

3.7.1 2D phantom

Computational grid

we used a computational grid with a size of 14×14 cm^2 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 $r = 6.8$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.

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 3.1(a), 3.1(b) and 3.1(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 $c_p = 1500$ ms^{-1}, $c_s = 0$ ms^{-1} and $\rho = 1000$ kgm^{-3}, and the red colour represents the skull with $c_p = 3000$ ms^{-1}, $c_s = 1500$ ms^{-1} and $\rho = 1850$ kgm^{-3}. These parameters were chosen following [25]. The absorption coefficients were set to $\alpha_{0,p} = 10$ dBMHz^{-y}cm^{-1} and $\alpha_{0,s} = 20$ dBMHz^{-y}cm^{-1} in the skull, and $\alpha_{0,p} = 0.75$ dBMHz^{-y}cm^{-1} and $\alpha_{0,s} = 0.5$ dBMHz^{-y}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 $y = 1.4$.

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 (3.21). 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 (3.21) 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.

Simulation setting

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

Scenario1: In general, the inverse problem in PAT is based on the assumption

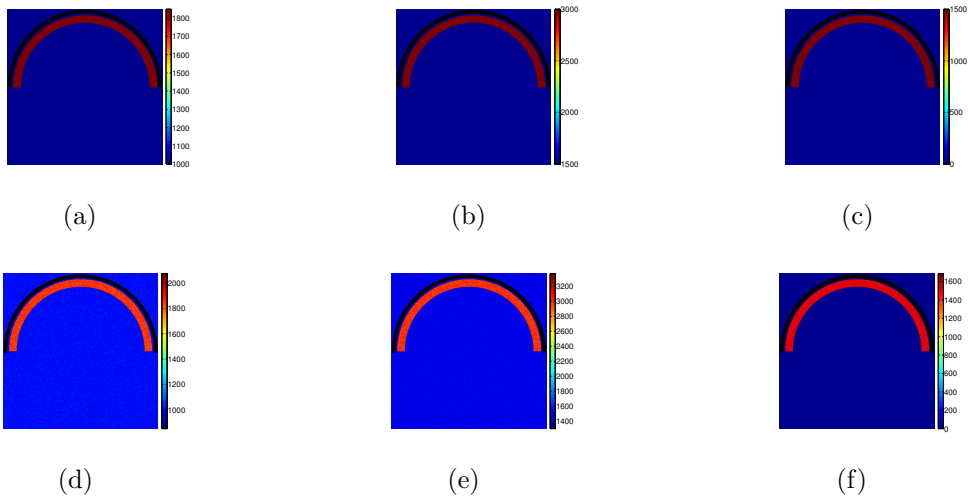
that the physical parameters of the medium are known. In our first experiment, we used the maps in figures 3.1(a), 3.1(b) and 3.1(c) as physical parameters for both data generation and image reconstruction. This implies that we have an exact knowledge of the physical parameters. Since this assumption 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 3.1(a), 3.1(b) and 3.1(c) with a 30dB Additive White Gaussian Noise (AWGN). The contaminated maps are displayed in figures 3.1(d), 3.1(e) and 3.1(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 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 3.2(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.

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



(a) (b) (c)

(d) (e) (f)

Figure 3.1: 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.

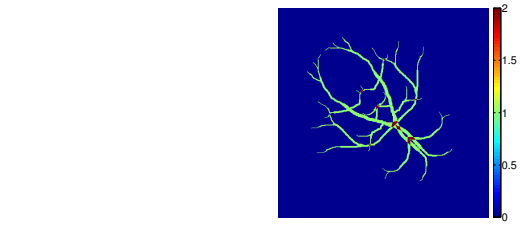
TR and ISTA, we applied TR optimistically on a non-absorbing medium with $\alpha_{0,p,s} = 0$. The images reconstructed by TR for scenarios 1 and 2 are displayed in figures 3.2(b) and 3.2(d), respectively. The computed forward and adjoint operators were then incorporated into the inverse solver discussed in section 3.6. 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 3.2(c) and 3.2(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:

(1) Relative Error (RE):

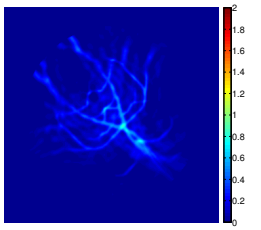
$$RE(P^k) = \frac{\left\| P^k - \tilde{P}_{\text{phantom}} \right\|_2}{\left\| \tilde{P}_{\text{phantom}} \right\|_2} \times 100, \tag{3.94}$$

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.

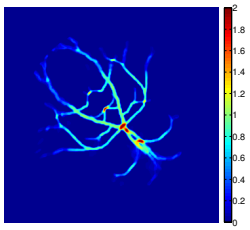
(2) Objective function ($F(P^k)$): (cf. section 3.6, (3.89)).



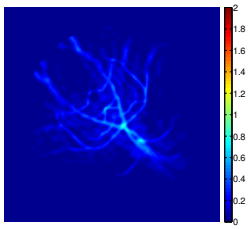
(a)



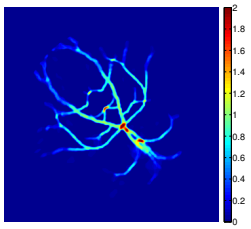
(b)



(c)



(d)



(e)

Figure 3.2: 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.

Observations

Figures 3.3(a) and 3.3(b) show RE and objective function values of iterates computed by ISTA versus the iteration number k, respectively. Figure 3.3(c) shows F from a large view 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 3.2(c). From figures 3.3(b) and 3.3(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 3.3(a) and 3.3(b), RE and F reached values of 44.29% and 93.52 at the final iteration, which corresponds to the image shown in figure 3.2(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.

3.7.2 3D phantom

Computational grid

This grid was created as a rectangular cuboid with a size of $14 \times 14 \times 3.5$ 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 8.75×10^{-2} 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

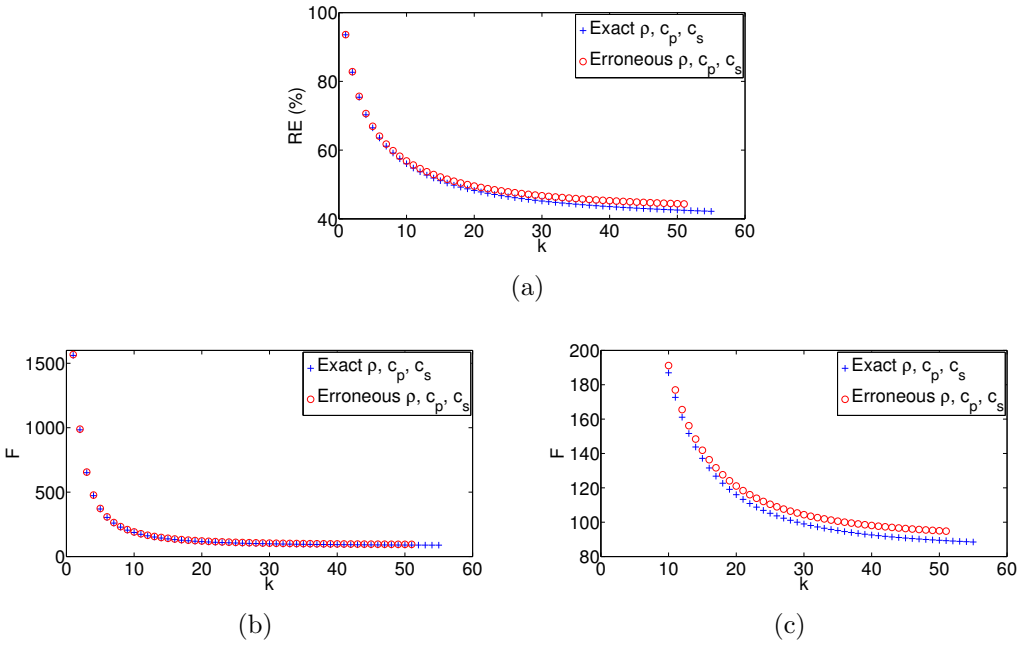


Figure 3.3: 2D phantom. (a) Relative Error (RE) (b) objective function (F) (c) F around the stopping point.

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.

Physical parameters

Figures 3.4(a), 3.4(b) and 3.4(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.

Validation of adjoint

We used an inner-product test based on (3.21) 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 (3.21) among 10 attempts was 2.21×10^{-4} and 3.47×10^{-5} for the grids used for image reconstruction and data generation, respectively.

Simulation setting

For image reconstruction, two scenarios were considered:
Scenario 1: The maps that are displayed in figures 3.4(a), 3.4(b) and 3.4(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 8.7514×10^5Hz 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 7.2018×10^5Hz and 7.4143×10^5Hz 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 6.8571×10^5Hz 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 3.5(a) shows the simulated phantom from a top view. Here, all 3D images

(a)

(b)

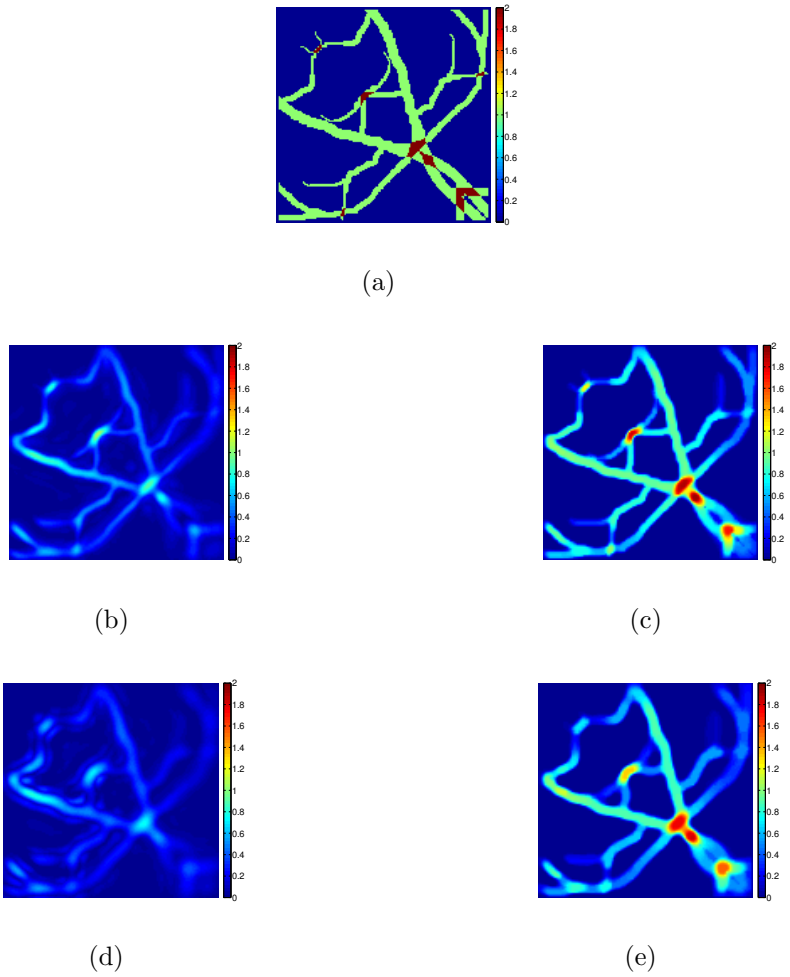
(c)

Figure 3.4: 3D phantom. Exact physical maps: (a) ρ (b) c_p (b) c_s.

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.

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 3.5(b) and 3.5(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 3.5(c) and 3.5(e) show an image of the final iterate



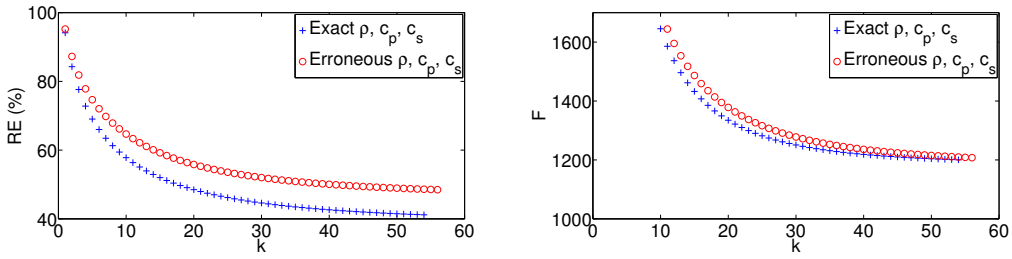
(a) (b) (c) (d) (e)

Figure 3.5: 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.

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.

Observations

Figure 3.6(a) shows the RE of the computed iterates versus iteration number. Additionally, figure 3.6(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.



(a) (b)

Figure 3.6: 3D phantom. (a) Relative Error (RE) (b) a large view of objective function F around the stopping point.

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 3.5(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 3.5(e), has an RE of 48.44% and an F of 1.20×10^3.

3.8 Discussion and conclusion

In this work, we derived the adjoint of the continuous map defined in (3.14) and (3.15), 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 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 (3.64) and (3.66).

From a numerical point of view, it was shown that this forward and adjoint pair satisfies the inner-product formula in (3.21). 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]. 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 3.2(e). However, for the 3D scenario, figure 3.5(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 3.2(c) and 3.2(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).

3.9 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 (3.19) into

(3.18)

$$\rho\frac{\partial \mathbf{v}_i^{p,s*}}{\partial t} = \sum_{p,s} q^{p,s}\left(\frac{\partial}{\partial x_i}\left(\lambda\sigma_{ll}^{p,s*}\right) + 2\frac{\partial}{\partial x_j}\left(\mu\sigma_{ij}^{p,s*}\right)\right) + \left(\sum_{p,s} q^{p,s}\sin(\pi y/2)L_{c_{p,s}}^{(y-1)*} - \cos(\pi y/2)L_{c_{p,s}}^{(y-2)*}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x_i}\left(\chi\sigma_{ll}^{p,s*}\right) + 2\frac{\partial}{\partial x_j}\left(\eta\sigma_{ij}^{p,s*}\right)\right) \tag{3.95}$$

Discretised adjoint: In (3.88), plugging the second line into the first line yields

$$\tilde{\mathbf{v}}_{n+1/2} = A_v\left(A_v\tilde{\mathbf{v}}_{n-1/2} + \Psi_{\text{dis}}^*\tilde{\sigma}_n - \frac{1}{\Delta t}\left(\left(\Psi_{\text{abs}}'^*\tilde{\sigma}_n\right)^{p,s} - A_v^2\left(\Psi_{\text{abs}}'^*\tilde{\sigma}_{n-1}\right)^{p,s}\right)\right). \tag{3.96}$$

The numerical computation of (3.95) is the same as (3.96), 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 (3.18) and (3.19) (resp. (3.88)) for the continuous (resp. discretised) adjoint, which are computed the same, as discussed in sections 3.4 and 3.5, respectively.

3.A Supplementary Materials

In this chapter, we considered the inverse problem of PAT for a reconstruction of the initial pressure distribution inside the skull from a set of time series of measured data that is measured outide the skull. To account for generation and propagation of shear waves within the skull, the forward operator was defined as a linear system of viso-elastic equations. We considered an isotropic heterogeneous lossy and elastic medium with an acoustic attenuation following a frequency power law.

Variational approaches often use an iterative implementation of the forward operator and its adjoint. We derived an adjoint using an *adjoint-then-discretise* approach (continuous adjoint). We used a k-space pseudospectral method for a discretisation of the forward operator and its adjoint. We showed that a discretisation of the continuous adjoint operator using a k-space pseudo-spectral method matches an algebraic adjoint of an associated discretised forward operator. Therefore, the numerical results in chapter 3 can be considered as derived by both *adjoint-then-discretise* and *discretise-then-adjoint* approaches.

3.A.1 Further details about the measurement operator

Continuous forward operator $\mathcal{M}$

From a physical point of view, the detectors have a limited access to the pressure field. They have access to the pressure field $p(x,t)$ only within a sequence of short intervals of times $t'_n \subset (0,T)$ ($n \in \{0,...,N_t-1\}$), which corresponds to the temporal response of detectors, and a small and finite volume of space $\Gamma \in \mathbb{R}^d$, which is a union of small volumes Γ_s around the position of detectors. We denote the position of the detector D_s by x_s ($s \in \{1,...,N_s\}$). Because of the temporal response of detectors, the detector measures a temporal average of $p(x,t)$ over the time periods t'_n. In this chapter, we denote the pressure field $p(x,t)$ that is temporally averaged over time periods $t_{n'}$ by $p(x,t_n)$.

Using this, we will provide further details about the measurement operator $\mathcal{M} : \mathcal{L}^2(\mathbb{R}^d) \to \mathbb{R}^{N_s}$, which at each time step n maps the pressure field $p(x,t_n) \in \mathcal{L}^2(\mathbb{R}^d)$ to $\hat{P}_n \in \mathbb{R}^{N_s}$, a vector of discretised measurements detected at N_s detectors at time step n. We define this operator in the form

$$\begin{aligned} &\mathcal{M} : \mathcal{L}^2\left(\mathbb{R}^d\right) \to \mathbb{R}^{N_s} \\ &\hat{P}_{s,n} = \left(\mathcal{M}[p]\right)_{s,n} = \int_{x\in\Gamma} b(x_s,x)\ p(x,t_n)dx. \end{aligned} \tag{3.97}$$

Here, the operator $\mathcal{M}$ is a weighted averaging operator, and averages the pressure field over the set Γ using the weighting coefficients $b(x_s,x)$. Also, $\hat{P}_{s,n}$ is a component of the vector $\hat{P}_n$ that is measured by the detector D_s (cf. (3.55)).

The weighting coefficients $b(x_s,x)$ are dependent on the method we use for modelling the detectors. As an example, if we simply average the pressure field over the set Γ_s, these coefficients are defined as

$$b(x_s,x) = \frac{1}{m(\Gamma_s)}\chi_{\Gamma_s}(x), \tag{3.98}$$

where,

$$\chi_{\Gamma_s}(x) = \begin{cases} 1, & x \in \Gamma_s \\ 0, & x \notin \Gamma_s. \end{cases} \tag{3.99}$$

Also, $m(\Gamma_s)$ is defined as

$$m(\Gamma_s) = \int_{\Gamma_s} dx. \tag{3.100}$$

Continuous adjoint operator $\mathcal{M}^*$

Having defined our continuous measurement operator $\mathcal{M}$, we now derive the adjoint of this operator. Using (3.97), together with (3.98), our adjoint measurement operator will be in the form

$$\begin{aligned} &\mathcal{M}^* : \mathbb{R}^{N_s} \to \mathcal{L}^2(\mathbb{R}^d) \\ &\mathcal{M}^*[\hat{P}_n] = \sum_{s=1}^{N_s} \frac{1}{m(\Gamma_s)} \chi_{\Gamma_s}(x) \hat{P}_{s,n}. \end{aligned} \tag{3.101}$$

The second line in (3.101) can be expressed as a *backprojection* of $\hat{P}_{n,s}$ using the weighting coefficient $\bar{b}(x_s, x)$ and a summation of the backprojected values over all detectors.

Discretised forward operator $\mathbb{M}$

Here, we explain the method we used for a numerical approximation of the measurement operator $\mathcal{M}$. As explained in step 5 in section 3.4.1, we used a trilinear interpolation method for an approximation of $\mathcal{M}$. We remind that we have used $P_{(\zeta;n)} \in \mathbb{R}^N$ for representing a discretisation of $p(x,t)$ (cf. section 3.4). Here, we also emphasise that throughout this thesis, following [2, 1], we assumed the geometry of the detectors as points. It is also worth mentioning that for a numerical approximation of $\mathcal{M}$, we used the k-Wave toolbox [38].

We start with a representation of the set Γ for our discretised operator. To do this, we choose a portion of the grid points that encompass small volumes in a neighborhood of the position x_s of the detectors. Using a trilinear interpolation, which we have used in this chapter, a triangulation is enforced on the chosen grid points. In a next step, for each detector, we choose a triangle (tetrahedron) that encompasses the point x_s, and denote it by T_s. We denote the vertices of T_s by x_l^s ($l \in \{1, .., d+1\}$). Here, we refer to a triangle (tetrahedron) as *voxel*.

Having defined a discretised representation for the set Γ, we now approximate the operator $\mathcal{M}$ (cf. equation (3.97)). If we assume that the volume of the set Γ_s is much smaller than the volume of the voxel T_s, $(\mathcal{M}[p])_{s,n}$ can be defined as an averaging operator over the set Γ_s in the form

$$(\mathcal{M}[p])_{s,n} = \frac{1}{m(\Gamma_s)} \int_{\Gamma_s} p(x, t_n) dx \approx p(x_s, t_n). \tag{3.102}$$

Here, we also assumed that $p(x, t_n)$ is not varying rapidly over each voxel. Using (3.102), it makes sense to approximate $(\mathcal{M}[p])_{s,n}$ by an interpolation from the grid points to the detector position x_s.

For an interpolation, we numerically calculated the Barycentric coordinates of x_s with respect to the vertices x_l^s, which correspond to the *voxel* T_s. We denote the Barycentric coordinate for the vertex l ($l = \{1, ..., d\}$) of the *voxel* T_s by b_l^s. Using the above definitions, the pressure measured by the detector D_s (at point x_s located inside the *voxel* T_s) at time step n is approximated using a linear and convex combination of the pressure values at the corresponding vertices x_l^s in the form

$$\hat{P}_{s,n} = \sum_{l=1}^{d+1} b_l^s P_{(\zeta_l^s;n)}, \tag{3.103}$$

where $b_l^s \geqslant 0$ and $\sum_{l=1}^{d+1} b_l^s = 1$. Using the above steps, we now define our discretised measurement operator $\mathbb{M}$, which has been defined in step 5 in section 3.4, and satisfies $\hat{P}_n = \mathbb{M} P_{(\zeta;n)}$ (cf. (3.55)).

Here, the operator $\mathbb{M}$ is a sparse matrix, and each row $\mathbb{M}_s \in \mathbb{R}^{1\times N}$ includes $d+1$ nonzero values, which are b_l^s ($l \in \{1, ..., d+1\}$), and are located at the columns corresponding to the indices of x_l^s in the computational grid.

Discretisation of the continuous measurement operator $\mathcal{M}^*$

Using an *adjoint-then-discretise* approach for a discretisation of $\mathcal{M}^*$, because the adjoint map should map to the dual space $\mathcal{L}^2(\mathbb{R}^d)^*$, it makes sense to discretise (3.101) in the same way as the forward measurement operator (3.97), which can also be considered as an element of $\mathcal{L}^2(\mathbb{R}^d)^*$. Using this fact, a discretisation of $\mathcal{M}^*$ gives the same formula as an algebraic adjoint $\mathbb{M}^*$, which is equivalent to using a *discretise-then-adjoint* approach. Therefore, in this chapter, we used $\mathbb{M}^*$ as an adjoint for the measurement operator using both *discretise-then-adjoint* and *adjoint-then-discretise* approaches.

Using a discretisation of our derived measurement operator ($\mathbb{M}$) and its adjoint ($\mathbb{M}^*$), in the next section, we will give further details about our algebaric adjoint acoustic operator, and how it gives the same formulae as a discretisation of our continuous adjoint acoustic operator.

3.A.2 Further details about our acoustic discretised adjoint

In the sequel, we will give further details about the formulae that have been derived for the algebraic adjoint (equation (3.88)), and show that this gives the same formulae as in section 3.4.2, i.e., the numercal computation of the continuous adjoint operator. We start with the first line in (3.88). Let us define the components of vectors $\tilde{v}_{n-1/2}$ and $\tilde{\sigma}_n$ in the form ${}_m\tilde{v}_{(i;\zeta;n-1/2)} \in \mathbb{R}^N$ and ${}_m\tilde{\sigma}_{(i;\zeta;n)} \in \mathbb{R}^N$, where $i, m \in \{1, ..., d\}$, ζ and n have been defined in section (3.4). Note that as opposed to components of vector $\bar{v}_{n-1/2}$ and $\bar{\sigma}_n$ (cf. the beginning of section 3.5), we have used the same notation for these vector and their components.

Considering the discretised k-space gradient operator in (3.76), together with equation (3.83), the first line of (3.88) yields

$$\begin{aligned} {}_m\tilde{v}_{(i;\zeta;n+\frac{1}{2})} = A_m \Bigg[& A_m \; {}_m\tilde{v}_{(i;\zeta;n-\frac{1}{2})} \\ & - \Delta t \Bigg[\sum_{p,s} \sum_{j=1}^{d} Q^{p,s} \Bigg[\frac{\partial_{p,s}}{\partial x_i^+} \big(\bar{\lambda} \; {}_i\tilde{\sigma}^{p,s}_{(jj;\zeta;n)} \big) + 2 \frac{\partial_{p,s}}{\partial x_j^{h(i,j)}} (\bar{\mu}_{ij} \; {}_j\tilde{\sigma}^{p,s}_{(ij;\zeta;n)}) \\ & + \bar{Y}_{\text{dis}} \; \bar{\tau}^{p,s}_{i,\text{dis}} \left(\frac{\partial_{p,s}}{\partial x_i^+} \big(\bar{\chi} \; {}_i\tilde{\sigma}^{p,s}_{(jj;\zeta;n)} \big) + 2 \frac{\partial_{p,s}}{\partial x_j^{h(i,j)}} (\bar{\eta}_{ij} \; {}_j\tilde{\sigma}^{p,s}_{(ij;\zeta;n)}) \right) \Bigg] \Bigg] \Bigg], \end{aligned} \tag{3.104}$$

where we remind that $h(i, j)$ has been defined in (3.47).

In the same way, a restriction of the second line in equation (3.88) to the component ${}_m\tilde{\mathbf{v}}^{p,s}_{(i;\zeta;n+\frac{1}{2})}$ gives

$$\begin{aligned} {}_m\tilde{\mathbf{v}}^{p,s}_{(i;\zeta;n+\frac{1}{2})} = {}_m \; & \tilde{v}^{p,s}_{(i;\zeta;n+\frac{1}{2})} \\ & + A_m \sum_{j=1}^{d} \bar{Y}_{\text{abs}} \; \bar{\tau}^{p,s}_{i,\text{abs}} \left(\frac{\partial_{p,s}}{\partial x_i^+} \big(\bar{\chi} \; {}_i\tilde{\sigma}^{p,s}_{(jj;\zeta;n)} \big) + 2 \frac{\partial_{p,s}}{\partial x_j^{h(i,j)}} (\bar{\eta}_{ij} \; {}_j\tilde{\sigma}^{p,s}_{(ij;\zeta;n)}) \right) \end{aligned} \tag{3.105}$$

where we have used (3.87).

