PAPER • OPEN ACCESS

Direct quantitative photoacoustic tomography for realistic acoustic media

To cite this article: Ashkan Javaherian and Sean Holman 2019 *Inverse Problems* **35** 084004

View the article online for updates and enhancements.

Recent citations

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Direct quantitative photoacoustic tomography for realistic acoustic media

Ashkan Javaherian[1] and Sean Holman[2]

[1] Department of Medical Physics and Biomedical Engineering, University College London, WC1E 6BT, London, United Kingdom
[2] School of Mathematics, University of Manchester, M13 9PL, Manchester, United Kingdom

E-mail: sean.holman@manchester.ac.uk and a.javaherian@ucl.ac.uk

CrossMark

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.

Keywords: photoacoustic tomography, single-stage, realistic acoustic media, heterogeneous sound speed and density, acoustic absorption

(Some figures may appear in colour only in the online journal)

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 [9, 28, 51, 52], 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 [9, 17, 23, 24, 54]. 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 [33, 39] or variational (optimisation) approaches [7, 15, 21, 25, 49]. 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 [19, 41, 42]. 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. [20, 35, 41]. 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 sufficiently 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 [19, 22, 32, 41, 42].

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, 18, 19]. 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 (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 [13, 40], 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 [45, 46]. 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 [13, 40].

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 [18, 20, 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 [18, 20, 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 [32, 41, 42] (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 [18, 19]. 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 [18, 19]. 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 [18, 19] for application on classical QPAT and direct QPAT, respectively.)

2. Direct QPAT on a continuous domain

In this section, we define our forward operator as a composite map on a continuous domain.

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

$$
\begin{aligned}
-\nabla \cdot \kappa(r)\nabla\phi(r) + \mu(r)\phi(r) = 0, \quad & r \in \Omega \\
\phi(r) + \frac{1}{2\gamma_d}\kappa(r)\frac{\partial\phi(r)}{\partial\hat{n}} = I/\gamma_d, \quad & r \in \partial\Omega,
\end{aligned}
\tag{1}
$$

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

$$
h[\kappa, \mu, I](r) = \mu(r)\phi[\kappa, \mu, I](r).
\tag{2}
$$

Because of thermo-elastic expansion effects, the induced spatially varying heating field causes an instant local increase in pressure that follows

$$
p_0(r) = \begin{cases} \Gamma(r)h(r), & r \in \Omega \\ 0, & r \in \mathbb{R}^d \backslash \Omega, \end{cases}
\tag{3}
$$

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].

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 [45, 46]. 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)

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

(4)

linearised equation of continuity (conservation of mass)

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

(5)

and equation of state

$$p(r,t) = c_0(r)^2 \left\{ 1 - \tau(r)\frac{\partial}{\partial t}(-\nabla^2)^{\frac{y}{2}-1} - \eta(r)(-\nabla^2)^{\frac{y-1}{2}} \right\} \rho(r,t)$$

(6)

with initial conditions

$$p(r,0) = p_0(r), \quad v(r,0) = 0.$$

(7)

Here, $\tau(r)$ and $\eta(r)$ are, respectively the absorption and dispersion proportionality coefficients, and are given by

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

(8)

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 1. For a fixed illumination I, the optical portion of the forward operator is a nonlinear map in the form

$$\Lambda_o : \mathcal{L}_+^\infty(\Omega) \times \mathcal{L}_+^\infty(\Omega) \to \mathcal{L}^2(\Omega)$$
$$\Lambda_o[\kappa, \mu](r) = h(r),$$

(9)

where $h(r)$ is given by (2). Also, the acoustic portion of the forward operator is a linear map in the form [1]

$$\Lambda_a : \mathcal{L}^2(\Omega) \to \mathbb{R}^{N_s N_t},$$
$$\Lambda_a[p_0](r,t) = \mathcal{M}p(r,t).$$

(10)

Using (9) and (10), the coupled opto-acoustic forward operator is defined by the composite map Λ, i.e.

$$\Lambda : \mathcal{L}_+^\infty(\Omega) \times \mathcal{L}_+^\infty(\Omega) \to \mathbb{R}^{N_s N_t},$$
$$\Lambda[\kappa, \mu] = \Lambda_a \big[\Lambda_o[\kappa, \mu] \big].$$

(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 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.

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]

$$\epsilon(\kappa, \mu) = \frac{1}{2} \sum_{q=1}^{N_q} \left\| \Lambda_q[\kappa, \mu] - \hat{P}_q \right\|_2^2, \tag{12}$$

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

$$D_x\epsilon = \sum_{q=1}^{N_q} D_x^*\Lambda_q \left(\Lambda_q[x] - \hat{P}_q \right). \tag{13}$$

Here, $D_x^*\Lambda$ denotes the adjoint of the Fréchet derivative of the forward operator, and is given by

$$D_x^*\Lambda[\hat{P}] = D_x^*\Lambda_o[\Lambda_a^*[\hat{P}]], \tag{14}$$

where Λ_a^* denotes the adjoint of the linear acoustic forward operator Λ_a (10) and $D_x^*\Lambda_o$ represents the adjoint of the Fréchet derivative of the optical forward operator (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 1. *Let us denote the solution of (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*

$$D[\kappa_0, \mu_0]\Lambda_o \begin{pmatrix} \delta\kappa \\ \delta\mu \end{pmatrix} = \delta\mu(r)\phi_0(r) + \mu_0(r)\delta\phi(r), \tag{15}$$

where $\delta\phi(r)$ satisfies

$$-\nabla \cdot \kappa_0(r)\nabla\delta\phi(r) + \mu_0(r)\delta\phi(r) = \nabla \cdot \delta\kappa(r)\nabla\phi_0(r) - \delta\mu(r)\phi_0(r), \quad r \in \Omega$$

$$\delta\phi(r) + \frac{1}{2\gamma_d}\kappa_0(r)\frac{\partial\delta\phi(r)}{\partial\hat{n}} = -\frac{1}{2\gamma_d}\delta\kappa(r)\frac{\partial\phi_0(r)}{\partial\hat{n}}, \quad r \in \partial\Omega. \tag{16}$$

The adjoint map $D_x^* \Lambda_o$ can then be calculated from

$$D[\kappa_0, \mu_0]^* \Lambda_o h(r) = \begin{pmatrix} \nabla \phi_0(r) \cdot \nabla \tilde{h}(r) \\ \phi_0(r) \tilde{h}(r) + \phi_0(r) h(r) \end{pmatrix}, \tag{17}$$

where the adjoint field $\tilde{h}(r)$ satisfies

$$-\nabla \cdot \kappa_0(r) \nabla \tilde{h}(r) + \mu_0(r) \tilde{h}(r) = -\mu_0(r) h(r), \quad r \in \Omega$$

$$\tilde{h}(r) + \frac{1}{2\gamma_d} \kappa_0(r) \frac{\partial \tilde{h}(r)}{\partial \hat{n}} = 0, \quad r \in \partial\Omega. \tag{18}$$

Lemma 1 can be proven using integration by parts.

3. Numerical computation

3.1. Numerical computation of the optical operators (Λ_o, $D_x \Lambda_o$ and $D_x^* \Lambda_o$)

3.1.1. 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 (1) is derived, i.e.

$$\int_\Omega \mu_0 \phi_0 \nu \, dr + \int_\Omega \kappa_0 \nabla \phi_0 \cdot \nabla \nu \, dr + 2\gamma_d \int_{\partial\Omega} \phi_0 \nu \, ds = \int_{\partial\Omega} 2 I_s \nu \, ds, \tag{19}$$

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 (16) is derived, i.e.

$$\int_\Omega \mu_0 \delta\phi \nu dr + \int_\Omega \kappa_0 \nabla\delta\phi \cdot \nabla\nu dr + 2\gamma_d \int_{\partial\Omega} \delta\phi \nu ds = -\int_\Omega \delta\kappa \nabla\phi_0 \cdot \nabla\nu dr - \int_\Omega \delta\mu\phi_0\nu dr. \tag{20}$$

In the same way, for an approximation of $D_x^* \Lambda_o$, a variational form of (18) is obtained, i.e.

$$\int_\Omega \mu_0 \tilde{h} \nu dr + \int_\Omega \kappa_0 \nabla \tilde{h} \cdot \nabla\nu dr + 2\gamma_d \int_{\partial\Omega} \tilde{h} \nu ds = -\int_\Omega \mu_0 h \nu dr. \tag{21}$$

3.1.2. 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 $\{\chi_j = 1_{t_j} \mid j = 1, ..., N_e\}$. Using this, the optical fields are approximated as [41]

$$\kappa_0(r) \approx \kappa^e(r) = \sum_{j=1}^{N_e} \hat{\kappa}_j \chi_j(r)$$

$$\mu_0(r) \approx \mu^e(r) = \sum_{j=1}^{N_e} \hat{\mu}_j \chi_j(r), \tag{22}$$

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 $\{\varphi_k \mid k = 1, ..., N_n\}$ in the form

$$\phi_0(r) \approx \phi_0^h(r) = \sum_{k=1}^{N_n} \Phi_{0,k} \varphi_k(r), \tag{23}$$

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

$$\tilde{h}(r) \approx \tilde{h}^h(r) = \sum_{k=1}^{N_n} \tilde{H}_k \varphi_k(r). \tag{24}$$

In the sequel, a field that is discretised at nodes as in (23) and (24) (resp. elements as in (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.

3.1.3. Matrix form of variational formulae. For a discretisation of the problem, a matrix form of the variational formulae in section 3.1.1 is derived [41, 43]. To do this, we define a system matrix $\boldsymbol{A}_o$ in the form

$$\boldsymbol{A}_o[\kappa^e, \mu^e] = K[\kappa^e] + C[\mu^e] + R \tag{25}$$

with

$$
\begin{aligned}
K_{kp}[\kappa^e] &= \int_\Omega \kappa^e \nabla \varphi_k \cdot \nabla \varphi_p \, \mathrm{d}r \\
C_{kp}[\mu^e] &= \int_\Omega \mu^e \varphi_k \varphi_p \, \mathrm{d}r \\
R_{kp} &= 2\gamma_d \int_{\partial\Omega} \varphi_k \varphi_p \, \mathrm{d}s \\
G_p &= \int_{\partial\Omega} 2I_s \varphi_p \, \mathrm{d}s,
\end{aligned}
\tag{26}
$$

where $p, k = 1,...,N_n$ denote nodal indices. We also define matrix $\boldsymbol{A}_{\delta,o}$ in the form

$$\boldsymbol{A}_{\delta,o}[\delta\kappa^e, \delta\mu^e] = K[\delta\kappa^e] + C[\delta\mu^e]. \tag{27}$$

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 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

$$
\begin{aligned}
&\mathbb{H}_o : \mathbb{R}^{N_e} \times \mathbb{R}^{N_e} \to \mathbb{R}^{N_e} \\
&H = \mathbb{H}_o[\hat{\kappa}, \hat{\mu}].
\end{aligned}
\tag{28}
$$

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

$$
\begin{aligned}
&\mathbb{J}_o : \mathbb{R}^{N_e} \times \mathbb{R}^{N_e} \to \mathbb{R}^{N_e} \\
&\delta H = \mathbb{J}_o[\hat{\kappa}, \hat{\mu}] \begin{pmatrix} \delta\hat{\kappa} \\ \delta\hat{\mu} \end{pmatrix}.
\end{aligned}
\tag{29}
$$

Now, we give further details on these operators.