For the third line in (3.88), considering (3.81) and the symmterisation operator in (3.78), a restriction of the third line in (3.88) to ${}_m\tilde{\sigma}^{p,s}_{(ij;\zeta;n)}$ gives

$$\begin{aligned} {}_m\tilde{\sigma}^{p,s}_{(ij;\zeta;n+1)} = A_m \Bigg[& A_m \; {}_m\tilde{\sigma}^{p,s}_{(ij;\zeta;n)} - \frac{\Delta t}{2} \Big(\frac{\partial_{p,s}}{\partial x_j^{-h(i,j)}} \frac{1}{\bar{\rho}_i} {}_j\tilde{\mathbf{v}}^{p,s}_{(i;\zeta;n+\frac{1}{2})} + \frac{\partial_{p,s}}{\partial x_i^{-h(i,j)}} \frac{1}{\bar{\rho}_j} {}_i\tilde{\mathbf{v}}^{p,s}_{(j;\zeta;n+\frac{1}{2})} \Big) \\ & + \mathbb{G}^T \mathbb{M}^T \hat{P}_{N_t-2-n} \Bigg] \; (n \in \{-1, ..., N_t - 2\}) \end{aligned} \tag{3.106}$$

where the last term has been derived from equation (3.69) in Corollary 3.1, together with the fact that the operator $\mathcal{M}\theta'$ (cf. the last term of the third line in (3.88)) is an implementation of equations (3.54) and (3.55).

For the last iteration, using an adjoint of the operator $\mathcal{S}_n$ (cf. (3.48)), the right-hand-side in equation (3.68) in Corollary 3.1 gives

$$p_{(\zeta;n=N_t-1)} = \mathbb{S}\left(-\frac{\delta_{ij}}{d}\left(\sum_{m=1}^{d} {}_m\tilde{\sigma}^p_{(ij;\zeta,n=N_t-1)}\right)\right) = -\mathbb{S}\left(\frac{1}{d}\sum_{l,m=1}^{d} {}_m\tilde{\sigma}^p_{(ll;\zeta;n=N_t-1)}\right), \tag{3.107}$$

together with a replacement of the last term in the right-hand-side of equation (3.106) with $\mathbb{G}^T\mathbb{M}^T\hat{p}^{\text{adj}}_{n+1/2}$. (This gives an additive source the same as (3.56) and (3.57) in section 3.4.2). Note that here, using the linearity of the operator $\mathbb{H}^*$, we have enforced the factor 1/2 in (3.48) and the summing operator in (3.68) on $\hat{P}$, rather than $\mathbb{H}^*$. Now, we enforce the replacement ${}_m\check{\mathbf{v}}^{p,s}_i = -\frac{1}{\bar{\rho}_i}\,{}_m\tilde{\mathbf{v}}^{p,s}_i$ ($m \in \{1,...,d\}$) in equations (3.104), (3.105) and (3.106). Using this replacement, (3.104) gives the same formula as (3.59), (3.105) is the same as (3.60), and (3.106) is equivalent to (3.61) and (3.62). Also, (3.107) is equivalent to (3.63). Therefore, an adjoint of our discretised operator gives the same formula as a discretisation of our continuous adjoint, as explained in section 3.4.2.

3.A.3 A connection to chapter 4

The main goal of PAT is a qunatitative reconstruction of a distribution of optical absorption coefficient. In chapter 2 and 3, we dealt with the pure acoustic portion of this inverse problem. This is a reconstruction of the initial pressure distribution from a set of time series of pressure data that is measured on the boundary, and is referred to here as photo-acoustic tomography (PAT). In chapter 2, we solved this inverse problem using a system of three-coupled first order wave equations that describe the propagation of photo-acoustic waves in inhomogeneous and lossy fluid media (cf. 1.7). In chapter 3, we solved the inverse problem of PAT using a system of coupled photo-acoustic wave equations that describe the propagation of acoustic waves in linear isotropic heterogeneous and lossy elastic media. For both systems, we modelled an absorption and physical dispersion following a frequency-power law using fractional Laplacian operators.

For the acoustic forward operator associated with lossy fluid media, we derived a continuous adjoint in chapter 2. We remind that a continuous adjoint can be considered as a generalised variant of a discretised (algebraic) adjoint, and is independent of the numerical scheme used for discretisation of the forward operator. In chapter 3, for an acoustic forward operator associated with lossy elestic media, we derived the same formulae for an associated adjoint using *adjoint-then-discretise* or *discretise-then-adjoint* methods.

In the next chapter, we extend the inverse problem of PAT for lossy heterogeneous and fluid media to a single-stage variant of qunatitative PAT, a direct reconstruction of optical attenuation coefficients from a set of acoustic data that is measured in time on the boundary. To do this, the forward problem will be defined using a composite opto-acoustic operator, which includes an optical operator ($\mathbb{H}_o$) and an acoustic operator ($\mathbb{H}_a$). The operator $\mathbb{H}_o$ will be formed as a discretisation of a Diffusion Approximation (DA) model (cf. section 1.9.2) using a Galerkin Finite element method. The operator $\mathbb{H}_a$ will be defined and discretised using the same method as chapter 2. In contrast to the pure acoustic problem (cf. chapter 2), for a single-stage inverse problem of QPAT, the optical forward operator ($\mathbb{H}_o$) is non-linear and ill-conditioned. Therefore, for practical studies, any errors in modelling the acoustic forward operator or an associated adjoint operator may grow during the inversion process. Motivated by these, we will extend this hybrid inverse problem to realistic acoustic media in order to minimise the modelling errors, and also we will use an algebraic acoustic adjoint, because this adjoint satisfies an adjoint test more accurately than a corresponding continuous adjoint. We emphasise that for the acoustic forward operator used in chapter 2 and 4, both *adjoint-then-discretise* [2, 1] and *discretise-then-adjoint* [16] approaches have been widely used for the inverse problem of PAT (the acoustic portion of QPAT).

Bibliography

[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. Mitsuhashi, J. Poudel, T. P. Mathethews, A. Gracia-Uribe, 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.

Chapter 4

Direct quantitative photoacoustic tomography for realistic acoustic media

At the time of submission of this thesis, the content of this chapter has been accepted for publication in the journal of *Inverse Problems, Special Issue on Variational Methods and Effective Algorithms for Imaging and Vision*. The authors of the accepted paper are A. Javaherian and S. Holman.

Abstract

Quantitative photo-acoustic tomography (QPAT) seeks to reconstruct a distribution of optical attenuation coefficients inside a sample from a set of time series of pressure data that is measured outside the sample. The associated inverse problems involve two steps, namely acoustic and optical, which can be solved separately or as a direct composite problem. We adopt the latter approach for realistic acoustic media that possess heterogeneous and often not accurately known distributions for sound speed and ambient density, as well as an attenuation following a frequency power law that is evident in tissue media. We use a Diffusion Approximation (DA) model for the optical portion of the problem. We solve the corresponding composite inverse problem using three total variation (TV) regularised optimisation approaches. Accordingly, we develop two Krylov-subspace inexact-Newton algorithms that utilise the Jacobian

matrix in a matrix-free manner in order to handle the computational cost. Additionally, we use a gradient-based algorithm that computes a search direction using the L-BFGS method, and applies a TV regularisation based on the Alternating Direction Method of Multipliers (ADMM) as a benchmark, because this method is popular for QPAT and direct QPAT. The results indicate the superiority of the developed inexact Newton algorithms over gradient-based Quasi-Newton approaches for a comparable computational complexity.

4.1 Introduction

Quantitative photo-acoustic Tomography (QPAT) is a steadily growing hybrid imaging paradigm that simultaneously takes advantage of the high spatial resolution provided by ultrasound imaging and the rich contrast attributed to optical imaging [48]. Typically, nanosecond-duration pulses of electromagnetic energy, in the visible or near-infrared ranges, are used to irridiate a sample. Depending on the optical properties of the sample, a fraction of the optical energy is absorbed, and converted into heat [50]. The generated heat induces a local increase in pressure via thermo-elastic expansion effects [50]. Because of the elasticity of soft tissues, the locally induced pressures propagate outwards as acoustic waves, and carry information about the optical properties of the sample to the surface. These acoustic waves are measured in time by ultrasound sensors located outside the surface of the sample [48]. Given a set of time series of data at the boundary, the objective in QPAT is to calculate a quantitative image of a distribution of optical absorption coefficients of the sample [48].

QPAT involves two distinct inverse problems, namely acoustic and optical. The acoustic inverse problem, often referred to as PAT, seeks to reconstruct a distribution of the spatially varying initial pressure from the boundary data. This is a linear inverse problem, for which a vast number of reconstruction methods are available [34]. For acoustically homogeneous media, there are exact inversion methods, e.g., back-projection [51, 52, 9, 28], or frequency-domain techniques [53, 55]. Time reversal (TR) is a less restrictive approach since it can be adapted to heterogeneous acoustic media and arbitrary detection surfaces [17, 54, 9, 23, 24]. Model-based iterative inversion approaches, in which the discrepancy between modeled data and the measured data is

iteratively reduced, are also widely used in ill-posed cases that arise from data incompleteness, modelling errors, finite sampling or noise [25]. The iterative methods can be categorised into convergent Neumann-series methods [39, 33] or variational (optimisation) approaches [49, 25, 15, 7, 21]. Further, the results of [39] show uniqueness and stability of recovery of the initial pressure distribution from data taken on an open subset of the boundary over a sufficiently long time. This result does not include the absorption and dispersion that we have included in our model. As far as we are aware, uniqueness for recovery of the initial pressure in the situation considered in this paper, including realistic absorption and dispersion, variable acoustic parameters, and data on only part of the boundary, is open.

The optical inverse problem is devoted to further reconstruction of an image of distributions of optical absorption and scattering coefficients from a recovered initial pressure distribution. This is a highly nonlinear and ill-posed inverse problem, and is commonly solved by iterative model-based approaches [41, 42, 19]. Using model-based approaches for solving these two inverse problems, an accurate, yet efficient, modelling of the associated physical processes is required [43].

A very accurate model for propagation of light is the Radiative Transfer Equation (RTE), which has been used for the optical portion of 2D QPAT, e.g. [41, 35, 20]. An analysis of the optical inverse problem of QPAT using RTE has been given in [3]. Since RTE is computationally very expensive, the applicability of this method for 3D QPAT is very limited. A Diffusion Approximation (DA) of the RTE is more efficient than RTE [43], and is thus more practical for 3D QPAT. In this study, we use the DA for modelling the optical portion of the inverse problem. The DA model is sufficienly accurate when the medium is highly scattering and the scattering is near-isotropic [43]. The DA has been widely used as a light propagation model for QPAT [41, 42, 32, 19, 22].

The DA model is defined as a function of optical absorption and scattering coefficients. It turns out that the recovery of both these coefficients from one optical source and wavelength is non-unique [6]. To have uniqueness for the inversion, three approaches are used, i.e., the assumption of the scattering coefficient as known [12], using more than one optical wavelength [5, 11, 32], or using more than one optical

source [4, 19, 18]. In this study the latter approach is used. We note that [4] established uniqueness of recovery, as well as a Hölder stability result, for the optical absorption and scattering, assuming the Gruneisen parameter (see equation (4.3) below) is known and some regularity requirements are satisfied, given two well-chosen optical sources. Furthermore, [4] also showed that even using additional sources it is not possible to determine all three of the optical absorption, scattering, and Gruneisen parameter.

For modelling the acoustic portion of the problem, the dependence of shape, spectrum and amplitude of propagating acoustic waves on properties of the medium [13], together with the highly nonlinear and ill-conditioned nature of the optical portion of the forward operator, motivates enriching the QPAT problem by simulation of tissue-realistic acoustic properties. To incorporate these effects, we use an acoustic model based on a linear system of three-coupled first-order wave equations which can be adapted to spatially varying sound speed and density [40, 13], and include two fractional Laplacian operators in order to account separately for acoustic absorption and dispersion following a frequency power law, which is evident in tissue media [46, 45]. For a numerical implementation of this acoustic model, we use a k-space pseudo-spectral method, which is popular for iterative PAT because of the high efficiency arising from a requirement of only two grid points per wave-length for defining a field, and a fast computation of the spatial gradient in the frequency domain [40, 13].

For realistic problems of QPAT, for example, when data are available merely on a part of the boundary, or when acoustic properties of the medium are not known exactly, the reconstruction of initial pressure distribution using the acoustic portion of the inverse problem is not sufficiently accurate to further serve as data for the optical inverse problem [18]. It turns out that a direct estimation of the optical coefficients from time series of measured data, referred to here as direct QPAT, is more stable than classical variants of QPAT, in which the two inverse problems are solved distinctly. The direct QPAT has thus recently received much attention, where the forward operator is defined using a composite opto-acoustic model [20, 18, 31]. Using direct QPAT, it will be possible to incorporate prior information about the optical parameters into the acoustic inverse problem [18], and mutually the optical inverse problem can utilise information about noise included in the boundary acoustic data [31]. Additionally, the acoustic

portion of the inverse problem can benefit from multi-source (resp. multi-wavelength) configurations, since the optical coefficients (resp. chromophore concentrations), as opposed to the initial pressure distribution, are independent of the changes in optical illumination (resp. wavelength) [16].

In [37, 38], a linear Born approximation of the DA model based on a Green's function approach is coupled with an acoustic model that uses a free-space Green's function method in order to directly reconstruct perturbations in the optical absorption and diffusion coefficients from time series of measured acoustic data. A simultaneous reconstruction of perturbations in optical absorption coefficient and sound speed given an optical scattering coefficient was also studied using a Born approximation [16]. To the best of our knowledge, existing studies for the direct variant of QPAT have been so far limited to homogeneous and non-attenuating acoustic media, and the acoustic portion of the forward operator and its adjoint are computed based on exact formulae using Green's function approaches [20, 18, 31].

The optimisation approaches for QPAT can be categorised into linearised (Jacobian-based) or nonlinear (gradient-based) approaches. A majority of Jacobian-based methods for the optical portion of QPAT can be fit into a class of Gauss-Newton methods [36]. For application of these methods to classical (resp. direct) QPAT, see [41, 42, 32] (resp. [31]). These studies utilise an explicit form of the Jacobian matrix. For direct QPAT, because of a very large size of time series of measured data, an explicit computation and storage of the Jacobian matrix is very expensive [31].

To avoid this problem, nonlinear gradient-based approaches have received much interest for the direct problem of QPAT [19, 18]. A majority of these approaches use a search direction based on Quasi-Newton methods which utilise only gradient information for an approximation of the Hessian matrix, and are thus memory-efficient [19, 18]. The computation of the gradient for direct QPAT is based on an *opto-acoustic* forward operator and an *acousto-optic* adjoint of the Fréchet derivative of the forward operator [20]. A memory-efficient Quasi-Newton method is Limited-memory BFGS (L-BFGS), for which the inverse of Hessian matrix is approximated using the gradient information stored in a user-adjusted number of iterations. Using a total variation (TV) regularisation approach based on an Alternating Direction Method of Multipliers (ADMM), L-BFGS has been used for computing an associated search direction for the

classical QPAT [19] and direct QPAT [18].

Contribution. Here, we develop two preconditioned inexact Newton (Newton-Krylov) algorithms for solving the direct problem of QPAT. In our first approach, the residual function is iteratively linearised, and each linearised subproblem is solved using a *subspace Krylov* method in a matrix-free manner, for which a total variation matrix is used as a preconditioner for accelerating the convergence of the Krylov method. We use the *Preconditioned Conjugate Gradient* (PCG) as the *Krylov* method, for which a TV-based preconditioning is applied using the *Lagged Diffusivity* (LD) method [47]. Our second approach uses two linearisations, the first of which is enforced to the nonlinear residual function, and the second is applied in order to handle the nonlinearity of a TV functional using a Primal-Dual Interior Point Method (PD-IPM). Using this approach, for each linearisation of the residual function, a sequence of normal equations is derived, and is solved using a *subspace Krylov* method in a matrix-free manner. We solve the arising normal equations using a PCG method, but we emphasise that an extension to other *Krylov* methods, e.g. a preconditioned variant of LSQR [2], is straightforward.

We implement our algorithms by assuming tissue realistic but erroneous properties for the acoustic medium with a limited-view setting for boundary measurements. We model the acoustic portion of the forward operator using a linear system of three-coupled first-order wave equations that can be adapted to heterogeneous media and account for acoustic absorption and dispersion following a frequency power law using two fractional Laplacians [45]. This acoustic model is very popular for PAT since it simulates an acoustic attenuation that is evident in many materials of interest including tissues [45]. To the best of our knowledge, this is the first study for direct QPAT that uses an acoustic model that accounts for tissue-realistic acoustic properties of the medium. For the acoustic portion of the problem, we include the action of perfectly matched layers (PMLs) in calculation of the adjoint operator. To the best of our knowledge, these effects have not been accounted for in existing studies (See [25]).

It is worth mentioning that a singular value decomposition (SVD) analysis on the acoustic forward operator we use has shown that as time steps increase some of the singular values of the forward operator become very small [46], and make the acoustic forward operator ill-conditioned. However, the use of an acoustically realistic forward operator may be necessary for direct QPAT, since the optical portion of the forward

operator is highly nonlinear and ill-conditioned, and thus errors in acoustic modelling may quickly grow in the inversion process, and dominate signal data.

Our numerical results show that the developed preconditioned Newton-Krylov optimisation algorithms perform much better than nonlinear Quasi-Newton methods that have been used in existing studies for direct QPAT. The algorithm we use as a benchmark utilises a TV regularisation based on an ADMM method, together with a search direction using an L-BFGS method. (See [19, 18] for application on classical QPAT and direct QPAT, respectively).

4.2 Direct QPAT on a continuous domain

In this section, we define our forward operator as a composite map on a continuous domain.

4.2.1 Modelling the optical portion of the problem

The time scale for propagation of acoustic waves is on the order of a micro-second, which is three orders of magnitude larger than the time-scale for illumination, propagation and absorption of light. Therefore, the generated pressure distribution is regarded as instantaneous for the acoustic problem, and is referred to as initial pressure distribution p_0. One way to define a forward operator for our QPAT problem is to combine the physics of the optical and acoustic portions of the problem using a simple composition of two maps, one modelling propagation and absorption of optical photons and the other modelling propagation of acoustic waves [20].

Accordingly, let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary $\partial\Omega$ and $d \in \{2, 3\}$ the spatial dimension. Additionally, let $\phi \in \mathcal{H}^1(\Omega)$ denote the photon density. For modelling the propagation of light, we use a time-independent variant of DA equations with the well-known Robin boundary condition [41]. This is written as

$$\begin{aligned} -\nabla \cdot \kappa(r)\nabla\phi(r) + \mu(r)\phi(r) &= 0, \quad r \in \Omega \\ \phi(r) + \frac{1}{2\gamma_d}\kappa(r)\frac{\partial\phi(r)}{\partial\hat{n}} &= I/\gamma_d, \quad r \in \partial\Omega, \end{aligned} \tag{4.1}$$

where $\mu(r), \kappa(r) \in \mathcal{L}^\infty_+(\Omega)$ denote the positive-valued optical absorption and diffusion coefficients, respectively. Here, r denotes the spatial position. Additionally, γ_d is a

dimension-dependent factor ($\gamma_2 = 1/\pi$ and $\gamma_3 = 1/4$), $\hat{n}$ is an outward unit normal, and I is an inward directed diffuse boundary current [29, 43]. Following the absorption of photons, a spatially varying heating field $h(r)$ is generated in the form

$$h[\kappa, \mu, I](r) = \mu(r)\phi[\kappa, \mu, I](r). \tag{4.2}$$

Because of thermo-elastic expansion effects, the induced spatially varying heating field causes an instant local increase in pressure that follows

$$p_0(r) = \begin{cases} \Gamma(r)h(r), & r \in \Omega \\ 0, & r \in \mathbb{R}^d \backslash \Omega, \end{cases} \tag{4.3}$$

where Γ is the Gruneisen parameter, and describes the efficiency of conversion of heat into pressure [20]. Here, we assume $\Gamma(r)$ constant and rescaled to 1, and thus $p_0(r) = h(r)$ [20].

4.2.2 Modelling the acoustic portion of the problem

We use a linear system of three-coupled first-order equations for describing the propagation of acoustic wavefields in an acoustically heterogeneous and lossy medium [46, 45]. To explain this, we define our fields, which are the acoustic pressure field $p(r,t)$, particle velocity vector $v(r,t)$ and acoustic density $\rho(r,t)$, where $r \in \mathbb{R}^d$ and $t \in [0, \infty)$ denote the spatial position and time. Additionally, we define the medium's acoustic parameters as sound speed $c_0(r)$, ambient density $\rho_0(r)$, attenuation coefficient $\alpha_0(r)$, and frequency power law exponent y. The acoustic wavefield propagation is now modeled by three equations, i.e., linearised equation of motion (conservation of momentum)

$$\frac{\partial v}{\partial t}(r,t) = -\frac{1}{\rho_0(r)}\nabla p(r,t), \tag{4.4}$$

linearised equation of continuity (conservation of mass)

$$\frac{\partial \rho}{\partial t}(r,t) = -\rho_0(r)\nabla \cdot v(r,t), \tag{4.5}$$

and equation of state

$$\begin{aligned} p(r,t) = c_0(r)^2 \Big\{ 1 - \tau(r)\frac{\partial}{\partial t}(-\nabla^2)^{\frac{y}{2}-1} \\ - \eta(r)(-\nabla^2)^{\frac{y-1}{2}} \Big\} \rho(r,t) \end{aligned} \tag{4.6}$$

with initial conditions

$$p(r,0) = p_0(r), \quad v(r,0) = 0. \tag{4.7}$$

Here, $\tau(r)$ and $\eta(r)$ are, respectively the absorption and dispersion proportionality coefficients, and are given by

$$\tau(r) = -2\alpha_0(r)c_0(r)^{y-1}, \quad \eta(r) = 2\alpha_0(r)c_0(r)^y \tan(\pi y/2). \tag{4.8}$$

4.2.3 Opto-acoustic forward operator

Having given the models for describing the optical and acoustic portions of the forward operator, we now define our *opto-acoustic* forward operator. To do this, we also require the measurement operator $\mathcal{M}$, which at each time step maps the pressure $p(r,t)$ to the measured data at sensors. The time series of acoustic data are denoted by $\hat{P} \in \mathbb{R}^{N_s N_t}$ with $N_s, N_t \in \mathbb{N}$ the number of detectors and the number of measurement time instants, respectively.

Definition 4.1. For a fixed illumination I, the optical portion of the forward operator is a nonlinear map in the form

$$\begin{aligned} &\Lambda_o : \mathcal{L}^\infty_+(\Omega) \times \mathcal{L}^\infty_+(\Omega) \to \mathcal{L}^2(\Omega) \\ &\Lambda_o[\kappa,\mu](r) = h(r), \end{aligned} \tag{4.9}$$

where $h(r)$ is given by (4.2). Also, the acoustic portion of the forward operator is a linear map in the form [1]

$$\begin{aligned} &\Lambda_a : \mathcal{L}^2(\Omega) \to \mathbb{R}^{N_s N_t}, \\ &\Lambda_a[p_0](r,t) = \mathcal{M}p(r,t). \end{aligned} \tag{4.10}$$

Using (4.9) and (4.10), the coupled opto-acoustic forward operator is defined by the composite map Λ, i.e.,

$$\begin{aligned} &\Lambda : \mathcal{L}^\infty_+(\Omega) \times \mathcal{L}^\infty_+(\Omega) \to \mathbb{R}^{N_s N_t}, \\ &\Lambda[\kappa,\mu] = \Lambda_a\big[\Lambda_o[\kappa,\mu]\big]. \end{aligned} \tag{4.11}$$

We note that, as already mentioned in the introduction, by the results of [39] and [4] if we consider a non-discretised version of the measurements, eliminate the absorption

and dispersion from our acoustic model, and add some regularity requirements on the parameters, then the forward map becomes injective. Injectivity in our situation, either with non-discretised measurements or discretised κ and μ is, to our knowledge, an open problem. However we will not consider this in the current paper, but rather focus on numerical methods.

4.2.4 Model-based approach for inverse problem

The inverse problem of direct QPAT is a simultaneous reconstruction of μ, κ from $\hat{P} = \Lambda[\kappa, \mu]$. Applying a model-based approach for solving this problem, the objective is to minimise an error functional, the sum of squared differences between modeled data and the measured data via an iterative adjustment of the unknown optical coefficients μ, κ. Using N_q optical illuminations, the error functional is defined by [18]

$$\epsilon(\kappa, \mu) = \frac{1}{2} \sum_{q=1}^{N_q} \left\| \Lambda_q[\kappa, \mu] - \hat{P}_q \right\|_2^2, \tag{4.12}$$

where, q indexes a set of N_q illuminations I_q with corresponding acoustic data $\hat{P}_q$ [18]. We combine the unknown parameters as $x = [\kappa, \mu] \in \mathcal{L}_+^\infty(\Omega) \times \mathcal{L}_+^\infty(\Omega)$.

Let $D_x\Lambda$ denote the Fréchet derivative of the forward operator at x. The Fréchet derivative of ϵ at x is given by

$$D_x\epsilon = \sum_{q=1}^{N_q} D_x^*\Lambda_q \left(\Lambda_q[x] - \hat{P}_q \right). \tag{4.13}$$

Here, $D_x^*\Lambda$ denotes the adjoint of the Fréchet derivative of the forward operator, and is given by

$$D_x^*\Lambda[\hat{P}] = D_x^*\Lambda_o[\Lambda_a^*[\hat{P}]], \tag{4.14}$$

where Λ_a^* denotes the adjoint of the linear acoustic forward operator Λ_a (4.10) and $D_x^*\Lambda_o$ represents the adjoint of the Fréchet derivative of the optical forward operator (4.9) both with respect to the $\mathcal{L}^2(\Omega)$ inner product. Formulae for Λ_a^* can be found in [26], while for $D_x^*\Lambda_o$ we have the next lemma.

Lemma 4.1. Let us denote the solution of (4.1) for the fixed illumination I, diffusion κ_0 and absorption μ_0 by ϕ_0. Then the Fréchet derivative of the optical portion of the

forward operator $D_x\Lambda_o$ at $x_0 = [\kappa_0, \mu_0]$ applied to the perturbations $\delta\kappa$ and $\delta\mu$ is given by

$$D[\kappa_0, \mu_0]\Lambda_o \begin{pmatrix} \delta\kappa \\ \delta\mu \end{pmatrix} = \delta\mu(r)\phi_0(r) + \mu_0(r)\delta\phi(r), \tag{4.15}$$

where $\delta\phi(r)$ satisfies

$$\begin{aligned} -\nabla\cdot\kappa_0(r)\nabla\delta\phi(r) + \mu_0(r)\delta\phi(r) &= \nabla\cdot\delta\kappa(r)\nabla\phi_0(r) - \delta\mu(r)\phi_0(r), \quad r\in\Omega \\ \delta\phi(r) + \frac{1}{2\gamma_d}\kappa_0(r)\frac{\partial\delta\phi(r)}{\partial\hat{n}} &= -\frac{1}{2\gamma_d}\delta\kappa(r)\frac{\partial\phi_0(r)}{\partial\hat{n}}, \quad r\in\partial\Omega. \end{aligned} \tag{4.16}$$

The adjoint map $D_x^*\Lambda_o$ can then be calculated from

$$D[\kappa_0, \mu_0]^*\Lambda_o h(r) = \begin{pmatrix} \nabla\phi_0(r)\cdot\nabla\tilde{h}(r) \\ \phi_0(r)\tilde{h}(r) + \phi_0(r)h(r) \end{pmatrix}, \tag{4.17}$$

where the adjoint field $\tilde{h}(r)$ satisfies

$$\begin{aligned} -\nabla\cdot\kappa_0(r)\nabla\tilde{h}(r) + \mu_0(r)\tilde{h}(r) &= -\mu_0(r)h(r), \quad r\in\Omega \\ \tilde{h}(r) + \frac{1}{2\gamma_d}\kappa_0(r)\frac{\partial\tilde{h}(r)}{\partial\hat{n}} &= 0, \quad r\in\partial\Omega. \end{aligned} \tag{4.18}$$

Lemma 4.1 can be proven using integration by parts.

4.3 Numerical computation

4.3.1 Numerical computation of the optical operators (Λ_o, $D_x\Lambda_o$ and $D_x^*\Lambda_o$)

Variational formulae

We use a first-order Galerkin finite element method (FEM) for approximation of the optical portion of our QPAT problem. For an approximation of Λ_o, a variational form of (4.1) is derived, i.e.,

$$\int_\Omega \mu_0\phi_0\nu\ dr + \int_\Omega \kappa_0\nabla\phi_0\cdot\nabla\nu\ dr + 2\gamma_d\int_{\partial\Omega}\phi_0\nu\ ds = \int_{\partial\Omega} 2I_s\nu\ ds, \tag{4.19}$$

where $\nu\in\mathcal{H}^1(\Omega)$ is a test function. Additionally, for an approximation of the Fréchet derivative operator $D_x\Lambda_o$ using FEM, a variational form of (4.16) is derived, i.e.,

$$\int_\Omega \mu_0\delta\phi\nu dr + \int_\Omega \kappa_0\nabla\delta\phi\cdot\nabla\nu dr + 2\gamma_d\int_{\partial\Omega}\delta\phi\nu ds = -\int_\Omega \delta\kappa\nabla\phi_0\cdot\nabla\nu dr - \int_\Omega \delta\mu\phi_0\nu dr. \tag{4.20}$$

In the same way, for an approximation of $D_x^*\Lambda_o$, a variational form of (4.18) is obtained, i.e.,

$$\int_\Omega \mu_0 \tilde{h}\nu dr + \int_\Omega \kappa_0 \nabla\tilde{h}\cdot\nabla\nu dr + 2\gamma_d \int_{\partial\Omega} \tilde{h}\nu ds = -\int_\Omega \mu_0 h\nu dr. \tag{4.21}$$

Discretisation of optical coefficients and fields

Let T denote a triangulation of Ω with N_e elements, i.e., $T = \{t_j \mid j = 1, ..., N_e\}$. Applying an approximation using FEM, we discretise the optical coefficients in a piecewise-constant basis $\{\chi_j = 1_{t_j} \mid j = 1, \dots, N_e\}$. Using this, the optical fields are approximated as [41]

$$\begin{aligned} \kappa_0(r) \approx \kappa^e(r) &= \sum_{j=1}^{N_e} \hat{\kappa}_j \chi_j(r) \\ \mu_0(r) \approx \mu^e(r) &= \sum_{j=1}^{N_e} \hat{\mu}_j \chi_j(r), \end{aligned} \tag{4.22}$$

where $\hat{\kappa}_j$ and $\hat{\mu}_j$ denote the discretised absorption and diffusion coefficients at element t_j. Additionally, $\phi_0(r)$ is approximated in a piecewise-linear basis $\{\varphi_k \mid k = 1, \dots, N_n\}$ in the form

$$\phi_0(r) \approx \phi_0^h(r) = \sum_{k=1}^{N_n} \Phi_{0,k}\varphi_k(r), \tag{4.23}$$

where $\Phi_{0,k}$ denotes the discretised photon density at node k, and N_n is the total number of nodes. We also approximate the adjoint field $\tilde{h}(r)$ in a piecewise-linear basis as

$$\tilde{h}(r) \approx \tilde{h}^h(r) = \sum_{k=1}^{N_n} \tilde{H}_k \varphi_k(r). \tag{4.24}$$

In the sequel, a field that is discretised at nodes as in (4.23) and (4.24) (resp. elements as in (4.22)) is called a nodal (resp. elemental) vector. In the same way as the continuous formulae, we use δ for signifying a perturbation in a discretised coefficient or field.

Matrix form of variational formulae

For a discretisation of the problem, a matrix form of the variational formulae in section 4.3.1 is derived. [43, 41]. To do this, we define a system matrix $\boldsymbol{A}_o$ in the form

$$\boldsymbol{A}_o[\kappa^e, \mu^e] = K[\kappa^e] + C[\mu^e] + R \tag{4.25}$$

with

$$\begin{aligned} K_{kp}[\kappa^e] &= \int_\Omega \kappa^e \nabla\varphi_k \cdot \nabla\varphi_p \; dr \\ C_{kp}[\mu^e] &= \int_\Omega \mu^e \varphi_k \varphi_p \; dr \\ R_{kp} &= 2\gamma_d \int_{\partial\Omega} \varphi_k \varphi_p \; ds \\ G_p &= \int_{\partial\Omega} 2I_s \varphi_p \; ds, \end{aligned} \tag{4.26}$$

where $p, k = 1, ..., N_n$ denote nodal indices. We also define matrix $\boldsymbol{A}_{\delta,o}$ in the form

$$\boldsymbol{A}_{\delta,o}[\delta\kappa^e, \delta\mu^e] = K[\delta\kappa^e] + C[\delta\mu^e]. \tag{4.27}$$

Using the above, we now define the discretised optical forward operators. We stress that this definition is setting the notation for the discretised operators, which will be described in detail below.

Definition 4.2. A discretisation of the optical forward operator Λ_o gives a map from a vector space of discretised (elemental) optical coefficients to a vector space of discretised (elemental) heating field coefficients in the form

$$\begin{aligned} &\mathbb{H}_o : \mathbb{R}^{N_e} \times \mathbb{R}^{N_e} \to \mathbb{R}^{N_e} \\ &H = \mathbb{H}_o[\hat{\kappa}, \hat{\mu}]. \end{aligned} \tag{4.28}$$

Additionally, a discretisation of the Fréchet derivative operator $D_x\Lambda_o$ at $X = [\hat{\kappa}, \hat{\mu}]$ applied on perturbation $\delta X = [\delta\hat{\kappa}, \delta\hat{\mu}]$ gives

$$\begin{aligned} &\mathbb{J}_o : \mathbb{R}^{N_e} \times \mathbb{R}^{N_e} \to \mathbb{R}^{N_e} \\ &\delta H = \mathbb{J}_o[\hat{\kappa}, \hat{\mu}] \begin{pmatrix} \delta\hat{\kappa} \\ \delta\hat{\mu} \end{pmatrix}. \end{aligned} \tag{4.29}$$

Now, we give further details on these operators.

Discretised forward operator $\mathbb{H}_o[\kappa_0, \mu_0]$

Plugging (4.23) into (4.19), together with taking $\nu(r)$ to be a basis function $\varphi_p(r)$, gives a linear system

$$\boldsymbol{A}_o[\kappa^e, \mu^e]\Phi_0 = G. \tag{4.30}$$

Using this, the heating field is approximated in a piecewise constant basis as [41, 19, 18]

$$H = \hat{\mu} \circ \mathbb{I}\Phi. \tag{4.31}$$

Here, $\circ$ denotes an element-wise product and $\mathbb{I}$ is a *node-to-element* map. For $\mathbb{I}$ we use the $\mathcal{L}^2$ orthogonal projection from the space of nodal representations to elemental representations. The action of $\mathbb{I}$ on a nodal vector θ restricted at element j is in the form

$$(\mathbb{I}\theta)_j = \frac{1}{d+1} \sum_{p \in \ell(j)} \theta_p \tag{4.32}$$

where p denotes the nodal index, and $\ell(j)$ is a set of $d+1$ nodes that correspond to element j. We also define $\mathbb{I}^+$ as a map from the space of elemental vectors to nodal vectors. The action of $\mathbb{I}^+$ on an elemental vector Θ restricted at node p is given by

$$\left(\mathbb{I}^+\Theta\right)_p = \sum_{j=1}^{N_e} \Theta_j \int_{t_j} \varphi_p(r) dr = \frac{1}{d+1} \sum_{j \in l(p)} S_j \Theta_j, \tag{4.33}$$

where S is the vector of volume of elements, and $l(p)$ is the set of elements that are connected to node p. Also note that from (4.32) and (4.33), $\mathbb{I}^+\Theta = \mathbb{I}^T (S \circ \Theta)$, where T denotes the transpose. In the sequel, we will use the notation $X = [\hat{\kappa}, \hat{\mu}]$ and $\delta X = [\delta\hat{\kappa}, \delta\hat{\mu}]$.

The matrix-free action of $\mathbb{J}_o[X]$ on perturbation δX

Here, we explain how the action of $\mathbb{J}_o[X]$ on a perturbation δX is approximated in a matrix-free manner. To do this, we approximate the perturbation field $\delta\phi(r)$ in a piecewise-linear basis in the same way as (4.23). Plugging the nodal vector $\delta\Phi$ into (4.20) gives a linear system for calculation of $\delta\Phi$ in the form

$$\boldsymbol{A}_o[\kappa^e, \mu^e]\delta\Phi = -\boldsymbol{A}_{\delta,o}[\delta\kappa^e, \delta\mu^e]\Phi_0, \tag{4.34}$$

where Φ_0 has been computed using (4.30). Finally, the perturbation in the heating field δH is computed as

$$\delta H = \delta\hat{\mu} \circ \mathbb{I}\Phi_0 + \hat{\mu} \circ \mathbb{I}\delta\Phi. \tag{4.35}$$

The matrix-free action of $\mathbb{J}_o^*[X]$ on H

Here we explain how the action of the adjoint of Fréchet derivative $\mathbb{J}_o^*$ on an elemental vector H can be approximated using an *adjoint-then-discretise* method. Plugging $\tilde{H}$ from (4.24) into the variational form of the adjoint formula (4.21) gives a linear system in the form

$$\boldsymbol{A}_0[\kappa^e, \mu^e]\tilde{H} = -\mathcal{G}, \tag{4.36}$$

where

$$\mathcal{G} = \mathbb{I}^+(\hat{\mu}_0 \circ H). \tag{4.37}$$

Finally, given the nodal vectors $\tilde{H}$ and Φ_0, we will approximate the action of the adjoint using (4.17). For this we must choose how to calculate the products that appear in (4.17), and then how to project onto the space of elemental vectors. To do this, we will also use the matrices

$$\frac{\partial \boldsymbol{A}_o}{\partial \hat{\kappa}_j} = \int_{t_j} \nabla\varphi_k \cdot \nabla\varphi_p \; dr \tag{4.38}$$

and

$$\frac{\partial \boldsymbol{A}_o}{\partial \hat{\mu}_j} = \int_{t_j} \varphi_k \varphi_p \; dr. \tag{4.39}$$

If the products in (4.17) are calculated by first multiplying the nodal functions, and then using an $\mathcal{L}^2$-orthogonal projection on the space of elemental functions, then for the discretisation of the adjoint we have

$$\mathbb{J}_o^*(H)_j = \begin{pmatrix} \frac{1}{S_j}\tilde{H}^T \frac{\partial \boldsymbol{A}_o}{\partial \hat{\kappa}_j}\Phi_0 \\ \frac{1}{S_j}\tilde{H}^T \frac{\partial \boldsymbol{A}_o}{\partial \hat{\mu}_j}\Phi_0 + H^T(\mathbb{I}\Phi_0) \end{pmatrix}. \tag{4.40}$$

Remark 4.1. Using a *discretise-then-adjoint method*, the adjoint of Fréchet derivative of the optical forward operator will be in the form (cf. [18], eq. (27))

$$\mathbb{J}_{o,\mathrm{dis}}^*(H)_j = \begin{pmatrix} \left[(\mathbb{I}\boldsymbol{A}_o^{-1})^T(\hat{\mu}_0 \circ H)\right]^T \left(-\frac{\partial \boldsymbol{A}_o}{\partial \hat{\kappa}_j}\right)\Phi_0 \\ \left[(\mathbb{I}\boldsymbol{A}_o^{-1})^T(\hat{\mu}_0 \circ H)\right]^T \left(-\frac{\partial \boldsymbol{A}_o}{\partial \hat{\mu}_j}\right)\Phi_0 + H^T(\mathbb{I}\Phi_0) \end{pmatrix}. \tag{4.41}$$

From (4.36), (4.37) and the fact that $\mathbb{I}^+\Theta = \mathbb{I}^T(S \circ \Theta)$ using (4.33), it can be shown that (4.41) matches (4.40).

Remark 4.2. From a theoretical point of view, it is perhaps more natural to make the projections from nodal to elemental representations in the $\mathcal{L}^\infty$ norm since $\kappa, \mu \in \mathcal{L}^\infty_+$. The κ component of the adjoint, i.e., $\nabla\phi_0(r) \cdot \nabla\tilde{h}(r)$, is constant at each element, and thus the $\mathcal{L}^\infty$ projection is the same as the $\mathcal{L}^2$ orthogonal projection. However, the two terms included in the μ component of the adjoint, i.e., $\phi_0(r)h(r) + \phi_0(r)\tilde{h}(r)$, will be different if $\mathcal{L}^\infty$ projection is used. Our simulation experiment showed us that an $\mathcal{L}^\infty$ projection gives almost the same reconstruction as the $\mathcal{L}^2$-orthogonal projection, but dramatically increases the computational cost. Therefore, we used the $\mathcal{L}^2$-orthogonal projection for approximation of the unknown optical coefficients.

4.3.2 Numerical computation of the acoustic operators (Λ_a and its adjoint)

Here, for a numerical computation of the acoustic portion of the problem, we used a k-space pseudo-spectral time-domain (PSTD) method [40, 13]. Applying this method, the spatial gradients are approximated in a frequency domain using a Fast Fourier Transform (FFT), and the temporal gradients are approximated using finite difference schemes [40, 13].

Discretisation of acoustic fields

We approximate the acoustic forward and adjoint operators on a uniform rectilinear grid staggered in space and time [40]. We denote the position of a given grid point in Cartesian coordinates by r_ζ, where $\zeta = (\zeta_1, ..., \zeta_d) \in \{1, \ ... \ , N_1\} \times ... \times \{1, ..., N_d\}$ with $N = \prod_{i=1}^{d} N_i$ the total number of grid points. We denote the grid separation along direction i by Δr_i. The time accessible to detectors, i.e., $t \in (0, T)$, is sampled with a temporal separation of Δt so that the time step n corresponds to the time instant $t_n = n\Delta t$. To accommodate a staggered spatial grid, we introduce operator $\mathbb{T}_i$, which shifts the point r by $\Delta r_i/2$ in coordinate i, and acts on a field as

$$\mathbb{T}_i f(r) = f(\mathbb{T}_i r). \tag{4.42}$$

To accommodate a staggered temporal grid, we shift the field at time t_n by $-\Delta t/2$, i.e., $t_{n-1/2} = n\Delta t - \Delta t/2$. Based on the above, the discretised particle velocity vector

at time step n is denoted by $v_{(i;\zeta;n-1/2)} \in \mathbb{R}_i^d \times \mathbb{R}_\zeta^N \times \mathbb{R}_n^{N_t+1}$. This approximates the actual velocity on the staggered spatial grid as

$$v_{(i;\zeta;n-1/2)} \approx \mathbb{T}_i v_i(r_\zeta, t_{n-1/2}). \tag{4.43}$$

We also approximate the acoustic density as $\rho_{(i;\zeta;n)} \approx \rho_i(r_\zeta, t_n) \in \mathbb{R}_i^d \times \mathbb{R}_\zeta^N \times \mathbb{R}_n^{N_t+1}$ and the scalar pressure as $p_{(\zeta;n)} \approx p(r_\zeta, t_n) \in \mathbb{R}_\zeta^N \times \mathbb{R}_n^{N_t+1}$. (Note that acoustic density is not a physical vector, but it is changed to a vector to accommodate the numerical method.)

Discretisation of medium's acoustic properties

We define a discretised variant of acoustic parameters as diagonal matrices of size $N \times N$ acting on the spatial index ζ of acoustic fields. We denote a discretised variant of $c_0(r)$ by $\bar{c}$. Also, an approximation of the ambient density $\rho_0(r)$ is denoted by $\bar{\rho}$. We also define $\rho_0(r)$ on a spatial grid staggered in coordinate i by $\bar{\rho}_i$.

Additionally, we define a discretised variant of the absorption and dispersion proportionality coefficients as $\bar{\tau}$ and $\bar{\eta}$, respectively. Also, a discretisation of the associated fractional Laplacian operators $N \times N$ is given by

$$Y_{\text{abs}} = F^{-1}\left\{k^{y-2}F\{\cdot\}\right\}, \quad Y_{\text{dis}} = F^{-1}\left\{k^{y-1}F\{\cdot\}\right\}, \tag{4.44}$$

where F and F^{-1} denote the FFT and its inverse, respectively. Applying the k-space pseudo-spectral method on a staggered spatial grid, the spatial gradient in coordinate i is in the form

$$\frac{\partial\{\cdot\}}{\partial r_i^{\pm}} = F^{-1}\left\{\mathbf{i}k_i e^{\pm \mathbf{i}k_i \Delta r_i/2}\text{sinc}\left(\bar{c}_{\text{ref}}k\Delta t/2\right)F\{\cdot\}\right\}, \tag{4.45}$$

where $\text{sinc}(\bar{c}_{\text{ref}}k\Delta t)/2$ enforces a k-space correction to the spatial gradient using a reference sound speed $\bar{c}_{\text{ref}}$ in order to minimise the numerical dispersion errors accumulated by the temporal integrations [40]. For further details on the k-space pseudo-spectral method, see [40, 13].

Because of the computation of the spatial derivatives via an FFT, wave wrapping may occur, i.e., the acoustic waves leaving one side of the grid reenter the opposite side. To avoid wave wrapping, an absorbing boundary condition, referred to as perfectly matched layer (PML), is added to each side of the grid. Here, the action of a PML on

a field in coordinate i is denoted by $A_i \in \mathbb{R}^{N\times N}$. This is defined as [40]

$$A_i = \text{diag}\left(e^{-\alpha_{a,i}\Delta t/2}\right), \tag{4.46}$$

where $\alpha_{a,i}$ is the attenuation enforced by the PML in coordinate i.

Having defined a discretisation of the optical forward operator, as well as acoustic fields and medium, we now complete the definition of our discretised composite operator.

Definition 4.3. A discretisation of the composite *opto-acoustic* forward operator Λ gives a map from optical coefficients to a set of time series of measured data in the form

$$\begin{aligned} &\mathbb{H} : \mathbb{R}^{N_e} \times \mathbb{R}^{N_e} \to \mathbb{R}^{N_s N_t} \\ &\hat{P} = \mathbb{H}_a \mathbb{H}_o[X], \end{aligned} \tag{4.47}$$

where $\mathbb{H}_o$ and $\mathbb{H}_a$ represent the discretised optical and acoustic forward operators, respectively. Also, an operator representing a discretisation of the Fréchet derivative of the forward operator $D_X\Lambda$ is defined by

$$\begin{aligned} &\mathbb{J} : \mathbb{R}^{N_e} \times \mathbb{R}^{N_e} \to \mathbb{R}^{N_s N_t} \\ &\delta\hat{P} = \mathbb{H}_a \mathbb{J}_o[X]\delta X. \end{aligned} \tag{4.48}$$

Using (4.48), we will use as a discretisation of the adjoint of Freć het derivative operator the map

$$\begin{aligned} &\mathbb{J}^* : \mathbb{R}^{N_s N_t} \to \mathbb{R}^{N_e} \times \mathbb{R}^{N_e} \\ &\partial X = \mathbb{J}_o^*[X]\mathbb{H}_a^*\left(\partial\hat{P}\right). \end{aligned} \tag{4.49}$$

Numerical computation of the acoustic forward operator $\mathbb{H}_a$

We now explain a discretisation of the acoustic forward operator ($\mathbb{H}_a$). We will use this later for calculation of the acoustic adjoint operator. For a numerical implementation of the acoustic forward operator given by (4.4), (4.5), (4.6) and (4.7) using a k-space pseudo-spectral method, we used an open-source code, which is freely available on the *k-Wave* website [44].

Using this code, the discretised initial pressure distribution at $t = 0$, denoted by P_0 is applied as an injection of mass, referred to as *additive source*. To do this, P_0 must

be split over two time steps $n = \{-1/2, +1/2\}$. (cf. [1], Appendix B, or the *k-Wave* manual [44]). This gives a source in coordinate i as [1]

$$s_{(i;\zeta;n+1/2)} = \begin{cases} \frac{1}{2d\Delta t\ \bar{c}^2}\mathbb{S}P_0 & n = -1, 0 \\ 0 & \text{otherwise,} \end{cases} \tag{4.50}$$

where $\mathbb{S}$ is a symmetric smoothing operator that is applied in order to mitigate unexpected oscillations of P_0. (cf. the *k-Wave* manual in [44]). Additionally, since s_i is added to the mass equation, a factor $\frac{1}{\Delta t\ \bar{c}^2}$ has been applied in order to account for the conversion of units from pressure to the time rate of density (cf. [1], Appendix B, or the *k-Wave* manual in [44]). Using the definitions given above, the calculation of $\mathbb{H}_a$ proceeds as follows.

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

1. Update the particle velocity vector field (conservation of momentum (4.4)*):*

$$v_{(i;\zeta;n+\frac{1}{2})} = A_i\left[A_i\ v_{(i;\zeta;n-\frac{1}{2})} - \frac{\Delta t}{\bar{\rho}_i}\frac{\partial}{\partial r_i^+}p_{(\zeta;n)}\right]. \tag{4.51}$$

2. Update the acoustic density field (conservation of mass (4.5)*) also adding source:*

$$\rho_{(i;\zeta;n+1)} = A_i\left[A_i\ \rho_{(i;\zeta;n)} - \Delta t\bar{\rho}\frac{\partial}{\partial r_i^-}v_{(i;\zeta;n+\frac{1}{2})}\right] +\ \Delta t\ s_{(i;\zeta;n+1/2)}. \tag{4.52}$$

3. Update the scalar pressure field (equation of state (4.6)*):*

$$p_{(\zeta;n+1)} = \bar{c}^2\left[(I_N - \bar{\eta}Y_{\text{dis}})\sum_{i=1}^{d}\rho_{(i;\zeta;n+1)} + \bar{\tau}Y_{\text{abs}}\sum_{i=1}^{d}A_i\bar{\rho}\frac{\partial}{\partial r_i^-}v_{(i;\zeta;n+\frac{1}{2})}\right] \tag{4.53}$$

where I_N denotes an identity matrix of size $N \times N$, and the last term is actually the action of $\bar{\tau}Y_{\text{abs}}$ on $\sum_{i=1}^{d}\frac{\partial}{\partial t}\rho_{(i;\zeta;n+1)}$, and is derived from (4.52).

4. Compute the measured pressure at detectors: This is defined using

$$\hat{P}_{n+1} = \mathbb{M}p_{(\zeta;n+1)}, \tag{4.54}$$

where $\mathbb{M}$ denotes a discretised variant of $\mathcal{M}$ (cf. section 4.2.3), and includes an interpolation operator for mapping the acoustic pressure field from grid points to position of ultrasound detectors. It is worth mentioning that $\mathbb{M}$ also depends on the size and properties of detectors. These effects are neglected in our study by assuming the detectors sample the pressure pointwise.

A *discretise-then-adjoint* method for derivation of the acoustic adjoint operator

Having defined a discretisation of the acoustic forward operator ($\mathbb{H}_a$), we now explain how to calculate the acoustic adjoint operator ($\mathbb{H}_a^*$) using a *discretise-then-adjoint* method. The derivation of $\mathbb{H}_a^*$ in this section is a modification to the study of [25] in the sense that the effects of the PMLs and an *additive source* are incorporated in calculation of the adjoint. Additionally, in contrast to [25], we will represent $\mathbb{H}_a^*$ as a discretised linear system of PDEs, the same as our representation for $\mathbb{H}_a$ [25].

To derive this adjoint, a matrix form of $\mathbb{H}_a$ must be derived using the details given in section 4.3.2. To do this, we start with definition of diagonal matrices $C = \bar{c}^2 \in \mathbb{R}^{N\times N}$ and $Q \in \mathbb{R}^{dN\times dN}$ with diagonal a d-times stack of diagonals of $\bar{\rho} \in \mathbb{R}^{N\times N}$. Also, we define a diagonal matrix $Q_s \in \mathbb{R}^{dN\times dN}$ with diagonal a stack of diagonals of $\bar{\rho}_i,\ (i \in \{1, ..., d\})$. We will also use $A \in \mathbb{R}^{dN\times dN}$ as a diagonal matrix with diagonal a stack of diagonals of $A_i\ \ (i \in \{1, ..., d\})$.

We now define a stack of coordinate-dependent particle velocity and acoustic density fields $v_{(i;\zeta,n)} \in \mathbb{R}^N,\ \rho_{(i;\zeta,n)} \in \mathbb{R}^N\ (i \in \{1, ..., d\})$ as $\bar{v}_{n-1/2} \in \mathbb{R}^{dN}$ and $\bar{\rho}_n \in \mathbb{R}^{dN}$, respectively. We also define $\bar{p}_n = p_{(\zeta,n)} \in \mathbb{R}^N$. A stack of all these vector fields yields $\bar{z}_n = \{\bar{v}_{n-1/2}^T\ \bar{\rho}_n^T\ \bar{p}_n^T\}^T \in \mathbb{R}^{(2d+1)N}$ at time step n.

Also, let $\mathcal{S} \in \mathbb{R}^{(2d+1)NN_t\times N}$ be the map from the discretised sought after initial pressure P_0 to an additive source S. In particular, at time step n, $S_{n+1/2} = \mathcal{S}_n P_0$, where

$$\mathcal{S}_n P_0 = \Delta t\ \mathcal{I}_s^z s_{(i;\zeta;n+1/2)} = \mathcal{I}_s^z 1_i \otimes \left(\frac{1}{2dC}\right) \mathbb{S} P_0 \quad \text{for } n = -1,\ 0, \tag{4.55}$$

and $\mathcal{S}_n P_0 = 0$ for $n \neq -1,\ 0$. Here, $s_{(i;\zeta;n+1/2)}$ has been derived from (4.50), and $1_i \otimes \in \mathbb{R}^{dN\times N}$ is the adjoint of $\sum_{i=1}^d \in \mathbb{R}^{N\times dN}$, which is the operator which sums over index i. Also, from (4.52) and (4.53), $\mathcal{I}_s^z \in \mathbb{R}^{(2d+1)N\times dN}$ is the map from $s_{(i;\zeta;n+1/2)}$ to $\bar{z}_n$ in the form

$$\mathcal{I}_s^z = \begin{bmatrix} 0_{dN\times dN} \\ I_{dN\times dN} \\ C\left((I_N - \bar{\eta} Y_{\text{dis}}) \sum_{i=1}^d\right)_{N\times dN} \end{bmatrix}. \tag{4.56}$$

We now introduce an operator $T \in \mathbb{R}^{(2d+1)N\times(2d+1)N}$ for defining a discretised formula

for the time sequence of fields as

$$\bar{z}_{n+1} = T\bar{z}_n + S_{n+1/2}. \tag{4.57}$$

Additionally, we introduce a measurement matrix $\boldsymbol{M} = \mathbb{M}\mathcal{I}_z^p$ with $\mathcal{I}_z^p \in \mathbb{R}^{N\times(2d+1)N}$ the projection from the space of $\bar{z}_n$ to the space of $\bar{p}_n$. Based on this, we now define our discretised acoustic forward operator.