3.1.4. Discretised forward operator $\mathbb{H}_o[\kappa_0, \mu_0]$. Plugging (23) into (19), together with taking $\nu(r)$ to be a basis function $\varphi_p(r)$, gives a linear system

$$\boldsymbol{A}_o[\kappa^e, \mu^e]\Phi_0 = G. \tag{30}$$

Using this, the heating field is approximated in a piecewise constant basis as [18, 19, 41]

$$H = \hat{\mu} \circ \mathbb{I}\Phi. \tag{31}$$

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

$$(\mathbb{I}\theta)_j = \frac{1}{d+1} \sum_{p \in \ell(j)} \theta_p \tag{32}$$

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

$$\left(\mathbb{I}^+\Theta\right)_p = \sum_{j=1}^{N_e} \Theta_j \int_{t_j} \varphi_p(r)\mathrm{d}r = \frac{1}{d+1} \sum_{j \in l(p)} S_j\Theta_j, \tag{33}$$

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 (32) and (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 $\delta X = [\delta\hat{\kappa}, \delta\hat{\mu}]$.

3.1.5. 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 (23). Plugging the nodal vector $\delta\Phi$ into (20) gives a linear system for calculation of $\delta\Phi$ in the form

$$\boldsymbol{A}_o[\kappa^e, \mu^e]\delta\Phi = -\boldsymbol{A}_{\delta,o}[\delta\kappa^e, \delta\mu^e]\Phi_0, \tag{34}$$

where Φ_0 has been computed using (30). Finally, the perturbation in the heating field δH is computed as

$$\delta H = \delta\hat{\mu} \circ \mathbb{I}\Phi_0 + \hat{\mu} \circ \mathbb{I}\delta\Phi. \tag{35}$$

3.1.6. 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 (24) into the variational form of the adjoint formula (21) gives a linear system in the form

$$\boldsymbol{A}_0[\kappa^e, \mu^e]\tilde{H} = -\mathcal{G}, \tag{36}$$

where

$$\mathcal{G} = \mathbb{I}^+(\hat{\mu}_0 \circ H). \tag{37}$$

Finally, given the nodal vectors $\tilde{H}$ and Φ_0, we will approximate the action of the adjoint using (17). For this we must choose how to calculate the products that appear in (17), and then how to project onto the space of elemental vectors. To do this, we will also use the matrices

$$\frac{\partial \boldsymbol{A}_o}{\partial \hat{\kappa}_j} = \int_{t_j} \nabla \varphi_k \cdot \nabla \varphi_p \, \mathrm{d}r \tag{38}$$

and

$$\frac{\partial \boldsymbol{A}_o}{\partial \hat{\mu}_j} = \int_{t_j} \varphi_k \varphi_p \, \mathrm{d}r. \tag{39}$$

If the products in (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

$$\mathbb{J}_o^*(H)_j = \begin{pmatrix} \frac{1}{S_j} \tilde{H}^T \frac{\partial \boldsymbol{A}_o}{\partial \hat{\kappa}_j} \Phi_0 \\ \frac{1}{S_j} \tilde{H}^T \frac{\partial \boldsymbol{A}_o}{\partial \hat{\mu}_j} \Phi_0 + H^T (\mathbb{I}\Phi_0) \end{pmatrix}. \tag{40}$$

Remark 1. Using a *discretise-then-adjoint method*, the adjoint of Fréchet derivative of the optical forward operator will be in the form (see [18], equation (27))

$$\mathbb{J}_{o,\mathrm{dis}}^*(H)_j = \begin{pmatrix} \left[(\mathbb{I}\boldsymbol{A}_o^{-1})^T (\hat{\mu}_0 \circ H) \right]^T \left(-\frac{\partial \boldsymbol{A}_o}{\partial \hat{\kappa}_j} \right) \Phi_0 \\ \left[(\mathbb{I}\boldsymbol{A}_o^{-1})^T (\hat{\mu}_0 \circ H) \right]^T \left(-\frac{\partial \boldsymbol{A}_o}{\partial \hat{\mu}_j} \right) \Phi_0 + H^T (\mathbb{I}\Phi_0) \end{pmatrix}. \tag{41}$$

From (36), (37) and the fact that $\mathbb{I}^+ \Theta = \mathbb{I}^T (S \circ \Theta)$ using (33), it can be shown that (41) matches (40).

Remark 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.

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 [13, 40]. 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 [13, 40].

3.2.1. 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

$$\mathbb{T}_i f(r) = f(\mathbb{T}_i r). \tag{42}$$

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

$$v_{(i;\zeta;n-1/2)} \approx \mathbb{T}_i v_i(r_\zeta, t_{n-1/2}). \tag{43}$$

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.)

3.2.2. 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

$$Y_{\text{abs}} = F^{-1} \left\{ k^{y-2} F\{\cdot\} \right\}, \quad Y_{\text{dis}} = F^{-1} \left\{ k^{y-1} F\{\cdot\} \right\}, \tag{44}$$

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

$$\frac{\partial \{\cdot\}}{\partial r_i^\pm} = F^{-1} \left\{ \mathbf{i} k_i e^{\pm \mathbf{i} k_i \Delta r_i/2} \text{sinc}\left(\bar{c}_{\text{ref}} k \Delta t/2\right) F\{\cdot\} \right\}, \tag{45}$$

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 [13, 40].

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]

$$A_i = \text{diag}\left(e^{-\alpha_{a,i}\Delta t/2}\right), \tag{46}$$

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 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

$$\mathbb{H} : \mathbb{R}^{N_e} \times \mathbb{R}^{N_e} \to \mathbb{R}^{N_s N_t}$$

$$\hat{P} = \mathbb{H}_a \mathbb{H}_o[X], \tag{47}$$

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

$$\mathbb{J} : \mathbb{R}^{N_e} \times \mathbb{R}^{N_e} \to \mathbb{R}^{N_s N_t}$$

$$\delta\hat{P} = \mathbb{H}_a \mathbb{J}_o[X]\delta X. \tag{48}$$

Using (48), we will use as a discretisation of the adjoint of Fréchet derivative operator the map

$$\mathbb{J}^* : \mathbb{R}^{N_s N_t} \to \mathbb{R}^{N_e} \times \mathbb{R}^{N_e}$$

$$\partial X = \mathbb{J}_o^*[X]\mathbb{H}_a^* \left(\partial\hat{P} \right). \tag{49}$$

3.2.3. 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)–(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 $n = \{-1/2, +1/2\}$. (See [1], appendix B, or the *k-Wave* manual [44]). This gives a source in coordinate i as [1]

$$s_{(i;\zeta;n+1/2)} = \begin{cases} \frac{1}{2d\Delta t \, \bar{c}^2} \mathbb{S}P_0 & n = -1, 0 \\ 0 & \text{otherwise,} \end{cases} \tag{50}$$

where $\mathbb{S}$ is a symmetric smoothing operator that is applied in order to mitigate unexpected oscillations of P_0. (See 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 (see [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)):

$$v_{(i;\zeta;n+\frac{1}{2})} = A_i \left[A_i \, v_{(i;\zeta;n-\frac{1}{2})} - \frac{\Delta t}{\bar{\rho}_i} \frac{\partial}{\partial r_i^+} p_{(\zeta;n)} \right]. \tag{51}$$

2. Update the acoustic density field (conservation of mass (5)) also adding source:

$$\rho_{(i;\zeta;n+1)} = A_i \left[A_i \, \rho_{(i;\zeta;n)} - \Delta t \bar{\rho} \frac{\partial}{\partial r_i^-} v_{(i;\zeta;n+\frac{1}{2})} \right] + \Delta t \, s_{(i;\zeta;n+1/2)}. \tag{52}$$

3. Update the scalar pressure field (equation of state (6)):

$$p_{(\zeta;n+1)} = \bar{c}^2 \left[(I_N - \bar{\eta}Y_{\text{dis}}) \sum_{i=1}^d \rho_{(i;\zeta;n+1)} + \bar{\tau}Y_{\text{abs}} \sum_{i=1}^d A_i \bar{\rho} \frac{\partial}{\partial r_i^-} v_{(i;\zeta;n+\frac{1}{2})} \right], \tag{53}$$

13

where I_N denotes an identity matrix of size $N \times N$, and the last term is actually the action of $\bar{\tau} Y_{\mathrm{abs}}$ on $\sum_{i=1}^{d} \frac{\partial}{\partial t} \rho_{(i;\zeta;n+1)}$, and is derived from (52).

4. *Compute the measured pressure at detectors:* This is defined using

$$\hat{P}_{n+1} = \mathbb{M} p_{(\zeta;n+1)}, \tag{54}$$

where $\mathbb{M}$ denotes a discretised variant of $\mathcal{M}$ (see section 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.