Definition 4.4. The discretised acoustic forward operator is defined by

$$\mathbb{H}_a : \mathbb{R}^N \to \mathbb{R}^{N_s N_t} \tag{4.58}$$

$$\mathbb{H}_a = \mathbb{H}_T \mathcal{S}, \tag{4.59}$$

where $\mathbb{H}_T : \mathbb{R}^{(2d+1)NN_t} \to \mathbb{R}^{N_s N_t}$ satisfies

$$\hat{P} = \mathbb{H}_T S, \quad \hat{P}_n = \boldsymbol{M}\bar{z}_n \quad (n \in \{0, ..., N_t - 1\}), \quad \hat{P} = \left[\hat{P}_{n+1}\right]_{n=-1}^{N_t-2}. \tag{4.60}$$

Here, $\bar{z}_n$ is determined by (4.57) with initial condition $\bar{z}_{-1} = 0$, and $\hat{P} \in \mathbb{R}^{N_s N_t}$ is a time-series stack of measured data at iterates $n \in \{0, ..., N_t - 1\}$.

Lemma 4.2. The action of the adjoint operator

$$\mathbb{H}_a^* : \mathbb{R}^{N_s N_t} \to \mathbb{R}^N, \tag{4.61}$$

$$\mathbb{H}_a^* = \mathcal{S}^* \mathbb{H}_T^*, \tag{4.62}$$

on $\hat{P}$, where $P_0 = \mathbb{H}_a^* \hat{P}$, is given by

$$P_0 = \sum_{n=-1}^{N_t-2} \mathcal{S}_n^* \bar{z}_{N_t-2-n}^*, \tag{4.63}$$

where $\bar{z}_n^*$ is determined by

$$\bar{z}_{-1}^* = 0, \quad \bar{z}_{n+1}^* = T^*\bar{z}_n^* + \boldsymbol{M}^*\hat{P}_{N_t-2-n} \quad (n \in \{-1,\ \dots\ , N_t - 2\}). \tag{4.64}$$

Proof. We have given the proof in [27], Lemma 2 and Corollary 1. □

Note that in the case we are considering the sum in (4.63) is actually just two terms. By commuting $n = -1, 0$ in (4.55) with $\boldsymbol{M}$ using the fact that the forward operator

is linear, we define a source for our acoustic adjoint operator as

$$s^{\text{adj}}_{n+1/2} = \mathbb{M}^* \begin{cases} \hat{P}_{N_t-1}, & n = -1 \\ \hat{P}_{N_t-n-1} + \hat{P}_{N_t-n-2}, & n = 0, ..., N_t - 2 \\ \hat{P}_0, & n = N_t - 1. \end{cases} \tag{4.65}$$

From (4.65), equation (4.64) in Lemma 4.2 can be modified as

$$\bar{z}^*_{-1} = 0, \quad \bar{z}^*_{n+1} = T^* \bar{z}^*_n + S^{\text{adj}}_{n+1/2} \quad (n \in \{-1, \ ... \ , N_t - 2\}), \tag{4.66}$$

where

$$S^{\text{adj}}_{n+1/2} = \left(\mathcal{I}^p_z\right)^* s^{\text{adj}}_{n+1/2}. \tag{4.67}$$

Now, we derive a matrix form of T and T^* using the forward model presented in section 4.3.2 to show how multiplication by each of them may be computed. We also denote a matrix form of the k-space spatial gradient in (4.45) as

$$\nabla^{\pm}_i = \frac{\partial}{\partial r^{\pm}_i} \in \mathbb{R}^{N \times N}. \tag{4.68}$$

Using this, we will use $\Phi \in \mathbb{R}^{dN \times dN}$ and $\Psi \in \mathbb{R}^{dN \times dN}$ composed of submatrices

$$(\Phi)_{ij} = -\Delta t \ \delta_{ij} \nabla^+_j \tag{4.69}$$

and

$$(\Psi)_{ij} = -\Delta t \ \delta_{ij} \nabla^-_j. \tag{4.70}$$

To make the notation more compact we also introduce matrices $D, E \in \mathbb{R}^{N \times dN}$ and $G \in \mathbb{R}^{N \times N}$ whose actions are given by

$$D\bar{v} = C \left(I_N - \bar{\eta} Y_{\text{dis}} - \bar{\tau} \frac{Y_{\text{abs}}}{\Delta t} \right) \sum_{i=1}^{d} (AQ\Psi A^2 \bar{v})_i \tag{4.71}$$

$$E\bar{\rho} = C \left(I_N - \bar{\eta} Y_{\text{dis}} \right) \sum_{i=1}^{d} (A^2 \bar{\rho})_i \tag{4.72}$$

$$G\bar{p} = C \left(I_N - \bar{\eta} Y_{\text{dis}} - \bar{\tau} \frac{Y_{\text{abs}}}{\Delta t} \right) \sum_{i=1}^{d} (AQ\Psi A Q_s^{-1} \Phi 1_i \otimes \bar{p})_i \tag{4.73}$$

The matrix T is then given, in block form, by

$$T = \begin{pmatrix} A^2 & 0 & AQ_s^{-1}\Phi 1_i\otimes \\ AQ\Psi A^2 & A^2 & AQ\Psi AQ_s^{-1}\Phi 1_i\otimes \\ D & E & G \end{pmatrix}. \tag{4.74}$$

From (4.74), the adjoint operator T^* will be in the form

$$T^* = \begin{pmatrix} A^2 & A^2\Psi^* QA & D^* \\ 0 & A^2 & E^* \\ \sum_{i=1}^d \Phi^* Q_s^{-1} A & \sum_{i=1}^d \Phi^* Q_s^{-1} A\Psi^* QA & G^* \end{pmatrix}. \tag{4.75}$$

From (4.75) and using (4.56), (4.66) and (4.67), as well as calculating the adjoints of D, E and G from (4.71), (4.72), and (4.73), the time sequence of adjoint fields is given iteratively by

$$\begin{aligned} \bar{\rho}_{n+1} &= A^2\left[\bar{\rho}_n + 1_i \otimes (I_N - Y_{\text{dis}}\bar{\eta})\, C\bar{p}_n\right] \\ \bar{v}_{n+1/2} &= A^2\left[\bar{v}_{n-1/2} + \Psi^* QA^{-1}\left(\bar{\rho}_{n+1} - A^2 1_i \otimes \frac{Y_{\text{abs}}}{\Delta t}\bar{\tau} C\bar{p}_n\right)\right] \\ \bar{p}_{n+1} &= \left(\sum_{i=1}^d \Phi^* Q_s^{-1} A^{-1}\bar{v}_{n+1/2}\right) + s^{\text{adj}}_{n+1/2} \\ \bar{p}^{\text{sol}} &= \boldsymbol{\vartheta}\left(\sum_{i=1}^d 1_i \otimes (I_N - Y_{\text{dis}}\bar{\eta})\, C\bar{p}_{n+1} + \bar{\rho}_{n+1}\right)\ (n = N_t - 2), \end{aligned} \tag{4.76}$$

where $\bar{p}^{\text{sol}} := \mathbb{H}_a^*(\hat{P})$, and $\boldsymbol{\vartheta} = \mathbb{S}\left(\frac{1}{2dC}\right)$ using (4.55). Now, applying the replacements $\hat{\boldsymbol{\rho}}_{n+1} = QA^{-1}\bar{\rho}_{n+1}$ and $\hat{\boldsymbol{v}}_{n+1} = (AQ_s)^{-1}\bar{v}_{n+1/2}$, together with $\Phi^* = -\Psi$ and $\Psi^* = -\Phi$, gives

$$\begin{aligned} \hat{\boldsymbol{\rho}}_{n+1} &= A\left[A\hat{\boldsymbol{\rho}}_n + Q1_i \otimes (I_N - Y_{\text{dis}}\bar{\eta})\, C\bar{p}_n\right] \\ \hat{\boldsymbol{v}}_{n+1/2} &= A\left[A\hat{\boldsymbol{v}}_{n-1/2} - Q_s^{-1}\Phi\left(\hat{\boldsymbol{\rho}}_{n+1} - QA1_i \otimes \frac{Y_{\text{abs}}}{\Delta t}\bar{\tau} C\bar{p}_n\right)\right] \\ \bar{p}_{n+1} &= \left(\sum_{i=1}^d -\Psi\hat{\boldsymbol{v}}_{n+1/2}\right) + s^{\text{adj}}_{n+1/2} \\ \bar{p}^{\text{sol}} &= \boldsymbol{\vartheta}\left(\sum_{i=1}^d 1_i \otimes (I_N - Y_{\text{dis}}\bar{\eta})\, C\bar{p}_{n+1} + AQ^{-1}\hat{\boldsymbol{\rho}}_{n+1}\right)\ (n = N_t - 2). \end{aligned} \tag{4.77}$$

Our numerical experiments showed that an operator $\mathbb{H}_a^*$ that is calculated using a *discretise-then-adjoint* method (4.77) satisfies an adjoint test with a higher accuracy than using an *adjoint-then-discretise* method used in [26]. Therefore, we used (4.77) for the acoustic adjoint operator.

4.4 Iterative model-based approaches for the direct problem of QPAT

Having defined a discretisation of opto-acoustic forward operator $\mathbb{H}$, the Fréchet derivative operator $\mathbb{J}$, and its adjoint $\mathbb{J}^*$, we now explain the iterative approaches we will use for minimisation of (4.12).

Considering (4.2), the dependence of heating field h on κ, μ is nonlinear. Furthermore, because of high scattering of light in tissue media, simultaneous reconstruction of κ, μ from h can be highly ill-posed. As a result, a minimisation of (4.12) is a non-convex, nonlinear and ill-posed inverse problem. It has been shown that simultaneous reconstruction of κ and μ using a single optical excitation, i.e., $N_q = 1$, does not have a unique solution [4]. However, the uniqueness and stability, in appropriate norms, of this inverse problem using $N_q > 1$ optical excitations under some geometric constraints has been established [4]. It is also worth mentioning that using our forward acoustic operator, which can be adapted to acoustically heterogeneous and lossy media, the direct problem of QPAT is more ill-posed than existing studies, for which spherical mean Radon transform or Green's function techniques have been used for solving the acoustic portion of the forward operator using an acoustically homogeneous and lossless medium, which does not hold in practice [20, 18, 31].

Remark 4.3. Since the magnitude of $\hat{\kappa}$ is often 1 to 2 orders of magnitude greater than $\hat{\mu}$, we follow [19, 18] and minimise ϵ (cf. (4.12)) with respect to scaled coefficients

$$\bar{X} = [\bar{\kappa}(\hat{\kappa}), \bar{\mu}(\hat{\kappa})]^T. \tag{4.78}$$

We will use two types of scaling. For the method described in section 4.5.1 we use a linear scaling (see (4.99)) while for the methods described in sections 4.5.2 and 4.5.2 we use logarithmic scaling (see (4.100)). In our numerical experiments the method of section 4.5.1 did not converge with logarithmic scaling.

Using these, we now consider a minimisation problem with respect to a scaled vector of optical coefficients $\bar{X}$ in the form

$$\bar{X}_* = \underset{\bar{X}_l \leqslant \bar{X} \leqslant \bar{X}_u}{\operatorname{argmin}} \ \epsilon(\bar{X}) = \underset{\bar{X}_l \leqslant \bar{X} \leqslant \bar{X}_u}{\operatorname{argmin}} \ \frac{1}{2} \sum_{q=1}^{N_q} \left\| \mathbb{H}_q[X[\bar{X}]] - \hat{P}_q \right\|_2^2, \tag{4.79}$$

where ϵ is a non-convex, nonlinear and smooth function. Also, $\bar{X}$ has been constrained by a lower bound $\bar{X}_l$ and an upper bound $\bar{X}_u$. Here, $\bar{X}_l$ and $\bar{X}_u$ are vectors with the same size as $\bar{X}$, and $\bar{X}_l \leqslant \bar{X} \leqslant \bar{X}_u$ indicates that $\left(\bar{X}\right)_j \geqslant \left(\bar{X}_l\right)_j$ and $\left(\bar{X}\right)_j \leqslant \left(\bar{X}_u\right)_j$ for all components j of these vectors. Note that here, we choose $\left(\bar{X}_l\right)_j$ and $\left(\bar{X}_u\right)_j$ fixed for all components j.

We will take two main approaches for solving (4.79), the first of which is a direct minimisation of the nonlinear objective function using a Quasi-Newton approach, and the second is to solve the minimsation problem as a sequence of convex and linearised subproblems using a matrix-free Jacobian-based method. We now explain these approaches.

A fixed point iteration arising from the optimality conditions of ϵ gives a sequence

$$\bar{X}_{k+1} = \bar{X}_k + \alpha_k d_k, \tag{4.80}$$

where d_k is a search direction for iteration k, and α_k is a step size along search direction d_k. (Throughout this manuscript, a subscript (resp. superscript) k indicates an iteration for an inner (resp. outer) loop.)

4.4.1 Newton's methods

From the second-order optimality condition for minimising ϵ, the *Newton search direction*, is derived using

$$d_k = -\mathcal{H}_k^{-1} \nabla \epsilon_k. \tag{4.81}$$

Here, $\nabla \epsilon_k$ is the first-order derivative of ϵ at $\bar{X}_k$, and is computed using

$$\nabla \epsilon_k = \sum_{q=1}^{N_q} \frac{\partial X}{\partial \bar{X}}[\bar{X}_k]\ \mathbb{J}_q^*[X_k] \left(\mathbb{H}_q[X_k] - \hat{P}_q\right), \tag{4.82}$$

and $\mathcal{H}_k$, is the second-order derivative of ϵ (the Hessian matrix). Using a *Gauss-Newton* method, the Hessian matrix is approximated using

$$\mathcal{H}_k = \sum_{q=1}^{N_q} \frac{\partial X}{\partial \bar{X}}[\bar{X}_k]\ \mathbb{J}_q^*[X_k] \mathbb{J}_q[X_k] \frac{\partial X}{\partial \bar{X}}[\bar{X}_k], \tag{4.83}$$

where a term including the second-order derivatives of $\mathbb{H}_q$ has been neglected as compared to the exact Hessian. A class of approaches that utilise (4.81) for minimisation

of ϵ are called *Newton's method*. Although *Newton's methods* benefit from a quadratic (optimal) rate of convergence, they pose some practical limitations for QPAT, i.e.,

a) The computation, storage and inversion of the Hessian matrix is expensive. To address this problem, an implicit inversion of $\mathcal{H}_k$ using an explicit form of $\mathbb{J}$ and $\mathbb{J}^*$ has been used in the context of diffuse optical tomography [36].

b) The size of the discretised heating field H, which represents the output of the optical portion of the discretised forward operator, is on the same order as the size of unknown parameters (cf. (4.28)). It has been shown that an explicit computation of the matrix $\mathbb{J}_o$ (resp. $\mathbb{J}$) requires at least $N_s + N_e$ times implementation of $\mathbb{H}_o$ (resp. $\mathbb{H}$) (See [19], page 11).

c) Considering the time series of measured data $\hat{P}$, which is of size $N_s N_t$, the Jacobian matrix $\mathbb{J}$ is dense, and thus a storage of $\mathbb{J}$ is impractical.

We will later explain how we have addressed these challenges using an inexact Newton method.

4.4.2 Nonlinear gradient-based methods

An alternative to Newton's method is using gradient-based Quasi-Newton approaches, for which $\mathcal{H}_k$ is not computed explicitly, but is approximated using solely information included in the first-order gradients $\nabla\epsilon$ possibly at previous steps. Additionally, an inversion of $\mathcal{H}_k$ can be avoided using a direct approximation of $\mathcal{H}_k^{-1}$, for which the so-called BFGS method is often used. Since $\mathcal{H}_k^{-1}$ is a dense matrix, and BFGS method poses challenges regarding memory, a limited-memory variant of BFGS (L-BFGS) method is used [19, 18]. Using L-BFGS, $\mathcal{H}_k^{-1}$ is updated using the most recent m pairs of $(\boldsymbol{s}, \boldsymbol{y})$ given by

$$\begin{aligned} \boldsymbol{s}_k &= \bar{X}_k - \bar{X}_{k-1} \\ \boldsymbol{y}_k &= \nabla\epsilon_k - \nabla\epsilon_{k-1}. \end{aligned} \tag{4.84}$$

(In our study, we empirically use $m = 5$.) We will also use $\boldsymbol{\rho}_k = 1/\boldsymbol{y}_k^T \boldsymbol{s}_k$, and an initial guess for the Hessian matrix in the form

$$\mathcal{H}_{k,0}^{-1} = \frac{\boldsymbol{s}_{k-1}^T \boldsymbol{y}_{k-1}}{\boldsymbol{y}_{k-1}^T \boldsymbol{y}_{k-1}} I, \tag{4.85}$$

where I is the identity matrix. By applying L-BFGS method to the constrained minimisation problem (4.79), the search direction d_k is computed using Algorithm 4.1. (See [19] and [18] for applications on QPAT and direct QPAT.)

Algorithm 4.1 L-BFGS (search direction): inner iteration k

Input: $\nabla\varepsilon_k, \mathcal{H}_{k,0}^{-1}$
Initialise: $\boldsymbol{q} = \nabla\varepsilon_k$
for $i = k-1, k-2, ..., k-m$ **do**
    $\boldsymbol{\alpha}_i = \boldsymbol{\rho}_i \boldsymbol{s}_i^T \boldsymbol{q}$
    $\boldsymbol{q} = \boldsymbol{q} - \boldsymbol{\alpha}_i \boldsymbol{y}_i$
end for
$\boldsymbol{r} = \mathcal{H}_{k,0}^{-1} \boldsymbol{q}$
for $i = k-m, k-m+1, ..., k-1$ **do**
    $\boldsymbol{r} = \boldsymbol{r} + \boldsymbol{s}_i \left(\boldsymbol{\alpha}_i - \boldsymbol{\rho}_i \boldsymbol{y}_i^T \boldsymbol{r}\right)$
end for
$\boldsymbol{d}_k = -\boldsymbol{r}$.

From the First-order Karush Kuhn Tucker (KKT) conditions associated with the constraint on $\bar{X}$, d_k is projected onto the feasible region using (See [18], equation (37))

$$d_k = \begin{cases} -\bar{X}_l, & \text{if } \bar{X}_k + \boldsymbol{d}_k \leqslant \bar{X}_l \\ \boldsymbol{d}_k, & \text{if } \bar{X}_l < \bar{X}_k + \boldsymbol{d}_k < \bar{X}_u \\ -\bar{X}_u, & \text{if } \bar{X}_k + \boldsymbol{d}_k \geqslant \bar{X}_u. \end{cases} \tag{4.86}$$

Using d_k given by (4.86), the step size in (4.80) is chosen by a standard backtracking line search satisfying the *Wolfe* conditions as well as the constraints

$$\begin{aligned} &\epsilon(\bar{X}_k + \alpha_k d_k) \leqslant \epsilon(\bar{X}_k) + c_1 \alpha_k d_k^T \nabla\epsilon(\bar{X}_k) \\ &d_k^T \nabla\epsilon(\bar{X}_k + \alpha_k d_k) \geqslant c_2 d_k^T \nabla\epsilon(\bar{X}_k), \end{aligned} \tag{4.87}$$

together with enforcing the bounds associated with the constraint, i.e.,

$$\bar{X}_l \leqslant \bar{X}_k + \alpha_k d_k \leqslant \bar{X}_u. \tag{4.88}$$

(See [19, 18].) In (4.87), $0 < c_1 < c_2 < 1$ are user-defined parameters. Applying these conditions, α_k is chosen using a backtracking line search, as given in Algorithm 4.2. Here, $\boldsymbol{\tau} < 1$ is a user-defined parameter.

Algorithm 4.2 Backtracking Line Search: inner iteration k

1: Input: $c_1, c_2, \boldsymbol{\tau}, \bar{X}_l, \bar{X}_u$
2: Initialise: $\alpha_0 = 1$
3: **while** (4.87) or (4.88) are not satisfied **do**
4: $\quad \alpha = \boldsymbol{\tau}\alpha$
5: **end while**
6: $\alpha_k = \alpha_*$.

4.5 Total Variation (TV) regularisation

To mitigate the ill-posedness of the problem, a regularisation functional must be added to the data fidelity term in (4.79) [19, 18]. This results in a minimisation problem in the form

$$\bar{X}_* = \underset{\bar{X}_l \leqslant \bar{X} \leqslant \bar{X}_u}{\arg\min} \left\{ \mathcal{F} := \epsilon[\bar{X}] + \lambda \mathcal{J}[\bar{X}] \right\}, \tag{4.89}$$

where $\mathcal{J}[\bar{X}]$ and λ, respectively denote the regularisation functional and the regularisation parameter, the latter of which makes a balance between a fidelity to the measured data $\hat{P}$ and to *a priori* knowledge about the true solution. In our study, based on an assumption that the optical coefficients are piecewise constant with sharp edges, we use $\mathcal{J}[\bar{X}] := \mathcal{R}[\bar{\kappa}] + \mathcal{R}[\bar{\mu}]$, where $\mathcal{R}[u]$ is a discretisation of the *Total-Variation (TV)* functional

$$\int_\Omega |\nabla u| dr \tag{4.90}$$

with u either $\bar{\kappa}$ or $\bar{\mu}$. Using (4.90),

$$\mathcal{J}[\bar{X}] = \|D\bar{X}\|_1 \tag{4.91}$$

with

$$D = \begin{bmatrix} D^{\bar{\kappa}} & 0_{N_l \times N_e} \\ 0_{N_l \times N_e} & D^{\bar{\mu}} \end{bmatrix}. \tag{4.92}$$

Here, $D^u \in \mathbb{R}^{N_l \times N_e}$ is a sparse matrix with N_e and N_l the total number of elements and the total number of internal edges between elements, respectively. Each row $D^u_l \in \mathbb{R}^{1 \times N_e}$ ($l \in \{1, ..., N_l\}$) has two nonzero components at indices j_1 and j_2, which correspond to two elements connected by the internal edge l. These have values a_l

and $-a_l$ with a_l the length (or area) of internal edge l [8]. Using this, the gradient of $\mathcal{J}(\bar{X})$ is a nonlinear operator in the form

$$\begin{aligned} &\nabla\mathcal{J} : \mathbb{R}^{2N_e} \to \mathbb{R}^{2N_e} \\ &\nabla\mathcal{J}[\bar{X}] = M[\bar{X}]\bar{X}, \end{aligned} \tag{4.93}$$

where $M(\bar{X})$ is given by

$$M[\bar{X}] = D^T C[\bar{X}] D \tag{4.94}$$

with $C(\bar{X})$ a diagonal matrix

$$C[\bar{X}] = \operatorname{diag}\left((|D\bar{X}|^2 + \beta)^{-1/2}\right). \tag{4.95}$$

Here, the smoothing parameter β is added in order to make $\nabla\mathcal{J}[\bar{X}]$ differentiable. Having defined our regularisation functional, we now explain the minimisation approaches we use for solving the direct problem of QPAT.

4.5.1 TV regularisation using Alternating Direction Method of Multipliers (ADMM)

Two major issues for minimisation of $\mathcal{F}$ is the nonlinearity of $\nabla\mathcal{J}[\bar{X}]$ and a loss of accuracy due to the smoothing parameter β. Note that a small value for β may deteriorate the convergence [47]. One way for addressing these difficulties is to use a slack variable for shifting the gradient of $\|D\bar{X}\|_1$ out of the non-differentiable region and penalising the applied shift. To do this, the Augmented Lagrangian is introduced which, following [19, 18], may be further rewritten as

$$\mathcal{F}_A(W, \bar{X}) = \varrho\left(\frac{1}{2}\|D\bar{X} - W + U_w\|_2^2 + \nu\|W\|_1\right) + \sum_{q=1}^{N_q}\frac{1}{2}\|\mathbb{H}_q[X(\bar{X})] - \hat{P}_q + U_{p,q}\|_2^2, \tag{4.96}$$

where ν and ϱ are constants and U_w and $U_{p,q}$ are rescaled Lagrange multipliers. Minimisation of $\mathcal{F}$ is then accomplished by alternating minimisation of (4.96) in W and $\bar{X}$, and updating of the Lagrange multipliers, using Algorithm 4.3. In Algorithm 4.3, Tol_{out} is a terminating threshold, and X^* denotes an optimal solution. The line 4 in Algorithm 4.3 is a minimisation of the first term in (4.96) with respect to W, and can

Algorithm 4.3 Gradient-based Quasi-Newton method using ADMM

1: Input: $\hat{P}_q,\ (q \in \{1, ..., N_q\})$
2: Initialise: $X^0, W^0 = 0$
3: **while** $\|\nabla_{\bar{X}} \mathcal{F}_A\left(\bar{X}^k, W^k\right)\| > Tol_{\text{out}}$ **do**
4: $\quad W^{k+1} = \arg\min_{W} \mathcal{F}_A\left(\bar{X}^k, W\right)$
5: $\quad \bar{X}^{k+1} = \arg\min_{\bar{X}_l \leqslant \bar{X} \leqslant \bar{X}_u} \mathcal{F}_A\left(\bar{X}, W^{k+1}\right)$
6: $\quad U_w^{k+1} = U_w^k + D\bar{X}^{k+1} - W^{k+1}$
7: $\quad U_{p,q}^{k+1} = U_{p,q}^k + \mathbb{H}_q[X^{k+1}] - \hat{P}_q\ (\forall q)$
8: $\quad$ Output: X^*
9: **end while**

be calculated exactly using a scalar-wise *Shrinkage formula* of the form

$$W^{k+1} = \max\left\{|D\bar{X}^k + U_w^k| - \nu, 0\right\} \operatorname{sgn}(D\bar{X}^k + U_w^k). \tag{4.97}$$

Additionally, the line 5 in this algorithm is a minimisation of $\mathcal{F}_A$ at W^{k+1} with respect to $\bar{X}$, and is done using L-BFGS algorithm, as explained in section 4.4.2. Note that we should replace ϵ by $\mathcal{F}_A$, and we hope this is not confusing for the reader. To do this, the first-order derivative of $\mathcal{F}_A$ with respect to $\bar{X}$ at inner iteration k is computed using

$$\nabla_{\bar{X}} \mathcal{F}_A = \varrho D^T\left(D\bar{X} - W + U_w\right) + \sum_{q=1}^{N_q} \frac{\partial X}{\partial \bar{X}}[\bar{X}]\mathbb{J}_q^*[X]\left(\mathbb{H}_q[X] - \hat{P}_q + U_{P_q}\right). \tag{4.98}$$

For this method we use a linear scaling

$$\bar{X} = \begin{bmatrix} \bar{\kappa} \\ \bar{\mu} \end{bmatrix} = \begin{bmatrix} \frac{\hat{\kappa}}{\text{mean}(\hat{\kappa}^0)} \\ \frac{\hat{\mu}}{\text{mean}(\hat{\mu}^0)} \end{bmatrix} \tag{4.99}$$

where the dominators are the mean value of initial guesses $\hat{\kappa}^0$ and $\hat{\mu}^0$.