3.2.4. 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 3.2.3. 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

$$\mathcal{S}_n P_0 = \Delta t \, \mathcal{I}_s^z s_{(i;\zeta;n+1/2)} = \mathcal{I}_s^z 1_i \otimes \left(\frac{1}{2dC}\right) \mathbb{S} P_0 \quad \text{for } n = -1, \, 0, \tag{55}$$

and $\mathcal{S}_n P_0 = 0$ for $n \neq -1, \, 0$. Here, $s_{(i;\zeta;n+1/2)}$ has been derived from (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 (52) and (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

$$\mathcal{I}_s^z = \begin{bmatrix} 0_{dN \times dN} \\ I_{dN \times dN} \\ C\left(\left(I_N - \bar{\eta} Y_{\mathrm{dis}}\right) \sum_{i=1}^{d}\right)_{N \times dN} \end{bmatrix}. \tag{56}$$

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

$$\bar{z}_{n+1} = T\bar{z}_n + S_{n+1/2}. \tag{57}$$

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. The discretised acoustic forward operator is defined by

$$\mathbb{H}_a : \mathbb{R}^N \to \mathbb{R}^{N_s N_t} \tag{58}$$

$$\mathbb{H}_a = \mathbb{H}_T \mathcal{S}, \tag{59}$$

where $\mathbb{H}_T : \mathbb{R}^{(2d+1)NN_t} \to \mathbb{R}^{N_s N_t}$ satisfies

$$\hat{P} = \mathbb{H}_T S, \quad \hat{P}_n = \boldsymbol{M}\bar{z}_n \quad (n \in \{0, ..., N_t - 1\}), \quad \hat{P} = \left[\hat{P}_{n+1}\right]_{n=-1}^{N_t-2}. \tag{60}$$

Here, $\bar{z}_n$ is determined by (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 2. *The action of the adjoint operator*

$$\mathbb{H}_a^* : \mathbb{R}^{N_s N_t} \to \mathbb{R}^N, \tag{61}$$

$$\mathbb{H}_a^* = \mathcal{S}^* \mathbb{H}_T^*, \tag{62}$$

on $\hat{P}$, where $P_0 = \mathbb{H}_a^* \hat{P}$, is given by

$$P_0 = \sum_{n=-1}^{N_t-2} \mathcal{S}_n^* \bar{z}_{N_t-2-n}^*, \tag{63}$$

where $\bar{z}_n^*$ is determined by

$$\bar{z}_{-1}^* = 0, \quad \bar{z}_{n+1}^* = T^* \bar{z}_n^* + \boldsymbol{M}^* \hat{P}_{N_t-2-n} \quad (n \in \{-1, ..., N_t - 2\}). \tag{64}$$

Proof. We have given the proof in [27], lemma 2 and corollary 1. $\qquad\square$

Note that in the case we are considering the sum in (63) is actually just two terms. By commuting $n = -1, 0$ in (55) with $\boldsymbol{M}$ using the fact that the forward operator is linear, we define a source for our acoustic adjoint operator as

$$s_{n+1/2}^{\text{adj}} = \mathbb{M}^* \begin{cases} \hat{P}_{N_t-1}, & n = -1 \\ \hat{P}_{N_t-n-1} + \hat{P}_{N_t-n-2}, & n = 0, ..., N_t - 2 \\ \hat{P}_0, & n = N_t - 1. \end{cases} \tag{65}$$

From (65), equation (64) in lemma 2 can be modified as

$$\bar{z}_{-1}^* = 0, \quad \bar{z}_{n+1}^* = T^* \bar{z}_n^* + S_{n+1/2}^{\text{adj}} \quad (n \in \{-1, ..., N_t - 2\}), \tag{66}$$

where

$$S_{n+1/2}^{\text{adj}} = \left(\mathcal{I}_z^p\right)^* s_{n+1/2}^{\text{adj}}. \tag{67}$$

Now, we derive a matrix form of T and T^* using the forward model presented in section 3.2.3 to show how multiplication by each of them may be computed. We also denote a matrix form of the k-space spatial gradient in (45) as

$$\nabla_i^{\pm} = \frac{\partial}{\partial r_i^{\pm}} \in \mathbb{R}^{N \times N}. \tag{68}$$

Using this, we will use $\Phi \in \mathbb{R}^{dN \times dN}$ and $\Psi \in \mathbb{R}^{dN \times dN}$ composed of submatrices

$$(\Phi)_{ij} = -\Delta t \, \delta_{ij} \nabla_j^+ \tag{69}$$

and

$$(\Psi)_{ij} = -\Delta t \, \delta_{ij} \nabla_j^- . \tag{70}$$

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

$$D\bar{v} = C \left(I_N - \bar{\eta} Y_{\text{dis}} - \bar{\tau} \frac{Y_{\text{abs}}}{\Delta t} \right) \sum_{i=1}^{d} (AQ\Psi A^2 \bar{v})_i \tag{71}$$

$$E\bar{\rho} = C \left(I_N - \bar{\eta} Y_{\text{dis}} \right) \sum_{i=1}^{d} (A^2 \bar{\rho})_i \tag{72}$$

$$G\bar{p} = C \left(I_N - \bar{\eta} Y_{\text{dis}} - \bar{\tau} \frac{Y_{\text{abs}}}{\Delta t} \right) \sum_{i=1}^{d} (AQ\Psi A Q_s^{-1} \Phi 1_i \otimes \bar{p})_i. \tag{73}$$

The matrix T is then given, in block form, by

$$T = \begin{pmatrix} A^2 & 0 & AQ_s^{-1}\Phi 1_i \otimes \\ AQ\Psi A^2 & A^2 & AQ\Psi A Q_s^{-1}\Phi 1_i \otimes \\ D & E & G \end{pmatrix}. \tag{74}$$

From (74), the adjoint operator T^* will be in the form

$$T^* = \begin{pmatrix} A^2 & A^2 \Psi^* Q A & D^* \\ 0 & A^2 & E^* \\ \sum_{i=1}^{d} \Phi^* Q_s^{-1} A & \sum_{i=1}^{d} \Phi^* Q_s^{-1} A \Psi^* Q A & G^* \end{pmatrix}. \tag{75}$$

From (75) and using (56), (66) and (67), as well as calculating the adjoints of D, E and G from (71)–(73), the time sequence of adjoint fields is given iteratively by

$$\bar{\rho}_{n+1} = A^2 \left[\bar{\rho}_n + 1_i \otimes (I_N - Y_{\text{dis}}\bar{\eta}) C\bar{p}_n \right]$$

$$\bar{v}_{n+1/2} = A^2 \left[\bar{v}_{n-1/2} + \Psi^* Q A^{-1} \left(\bar{\rho}_{n+1} - A^2 1_i \otimes \frac{Y_{\text{abs}}}{\Delta t} \bar{\tau} C\bar{p}_n \right) \right]$$

$$\bar{p}_{n+1} = \left(\sum_{i=1}^{d} \Phi^* Q_s^{-1} A^{-1} \bar{v}_{n+1/2} \right) + s_{n+1/2}^{\text{adj}}$$

$$\bar{p}^{\text{sol}} = \vartheta \left(\sum_{i=1}^{d} 1_i \otimes (I_N - Y_{\text{dis}}\bar{\eta}) C\bar{p}_{n+1} + \bar{\rho}_{n+1} \right) \quad (n = N_t - 2), \tag{76}$$

where $\bar{p}^{\text{sol}} := \mathbb{H}_a^*(\hat{P})$, and $\vartheta = \mathbb{S}\left(\frac{1}{2dC}\right)$ using (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

$$\hat{\boldsymbol{\rho}}_{n+1} = A \left[A\hat{\boldsymbol{\rho}}_n + Q1_i \otimes (I_N - Y_{\mathrm{dis}}\bar{\eta}) \, C\bar{p}_n \right]$$

$$\hat{\boldsymbol{v}}_{n+1/2} = A \left[A\hat{\boldsymbol{v}}_{n-1/2} - Q_s^{-1} \Phi \left(\hat{\boldsymbol{\rho}}_{n+1} - QA1_i \otimes \frac{Y_{\mathrm{abs}}}{\Delta t} \bar{\tau} C\bar{p}_n \right) \right]$$

$$\bar{p}_{n+1} = \left(\sum_{i=1}^{d} -\Psi \hat{\boldsymbol{v}}_{n+1/2} \right) + s_{n+1/2}^{\mathrm{adj}}$$

$$\bar{p}^{\mathrm{sol}} = \boldsymbol{\vartheta} \left(\sum_{i=1}^{d} 1_i \otimes (I_N - Y_{\mathrm{dis}}\bar{\eta}) \, C\bar{p}_{n+1} + AQ^{-1}\hat{\boldsymbol{\rho}}_{n+1} \right) \quad (n = N_t - 2).$$

$$(77)$$

Our numerical experiments showed that an operator $\mathbb{H}_a^*$ that is calculated using a *discretise-then-adjoint* method (77) satisfies an adjoint test with a higher accuracy than using an *adjoint-then-discretise* method used in [26]. Therefore, we used (77) for the acoustic adjoint operator.

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 (12).

Considering (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 (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 [18, 20, 31].

Remark 3. Since the magnitude of $\hat{\kappa}$ is often 1–2 orders of magnitude greater than $\hat{\mu}$, we follow [18, 19] and minimise ϵ (see (12)) with respect to scaled coefficients

$$\bar{X} = [\bar{\kappa}(\hat{\kappa}), \bar{\mu}(\hat{\kappa})]^T. \tag{78}$$

We will use two types of scaling. For the method described in section 5.1 we use a linear scaling (see (99)) while for the methods described in sections 5.2.1 and 5.2.2 we use logarithmic scaling (see (100)). In our numerical experiments the method of section 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

$$\bar{X}_* = \underset{\bar{X}_l \leqslant \bar{X} \leqslant \bar{X}_u}{\operatorname{argmin}} \ \epsilon(\bar{X}) = \underset{\bar{X}_l \leqslant \bar{X} \leqslant \bar{X}_u}{\operatorname{argmin}} \frac{1}{2} \sum_{q=1}^{N_q} \left\| \mathbb{H}_q[X[\bar{X}]] - \hat{P}_q \right\|_2^2, \tag{79}$$

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 $(\bar{X})_j \geqslant (\bar{X}_l)_j$ and $(\bar{X})_j \leqslant (\bar{X}_u)_j$ for all components j of these vectors. Note that here, we choose $(\bar{X}_l)_j$ and $(\bar{X}_u)_j$ fixed for all components j.

We will take two main approaches for solving (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

$$\bar{X}_{k+1} = \bar{X}_k + \alpha_k d_k, \tag{80}$$

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.)