4.5.2 Linearised matrix-free Jacobian-based method

In this section, we explain two methods that we use for solving the problem (4.79) as a sequence of linearised subproblems. Here, we will make a balance between reconstruction of $\hat{\kappa}$ and $\hat{\mu}$ by using the logarithmic rescaling

$$\bar{X} = \log \frac{X}{X^0}, \tag{4.100}$$

which also implicitly enforces positivity on $\hat{\kappa}$ and $\hat{\mu}$. Note that in (4.100), the division X/X^0 and the Logarithm operator are understood elementwise. Additionaly, we set

$\bar{X}_l = -\infty$ and $\bar{X}_u = +\infty$. This gives a non-constrained form of the problem. Our numerical results showed that enforcing bounds on the solutions is not required because of a good stability provided by these approaches. Also, X^0 denotes an initial guess.

Accordingly, given an iterate $\bar{X}^k$ and a point $\bar{X}$ in a neighborhood of $\bar{X}^k$, the forward operator $\mathbb{H}$ is linearised using an approximation

$$\mathbb{H}(\bar{X}) \approx \mathbb{H}(\bar{X}^k) + \mathbb{J}[X^k]\frac{\partial X}{\partial \bar{X}}[\bar{X}^k](\bar{X} - \bar{X}^k). \tag{4.101}$$

Applying the approximation (4.101) on the problem (4.79) yields the minimisation problem

$$d^k = \underset{d}{\operatorname{argmin}} \ \frac{1}{2}\sum_{q=1}^{N_q} \left\| \mathbb{J}_q[X^k]\frac{\partial X}{\partial \bar{X}}[\bar{X}^k]d - \left(\hat{P}_q - \mathbb{H}_q[X^k]\right)\right\|_2^2, \tag{4.102}$$

where we have changed from $\bar{X}$ to $d = \bar{X} - \bar{X}^k$ in the minimisation. Note that we have used k as a superscript in order to indicate a linearised subproblem (outer iteration), as opposed to a subscript in (4.81) that indicates an inner iteration. The k-th linearised subproblem (4.102) gives a *normal equation* in the form of

$$\mathcal{H}^k d^k = -\nabla\epsilon^k, \tag{4.103}$$

where $\nabla\epsilon^k$ and $\mathcal{H}^k$ are obtained from (4.82) and (4.83), respectively. The normal equation (4.103) is a variant of (4.81), for which $\mathcal{H}^k$ is approximated using a *Gauss-Newton* method. Here, to avoid a storage of $\mathbb{J}$, we solve each linearised subproblem (4.103) using a *Krylov subspace* method in a matrix-free manner, for which implicit forms of operators $\mathbb{J}$ and $\mathbb{J}^*$ are used.

As discussed in section 4.4.1, *Newton's methods* converge rapidly, but solving a normal equation with a high accuracy for each linearisation is very expensive. From a theoretical point of view, using *Krylov* methods, the total number of iterations for reaching a minimiser is on the same order of the number of unknowns. Therefore, we solve (4.103) roughly using a loose stopping tolerance, i.e.,

$$\mathcal{H}^k \tilde{d}^k = -\nabla\epsilon^k + v_k, \quad \|v^k\|_2/\|\nabla\epsilon^k\|_2 \leqslant \eta^k, \tag{4.104}$$

where $\tilde{d}^k$ denotes a rough solution. It has been shown that under assumptions that $\mathcal{H}^k$ is symmetric and positive definite, the solutions $\tilde{d}^k$ are sufficiently small and $\eta^k < 1$, the local convergence is guaranteed using a step size $\alpha^k = 1$ (cf. [14], section 2).

A class of approaches that use (4.104) for minimisation of (4.79) are called *Inexact Newton's methods* [14]. In the sequel, we use d^k, rather than $\tilde{d}^k$, for indicating a rough solution of (4.104), and we hope this is not confusing for the readers.

Lagged diffusivity (LD) method with priorconditioning

As discussed above, for our direct QPAT problem, the Hessian matrices $\mathcal{H}^k$ are ill-conditioned. Therefore, a regularisation functional must be added in order to stabilise the problem. Our first approach for an inclusion of TV regularisation in (4.104) is based on solving (4.89) via an iterative linearisation of an associated objective function $\mathcal{F}$. Strictly speaking, we first add the regularisation functional to an original nonlinear problem, and then the linearisations are applied to a regularised form of a nonlinear objective function. The k-th linearisation of the data fidelity term ϵ using (4.102), together with $\nabla\mathcal{J}[\bar{X}]$ defined by (4.93), gives a TV regularised variant of (4.104) in the form

$$\mathcal{H}^k d^k + \lambda M[\bar{X}]\bar{X} = -\nabla\epsilon^k. \tag{4.105}$$

Let us denote an initial guess for k-th subproblem by $\bar{X}_0^k$, which is calculated using the previous linearised subproblem $k-1$. One way for addressing the nonlinearity of $M[\bar{X}]\bar{X}$ (cf. (4.94)) is replacing $M[\bar{X}]$ by $M^k = M[\bar{X}_0^k]$. This gives a normal equation in the form

$$\left(\mathcal{H}^k + \lambda M^k\right) d^k = -\nabla\epsilon^k - \lambda M^k \bar{X}_0^k. \tag{4.106}$$

Linearisation of M using the above equation is called the *Lagged Diffusivity* (LD) method [47]. (See [22] for an application of LD on the purely optical problem of QPAT.) Our method now is to follow [2] in order to convert (4.106) into a similar problem in which the regularisation is obtained by early termination of an iterative method rather than tuning of parameter λ. First, since M^k may only be positive semi-definite we approximate it by $M_\gamma^k = M^k + \gamma I$. Replacing M^k by M_γ^k in (4.106) we next multiply by $(M_\gamma^k)^{-1}$ so that (4.106) becomes

$$\left((M_\gamma^k)^{-1}\mathcal{H}^k + \lambda I\right) d^k = -(M_\gamma^k)^{-1}\nabla\epsilon^k - \lambda\bar{X}_0^k \tag{4.107}$$

Applying *Krylov* methods for solving (4.106), the iterates lie in a subspace [2]

$$\mathcal{K}^{(M_\gamma^k)^{-1}\mathcal{H}^k+\lambda I} = \mathcal{K}^{(M_\gamma^k)^{-1}\mathcal{H}^k} = \text{span}\left\{ -\left((M_\gamma^k)^{-1}\mathcal{H}^k\right)^{i\in\left\{0,\dots,i_{\max}^k-1\right\}} \left((M_\gamma^k)^{-1}\nabla\epsilon^k + \lambda\bar{X}_0^k\right)\right\}, \tag{4.108}$$

where $i_{\max}^k$ is the maximum number of iterations for the *Krylov* method. Our numerical experience shows that using small values for λ, the term $\lambda\bar{X}_0^k$ will have a small effect, and indeed we drop this in our method by taking $\lambda = 0$.

Using this approach for applying regularisation on a normal equation is called *priorconditioning*. Note that in this approach we adjust the regularisation by $i_{\max}$. In contrast to an empirical choice for the regularisation parameter λ, which requires a recomputation of the problem, $i_{\max}$ can be implicitly controlled by a stopping tolerance [2]. Using the above, our subproblem is to solve a *priorconditioned* variant of (4.106) in the form

$$(M_\gamma^k)^{-1}\mathcal{H}^k d^k = -(M_\gamma^k)^{-1}\nabla\epsilon^k. \tag{4.109}$$

It turns out that (4.109) provides a better convergence than (4.106) since the structure of the prior is directly included in the Jacobian matrix [2]. Here, we solve (4.109) using the *Preconditioned Conjugate Gradient* (PCG) method, as outlined in Algorithm 4.4.

Algorithm 4.4 PCG algorithm for solving linearised subproblem k

Input: $\nabla\epsilon^k, \mathcal{H}^k, M_\gamma^k$
Initialise: $i = 0, \bar{X}_0 = 0$
$\boldsymbol{r}_0 = -\nabla\epsilon^k$
Solve $M_\gamma^k \boldsymbol{z}_0 = \boldsymbol{r}_0$
$\boldsymbol{d}_0 = \boldsymbol{z}_0$
while $i < i_{\max} \ \cap \ \left(i < i_m \ \cup \ 1 - \frac{\boldsymbol{r}_i^T\boldsymbol{z}_i}{\boldsymbol{r}_{i-i_m}^T\boldsymbol{z}_{i-i_m}} > Tol_{\text{in}}\right)$ **do**
    $\boldsymbol{\alpha}_i = \frac{\boldsymbol{r}_i^T\boldsymbol{z}_i}{\boldsymbol{d}_i^T\mathcal{H}^k\boldsymbol{d}_i}$
    $\bar{X}_{i+1} = \bar{X}_i + \boldsymbol{\alpha}_i\boldsymbol{d}_i$
    $\boldsymbol{r}_{i+1} = \boldsymbol{r}_i - \boldsymbol{\alpha}_i\mathcal{H}^k\boldsymbol{d}_i$
    Solve $M_\gamma^k\boldsymbol{z}_{i+1} = \boldsymbol{r}_{i+1}$
    $\boldsymbol{\beta}_i = \frac{\boldsymbol{r}_{i+1}^T\boldsymbol{z}_{i+1}}{\boldsymbol{r}_i^T\boldsymbol{z}_i}$
    $\boldsymbol{d}_{i+1} = \boldsymbol{z}_{i+1} + \boldsymbol{\beta}_i\boldsymbol{d}_i$
end while
Output: d^k

We will terminate Algorithm 4.4 if $i > i_m$ with i_m a user-adjusted number of inner iterations, and a relative reduction in $\boldsymbol{r}_i^T\boldsymbol{z}_i$ during i_m inner iterations becomes less than

a user-adjusted threshold Tol_{in}. Also, the PCG algorithm is unconditionally stopped whenever $i > i_{\max}$.

Remark 4.4. Each inner iteration of the PCG loop involves an implicit inversion of the sparse matrix M_γ^k (See [2]). For the direct QPAT, the cost of an inversion of M_γ^k is negligible, compared to an implementation of the Jacobian and its transpose.

Using the LD method, together with a logarithmic scaling, our inexact Newton algorithm is outlined in Algorithm 4.5.

Algorithm 4.5 Inexact Newton method using LD

1: Input: $\hat{P}_q$ $(q \in \{1, ..., N_q\})$
2: Initialise: $k = 0, X^0$
3: **while** $k = 0 \ \cup \ 1 - \frac{\epsilon^k}{\epsilon^{k-1}} > Tol_{\text{out}}$ **do**
4: Apply linearisation on (4.89)
5: Compute d^k from (4.109) using Algorithm 4.4
6: Compute $\bar{X}^{k+1}$ using (4.80) and $\alpha^k = 1$
7: $X^{k+1} = X^0 e^{\bar{X}^{k+1}}$
8: **end while**
9: Output: X^*

Primal-Dual Interior-Point-Method (PD-IPM)

A technique for linearisation of M was developed using a primal-dual method, and was shown to give better convergence than the LD method, especially for small values of β [10]. In [10], the PD-IPM technique was used for enforcing TV regularisation when inverting a linear blurring operator. In contrast to our first approach using the LD method, here we first linearise the data fidelity function, and then add a TV regularisation function to each linearised subproblem using the PD-IPM approach. Using this, we iteratively solve a TV regularised variant of the linearised subproblem (4.102) in the form

$$d^k = \underset{d}{\operatorname{argmin}} \ \frac{1}{2} \sum_{q=1}^{N_q} \left\| \mathbb{J}_q[X_0^k] \frac{\partial X}{\partial \bar{X}}[\bar{X}_0^k] d - \left(\hat{P}_q - \mathbb{H}_q[X_0^k] \right) \right\|_2^2 + \lambda \|Dd\|_1. \tag{4.110}$$

The main idea for linearisation of the nonlinear M (cf. (4.94) and (4.95)) using the PD-IPM approach is introducing a dual parameter $\chi = C[d]Dd \in \mathbb{R}^{2N_l}$. This gives a

system of coupled nonlinear PDEs for subproblem k in the form

$$\begin{aligned} \lambda D^T\chi + \mathcal{H}d - \sum_{q=1}^{N_q} \frac{\partial X}{\partial \bar{X}}[\bar{X}_0]\mathbb{J}_q^*[X_0]\left(\hat{P}_q - \mathbb{H}_q[X_0]\right) = g(\chi, d) = 0 \\ C^{-1}[d]\chi - Dd = f(\chi, d) = 0, \end{aligned} \tag{4.111}$$

where we have removed the superscripts indicating subproblem k for brevity. A linearisation of this system with respect to (χ, d) gives

$$\begin{bmatrix} C^{-1}[d] & -\Big(I - C[d]\chi(Dd)^T\Big)D \\ \lambda D^T & \mathcal{H} \end{bmatrix} \begin{bmatrix} \delta\chi \\ \delta d \end{bmatrix} = \begin{bmatrix} -f(\chi, d) \\ -g(\chi, d) \end{bmatrix} \tag{4.112}$$

Here, we have used the derivative of the terms in the left-hand-sides of (4.111) with respect to χ and d, and the fact that $C^{-1}[d] = \text{diag}\left((|Dd|^2 + \beta)^{1/2}\right)$ from (4.95). If we make the replacement $\chi = C[d]Dd$, then the above linearised system gives decoupled equations

$$\begin{aligned} &\Big[\lambda D^T C[d_{k'}]\Big(I - C[d_{k'}]\chi_{k'}(Dd_{k'})^T\Big)D + \mathcal{H}\Big]\delta d_{k'} \\ =& -\lambda M(d_{k'}) + \sum_{q=1}^{N_q} \frac{\partial X}{\partial \bar{X}}[\bar{X}_0]\mathbb{J}_q^*[X_0]\left(\hat{P}_q - \mathbb{H}_q[X_0] - \mathbb{J}_q[X_0]\frac{\partial X}{\partial \bar{X}}[\bar{X}_0]d_{k'}\right) \end{aligned} \tag{4.113}$$

and

$$\delta\chi_{k'} = C[d_{k'}]\Big(I - C[d_{k'}]\chi_{k'}(Dd_{k'})^T\Big)D\delta d_{k'} - \chi_{k'} + C[d_{k'}]Dd_{k'}, \tag{4.114}$$

where the subscript k' indicates an inner sub-subproblem. The same as the LD method, we solve a priorconditioned form of (4.113) with $\lambda = 0$, i.e.,

$$\tilde{M}_{\gamma,k'}^{-1}\mathcal{H}\delta d_{k'} = -\tilde{M}_{\gamma,k'}^{-1}\nabla\tilde{\epsilon}_{k'}, \tag{4.115}$$

where

$$\tilde{M}_{\gamma,k'} = D^T C[d_{k'}]\Big(I - C[d_{k'}]\chi_{k'}(Dd_{k'})^T\Big)D + \gamma I, \tag{4.116}$$

and $-\nabla\tilde{\epsilon}_{k'}$ is actually the second term in the right-hand-side of (4.113). Here, the subscript k' indicates the fact that for each linearised subproblem k (superscript), we solve a sequence of sub-subproblems (4.114) and (4.115) using an update of $d_{k'}$, $M_{\gamma,k'}$ and $\nabla\tilde{\epsilon}_{k'}$. The developed inexact Newton method using a TV regularisation based on the PD-IPM approach is outlined in Algorithm 4.6. We use step sizes $\alpha_{k'} = 1$ and

Algorithm 4.6 Inexact Newton method using PD-IPM

1: Input: $\hat{P}_q$ $(q \in \{1, ..., N_q\})$
2: Initialise: $k = 0, X^0$
3: **while** $k = 0 \ \cup \ \ 1 - \frac{\epsilon^k}{\epsilon^{k-1}} > Tol_{\text{out}}$ **do**
4: Apply linearisation and derive the objective function in (4.110)
5: Initialise: $\bar{X}_0^k, d_0 = 0, \chi_0 = 0$
6: **while** $k' < k'_{\max} \ \cap \ \left(k' = 0 \ \cup \ 1 - \frac{\boldsymbol{r}_{k',*}}{\boldsymbol{r}_{k'-1,*}} > Tol_{\text{med}}\right)$ **do**
7: Compute $\delta d_{k'}$ from (4.115) using Algorithm 4.4
8: $d_{k'+1} = d_{k'} + \alpha_{k'} \delta d_{k'}$
9: Compute $\delta \chi_{k'}$ using (4.114)
10: $\chi_{k'+1} = \chi_{k'} + s_{k'} \delta \chi_{k'}$
11: **end while**
12: Compute $\bar{X}^{k+1}$ using (4.80) and $\alpha^k = 1$
13: $X^{k+1} = X^0 e^{\bar{X}^{k+1}}$
14: **end while**
15: Output: X^*

$\alpha^k = 1$ for all k' and k, respectively. The step size $s_{k'}$ is described below the algorithm (see (4.117)).

Here, each iteration k' amounts to solving a PCG loop. The optimal $\boldsymbol{r}$ provided by each PCG loop k' is denoted by $\boldsymbol{r}_{k',*}$. Using this, we terminate each outer subproblem k using a stopping criterion given in line 6 in Algorithm 4.6. This stopping criterion uses a stopping threshold Tol_{med}.

Additionally, following [8], we choose $s_{k'}$ using a *step length rule*. Using this approach, we choose

$$s_{k'} = \min\left(1, \varphi_*\right) \delta \chi_{k'}. \tag{4.117}$$

Here, φ_* is the largest φ that satisfies a feasibility condition

$$\left(|\chi_{k'} + \varphi \delta \chi_{k'}|\right)_j \leqslant 1, \ \forall j = 1, ..., 2N_l, \tag{4.118}$$

where j denotes the index of components of χ.

4.6 Numerical results

The TV regularised minimisation approaches that have been explained in section 4.5, i.e., ADMM, LD and PD-IPM, were used for a simultaneous reconstruction of images of optical absorption coefficient μ and diffusion coefficient κ for 2D and 3D phantoms.

4.6.1 2D phantom

The 2D simulation was performed on a square domain $[-5, +4.92] \times [-5, +4.92]$ mm^2.

Optical excitation

Four different optical excitation patterns were used, i.e., $N_q = 4$. For each optical excitation q, we used a discretisation of an inward directed diffuse boundary current $I_{s,q}$ (J/mm) that obeys

$$I_{s,q}(r) = \begin{cases} 1, & r \in \iota_q \\ 0, & r \in \partial\Omega \backslash \iota_q, \end{cases} \tag{4.119}$$

where $\iota_q \subset \partial\Omega$ denotes the source position for optical excitation q, and was set each side of the square for each optical excitation (cf. the second line in equation (4.1)).

Discretisation for data generation

For generation of time series of boundary data, the square domain was discretised using a grid with 128×128 nodes and an even separation distance of 7.81×10^{-2}mm along both Cartesian coordiantes. For the optical portion of the problem, a triangulation was applied so that each two finite elements form a pixel, and the centre of the pixel matches an associated node on the acoustic grid. For the acoustic portion of the problem, to mitigate wave wrapping [40], a perfectly matched layer (PML) having a thickness of 20 grid points and a maximum attenuation coefficient of 2 nepers per grid point was added to each side of the grid. The propagated wavefield was detected in 1017 time steps using 158 detectors that are equidistantly placed on the left and top sides of the computational grid, as shown in figure 4.1. A 30 dB Additive White Gaussian Noise (AWGN) was then added to the simulated data.

Discretisation for image reconstruction

To avoid an inverse crime for discretisation, the image reconstruction was done using a computational grid made up of 80×80 nodes with an even separation distance of 1.25×10^{-1}mm.

Table 4.1: The minimal and maximal values for acoustic properties of the 2D phantom.

	$c_0(\mathrm{ms}^{-1})$		$\rho_0(\mathrm{kgm}^{-3})$	
	min	max	min	max
Data generation	1.129×10^3	1.902×10^3	0.620×10^3	1.390×10^3
Image reconstruction	1.276×10^3	1.725×10^3	0.750×10^3	1.250×10^3

Acoustic properties

To the best of our knowledge, our manuscript reports the first results on the direct QPAT for realistic acoustic media, for which acoustic characteristics of tissue media such as heterogeneity and attenuation are taken into account.

In addition, PAT and QPAT use an assumption that the acoustic properties of the medium are known. This assumption does not hold in practical cases. For example, it has been shown that the acoustic properties of the breast vary up to 15 %. These variations are often not exactly known for reconstruction. As a result, using the same acoustic properties for data generation and image reconstruction may be an inverse crime. To avoid this, for data generation, we corrupted the acoustic properties of the medium with 30 dB AWGN noise. Figures 4.1(a) and 4.1(b) show the contaminated distributions of sound speed (c_0) and ambient density (ρ_0) for data generation, respectively. The minimal and maximal values of these maps are given in the top row of Table 4.1. Using these values, the computational grid for data generation supports a maximal frequency up to 7.223 MHz. (See the *k-Wave* manual [44].) Also, in figures 4.1(a) and 4.1(b), the position of detectors is shown by the black circles matching the left and top sides of the grid.

For image reconstruction, we used the clean maps that are shown in figures 4.1(c) and 4.1(d). The minimal and maximal values of these maps are given in the bottom row of Table 4.1. Using these acoustic maps, the grid for image reconstruction supports a maximal frequency up to 5.101 MHz. From Table 4.1, the incorporated noise has provided a 10-15% relative discrepancy between the acoustic properties used for data generation and image reconstruction.

Furthermore, the acoustic medium was assumed attenuating, where acoustic absorption and dispersion follow a frequency power law [46]. Accordingly, we used a constant attenuation coefficient $\alpha_0 = 0.75$ dB $\mathrm{MHz}^{-y}\,\mathrm{cm}^{-1}$ and an exponent factor $y = 1.5$ for both data generation and image reconstruction (cf. (4.8)). These values

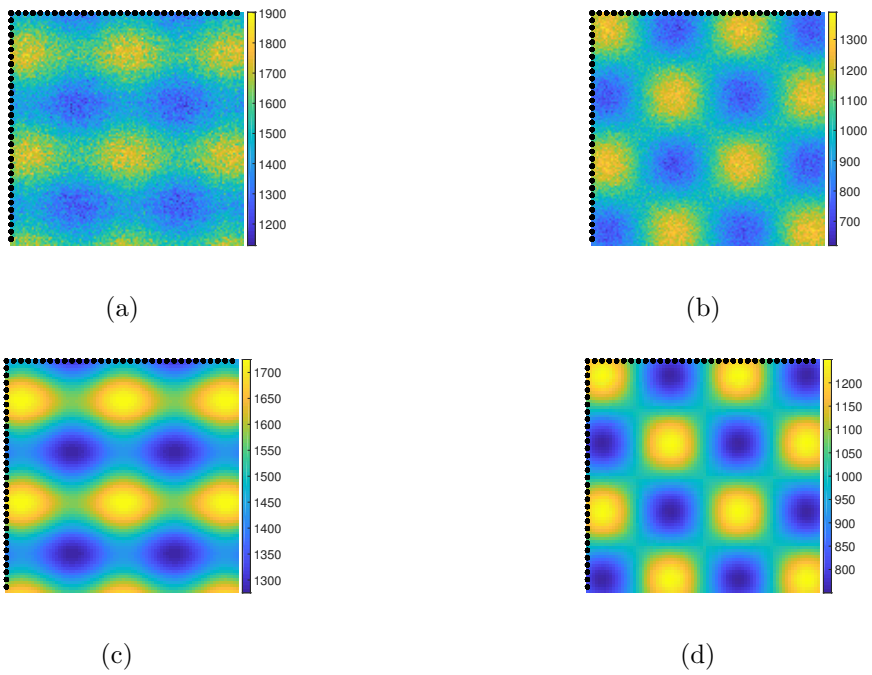


Figure 4.1: Acoustic properties for 2D case. Data generation: (a) c_0 (b) ρ_0, and image reconstruction: (c) c_0 (d) ρ_0.

were chosen so that they approximately simulate the acoustic attenuation properties of the breast.

Optical phantom

We simulate the distributions of optical coefficients so that they follow the optical properties of soft tissues in the sense that they often possess a broad range of values for the optical coefficients. Our numerical experience showed that this is a challenge for image reconstruction, although this issue has been neglected in many of studies of QPAT. The reader is referred to [31] for a study on the direct QPAT, in which this issue has been considered. Accordingly, we simulate a distribution for optical absorption coefficient with 20 values within a range $\mu \in [0.025, 0.325]$ mm^{-1} with a background 0.075 mm^{-1}. Also, a distribution for the diffusion coefficient is simulated so that it has 6 values within a range $\kappa \in [0.2, 0.4]$ mm^{-1} with a background 0.3 mm^{-1}. Figures 4.2(a) and 4.2(b) show the map for μ and κ, respectively.

Image reconstruction

We initialised all algorithms using values 1.2 times more than the mean of optical coefficients for the associated phantoms.

ADMM. The ADMM approach (cf. Algorithm 4.3) was applied for a simultaneous reconstruction of μ and κ. Since ADMM is our benchmark method, the associated parameters were chosen very carefully in order to obtain the best possible image. The line 5 in Algorithm 4.3 is a minimisation of $\mathcal{F}_A$ (cf. (4.96)) with respect to $\bar{X}$ using the L-BFGS algorithm given in Algorithm 4.1. The L-BFGS algorithm uses a backtracking line search given in Algorithm 4.2 with $c_1 = 1 \times 10^{-4}$, $c_2 = 0.9$ and $\tau = 0.25$. Also, we set $\varrho = 1$ (cf.(4.96)), and $\nu = 1 \times 10^{-3}$ for the *Shrinkage* operator (4.97). We terminated the ADMM algorithm using $Tol_{\text{out}} = 1 \times 10^{-2}$ (cf. Algorithm 4.3).

LD. The LD method was applied using Algorithm 4.5. Using this algorithm, each linearised subproblem is solved using a PCG loop (cf. Algorithm 4.4) by setting $i_{\max} = 30$, $i_m = 5$, and $Tol_{\text{in}} = 0$. The latter parameter implies that we terminate each PCG loop, if

$$i > i_{\max} \ \cup \ \left(i > i_m \ \cap \ \boldsymbol{r}_i^T \boldsymbol{z}_i > \boldsymbol{r}_{i-i_m}^T \boldsymbol{z}_{i-i_m}\right). \tag{4.120}$$

Note that we observed a nonmonotone convergence for iterates of each PCG loop (inner iterations) in regions close to an optimal solution X^*, but the sequence $\epsilon(X^k)$ always monotonically converged to $\epsilon(X^*)$ using $\alpha^k = 1$. (We suggest using an Armijo condition for the outer iterations, although we observed that a nonmonotonic reduction in iterates of the PCG loop associated with outer iteration k is sufficient for providing a descent search direction, i.e., $\epsilon(X^{k+1}) < \epsilon(X^k)$). The TV preconditioner was applied using $\gamma = 1 \times 10^{-6}$ and $\beta = 2 \times 10^{-5}$. Our LD algorithm was stopped using $Tol_{\text{out}} = 1 \times 10^{-3}$ (cf. Algorithm 4.5).

PD-IPM. The PD-IPM technique was applied using Algorithm 4.6. Because of applying two layers of linearisation, each of the outer linearised problems are solved using a sequence of inner linearised subproblems. For solving a normal equation associated with each inner linearised subproblem, we terminated each PCG algorithm using the same parameters as in the LD method. The TV preconditioner was applied using $\gamma = 1 \times 10^{-6}$ and $\beta = 1 \times 10^{-6}$.

We terminated each outer linearised suproblem using a threshold $Tol_{med} = 1 \times 10^{-3}$

and $k'_{\max} = 20$ (cf. Algorithm 4.6). Our PD-IPM was terminated using a stopping threshold $Tol_{out} = 1 \times 10^{-3}$.

Evaluation of image reconstruction

The criterion that we use for measuring the convergence of sequence X^k to a ground truth image (phantom) is Relative Error (RE), which is calculated as

$$RE(u^k) = 100 \times \frac{\|u^k - u^{phantom}\|_2}{\|u^{phantom}\|_2}. \tag{4.121}$$

Here, u^k is the solution for either κ or μ at outer iteration k that is interpolated back to the grid for data generation (phantom). Also, the superscript *phantom* indicates the distribution of optical coeffcients for the phantom.

We also consider $\epsilon(X^k)$ as the second criterion for convergence (cf. (4.79)).

Observations

ADMM. Using the parameters given in section 4.6.1, the ADMM algorithm was stopped after outer iteration 6. The final reconstructed images for the optical absorption coefficient μ and diffusion coefficient κ are shown in figures 4.2(c) and 4.2(d), respectively. Figure 4.3(a) shows the RE of sequence computed by the ADMM algorithm at outer iterations.

LD. For the LD algorithm, the associated stopping criterion was satisfied after outer iteration 30. Figures 4.2(e) and 4.2(f) show the final reconstructed images for μ and κ, respectively. Figure 4.3(b) shows the RE of the iterates provided by LD for outer iterations k. The computed values for $\epsilon(X^k)$ are shown in figure 4.4(a). This figure is shown from an enlarged view around the optimal solution in figure 4.4(b). As shown in these figures, ϵ monotonically converges to a minimiser for all outer iterations using our choice for the step size $\alpha^k = 1$ $(\forall k)$.

PD-IPM. The stopping criterion for the PD-IPM algorithm was satisfied after 4 outer iterations. The final reconstructed images for μ and κ are shown in figures 4.2(g) and 4.2(h), respectively. Figure 4.3(c) shows the RE of solutions (optical coefficients) computed by the PD-IPM algorithm for outer iterations k. Figure 4.4(c) shows the obtained values for $\epsilon(X^k)$. This figure is shown from an enlarged view around the optimal solution in figure 4.4(d). As shown in these figures, our choices for the step

sizes associated with outer (resp. inner) subproblems, which are $\alpha^k = 1$ (resp. $\alpha_{k'} = 1$), provided a monotonic reduction for values of ϵ^k (resp. $\epsilon_{k'}$) for all iterations.

Remark 4.5. Using PD-IPM, for both outer and inner linearised subproblems, the first step is a compuation of associated ϵ and the gradient $\nabla\epsilon$ (cf. Algorithm 4.6). As a result, using a line search for compuation of α^k and $\alpha_{k'}$, e.g., a backtracking line search using Wolfe conditions, is straightforward, and does not impose additional computational cost. However, our numerical experience showed that a reduction in $\boldsymbol{r}_i^T \boldsymbol{z}_i$ provided by the PCG loops (cf. Algorithm 4.4) is sufficient for a monotonic reduction of the objective function without using a line search.

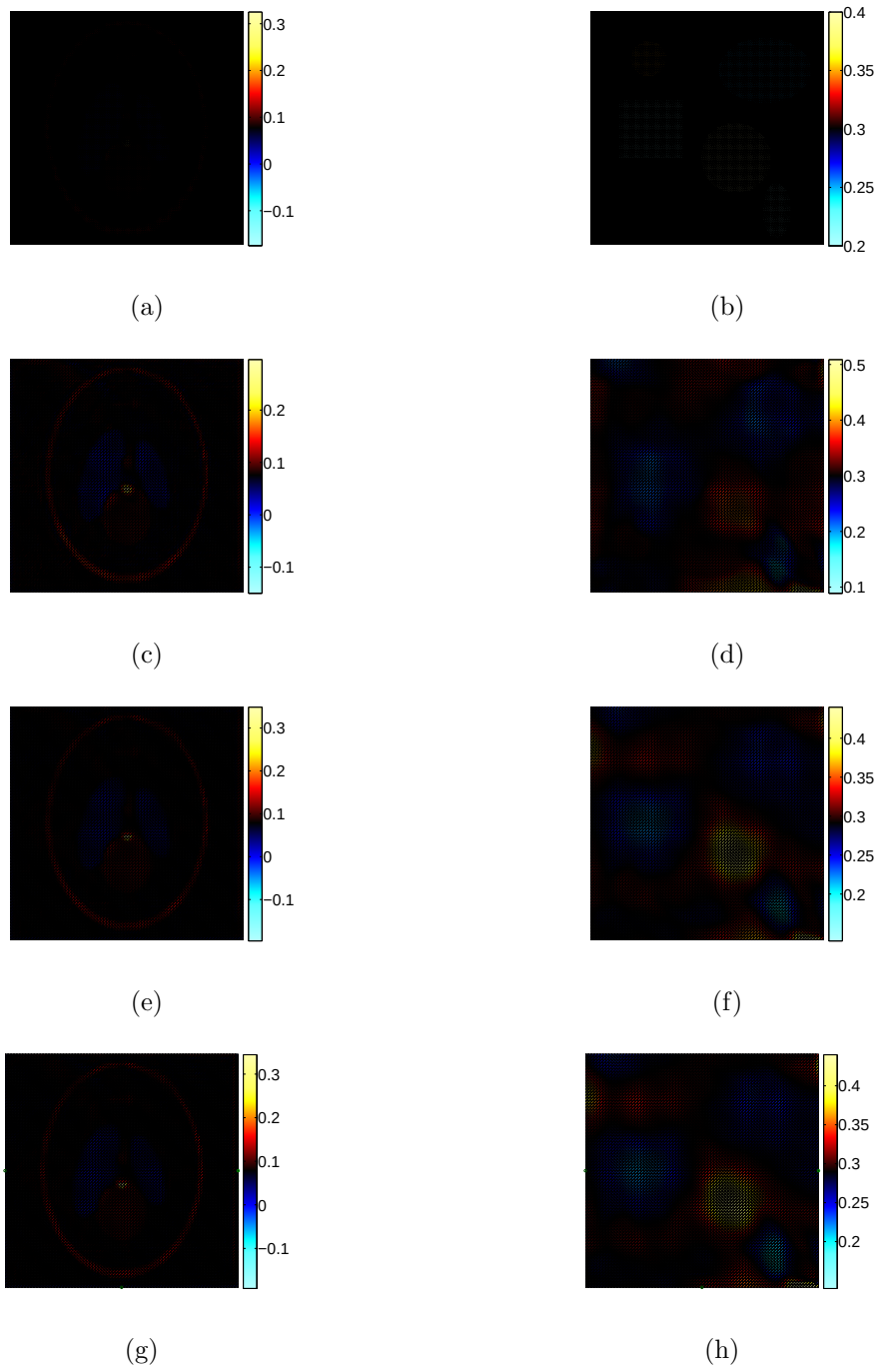


Figure 4.2: Optical coefficients for 2D case. Phantom: (a) absorption coefficient μ (b) diffusion coefficient κ. The images reconstructed by ADMM: (c) μ (d) κ, LD: (e) μ (f) κ, and PD-IPM: (g) μ (h) κ.

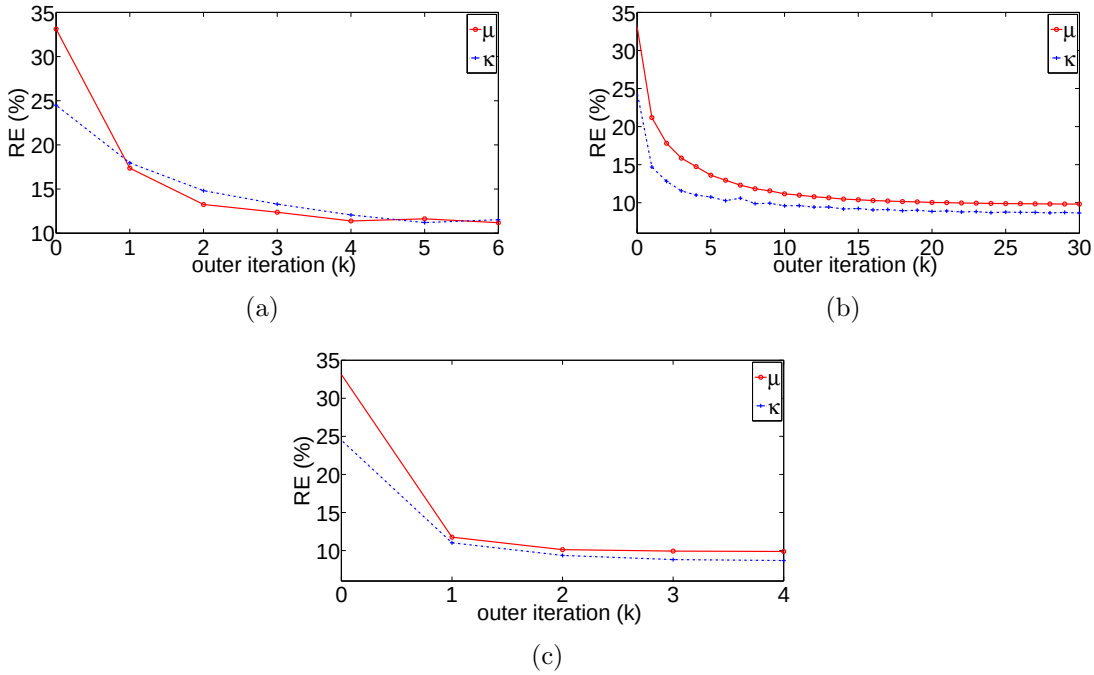


Figure 4.3: RE versus outer iteration k for 2D case: (a) ADMM (b) LD (c) PD-IPM.

Table 4.2: RE(%) of the final reconstructed images for 2D case.

Methods	μ	κ
ADMM	11.1882	11.5316
LD	9.8799	8.6899
PD-IPM	9.8513	8.6547

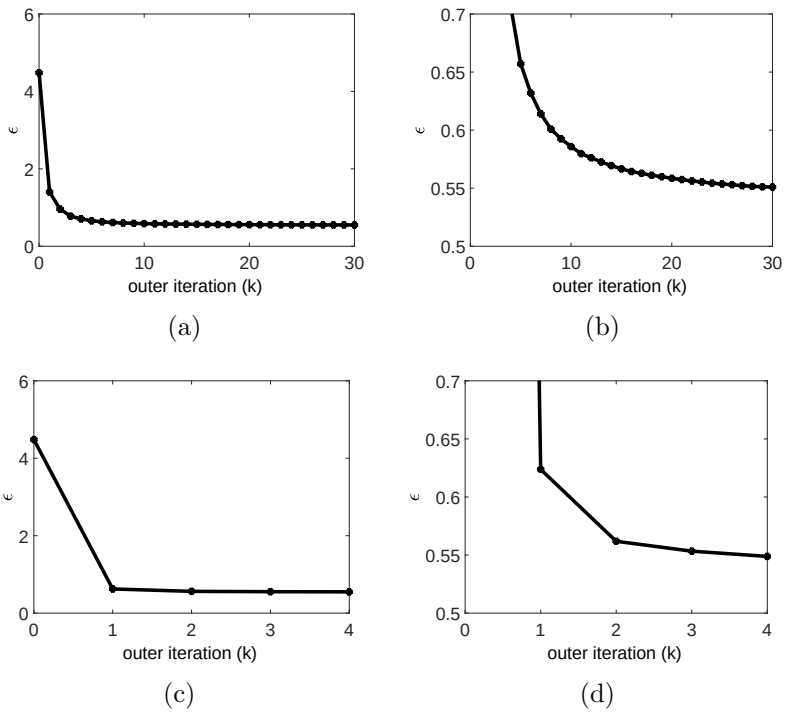
Table 4.2 shows the RE values for the final reconstructed images shown in figure 4.2.

4.6.2 3D phantom

We performed our 3D simulation on a cubic domain $[-5,5]\times[-5,5]\times[-5,5]$ mm^3.

Optical excitation

We used three optical excitation patterns, i.e., $N_q = 3$. For each optical excitation q, we used a discretisation of an inward directed diffuse boundary current $I_{s,q}$ (J/mm^2) that obeys (4.119) with $\iota_q \subset \partial\Omega$ two confronting faces of the grid, i.e. the left-right, posterior-anterior, and bottom-top faces.



(a) (b) (c) (d)

Figure 4.4: ϵ versus outer iteration k for 2D case. (a) LD (b) LD from an enlarged view around the optimal point (c) PD-IPM (d) PD-IPM from an enlarged view around the optimal point.

Discretisation for data generation

For data generation, the cubic domain was discretised using a grid with $37 \times 37 \times 37$ nodes and an even separation distance of 2.78×10^{-1}mm along all Cartesian coordinates. For the optical portion of the problem, an FE mesh was simulated so that each set of six tetrahedral voxels forms a cubic pixel with a centre matching an associated node on the acoustic grid. For the acoustic portion of the problem, we added a perfectly matched layer (PML) with a thickness of 8 grid points and a maximum attenuation coefficient of 2 nepers per grid point. The acoustic wavefield was detected in 292 time instants using 2145 detectors that are equidistantly placed on two (the left and posterior) faces of the grid. A 30 dB Additive White Gaussian Noise (AWGN) was then added to the simulated data.

Discretisation for image reconstruction

For image reconstruction, we avoided an inverse crime for discretisation by using a grid made up of $33 \times 33 \times 33$ nodes with a homogeneous separation distance of 3.125×10^{-1}mm along all Cartesian coordinates.

Table 4.3: The minimal and maximal values for acoustic properties of the 3D phantom.

	$c_0(\mathrm{ms}^{-1})$		$\rho_0(\mathrm{kgm}^{-3})$	
	min	max	min	max
Data generation	1.186×10^3	2.077×10^3	0.853×10^3	1.374×10^3
Image reconstruction	1.390×10^3	1.897×10^3	0.956×10^3	1.252×10^3

Acoustic properties

As discussed in section 4.6.1, to avoid an inverse crime for acoustic properties of the medium, we corrupted the sound speed and ambient density with 30 dB AWGN noise for simulation of data, whereas we used the clean acoustic maps for image reconstruction. Table 4.3 shows the minimial and maximal values for these maps. Using this table, the grid for data generation (resp. image reconstruction) supports a maximal frequency of 2.0766 MHz (resp. 2.156 MHz). The distributions of the sound speed and ambient density for the 3D phantom for data generation (resp. image reconstruction) are shown from a top view in figures 4.5(a) and 4.5(b) (resp. 4.5(c) and 4.5(d)), respectively. Additionally, the acoustic attenuation coefficient and the associated exponent factor (cf. equation (4.8)) was simulated the same as the 2D phantom.

Optical phantom

The left columns in figures 4.6(a) and 4.6(b) show the distributions of μ and κ for the 3D phantom, respectively. The images are obtained in horizontal planes (slices) $z = \{3, 2, 1, 0, -1, -2, -3\}$ mm, and the colorbars are shown to the right of images.

Image reconstruction

All algorithms were initialised using values 1.2 times more than the mean of optical coefficients for the 3D phantoms.

ADMM. The parameters for an implementation of the ADMM algorithm were chosen carefully in order to obtain almost the best possible image. All these parameres match our choices for the 2D phantom.

LD. For an implementation of the LD algorithm, the arising linearised subproblems were solved using a PCG algorithm with the same parameters as the 2D phantom, except that for stopping each PCG loop, we used $i_m = 3$. The TV preconditioner M was applied using $\gamma = 1 \times 10^{-8}$ and $\beta = 1 \times 10^{-6}$. We stopped our LD algorithm using

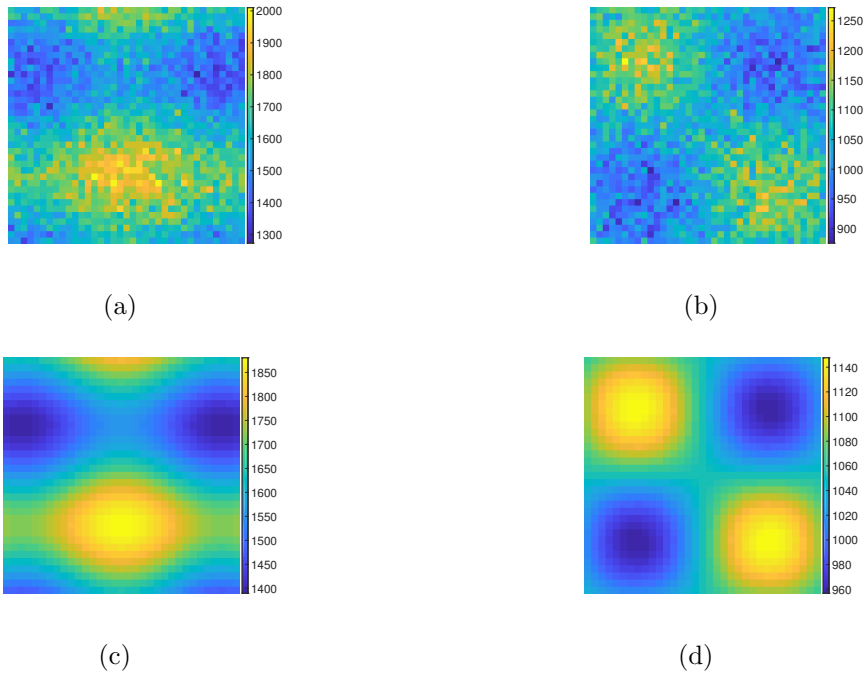


Figure 4.5: Acoustic properties for 3D case from a top view. Data generation: (a) c_0 (b) ρ_0, and image reconstruction: (c) c_0 (d) ρ_0.

$Tol_{\text{out}} = 1 \times 10^{-3}$.

PD-IPM. For an implementation of the PD-IPM algorithm, each outer linearised problem was solved using a sequence of inner linearised subproblems. The TV preconditioner was applied using $\gamma = 1 \times 10^{-8}$ and $\beta = 1 \times 10^{-8}$. For termination of each PCG loop associated with each inner linearised problem, we used $i_m = 2$ and $i_{\max} = 30$ (cf. Algorithm 4.4). We also terminated each outer linearised subproblem using $k'_{\max} = 25$ and $Tol_{\text{med}} = 1 \times 10^{-3}$. Also, our PD-IPM algorithm was terminated using a stopping threshold $Tol_{out} = 1 \times 10^{-3}$.

Observations

ADMM. The stopping criterion for the ADMM algorithm was satisfied after outer iteration 7. In figures 4.6(a) and 4.6(b), the second columns (from the left side) show the final reconstructed images for μ and κ, respectively. These images are shown using horizonal slices the same as the first columns for the phantom. Figure 4.7(a) shows the RE of solutions computed by the ADMM algorithm for μ and κ at outer iterations.

LD. The LD algorithm was terminated after 10 iterations. The final reconstructed

Table 4.4: RE(%) of the final reconstructed images for 3D case.

Methods	μ	κ
ADMM	14.0210	11.3598
LD	11.4718	8.1426
PD-IPM	11.0694	7.9647

images for μ and κ are shown in the 3rd columns of figures 4.6(a) and 4.6(b), respectively. The images are shown in the same way as the ADMM method. Figure 4.7(b) shows the RE of the solutions computed by LD for outer iterations k. The computed values for $\epsilon(X^k)$ are shown in figure 4.8(a), and 4.8(b) from an enlarged view around the optimal point. As shown in these figures, ϵ monotonically converges to a minimiser for all outer iterations.

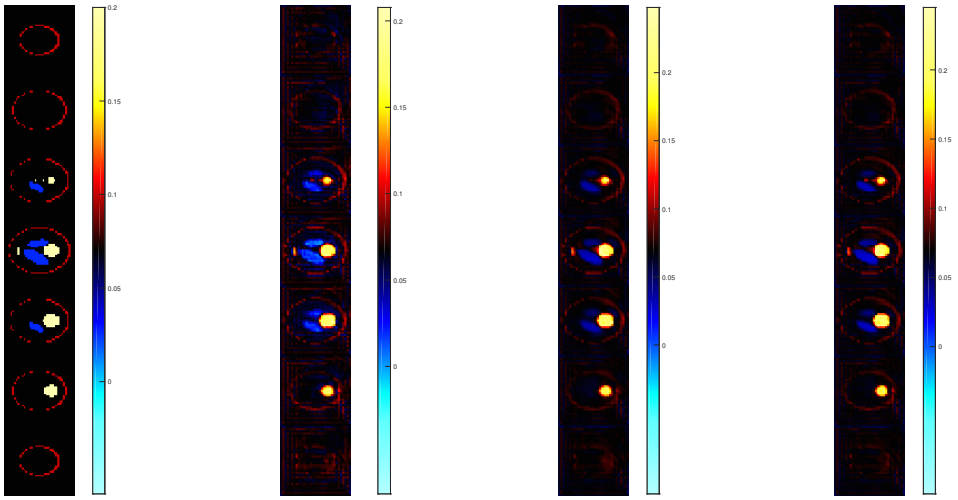
PD-IPM. The stopping criterion for the PD-IPM algorithm was satisfied after 6 outer iterations. The final reconstructed images for μ and κ are shown in the 4th columns of figures 4.6(a) and 4.6(b), respectively. Figure 4.7(c) shows the RE of solutions (optical coefficients) computed by the PD-IPM algorithm for outer iterations k. Figures 4.8(c) and 4.8(d) show the obtained values for $\epsilon(X^k)$. As shown in these figures, our choices for the step sizes associated with outer (resp. inner) subproblems, i.e., $\alpha^k = 1$ (resp. $\alpha_{k'} = 1$), provided a monotonic reduction for ϵ^k (resp. $\epsilon_{k'}$).

Table 4.4 shows the RE values for the final reconstructed images shown in figure 4.6.

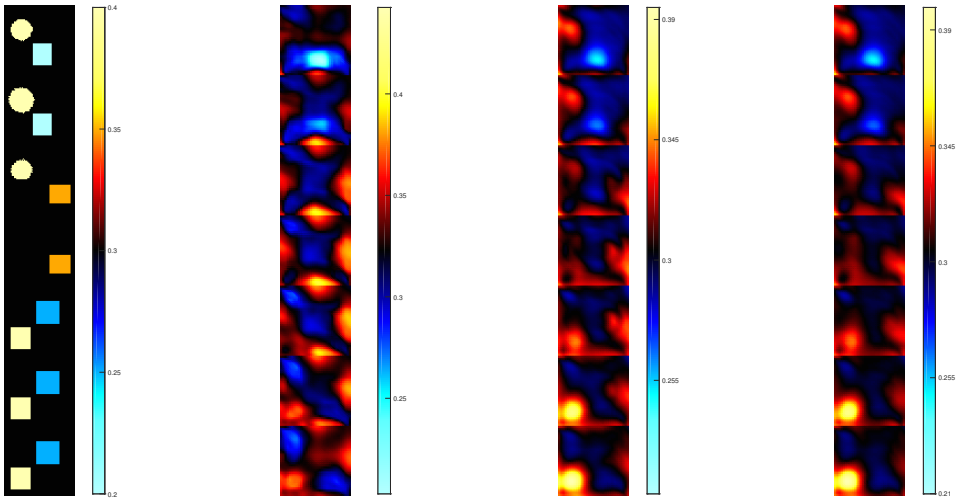
4.7 Discussion

In the previous section, we numerically evaluated the performance of the used iterative algorithms for a direct problem of QPAT for realistic acoustic media. In this section, we give further details on our numerical results.

The quality of images reconstructed by the ADMM algorithm was sensitive to choices for ϱ and Tol_{out}, and thus these parameters were chosen very carefully. This leads to several repetition of the entire reconstruction. For example, by using a smaller Tol_{out} (further proceeding of the iterations), the RE values started to increase. One way for avoiding this may be increasing the amount of regularisation via an increase in ϱ, but our numerical experience shows that choosing greater values for ϱ negatively affects the convergence of the algorithm. As shown in the second columns in figure 4.6,



(a)



(b)

Figure 4.6: Optical coefficients for 3D case. (a) μ, from the left to right: phantom, ADMM, LD and PD-IPM. (b) κ, from the left to right: phantom, ADMM, LD and PD-IPM.

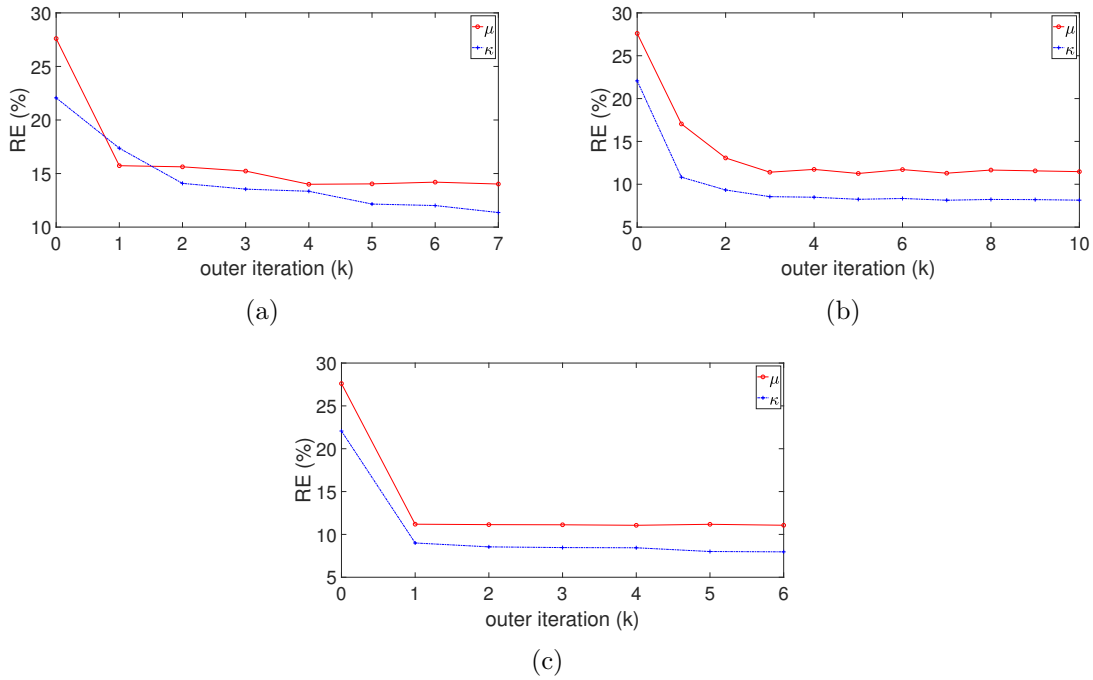


Figure 4.7: RE versus outer iteration k for 3D case. (a) ADMM (b) LD (c) PD-IPM.