4.1. Newton's methods

From the second-order optimality condition for minimising ϵ, the *Newton search direction*, is derived using

$$d_k = -\mathcal{H}_k^{-1} \nabla \epsilon_k. \tag{81}$$

Here, $\nabla \epsilon_k$ is the first-order derivative of ϵ at $\bar{X}_k$, and is computed using

$$\nabla \epsilon_k = \sum_{q=1}^{N_q} \frac{\partial X}{\partial \bar{X}} [\bar{X}_k] \, \mathbb{J}_q^*[X_k] \left(\mathbb{H}_q[X_k] - \hat{P}_q \right), \tag{82}$$

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

$$\mathcal{H}_k = \sum_{q=1}^{N_q} \frac{\partial X}{\partial \bar{X}} [\bar{X}_k] \, \mathbb{J}_q^*[X_k] \mathbb{J}_q[X_k] \frac{\partial X}{\partial \bar{X}} [\bar{X}_k], \tag{83}$$

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 (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 (see (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], p 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.

Table 1. The minimal and maximal values for acoustic properties of the 2D phantom.

	c_0 (m s^{-1})		ρ_0 (kg m^{-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

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 [18, 19]. Using L-BFGS, $\mathcal{H}_k^{-1}$ is updated using the most recent m pairs of (s, y) given by

$$
\begin{aligned}
s_k &= \bar{X}_k - \bar{X}_{k-1} \\
y_k &= \nabla \epsilon_k - \nabla \epsilon_{k-1}.
\end{aligned}
\tag{84}
$$

(In our study, we empirically use $m = 5$.) We will also use $\rho_k = 1/y_k^T s_k$, and an initial guess for the Hessian matrix in the form

$$
\mathcal{H}_{k,0}^{-1} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} I,
\tag{85}
$$

where I is the identity matrix. By applying L-BFGS method to the constrained minimisation problem (79), the search direction d_k is computed using algorithm 1. (See [19] and [18] for applications on QPAT and direct QPAT.)

Algorithm 1. L-BFGS (search direction): inner iteration k.

1: Input: $\nabla \varepsilon_k, \mathcal{H}_{k,0}^{-1}$
2: Initialise: $q = \nabla \varepsilon_k$
3: **for** $i = k - 1, k - 2, ..., k - m$ **do**
4:      $\alpha_i = \rho_i s_i^T q$
5:      $q = q - \alpha_i y_i$
6: **end for**
7: $r = \mathcal{H}_{k,0}^{-1} q$
8: **for** $i = k - m, k - m + 1, ..., k - 1$ **do**
9:      $r = r + s_i \left(\alpha_i - \rho_i y_i^T r \right)$
10: **end for**
11: $d_k = -r$.

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))

$$
d_k = \begin{cases}
-\bar{X}_l, & \text{if } \bar{X}_k + d_k \leqslant \bar{X}_l \\
d_k, & \text{if } \bar{X}_l < \bar{X}_k + d_k < \bar{X}_u \\
-\bar{X}_u, & \text{if } \bar{X}_k + d_k \geqslant \bar{X}_u.
\end{cases}
\tag{86}
$$

Using d_k given by (86), the step size in (80) is chosen by a standard backtracking line search satisfying the *Wolfe* conditions as well as the constraints

$$
\begin{aligned}
\epsilon(\bar{X}_k + \alpha_k d_k) &\leqslant \epsilon(\bar{X}_k) + c_1 \alpha_k d_k^T \nabla \epsilon(\bar{X}_k) \\
d_k^T \nabla \epsilon(\bar{X}_k + \alpha_k d_k) &\geqslant c_2 d_k^T \nabla \epsilon(\bar{X}_k),
\end{aligned}
\tag{87}
$$

together with enforcing the bounds associated with the constraint, i.e.

$$
\bar{X}_l \leqslant \bar{X}_k + \alpha_k d_k \leqslant \bar{X}_u.
\tag{88}
$$

(See [18, 19].) In (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 2. Here, $\tau < 1$ is a user-defined parameter.

Algorithm 2. Backtracking line search: inner iteration k.

1: Input: $c_1, c_2, \tau, \bar{X}_l, \bar{X}_u$
2: Initialise: $\alpha_0 = 1$
3: **while** (87) or (88) are not satisfied **do**
4:     $\alpha = \tau \alpha$
5: **end while**
6: $\alpha_k = \alpha_*$.

---

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 (79) [18, 19]. This results in a minimisation problem in the form

$$
\bar{X}_* = \underset{\bar{X}_l \leqslant \bar{X} \leqslant \bar{X}_u}{\arg \min} \left\{ \mathcal{F} := \epsilon[\bar{X}] + \lambda \mathcal{J}[\bar{X}] \right\},
\tag{89}
$$

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

$$
\int_\Omega |\nabla u| \mathrm{d}r
\tag{90}
$$

with u either $\bar{\kappa}$ or $\bar{\mu}$. Using (90),

$$
\mathcal{J}[\bar{X}] = \|D\bar{X}\|_1
\tag{91}
$$

with

$$
D = \begin{bmatrix} D^{\bar{\kappa}} & 0_{N_l \times N_e} \\ 0_{N_l \times N_e} & D^{\bar{\mu}} \end{bmatrix}.
\tag{92}
$$

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 $D_l^u \in \mathbb{R}^{1 \times N_e}$ ($l \in \{1, ..., N_l\}$) 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

Table 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

$$\nabla \mathcal{J} : \mathbb{R}^{2N_e} \to \mathbb{R}^{2N_e}$$
$$\nabla \mathcal{J}[\bar{X}] = M[\bar{X}]\bar{X}, \tag{93}$$

where $M(\bar{X})$ is given by

$$M[\bar{X}] = D^T C[\bar{X}] D \tag{94}$$

with $C(\bar{X})$ a diagonal matrix

$$C[\bar{X}] = \text{diag}\left((|D\bar{X}|^2 + \beta)^{-1/2} \right). \tag{95}$$

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.

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 [18, 19], may be further rewritten as

$$\mathcal{F}_A(W, \bar{X}) = \varrho\left(\frac{1}{2}\|D\bar{X} - W + U_w\|_2^2 + \nu\|W\|_1 \right) + \sum_{q=1}^{N_q} \frac{1}{2}\|\mathbb{H}_q[X(\bar{X})] - \hat{P}_q + U_{p,q}\|_2^2, \tag{96}$$

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 (96) in W and $\bar{X}$, and updating of the Lagrange multipliers, using algorithm 3. In algorithm 3, Tol_{out} is a terminating threshold, and X^* denotes an optimal solution. The line 4 in algorithm 3 is a minimisation of the first term in (96) with respect to W, and can be calculated exactly using a scalar-wise *Shrinkage formula* of the form

$$W^{k+1} = \max\left\{ |D\bar{X}^k + U_w^k| - \nu, 0 \right\} \text{sgn}(D\bar{X}^k + U_w^k). \tag{97}$$

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 4.2. 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

$$\nabla_{\bar{X}}\mathcal{F}_A = \varrho D^T \left(D\bar{X} - W + U_w \right) + \sum_{q=1}^{N_q} \frac{\partial X}{\partial \bar{X}}[\bar{X}]\mathbb{J}_q^*[X] \left(\mathbb{H}_q[X] - \hat{P}_q + U_{P_q} \right). \tag{98}$$

Table 3. The minimal and maximal values for acoustic properties of the 3D phantom.

	c_0 (m s^{-1})			ρ_0 (kg m^{-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

For this method we use a linear scaling:

$$\bar{X} = \begin{bmatrix} \bar{\kappa} \\ \bar{\mu} \end{bmatrix} = \begin{bmatrix} \frac{\hat{\kappa}}{\mathrm{mean}(\hat{\kappa}^0)} \\ \frac{\hat{\mu}}{\mathrm{mean}(\hat{\mu}^0)} \end{bmatrix}, \tag{99}$$

where the dominators are the mean value of initial guesses $\hat{\kappa}^0$ and $\hat{\mu}^0$.

Algorithm 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** $\|\nabla_{\bar{X}} \mathcal{F}_A (\bar{X}^k, W^k)\| > Tol_{\mathrm{out}}$ **do**
4:　　$W^{k+1} = \arg \min_{W} \mathcal{F}_A (\bar{X}^k, W)$
5:　　$\bar{X}^{k+1} = \arg \min_{\bar{X}_l \leqslant \bar{X} \leqslant \bar{X}_u} \mathcal{F}_A (\bar{X}, W^{k+1})$
6:　　$U_w^{k+1} = U_w^k + D\bar{X}^{k+1} - W^{k+1}$
7:　　$U_{p,q}^{k+1} = U_{p,q}^k + \mathbb{H}_q[X^{k+1}] - \hat{P}_q \ (\forall q)$
8:　　Output: X^*
9: **end while**

5.2. Linearised matrix-free Jacobian-based method

In this section, we explain two methods that we use for solving the problem (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

$$\bar{X} = \log \frac{X}{X^0}, \tag{100}$$

which also implicitly enforces positivity on $\hat{\kappa}$ and $\hat{\mu}$. Note that in (100), the division X/X^0 and the logarithm operator are understood elementwise. Additionally, 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:

$$\mathbb{H}(\bar{X}) \approx \mathbb{H}(\bar{X}^k) + \mathbb{J}[X^k]\frac{\partial X}{\partial \bar{X}}[\bar{X}^k](\bar{X} - \bar{X}^k). \tag{101}$$

Applying the approximation (101) on the problem (79) yields the minimisation problem

$$d^k = \underset{d}{\operatorname{argmin}} \ \frac{1}{2} \sum_{q=1}^{N_q} \left\| \mathbb{J}_q[X^k]\frac{\partial X}{\partial \bar{X}}[\bar{X}^k]d - \left(\hat{P}_q - \mathbb{H}_q[X^k]\right) \right\|_2^2, \tag{102}$$

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 (81) that indicates an inner iteration. The kth linearised subproblem (102) gives a *normal equation* in the form of

$$\mathcal{H}^k d^k = -\nabla \epsilon^k, \tag{103}$$

where $\nabla \epsilon^k$ and $\mathcal{H}^k$ are obtained from (82) and (83), respectively. The normal equation (103) is a variant of (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 (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 4.1, *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 (103) roughly using a loose stopping tolerance, i.e.

$$\mathcal{H}^k \tilde{d}^k = -\nabla \epsilon^k + v_k, \quad \|v^k\|_2 / \|\nabla \epsilon^k\|_2 \leqslant \eta^k, \tag{104}$$

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 $\alpha^k = 1$ (see [14], section 2). A class of approaches that use (104) for minimisation of (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 (104), and we hope this is not confusing for the readers.

5.2.1. 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 (104) is based on solving (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 kth linearisation of the data fidelity term ϵ using (102), together with $\nabla \mathcal{J}[\bar{X}]$ defined by (93), gives a TV regularised variant of (104) in the form

$$\mathcal{H}^k d^k + \lambda M[\bar{X}] \bar{X} = -\nabla \epsilon^k. \tag{105}$$

Let us denote an initial guess for kth subproblem by $\bar{X}_0^k$, which is calculated using the previous linearised subproblem $k - 1$. One way for addressing the nonlinearity of $M[\bar{X}]\bar{X}$ (see (94)) is replacing $M[\bar{X}]$ by $M^k = M[\bar{X}_0^k]$. This gives a normal equation in the form

$$\left(\mathcal{H}^k + \lambda M^k\right) d^k = -\nabla \epsilon^k - \lambda M^k \bar{X}_0^k. \tag{106}$$

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 (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 (106) we next multiply by $(M_\gamma^k)^{-1}$ so that (106) becomes

$$\left((M_\gamma^k)^{-1}\mathcal{H}^k + \lambda I\right) d^k = -(M_\gamma^k)^{-1}\nabla \epsilon^k - \lambda \bar{X}_0^k. \tag{107}$$

Table 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

Applying *Krylov* methods for solving (106), the iterates lie in a subspace [2]

$$\mathcal{K}^{(M_\gamma^k)^{-1}\mathcal{H}^k+\lambda I} = \mathcal{K}^{(M_\gamma^k)^{-1}\mathcal{H}^k} = \text{span}\left\{ -\left((M_\gamma^k)^{-1}\mathcal{H}^k\right)^{i\in\left\{0,\ldots,i_{\max}^k-1\right\}}\left((M_\gamma^k)^{-1}\nabla\epsilon^k + \lambda\bar{X}_0^k\right)\right\}, \quad (108)$$

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 (106) in the form

$$(M_\gamma^k)^{-1}\mathcal{H}^k d^k = -(M_\gamma^k)^{-1}\nabla\epsilon^k. \quad (109)$$

It turns out that (109) provides a better convergence than (106) since the structure of the prior is directly included in the Jacobian matrix [2]. Here, we solve (109) using the *Preconditioned Conjugate Gradient* (PCG) method, as outlined in algorithm 4.

Algorithm 4. PCG algorithm for solving linearised subproblem k.

1: Input: $\nabla\epsilon^k, \mathcal{H}^k, M_\gamma^k$
2: Initialise: $i = 0, \bar{X}_0 = 0$
3: $\boldsymbol{r}_0 = -\nabla\epsilon^k$
4: Solve $M_\gamma^k \boldsymbol{z}_0 = \boldsymbol{r}_0$
5: $\boldsymbol{d}_0 = \boldsymbol{z}_0$
6: **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**
7:     $\boldsymbol{\alpha}_i = \frac{\boldsymbol{r}_i^T \boldsymbol{z}_i}{\boldsymbol{d}_i^T \mathcal{H}^k \boldsymbol{d}_i}$
8:     $\bar{X}_{i+1} = \bar{X}_i + \boldsymbol{\alpha}_i \boldsymbol{d}_i$
9:     $\boldsymbol{r}_{i+1} = \boldsymbol{r}_i - \boldsymbol{\alpha}_i \mathcal{H}^k \boldsymbol{d}_i$
10:     Solve $M_\gamma^k \boldsymbol{z}_{i+1} = \boldsymbol{r}_{i+1}$
11:     $\boldsymbol{\beta}_i = \frac{\boldsymbol{r}_{i+1}^T \boldsymbol{z}_{i+1}}{\boldsymbol{r}_i^T \boldsymbol{z}_i}$
12:     $\boldsymbol{d}_{i+1} = \boldsymbol{z}_{i+1} + \boldsymbol{\beta}_i \boldsymbol{d}_i$
13: **end while**
14: Output: d^k

We will terminate algorithm 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. 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 5.

Algorithm 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 (89)
5:     Compute d^k from (109) using algorithm 4
6:     Compute $\bar{X}^{k+1}$ using (80) and $\alpha^k = 1$
7:     $X^{k+1} = X^0 e^{\bar{X}^{k+1}}$
8: **end while**
9: Output: X^*

5.2.2. 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 (102) in the form

$$d^k = \underset{d}{\text{argmin}} \ \frac{1}{2} \sum_{q=1}^{N_q} \left\| \mathbb{J}_q[X_0^k] \frac{\partial X}{\partial \bar{X}}[\bar{X}_0^k] d - \left(\hat{P}_q - \mathbb{H}_q[X_0^k] \right) \right\|_2^2 + \lambda \|Dd\|_1. \quad (110)$$

The main idea for linearisation of the nonlinear M (see (94) and (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

$$\lambda D^T \chi + \mathcal{H}d - \sum_{q=1}^{N_q} \frac{\partial X}{\partial \bar{X}}[\bar{X}_0] \mathbb{J}_q^*[X_0] \left(\hat{P}_q - \mathbb{H}_q[X_0] \right) = g(\chi, d) = 0$$

$$C^{-1}[d]\chi - Dd = f(\chi, d) = 0, \quad (111)$$

where we have removed the superscripts indicating subproblem k for brevity. A linearisation of this system with respect to (χ, d) gives

$$\begin{bmatrix} C^{-1}[d] & -\left(I - C[d]\chi(Dd)^T\right)D \\ \lambda D^T & \mathcal{H} \end{bmatrix} \begin{bmatrix} \delta\chi \\ \delta d \end{bmatrix} = \begin{bmatrix} -f(\chi, d) \\ -g(\chi, d) \end{bmatrix} \quad (112)$$

Here, we have used the derivative of the terms in the left-hand-sides of (111) with respect to χ and d, and the fact that $C^{-1}[d] = \text{diag}\left((|Dd|^2 + \beta)^{1/2}\right)$ from (95). If we make the replacement $\chi = C[d]Dd$, then the above linearised system gives decoupled equations

$$\left[\lambda D^T C[d_{k'}]\left(I - C[d_{k'}]\chi_{k'}(Dd_{k'})^T\right)D + \mathcal{H}\right]\delta d_{k'}$$

$$= -\lambda M(d_{k'}) + \sum_{q=1}^{N_q}\frac{\partial X}{\partial \overline{X}}[\overline{X}_0]\mathbb{J}_q^*[X_0]\left(\hat{P}_q - \mathbb{H}_q[X_0] - \mathbb{J}_q[X_0]\frac{\partial X}{\partial \overline{X}}[\overline{X}_0]d_{k'}\right) \tag{113}$$

and

$$\delta\chi_{k'} = C[d_{k'}]\left(I - C[d_{k'}]\chi_{k'}(Dd_{k'})^T\right)D\delta d_{k'} - \chi_{k'} + C[d_{k'}]Dd_{k'}, \tag{114}$$

where the subscript k' indicates an inner sub-subproblem. The same as the LD method, we solve a priorconditioned form of (113) with $\lambda = 0$, i.e.

$$\tilde{M}_{\gamma,k'}^{-1}\mathcal{H}\delta d_{k'} = -\tilde{M}_{\gamma,k'}^{-1}\nabla\tilde{\epsilon}_{k'}, \tag{115}$$

where

$$\tilde{M}_{\gamma,k'} = D^T C[d_{k'}]\left(I - C[d_{k'}]\chi_{k'}(Dd_{k'})^T\right)D + \gamma I, \tag{116}$$

and $-\nabla\tilde{\epsilon}_{k'}$ is actually the second term in the right-hand-side of (113). Here, the subscript k' indicates the fact that for each linearised subproblem k (superscript), we solve a sequence of sub-subproblems (114) and (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 6. We use step sizes $\alpha_{k'} = 1$ and $\alpha^k = 1$ for all k' and k, respectively. The step size $s_{k'}$ is described below the algorithm (see (117)).

Algorithm 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 (110)
5:     Initialise: $\overline{X}_0^k, d_0 = 0, \chi_0 = 0$

6:     **while** $k' < k'_{\text{max}}$ $\cap$ $\left(k' = 0 \cup 1 - \frac{r_{k',*}}{r_{k'-1,*}} > Tol_{\text{med}}\right)$ **do**
7:        Compute $\delta d_{k'}$ from (115) using algorithm 4
8:        $d_{k'+1} = d_{k'} + \alpha_{k'}\delta d_{k'}$
9:        Compute $\delta\chi_{k'}$ using (114)
10:       $\chi_{k'+1} = \chi_{k'} + s_{k'}\delta\chi_{k'}$
11:     **end while**
12:     Compute $\overline{X}^{k+1}$ using (80) and $\alpha^k = 1$
13:     $X^{k+1} = X^0 e^{\overline{X}^{k+1}}$
14: **end while**
15: Output: X^*

Here, each iteration k' amounts to solving a PCG loop. The optimal r provided by each PCG loop k' is denoted by $r_{k',*}$. Using this, we terminate each outer subproblem k using a stopping criterion given in line 6 in algorithm 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

$$s_{k'} = \min{(1, \varphi_*)}\,\delta\chi_{k'}.\tag{117}$$

Here, φ_* is the largest φ that satisfies a feasibility condition

$$(|\chi_{k'} + \varphi\delta\chi_{k'}|)_j \leqslant 1,\ \forall j = 1, ..., 2N_l,\tag{118}$$

where j denotes the index of components of χ.

6. Numerical results

The TV regularised minimisation approaches that have been explained in section 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.

6.1. 2D phantom

The 2D simulation was performed on a square domain $[-5, +4.92] \times [-5, +4.92]$ mm^2.

6.1.1. 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^{-1}) that obeys

$$I_{s,q}(r) = \begin{cases} 1, & r \in \iota_q \\ 0, & r \in \partial\Omega \backslash \iota_q, \end{cases}\tag{119}$$

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 (see the second line in equation (1)).

6.1.2. 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 1. A 30 dB additive white Gaussian noise (AWGN) was then added to the simulated data.

6.1.3. 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.