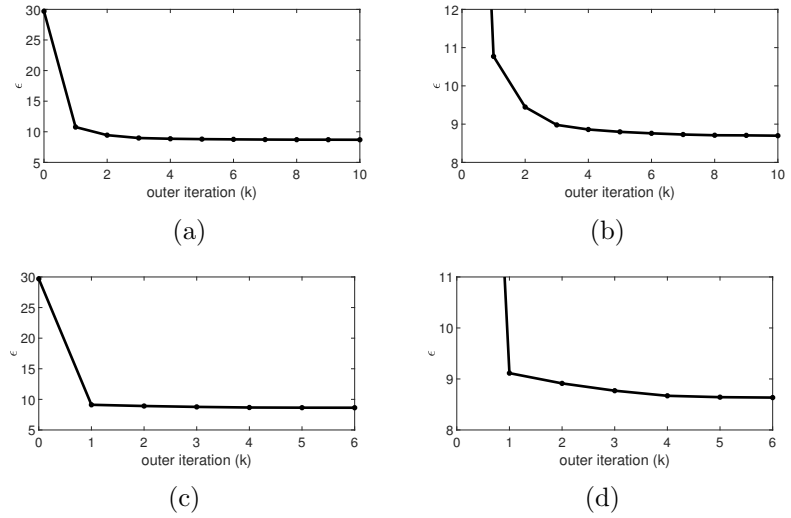


Figure 4.8: ϵ versus outer iteration k for 3D case. (a) LD (b) LD from an enlarged view around the optimal point (c) PD-IPM (d) PD-IPM from an enlarged view around the optimal point.

the reconstructed image for μ includes a high level of artifact, and also the ADMM algorithm failed to produce an accurate image for κ. (Note that better images were reconstructed using ADMM for 2D case.) The poor performance of the ADMM algorithm, especially in 3D case, may be because of assuming a high level of errors in the acoustic properties, as shown in figures 4.1 and 4.5.

The LD algorithm was not sensitive to γ, Tol_{in} and Tol_{out}. Both ϵ and RE will proceed with a monotonic reduction, if we choose smaller values for Tol_{out}. This is indicated by figures 4.4(b) and 4.8(b). However, the convergence of the LD algorithm will be deteriorated, if we use very small values for β, for example the values that we used for the PD-IPM algorithm.

Our numerical experience showed that the PD-IPM algorithm was also not sensitive to choices for γ, Tol_{in}, Tol_{med} and Tol_{out}. (For example, by choosing $Tol_{\text{med}} = 0$, the performance of the algorithm is almost the same.) Additionally, this algorithm converged well using very small values for β.

Regarding the computational cost, the ADMM algorithm reconstructed final images using 100-150 gradient-based iterations for both 2D and 3D cases. (Note that each iteration involves at least an implementation of the forward operator and the adjoint of the Jacobian matrix.) Note also that for ADMM, as discussed in the first paragraph of this section, using a smaller stopping threshold deteriorates the quality of reconstructed images. The LD and PD-IPM approaches produced the final images in 150-200 inner iterations. Note that the major cost of each inner iteration for the PCG loop involves an implicit implementation of the Jacobian matrix and its transpose, and has almost the same computational cost as the gradient.

For a direct problem of QPAT for realistic acoustic media with an error in estimation of acoustic properties, our numerical results show that the developed matrix-free Jacobian-based inexact-Newton methods outperform gradient-based approaches that utilise a search direction using Quasi-Newton approaches like the L-BFGS method [19, 18], at the same time does not impose a large computational cost due to an explicit construction of the Jacobian matrix [31].

Our next goal is an extension of multi-source QPAT to multi-spectral QPAT [5], which is more practical for biomedical cases, especially when a limited view is accessible for optical excitations. A simultaneous reconstruction of the optical coefficients

and the sound speed using adjunct information obtained from ultrasound computed tomography may be promising for improving the quality of reconstructed images [30].

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). There were no data used or collected for this research.

4.A Supplementary materials

In this section, we will provide further details about some formulae in this chapter. These details have not been embedded in the paper because of space constraints.

4.A.1 A proof for Lemma 4.1

Here, we prove that the adjoint of the Fréchet derivative of the optical forward operator satisfies equations (4.17) and (4.18) in Lemma 4.1.

Proof. Here, for brevity, the dependence on r is neglected. Applying integration-by-parts using the divergence theorem to the first line in (4.17) gives

$$\begin{aligned}
&\int_\Omega \begin{pmatrix} \nabla\phi_0 \cdot \nabla\tilde{h} \\ \phi_0\tilde{h} + \phi_0 h \end{pmatrix} \cdot \begin{pmatrix} \delta\kappa \\ \delta\mu \end{pmatrix} \, \mathrm{d}r \\
&= \int_\Omega -(\nabla \cdot \delta\kappa\nabla\phi_0)\ \tilde{h} + \phi_0\ \tilde{h}\ \delta\mu + \phi_0\ h\ \delta\mu \, \mathrm{d}r \\
&+ \int_{\partial\Omega} \delta\kappa\ \tilde{h}\ \frac{\partial\phi_0}{\partial\hat{n}} \, \mathrm{d}s
\end{aligned} \tag{4.122}$$

By the assumption that $\delta\phi$ satisfies (4.16), the right-hand-side of the above equation can be rewritten as

$$\int_\Omega (\nabla \cdot \kappa_0\nabla\delta\phi - \mu_0\ \delta\phi)\ \tilde{h} + \phi_0\ h\ \delta\mu \, \mathrm{d}r + \int_{\partial\Omega} \delta\kappa\ \tilde{h}\ \frac{\partial\phi_0}{\partial\hat{n}} \, \mathrm{d}s. \tag{4.123}$$

Applying integration-by-parts two times using the divergence theorem shows that (4.123) is equal to

$$\begin{aligned}&\int_{\Omega}(\nabla\cdot\kappa_0\nabla\tilde{h}-\mu_0\ \tilde{h})\ \delta\phi+\phi_0\ h\ \delta\mu\ \mathrm{d}r\\&+\int_{\partial\Omega}\delta\kappa\ \tilde{h}\ \frac{\partial\phi_0}{\partial\hat{n}}+\kappa_0\ \tilde{h}\ \frac{\partial\delta\phi}{\partial\hat{n}}-\kappa_0\ \delta\phi\ \frac{\partial\tilde{h}}{\partial\hat{n}}\ \mathrm{d}s.\end{aligned}\tag{4.124}$$

By the assumption that $\tilde{h}$ satisfies (4.18), (4.124) is equal to

$$\begin{aligned}&\int_{\Omega}(\mu_0\ \delta\phi+\phi_0\ \delta\mu)\ h\ \mathrm{d}r\\&+\int_{\partial\Omega}\delta\kappa\ \tilde{h}\ \frac{\partial\phi_0}{\partial\hat{n}}+\kappa_0\ \tilde{h}\ \frac{\partial\delta\phi}{\partial\hat{n}}-\kappa_0\ \delta\phi\ \frac{\partial\tilde{h}}{\partial\hat{n}}\ \mathrm{d}s.\end{aligned}\tag{4.125}$$

Using the boundary condition in (4.16), the integrand in the second line in (4.125) yields

$$-2\gamma_d\delta\phi\left(\tilde{h}+\frac{1}{2\gamma_d}\kappa_0\frac{\partial\tilde{h}}{\partial\hat{n}}\right),\tag{4.126}$$

which is equal to zero using the boundary condition in (4.18). This proves the adjoint is as claimed. □

4.A.2 Discretisation of $\mathbb{J}_o^*$ using an adjoint-then-discretise approach

Here, we provide further details about a discretisation of the adjoint of the Fréchet derivative of the optical forward operator using an *adjoint-then-discretise* approach (cf. (4.40)). We will prove that using an *adjoint-then-discretise* approach, if we approximate the products in (4.17) by first multiplying the nodal functions and then using an $\mathcal{L}^2$-orthogonal projection on the space of elemental functions, then a discretisation of the adjoint will give equation (4.40). We also remind that a discretisation of the adjoint of the Fréchet derivative of the optical forward operator using the *adjoint-then-discretise* approach given in this chapter (cf. (4.40)) gives the same formula as a corresponding discretised adjoint (cf. (4.41)) [18].

Proof. For any $f\in\mathcal{L}^2(\Omega)$, the $\mathcal{L}^2$-orthogonal projection of f onto the span of $\{\chi_j\}$ is defined by

$$f\approx f^e=\sum_{j=1}^{N_e}\hat{f}_j\chi_j.$$

Then since $\{\chi_j/S_j^{1/2}\}$ is an $\mathcal{L}^2$-orthonormal set, the coefficients $\hat{f}_j$ are given by

$$\hat{f}_j = \frac{1}{S_j}\int_{t_j} f(r)\ dr, \tag{4.127}$$

where t_j is the element j. Now, let us consider the first line of the adjoint given by formula (4.17), which is $\nabla\phi_0 \cdot \nabla\tilde{h}$. In the discretization, ϕ_0 (resp. $\tilde{h}$) is approximated in a piecewise linear basis as ϕ^h (resp. $\tilde{h}^h$), as shown in (4.23) (resp.(4.24)). Using these, the coefficients associated with the $\mathcal{L}^2$-orthonormal projection of $\nabla\phi_0^h \cdot \nabla\tilde{h}^h$ onto the span of $\{\chi_j\}$ will be

$$\begin{aligned}(\widehat{\nabla\phi_0^h \cdot \nabla\tilde{h}^h})_j &= \frac{1}{S_j}\int_{t_j} \nabla\phi_0^h(r) \cdot \nabla\tilde{h}^h(r)dr \\ &= \frac{1}{S_j}\sum_{k,p=1}^{N_n}\int_{t_j} \nabla\varphi_k(r) \cdot \nabla\varphi_p(r)dr \quad \tilde{H}_k\Phi_{0,p},\end{aligned} \tag{4.128}$$

Plugging (4.38) into (4.128) gives

$$(\widehat{\nabla\phi_0^h \cdot \nabla\tilde{h}^h})_j = \frac{1}{S_j}\sum_{k,p=1}^{N_n}\frac{\partial \boldsymbol{A}_{o,kp}}{\partial\hat{\kappa}_j}\tilde{H}_k\Phi_{0,p}. \tag{4.129}$$

This proves that the first component of the adjoint is as claimed.

Now let us consider the first term in the second line of the adjoint in (4.17). As above, a projection of $\phi_0^h \cdot \tilde{h}^h$ onto the span of $\{\chi_j\}$ yields

$$\begin{aligned}(\widehat{\phi_0^h \cdot \tilde{h}^h})_j &= \frac{1}{S_j}\int_{t_j} \phi_0^h(r)\tilde{h}^h(r)\ dr \\ &= \frac{1}{S_j}\sum_{k,p=1}^{N_n}\int_{t_j} \varphi_k(r)\varphi_p(r)\ dr \quad \tilde{H}_k\Phi_{0,p}.\end{aligned} \tag{4.130}$$

Plugging (4.39) into (4.130) gives

$$(\widehat{\phi_0^h \cdot \tilde{h}^h})_j = \frac{1}{S_j}\sum_{k,p=1}^{N_n}\frac{\partial \boldsymbol{A}_{o,kp}}{\partial\hat{\mu}_j}\tilde{H}_k\Phi_{0,p}. \tag{4.131}$$

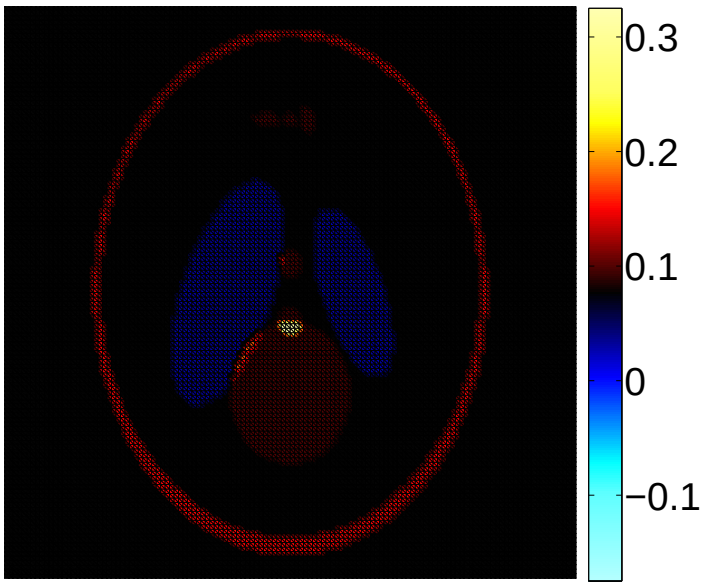
For the second term in the second line of the adjoint, having defined the elemental vector H, together with an approximation of ϕ_0 at element j, gives

$$\begin{aligned}H_j(\widehat{\phi_0^h})_j &= H_j\frac{1}{S_j}\int_{t_j} \phi_0^h(r)\ dr \\ &= H_j\frac{1}{S_j}\sum_{k=1}^{N_n}\int_{t_j} \varphi_k(r)\ dr\ \Phi_k \\ &= H_j\left(\mathbb{I}\Phi\right)_j,\end{aligned} \tag{4.132}$$

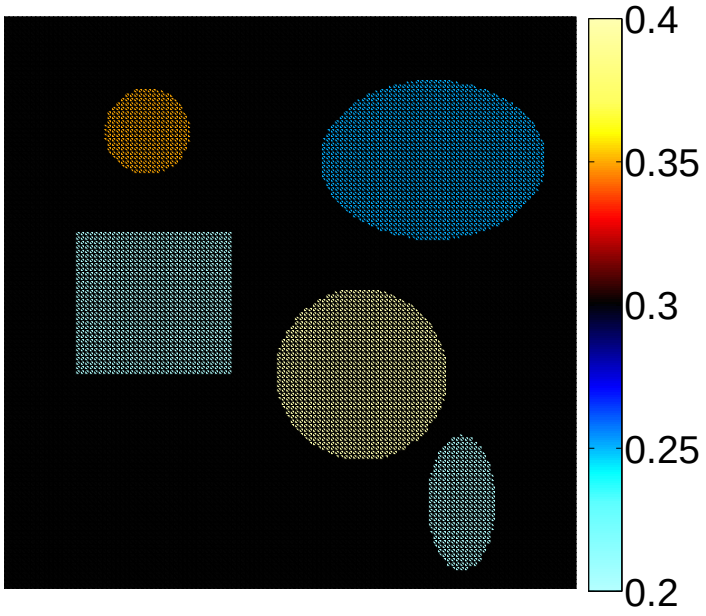
and hence the proof is completed. □

4.A.3 The reconstructed images from a larger view

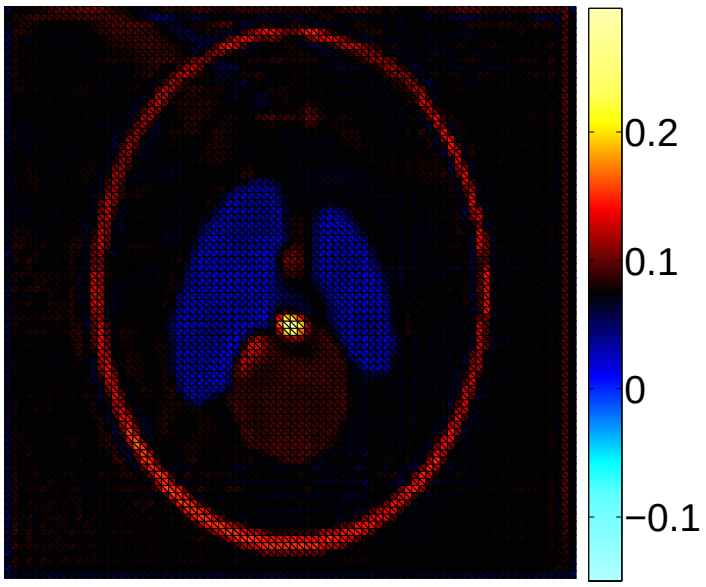
Because a visual comparison between the reconstructed images shown in figures 4.2 and 4.6 may be difficult for the reader, we visualise these reconstructed images from a larger view. Accordingly, figures 4.9 and 4.10 show the reconstructed images in figures 4.2 and 4.6 from a larger view.



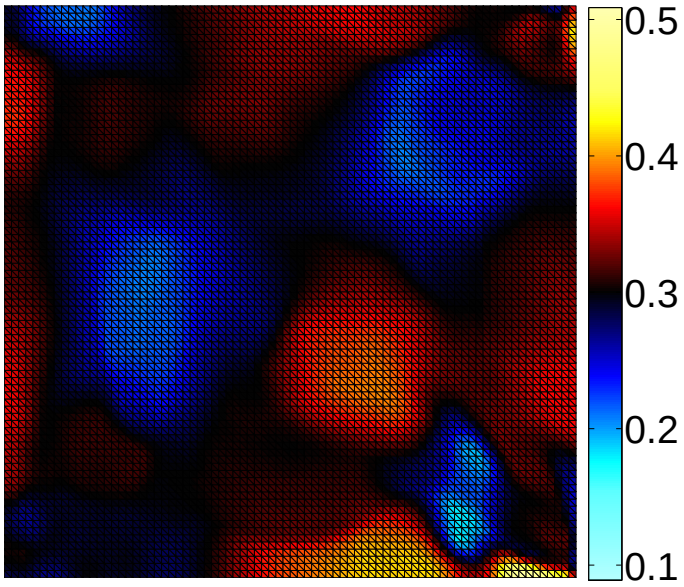
(a)



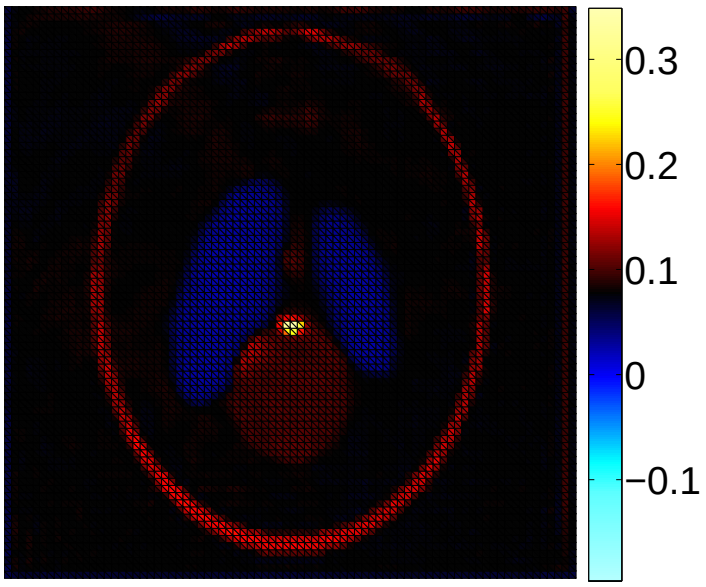
(b)



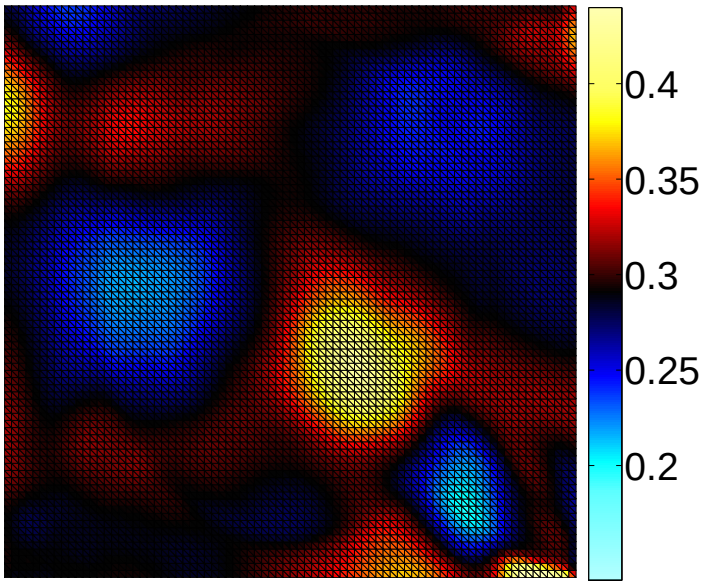
(c)



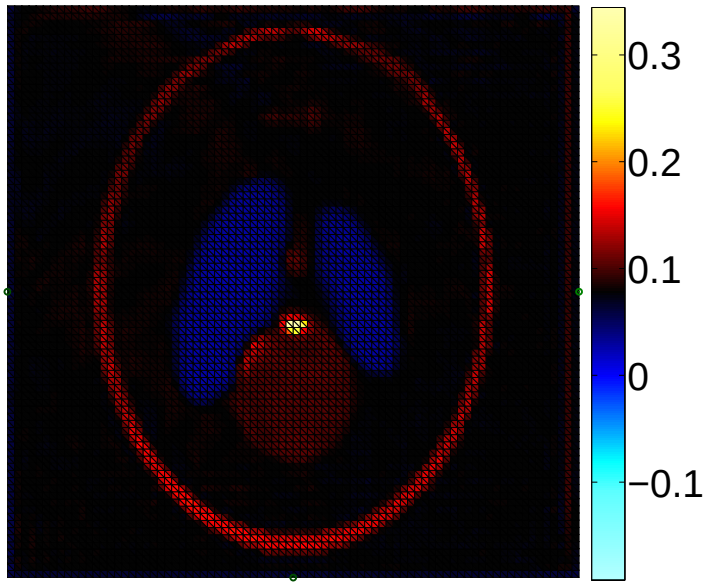
(d)



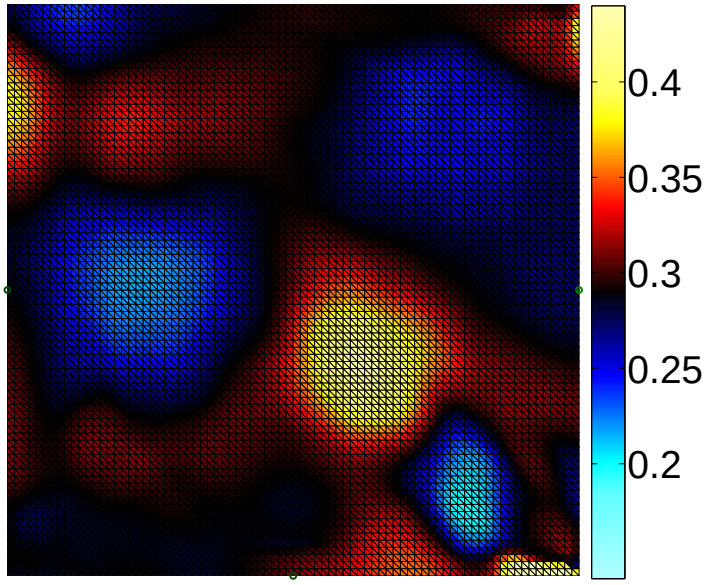
(e)



(f)

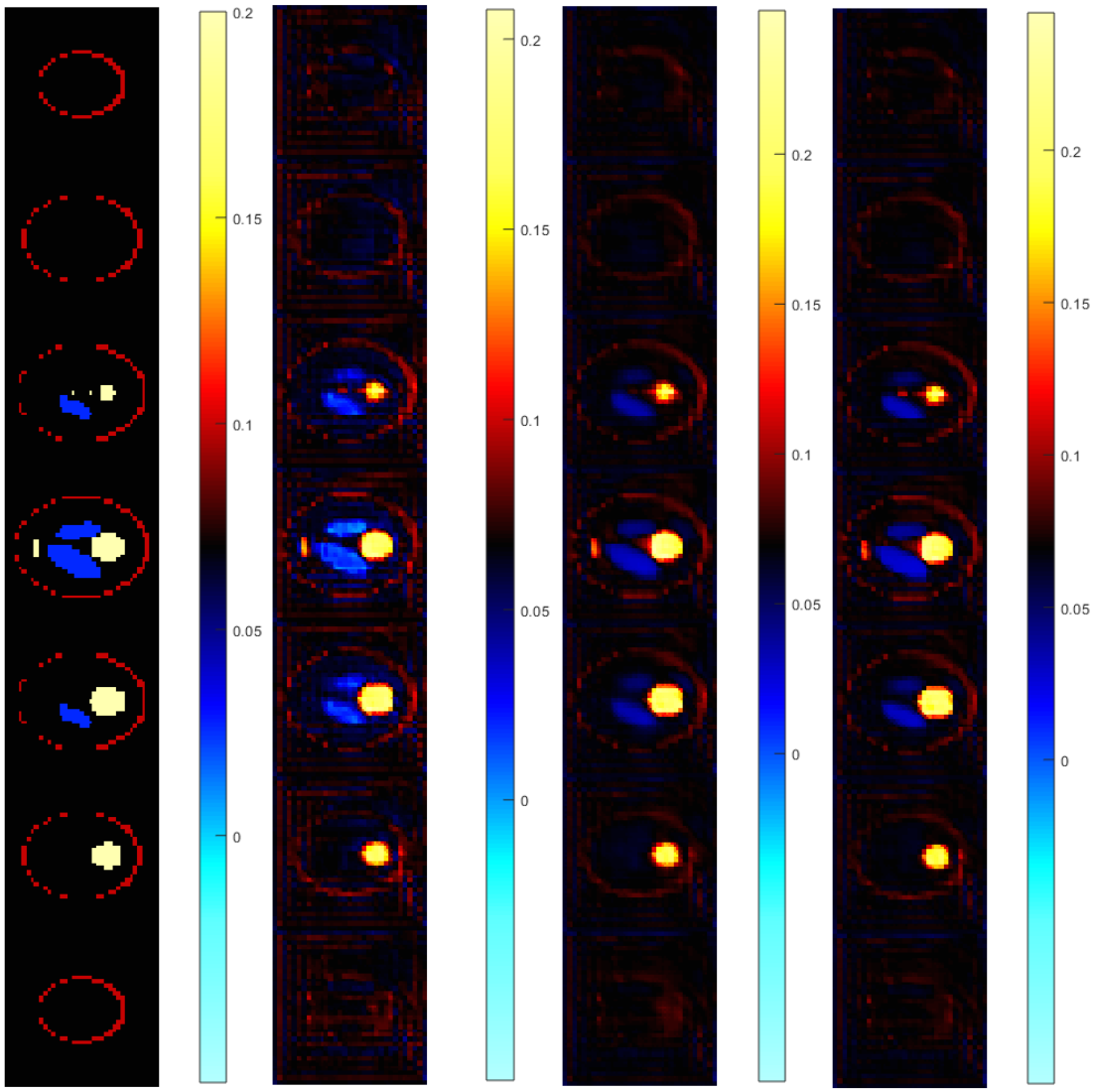


(g)

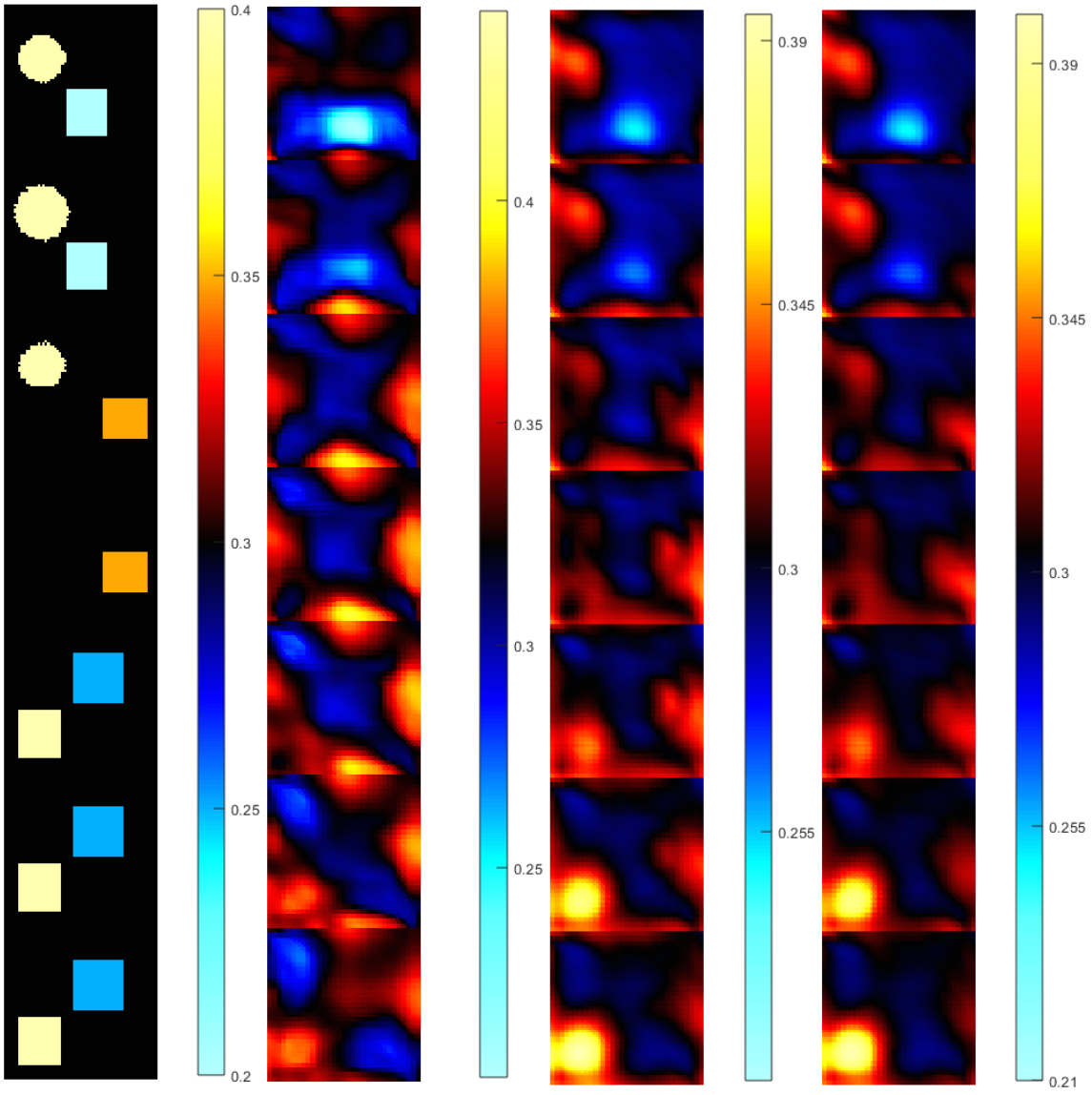


(h)

Figure 4.9: A visualisation of figure 4.2 from a larger view. Optical coefficients for 2D case. Phantom: (a) absorption coefficient μ (b) diffusion coefficient κ. The images reconstructed by ADMM: (c) μ (d) κ, LD: (e) μ (f) κ, and PD-IPM: (g) μ (h) κ.