6.1.4. 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
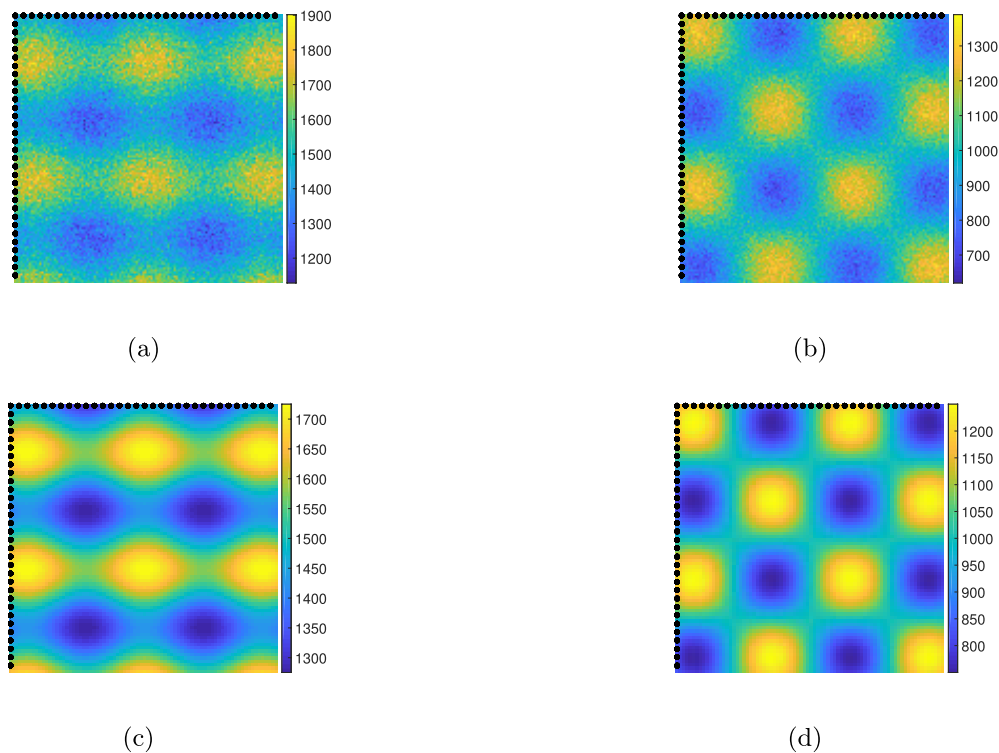
Figure 1. Acoustic properties for 2D case. Data generation: (a) c_0 (b) ρ_0, and image reconstruction: (c) c_0 (d) ρ_0.

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 1(a) and (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 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 1(a) and (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 1(c) and (d). The minimal and maximal values of these maps are given in the bottom row of table 1. Using these acoustic maps, the grid for image reconstruction supports a maximal frequency up to 5.101 MHz. From table 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 MHz^{-y} cm^{-1} and an exponent factor $y = 1.5$ for both data generation and image reconstruction (see (8)). These values were chosen so that they approximately simulate the acoustic attenuation properties of the breast.

6.1.5. 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 six values within a range $\kappa \in [0.2, 0.4]$ mm^{-1} with a background 0.3 mm^{-1}. Figures 2(a) and (b) show the map for μ and κ, respectively.

6.1.6. 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 (see algorithm 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 3 is a minimisation of $\mathcal{F}_A$ (see (96)) with respect to $\bar{X}$ using the L-BFGS algorithm given in algorithm 1. The L-BFGS algorithm uses a backtracking line search given in algorithm 2 with $c_1 = 1 \times 10^{-4}$, $c_2 = 0.9$ and $\tau = 0.25$. Also, we set $\varrho = 1$ (see (96)), and $\nu = 1 \times 10^{-3}$ for the *Shrinkage* operator (97). We terminated the ADMM algorithm using $\text{Tol}_{\text{out}} = 1 \times 10^{-2}$ (see algorithm 3).

LD. The LD method was applied using algorithm 5. Using this algorithm, each linearised subproblem is solved using a PCG loop (see algorithm 4) by setting $i_{\max} = 30$, $i_m = 5$, and $\text{Tol}_{\text{in}} = 0$. The latter parameter implies that we terminate each PCG loop, if

$$i > i_{\max} \ \cup \ \left(i > i_m \ \cap \ \boldsymbol{r}_i^T \boldsymbol{z}_i > \boldsymbol{r}_{i-i_m}^T \boldsymbol{z}_{i-i_m} \right). \tag{120}$$

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 $\text{Tol}_{\text{out}} = 1 \times 10^{-3}$ (see algorithm 5).

PD-IPM. The PD-IPM technique was applied using algorithm 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 $\text{Tol}_{\text{med}} = 1 \times 10^{-3}$ and $k'_{\max} = 20$ (see algorithm 6). Our PD-IPM was terminated using a stopping threshold $\text{Tol}_{\text{out}} = 1 \times 10^{-3}$.

6.1.7. 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

$$\text{RE}(u^k) = 100 \times \frac{\|u^k - u^{\text{phantom}}\|_2}{\|u^{\text{phantom}}\|_2}. \tag{121}$$
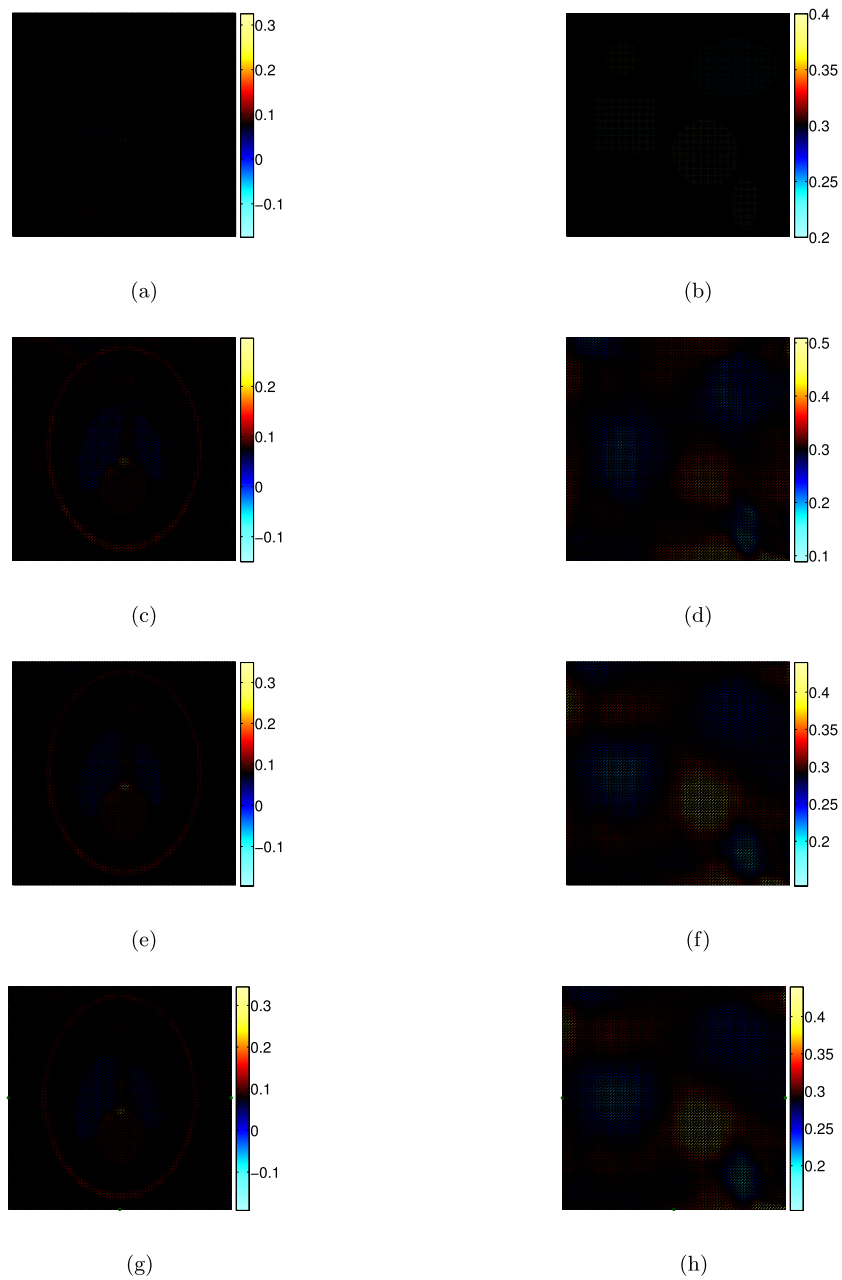
Figure 2. Optical coefficients for 2D case. Phantom: (a) absorption coefficient μ (b) diffusion coefficient κ. The images reconstructed by ADMM: (c) μ (d) κ, LD: (e) μ (f) κ, and PD-IPM: (g) μ (h) κ.

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 coefficients for the phantom.

We also consider $\epsilon(X^k)$ as the second criterion for convergence (see (79)).

6.1.8. Observations.

ADMM. Using the parameters given in section 6.1.6, 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 2(c) and (d), respectively. Figure 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 2(e) and (f) show the final reconstructed images for μ and κ, respectively. Figure 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(a). This figure is shown from an enlarged view around the optimal solution in figure 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 four outer iterations. The final reconstructed images for μ and κ are shown in figures 2(g) and (h), respectively. Figure 3(c) shows the RE of solutions (optical coefficients) computed by the PD-IPM algorithm for outer iterations k. Figure 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(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 5. Using PD-IPM, for both outer and inner linearised subproblems, the first step is a compuation of associated ϵ and the gradient $\nabla\epsilon$ (see algorithm 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 $r_i^T z_i$ provided by the PCG loops (see algorithm 4) is sufficient for a monotonic reduction of the objective function without using a line search.

Table 2 shows the RE values for the final reconstructed images shown in figure 2.

6.2. 3D phantom

We performed our 3D simulation on a cubic domain $[-5, 5] \times [-5, 5] \times [-5, 5]$ mm^3.

6.2.1. 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 (119) with $\iota_q \subset \partial\Omega$ two confronting faces of the grid, i.e. the left–right, posterior–anterior, and bottom–top faces.

6.2.2. 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
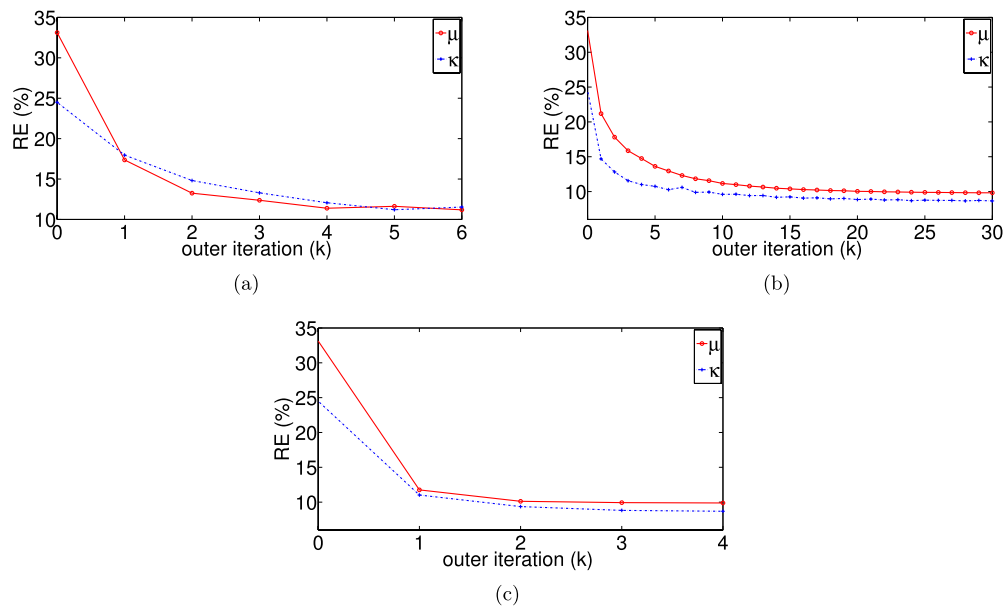
Figure 3. RE versus outer iteration *k* for 2D case: (a) ADMM, (b) LD, (c) PD-IPM.
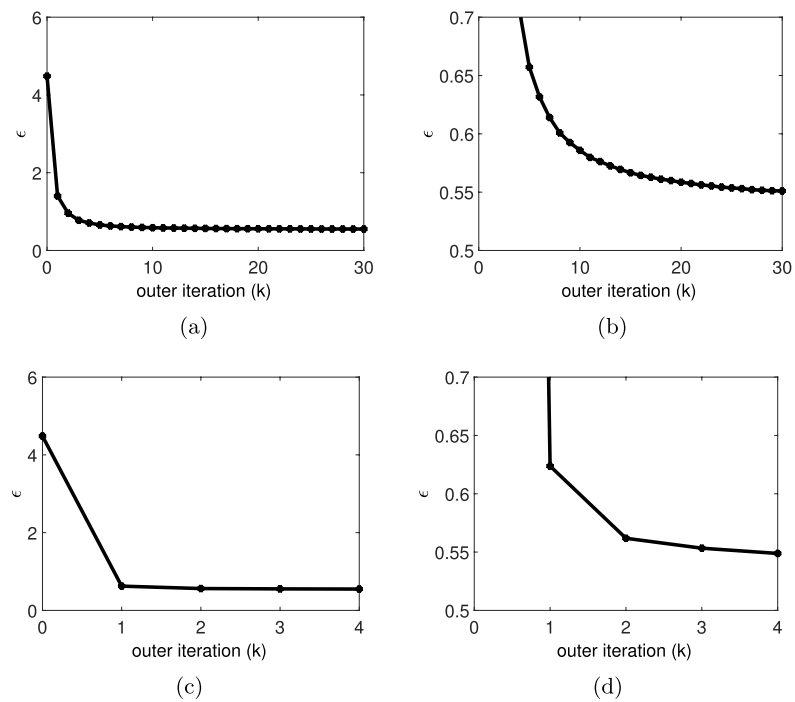

Figure 4. ϵ versus outer iteration *k* for 2D case. (a) LD, (b) LD from an enlarged view around the optimal point, (c) PD-IPM, (d) PD-IPM from an enlarged view around the optimal point.

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 AWGN was then added to the simulated data.

6.2.3. 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.

6.2.4. Acoustic properties. As discussed in section 6.1.4, 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 3 shows the minimal 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 5(a) and (b) (resp. figures 5(c) and (d)), respectively. Additionally, the acoustic attenuation coefficient and the associated exponent factor (see equation (8)) was simulated the same as the 2D phantom.

6.2.5. Optical phantom. The left columns in figures 6(a) and (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.

6.2.6. 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 $\text{Tol}_{\text{out}} = 1 \times 10^{-3}$.

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 $i_{\max} = 30$ (see algorithm 4). We also terminated each outer linearised subproblem using $k'_{\max} = 25$ and $\text{Tol}_{\text{med}} = 1 \times 10^{-3}$. Also, our PD-IPM algorithm was terminated using a stopping threshold $\text{Tol}_{\text{out}} = 1 \times 10^{-3}$.

6.2.7. Observations.

ADMM. The stopping criterion for the ADMM algorithm was satisfied after outer iteration 7. In figures 6(a) and (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 7(a) shows the RE of solutions computed by the ADMM algorithm for μ and κ at outer iterations.
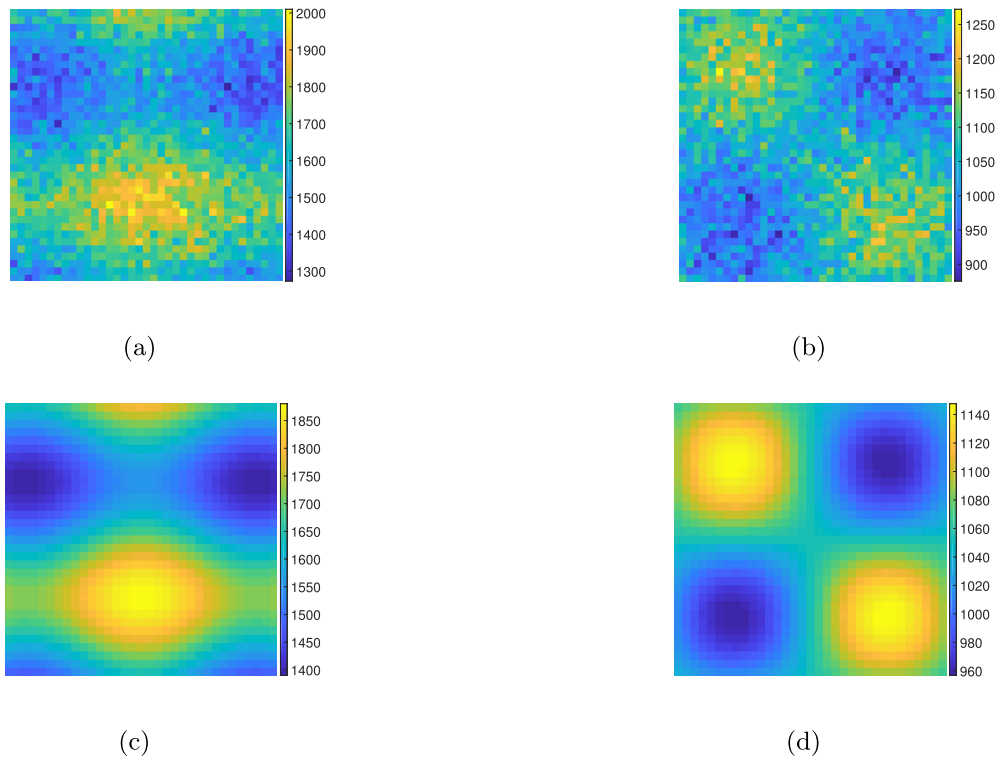
Figure 5. Acoustic properties for 3D case from a top view. Data generation: (a) c_0, (b) ρ_0; and image reconstruction: (c) c_0, (d) ρ_0.

LD. The LD algorithm was terminated after ten iterations. The final reconstructed images for μ and κ are shown in the 3rd columns of figures 6(a) and (b), respectively. The images are shown in the same way as the ADMM method. Figure 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 figures 8(a), and (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 six outer iterations. The final reconstructed images for μ and κ are shown in the 4th columns of figures 6(a) and (b), respectively. Figure 7(c) shows the RE of solutions (optical coefficients) computed by the PD-IPM algorithm for outer iterations k. Figures 8(c) and (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. $\epsilon_{k'}$).
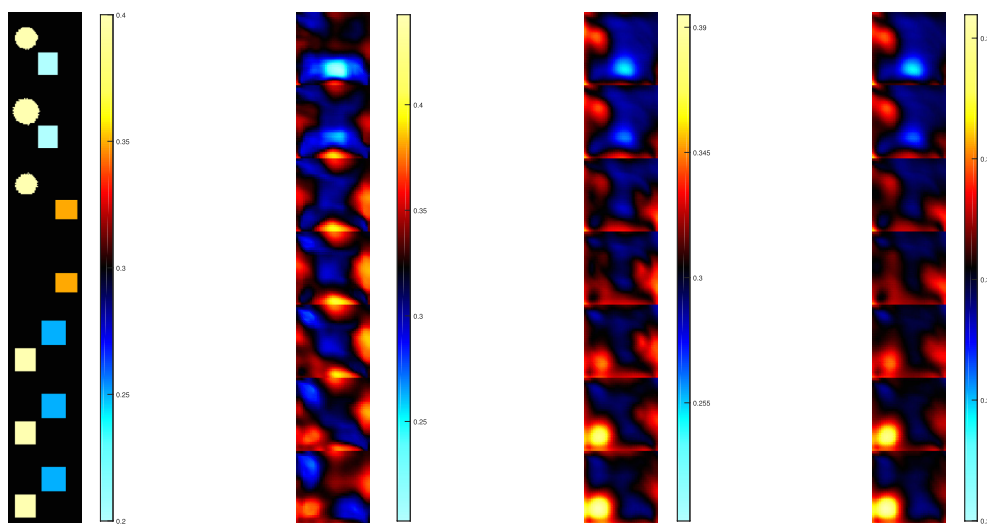
Table 4 shows the RE values for the final reconstructed images shown in figure 6.