(a)



(b)

Figure 4.10: A visualisation of figure 4.6 from a larger view. Optical coefficients for 3D case. (a) μ, from the left to right: phantom, ADMM, LD and PD-IPM. (b) κ, from the left to right: phantom, ADMM, LD and PD-IPM.

Bibliography

[1] 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.

[2] S. R. Arridge, M. M. Betcke, and L. Harhanen. Iterated preconditioned lsqr method for inverse problems on unstructured grids. *Inverse Problems*, 30(7):075009, 2014.

[3] G. Bal, A. Jollivet, and V. Jugnon. Inverse transport theory of photoacoustics. *Inverse Problems*, 26(2):025011, 2010.

[4] G. Bal and K. Ren. Multi-source quantitative photoacoustic tomography in a diffusive regime. *Inverse Problems*, 27(7):075003, 2011.

[5] G. Bal and K. Ren. On multi-spectral quantitative photoacoustic tomography in diffusive regime. *Inverse Problems*, 28(2):025010, 2012.

[6] G. Bal and G. Uhlmann. Inverse diffusion theory of photoacoustics. *Inverse Problems*, 26(8):085010, 2010.

[7] Z. Belhachmi, T. Glatz, and O. Scherzer. A direct method for photoacoustic tomography with inhomogeneous sound speed. *Inverse Problems*, 32(4):045005, 2016.

[8] A. Borsic, B. M. Graham, A. Adler, and W. R. B. Lionheart. In vivo impedance imaging with total variation regularization. *IEEE Transactions on Medical Imaging*, 29(1):44–54, Jan 2010.

[9] P. Burgholzer, G. J. Matt, M. Haltmeier, and G. Paltauf. Exact and approximative imaging methods for photoacoustic tomography using an arbitrary detection surface. *Phys. Rev. E*, 75:046706, Apr 2007.

[10] T. Chan, G. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing*, 20(6):1964–1977, 1999.

[11] B. Cox, J. G. Laufer, S. R. Arridge, and P. C. Beard. Quantitative spectroscopic photoacoustic imaging: a review. *Journal of Biomedical Optics*, 17(6):061202–1–061202–22, 2012.

[12] B. T. Cox, S. R. Arridge, K. P. Köstli, and P. C. Beard. Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method. *Appl. Opt.*, 45(8):1866–1875, Mar 2006.

[13] 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.

[14] R. Dembo, S. Eisenstat, and T. Steihaug. Inexact newton methods. *SIAM Journal on Numerical Analysis*, 19(2):400–408, 1982.

[15] L. Ding, X. L. Deán-Ben, C. Lutzweiler, D. Razansky, and V. Ntziachristos. Efficient non-negative constrained model-based inversion in optoacoustic tomography. *Phys. Med. Biol*, 60(17):6733–6750, 2015.

[16] T. Ding, K. Ren, and S. Vallélian. A one-step reconstruction algorithm for quantitative photoacoustic imaging. *Inverse Problems*, 31(9):095005, 2015.

[17] D. Finch and R. Sarah K. Patch. Determining a function from its mean values over a family of spheres. *SIAM Journal on Mathematical Analysis*, 35(5):1213–1240, 2004.

[18] H. Gao, J. Feng, and L. Song. Limited-view multi-source quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065004, 2015.

[19] H. Gao, S. Osher, and H. Zhao. *Quantitative Photoacoustic Tomography*, pages 131–158. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[20] M. Haltmeier, L. Neumann, and S. Rabanser. Single-stage reconstruction algorithm for quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065005, 2015.

[21] 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.

[22] A. Hannukainen, N. Hyvönen, H. Majander, and T. Tarvainen. Efficient inclusion of total variation type priors in quantitative photoacoustic tomography. *SIAM Journal on Imaging Sciences*, 9(3):1132–1153, 2016.

[23] Y. Hristova. Time reversal in thermoacoustic tomography—an error estimate. *Inverse Problems*, 25(5):055008, 2009.

[24] 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.

[25] 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.

[26] A. Javaherian and S. Holman. A multi-grid iterative method for photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 36(3):696–706, March 2017.

[27] A. Javaherian and S. Holman. A continuous adjoint for photo-acoustic tomography of the brain. *Inverse Problems*, 34(8):085003, 2018.

[28] L. A. Kunyansky. Explicit inversion formulae for the spherical mean radon transform. *Inverse Problems*, 23(1):373, 2007.

[29] S. M., A. S. R., H. M., and D. D. T. The finite element method for the propagation of light in scattering media: Boundary and source conditions. *Medical Physics*, 22(11):1779–1792.

[30] T. P. Matthews and M. A. Anastasio. Joint reconstruction of the initial pressure and speed of sound distributions from combined photoacoustic and ultrasound tomography measurements. *Inverse Problems*, 33(12):124002, 2017.

[31] A. Pulkkinen, B. T. Cox, S. R. Arridge, H. Goh, J. P. Kaipio, and T. Tarvainen. Direct estimation of optical parameters from photoacoustic time series in

quantitative photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 35(11):2497–2508, Nov 2016.

[32] A. Pulkkinen, B. T. Cox, S. R. Arridge, J. P. Kaipio, and T. Tarvainen. A bayesian approach to spectral quantitative photoacoustic tomography. *Inverse Problems*, 30(6):065012, 2014.

[33] 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.

[34] A. Rosenthal, V. Ntziachristos, and D. Razansky. Acoustic inversion in optoacoustic tomography: A review. *Current Medical Imaging Reviews*, 9(4):318–336, 2013.

[35] T. Saratoon, T. Tarvainen, B. T. Cox, and S. R. Arridge. A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation. *Inverse Problems*, 29(7):075006, 2013.

[36] M. Schweiger, S. R. Arridge, and I. Nissilä. Gauss–newton method for image reconstruction in diffuse optical tomography. *Physics in Medicine & Biology*, 50(10):2365, 2005.

[37] P. Shao, T. Harrison, and R. J. Zemp. Iterative algorithm for multiple illumination photoacoustic tomography (mipat) using ultrasound channel data. *Biomed. Opt. Express*, 3(12):3240–3249, Dec 2012.

[38] N. Song, C. Deumié, and A. D. Silva. Considering sources and detectors distributions for quantitative photoacoustic tomography. *Biomed. Opt. Express*, 5(11):3960–3974, Nov 2014.

[39] P. Stefanov and G. Uhlmann. Thermoacoustic tomography with variable sound speed. *Inverse Problems*, 25(7):075011, 2009.

[40] 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.

[41] T. Tarvainen, B. T. Cox, J. P. Kaipio, and S. R. Arridge. Reconstructing absorption and scattering distributions in quantitative photoacoustic tomography. *Inverse Problems*, 28(8):084009, 2012.

[42] T. Tarvainen, A. Pulkkinen, B. T. Cox, J. P. Kaipio, and S. R. Arridge. Bayesian image reconstruction in quantitative photoacoustic tomography. *IEEE Trans. Med. Imag.*, 32(12):2287–2298, 2013.

[43] T. Tarvainen, M. Vauhkonen, V. Kolehmainen, and J. P. Kaipio. Finite element model for the coupled radiative transfer equation and diffusion approximation. *International Journal for Numerical Methods in Engineering*, 65(3):383–405.

[44] 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.

[45] 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.

[46] 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.

[47] C. Vogel and M. Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.

[48] K. Wang and M. A. Anastasio. *Photoacoustic and Thermoacoustic Tomography: Image Formation Principles*, pages 781–815. Springer New York, New York, NY, 2011.

[49] K. Wang, R. Su, A. A. Oraevsky, and M. A. Anastasio. Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography. *Phys. Med. Biol*, 57(17):5399–5423, 2012.

[50] L. V. Wang and J. Yao. A practical guide to photoacoustic tomography in the life sciences. *Nature Methods*, 13:627–638, 2016.

[51] M. Xu and L. V. Wang. Time-domain reconstruction for thermoacoustic tomography in a spherical geometry. *IEEE Trans. Med. Imag.*, 21(7):814–822, 2002.

[52] M. Xu and L. V. Wang. Universal back-projection algorithm for photoacoustic computed tomography. *Phys. Rev. E*, 71(1):016706, 2005.

[53] Y. Xu, D. Feng, and L. V. Wang. Exact frequency-domain reconstruction for thermoacoustic tomography. i. planar geometry. *IEEE Transactions on Medical Imaging*, 21(7):823–828, July 2002.

[54] Y. Xu and L. V. Wang. Time reversal and its application to tomography with diffracting sources. *Phys. Rev. Lett.*, 92:033902, Jan 2004.

[55] Y. Xu, M. Xu, and L. V. Wang. Exact frequency-domain reconstruction for thermoacoustic tomography. ii. cylindrical geometry. *IEEE Transactions on Medical Imaging*, 21(7):829–833, July 2002.

Chapter 5

Conclusion, discussion and future works

In this thesis, we dealt with variational approaches for photo-acoustic tomography (PAT) and quantitative photo-acoustic tomography (QPAT). Variational approaches are among the robust methods for solving the inverse problem of PAT because of their ability in coping with the ill-posedness of the problem. Considering PAT (the acoustic portion of QPAT), an ill-posedness may arise, for example because of errors in estimation of medium's acoustic properties or a few-view or limited-view detection surface [17]. Among variational approaches that have been used for the pure acoustic problem of PAT, gradient-based methods that use forward-backward splitting methods, are very popular, because they are cheaper than second-order methods regarding memory space and computational cost [17, 2, 1].

Because of the compartmentalised distribution of light absorbing molecules in tissue media, the generated ultrasound waves are broadband [27]. For modelling generation and propagation of such broadband acoustic waves, very dense computational grids must be used. This makes an application of variational methods for PAT very challenging, because the associated algorithms are iterative.

In chapter 2, to mitigate the computational cost of variational approaches for the acoustic inverse problem of PAT, we proposed a line-search multi-grid (MG) algorithm using the well-known Nash's method [21]. The objective function we considered is a sum of a data fidelity function, which is smooth, and is an $\mathcal{L}^2$ norm of the discrepancy between a pair of modelled and measured sets of times series of boundary data, and

a total variation (TV) functional, which is nonsmooth [9, 8]. We also enforced a non-negativity constraint using the fact that the initial pressure distribution cannot be negative [17].

At each iteration, our MG algorithm decides between a direct search direction on a finer (target) grid, or alternatively a recursive search direction that is computed via a minimisation of the objective function on a coarser level. When the algorithm decides to compute a recursive search direction, the objective function is transferred to a coarser level using the Nash's technique [21]. The Nash's method was originally developed for minimisation of smooth functions, so we approximated the TV functional (the non-smooth part of the objective function) with a smoothed variant. We also used a method proposed in [20] for transferring the non-negativity constraint to a coarser level. Our developed MG algorithm was adapted to FBS methods, i.e., ISTA and FISTA [9, 8]. Using both algorithms, an application of the proposed MG method significantly improved the speed of image reconstruction [18].

Our MG algorithm is limited to FBS algorithms. An extension of MG optimisation methods to other popular classes of variational approaches for PAT, for example subspace-Krylov methods like PCG algorithm [16], may need further attention. In addition, our MG algorithm is limited to the acoustic inverse problem of PAT, but the objective of PAT is a reconstruction of optical attenuation coefficients (QPAT). Therefore, an extension of MG algorithms to a single-stage variant of QPAT may be promising, because an inverse problem of QPAT for biomedical applications are often very large-scale. For a single-stage variant of QPAT, the forward operator is a composite operator including a nonlinear and ill-conditioned operator for modelling the optical portion of the problem and a linear operator for modelling the acoustic portion of the problem [15]. This makes an application of MG optimisation approaches for QPAT more challenging than PAT.

In chapter 3, we used variational approaches for photo-acoustic tomography of the brain. An associated inverse problem is a reconstruction of the initial pressure distribution inside the skull from a set of time series of pressure data that is measured outside the skull. The photo-acoustic forward operator that has been used in chapter 2 describes the propagation of acoustic waves solely for lossy fluid media, and does not account for generation and propagation of shear waves within solid media like the

skull. Therefore, to account for propagation of shear waves, we used a system of photo-acoustic wave equations that describe the propagation of acoustic waves in linear, isotropic and elastic media, and consider an acoustic absorption and dispersion that obey a frequency power law using fractional Laplacians [26], the same as in chapter 2 [25]. We derived a corresponding adjoint using an *adjoint-then-discretise* method [19]. We used a k-space pseudospectral method for discretisation of the forward and adjoint operators [24, 10, 11], as described in section 1.7. Using this discretisation approach, we analytically showed that a discretisation of our derived continuous adjoint matches an algebraic adjoint of a corresponding discretised forward operator [19]. From a numerical point of view, we verified a discretisation of our derived adjoint operator using a discretisation of an inner product formula, which was already used for an analytic calculation of our continuous adjoint operator.

For modelling the acoustic absorption and dispersion, a computation of the fractional Laplacian operators using a k-space pseudospectral method is more efficient than solving the associated fractional temporal derivatives. However, fractional Laplacian operators use a dispersion formula, i.e., $w \approx ck$ with w the angular frequency, c the sound speed and k the spatial wavenumber [26]. Since the compressional and shear parts of acoustic waves propagate at different sound speeds c, different dispersion formulae must be used. This increases the computational cost for computation of the forward and adjoint operators almost twice.

Since photo-acoustic waves are highly broadband, very dense grids must be used for modelling the high frequency components of the generated ultrasound waves. As a result, developing more efficient methods for describing the attenuation of acoustic waves in lossy elastic media can be useful for practical cases [14]. One advantage of using a continuous adjoint is that an inclusion of any modifications applied on the forward problem in the adjoint operator will be straightforward, and does not require a recalculation of the adjoint operator [19]. Also, a continuous adjoint is independent of the method used for discretisation of the forward and adjoint operators [2]. Therefore, the derived continuous adjoint may be valid for any discretisation schemes that will be developed in future.

We tested the performance of our forward and adjoint operators using a classical algorithm like ISTA. More efficient and tolerant algorithms can be used for solving

the inverse problem of PAT in lossy elastic media. For example, an inclusion of our developed forward and adjoint pair in accelerated variants of FBS algorithms like FISTA [9, 8], or Krylov subspace methods, e.g., PCG [16] or LSQR [3], may be useful for improving the speed of image reconstruction.

In addition, the geometry and acoustic properties of the skull may not be known precisely for real cases. To account for this challenge in our simulations, in addition to avoid a classical inverse crime for discretisation of the domain, we used different acoustic properties for data generation and image reconstruction, together with a shift in soft tissue-skull interfaces. Further studies are required for a minimisation of errors in modelling acoustic properties. For example, using adjunct imaging modalities like ultrasound computed tomography (USCT) for a reconstruction of the sound speed before or during the image reconstruction for PAT, or using x-ray computed tomography for a determination of the geometry of the skull may be promising.

In chapter 4, we dealt with variational approaches for quantitative photo-acoustic tomography (QPAT). QPAT involves two inverse problems, namely acoustic and optical. In a classical variant of QPAT, these two inverse problems are solved separately [23, 13]. Using this class of approaches, an acoustic portion of the inverse problem is first solved in order to reconstruct an initial pressure distribution from a set of time series of boundary data, and the solution is used as data for the optical portion of the inverse problem. For cases the acoustic portion of QPAT is ill-posed, for example because of errors in modelling the acoustic properties or a limited-view detection surface, the solution of the acoustic inverse problem will not be accurate sufficiently to act as data for the optical portion of the inverse problem [15, 12]. Because the optical portion of the forward operator is nonlinear and ill-posed, this error may grow during the inversion process.

In recent years, a single-stage reconstruction of the optical attenuation coefficients from the boundary pressure data has attracted much attention [15]. One advantage of using a single-stage approach is that the acoustic portion of the inverse problem benefits from a priori knowledge about the optical attenuation coefficients, the same as the optical portion. Additionally, the optical portion of the inverse problem benefits from approaches for modelling noise in the boundary data, the same as the acoustic inverse problem. Furthermore, to mitigate the ill-posedness of the inverse problem

of QPAT, multi-source [5] or multi-wavelength [6] settings are often used. Because applying any changes in the optical sources or wavelengths will change the induced initial pressure distribution, the acoustic portion of the inverse problem will fail to benefit from using muti-source or multi-wavelength settings, if the acoustic and optical portions of the inverse problem are solved separately, as opposed to a single-stage variant of QPAT [15, 12].

In chapter 4, we solved a single-stage variant of QPAT for heterogeneous and lossy acoustic media. For the optical portion of the forward operator, we used a Diffusion Approximation (DA) [7] to Radiative Transfer Equation (RTE) [4], because this model is more efficient than RTE regarding the computational cost. For the acoustic portion of the forward operator, we used a model that has been used in chapter 2 for our MG algorithm. To the best of our knowledge, existing studies for the single-stage problem of QPAT has been limited to homogeneous and lossless media, and have been solved by explicit inversion formulae [15, 12, 22]. Assuming a simplified acoustic medium for a biomedical application of direct QPAT may not be accurate, because this neglects a dependence of the propagation of acoustic waves on the characteristic properties of tissue media [10]. The arising error may not deteriorate the solution for a well-posed acoustic inverse problem of PAT, but for direct QPAT, this error may be significantly amplified by the optical portion of the forward operator, which is nonlinear and ill-conditioned. Motivated by this, we included the system of coupled first-order wave equations that has already been used in chapter 2 in our composite forward operator for direct QPAT. Using this, we simulated a heterogeneous and lossy acoustic medium with an absorption and physical dispersion that follow a frequency power law [25]. Since the heterogeneous acoustic properties of tissue media are not precisely accessible for image reconstruction, we avoided an inverse crime for medium's properties via an inclusion of a 30 dB noise in the acoustic properties for data generation.

The same as in chapter 2, we used a k-space pseudospectral method for the discretisation of the acoustic forward operator (cf. section 1.7). We also used a first-order finite element method for discretisation of a DA approximation that has been used for modelling the optical portion of the problem. To mitigate the ill-posedness of the problem, we used a multi-source setting for optical illuminations. Accordingly, we used four illuminations for a 2D phantom and three illuminations for a 3D phantom.

Using realistic and erroneous acoustic properties makes a single-stage inverse problem of QPAT more ill-posed than in [12] and [22]. A gradient-based quasi-Newton approach that uses an LBFGS method for approximation of the inverse of the Hessian matrix, and is regularised using a total variation approach in a framework based on an Alternating Direction Method of Multipliers (ADMM) algorithm, has been successfully applied for a 2D problem of single-stage QPAT [12]. We used this algorithm as a benchmark for our study. Although this algorithm converges well for the phantom used in [12], our numerical results show that this approach is not sufficiently accurate, when we use our acoustic forward operator for modelling propagation of acoustic waves in a medium with heterogeneous and erroneous sound speed and ambient density.

Correspondingly, we developed two inexact Newton algorithms for a single-stage problem of QPAT for realistic acoustic media. Our numerical results showed that our developed algorithms reconstruct more accurate images than nonlinear gradient-based algorithms using ADMM approach (our benchmark), for heterogeneous and lossy acoustic media. Note that our considered inverse problem is more ill-posed than existing studies, because we included an error in estimation of the acoustic properties, and also we used a limited view for detection of pressure data on the boundary. We applied these in order to make our simulations closer to a practical problem.

In our first developed approach, the residual function is iteratively linearised, and each linearised subproblem is solved using a linear preconsitioned conjugate gradient (PCG) algorithm, for which we used a matrix-free Jacobian based method. Also, a discretisation of a total variation functional that is updated using a Lagged Diffusivity (LD) approach [28] was used as a preconditioner [3]. The Jacobian matrix and an associated adjoint are computed implicitly as a discretisation of the Fréchet derivative of our composite forward operator and its adjoint. Our second proposed approach is an extension to our first approach, for which the objective function is linearised in two steps, the first of which is applied on the data fidelity function, and the second is used in order to handle the nonlinearity of a TV function using a PD-IPM approach.

In this thesis, we used PCG for solving each linearised subproblem for simplicity, but using more advanced krylov subspace methods such as LSQR [3] may be promising. As shown in section 4.6, the developed algorithms produced more accurate images than an LBFGS-based ADMM method (our benchmark). Also, we showed that our

developed preconditioned inexact Newton algorithms are less sensitive to changes in the reconstruction parameters, which are often chosen heuristically. Indeed, a high sensitivity of the solution to the *ad-hoc* reconstruction parameters may lead to inaccurate images, and thus a recalculation of the inverse problem may be required.

In addition, in our study, we stabilised the inverse problem of direct QPAT using a multi-source setting for optical illuminations, but using a multi-wavelength setting may be more practical for real cases, because the whole boundary may not be accessible for optical illuminations in medical applications. As a result, an extension of our developed algorithms to a muti-frequency variant of a single-stage inverse problem of QPAT looks promising [6].

Bibliography

[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. pages 1–19, 2016. arXiv:1605.00133v1.

[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] S. R. Arridge, M. M. Betcke, and L. Harhanen. Iterated preconditioned lsqr method for inverse problems on unstructured grids. *Inverse Problems*, 30(7):075009, 2014.

[4] G. Bal, A. Jollivet, and V. Jugnon. Inverse transport theory of photoacoustics. *Inverse Problems*, 26(2):025011, 2010.

[5] G. Bal and K. Ren. Multi-source quantitative photoacoustic tomography in a diffusive regime. *Inverse Problems*, 27(7):075003, 2011.

[6] G. Bal and K. Ren. On multi-spectral quantitative photoacoustic tomography in diffusive regime. *Inverse Problems*, 28(2):025010, 2012.

[7] G. Bal and G. Uhlmann. Inverse diffusion theory of photoacoustics. *Inverse Problems*, 26(8):085010, 2010.

[8] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.*, 18(11):2419–2434, 2009.

[9] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci.*, 2(1):183–202, 2009.

[10] 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.

[11] 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.

[12] H. Gao, J. Feng, and L. Song. Limited-view multi-source quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065004, 2015.

[13] H. Gao, S. Osher, and H. Zhao. *Quantitative Photoacoustic Tomography*, pages 131–158. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[14] M. Haltmeier, R. Kowar, and L. V. Nguyen. Iterative methods for photoacoustic tomography in attenuating acoustic media. *Inverse Problems*, 33(11):115009, 2017.

[15] M. Haltmeier, L. Neumann, and S. Rabanser. Single-stage reconstruction algorithm for quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065005, 2015.

[16] 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.

[17] 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.

[18] A. Javaherian and S. Holman. A multi-grid iterative method for photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 36(3):696–706, March 2017.

[19] A. Javaherian and S. Holman. A continuous adjoint for photo-acoustic tomography of the brain. *Inverse Problems*, 34(8):085003, 2018.

[20] M. Kocvara and S. Mohammad. A first-order multigrid method for bound-constrained convex optimization. *Optim. Method Softw.*, 31(3):622–644, 2016.

[21] S. G. Nash. A multigrid approach to discretized optimization problems. *Optim. Method Softw.*, 14:99–116, 2000.

[22] A. Pulkkinen, B. T. Cox, S. R. Arridge, H. Goh, J. P. Kaipio, and T. Tarvainen. Direct estimation of optical parameters from photoacoustic time series in quantitative photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 35(11):2497–2508, Nov 2016.

[23] A. Pulkkinen, B. T. Cox, S. R. Arridge, J. P. Kaipio, and T. Tarvainen. A bayesian approach to spectral quantitative photoacoustic tomography. *Inverse Problems*, 30(6):065012, 2014.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] C. Vogel and M. Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.