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.
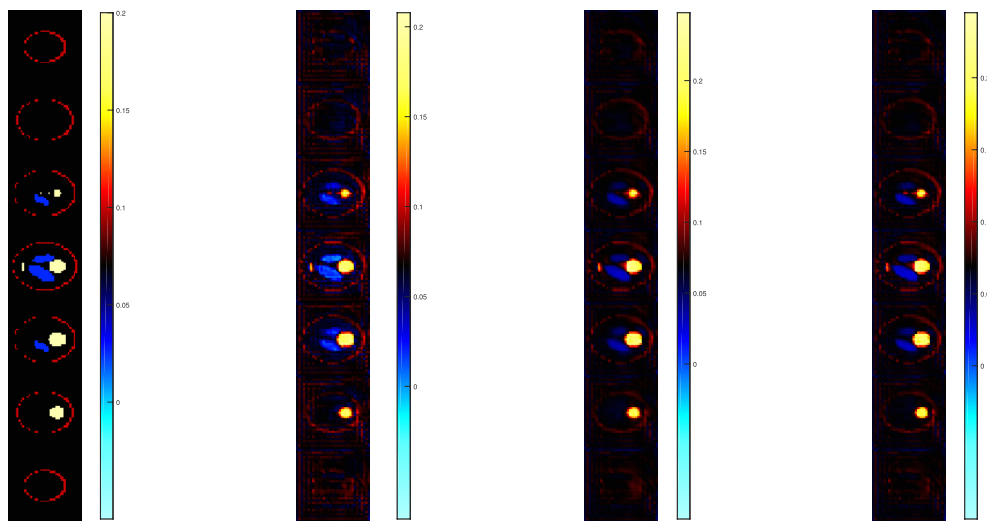
(a)

(b)

Figure 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.
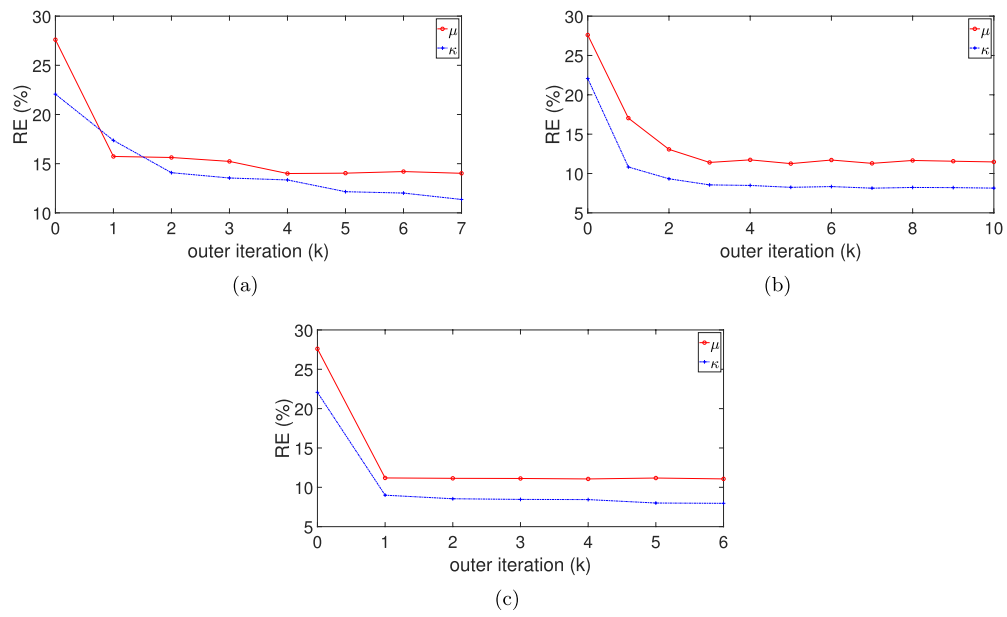
Figure 7. RE versus outer iteration k for 3D case. (a) ADMM, (b) LD, (c) PD-IPM.
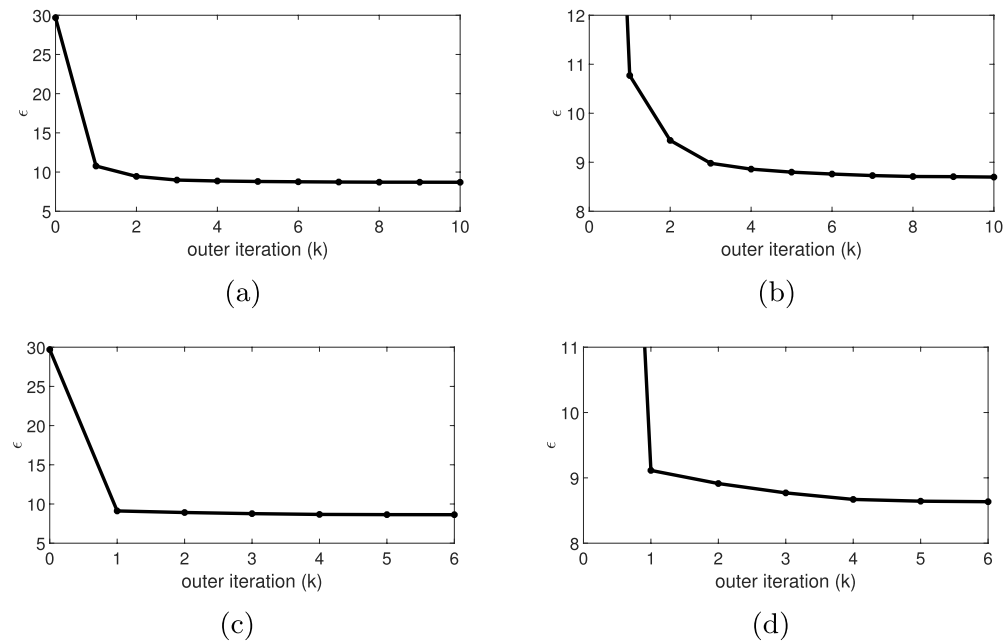

Figure 8. ϵ versus outer iteration k for 3D case. (a) LD, (b) LD from an enlarged view around the optimal point, (c) PD-IPM, (d) PD-IPM from an enlarged view around the optimal point.

The 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 6, 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 1 and 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(b) and 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 $\text{Tol}_{\text{med}} = 0$, 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 [18, 19], 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].

Acknowledgment

ORCID iDs

Ashkan Javaherian ![ORCID] https://orcid.org/0000-0003-0798-3196
Sean Holman ![ORCID] https://orcid.org/0000-0001-8050-2585

References

[1] Arridge S R, Betcke M M, Cox B T, Lucka F and Treeby B E 2016 On the adjoint operator in photoacoustic tomography *Inverse Problems* **32** 115012

[2] Arridge S R, Betcke M M and Harhanen L 2014 Iterated preconditioned lsqr method for inverse problems on unstructured grids *Inverse Problems* **30** 075009

[3] Bal G, Jollivet A and Jugnon V 2010 Inverse transport theory of photoacoustics *Inverse Problems* **26** 025011

[4] Bal G and Ren K 2011 Multi-source quantitative photoacoustic tomography in a diffusive regime *Inverse Problems* **27** 075003

[5] Bal G and Ren K 2012 On multi-spectral quantitative photoacoustic tomography in diffusive regime *Inverse Problems* **28** 025010

[6] Bal G and Uhlmann G 2010 Inverse diffusion theory of photoacoustics *Inverse Problems* **26** 085010

[7] Belhachmi Z, Glatz T and Scherzer O 2016 A direct method for photoacoustic tomography with inhomogeneous sound speed *Inverse Problems* **32** 045005

[8] Borsic A, Graham B M, Adler A and Lionheart W R B 2010 *In vivo* impedance imaging with total variation regularization *IEEE Trans. Med. Imaging* **29** 44–54

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

[10] Chan T, Golub G and Mulet P 1999 A nonlinear primal-dual method for total variation-based image restoration *SIAM J. Sci. Comput.* **20** 1964–77

[11] Cox B, Laufer J G, Arridge S R and Beard P C 2012 Quantitative spectroscopic photoacoustic imaging: a review *J. Biomed. Opt.* **17** 061202

[12] Cox B T, Arridge S R, Köstli K P and Beard P C 2006 Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method *Appl. Opt.* **45** 1866–75

[13] Cox B T, Kara S, Arridge S R and Beard P C 2007 k-space propagation models for acoustically heterogeneous media: application to biomedical photoacoustics *J. Acoust. Soc. Am.* **121** 3453–64

[14] Dembo R, Eisenstat S and Steihaug T 1982 Inexact newton methods *SIAM J. Numer. Anal.* **19** 400–8

[15] Ding L, Deán-Ben X L, Lutzweiler C, Razansky D and Ntziachristos V 2015 Efficient non-negative constrained model-based inversion in optoacoustic tomography *Phys. Med. Biol.* **60** 6733–50

[16] Ding T, Ren K and Vallélian S 2015 A one-step reconstruction algorithm for quantitative photoacoustic imaging *Inverse Problems* **31** 095005

[17] Finch D and Patch R S K 2004 Determining a function from its mean values over a family of spheres *SIAM J. Math. Anal.* **35** 1213–40

[18] Gao H, Feng J and Song L 2015 Limited-view multi-source quantitative photoacoustic tomography *Inverse Problems* **31** 065004

[19] Gao H, Osher S and Zhao H 2012 Quantitative Photoacoustic Tomography *Mathematical Modeling in Biomedical Imaging II.* (*Lecture Notes in Mathematics* vol 2035) ed H Ammari (Berlin: Springer) pp 131–58

[20] Haltmeier M, Neumann L and Rabanser S 2015 Single-stage reconstruction algorithm for quantitative photoacoustic tomography *Inverse Problems* **31** 065005

[21] Haltmeier M and Nguyen L V 2017 Analysis of iterative methods in photoacoustic tomography with variable sound speed *SIAM J. Imaging Sci.* **10** 751–81

[22] Hannukainen A, Hyvonen N, Majander H and Tarvainen T 2016 Efficient inclusion of total variation type priors in quantitative photoacoustic tomography *SIAM J. Imaging Sci.* **9** 1132–53

[23] Hristova Y 2009 Time reversal in thermoacoustic tomography—an error estimate *Inverse Problems* **25** 055008

[24] Hristova Y, Kuchment P and Nguyen L 2008 Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media *Inverse Problems* **24** 055006

[25] Huang C, Wang K, Nie L, Wang L V and Anastasio M A 2013 Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media *IEEE Trans. Med. Imaging* **32** 1097–110

[26] Javaherian A and Holman S 2017 A multi-grid iterative method for photoacoustic tomography *IEEE Trans. Med. Imaging* **36** 696–706

[27] Javaherian A and Holman S 2018 A continuous adjoint for photo-acoustic tomography of the brain *Inverse Problems* **34** 085003

[28] Kunyansky L A 2007 Explicit inversion formulae for the spherical mean radon transform *Inverse Problems* **23** 373

[29] Schweiger M, Arridge S R, Hiraoka M and Delpy D T 1995 The finite element method for the propagation of light in scattering media: boundary and source conditions *Med. Phys.* **22** 1779–92

[30] Matthews T P and Anastasio M A 2017 Joint reconstruction of the initial pressure and speed of sound distributions from combined photoacoustic and ultrasound tomography measurements *Inverse Problems* **33** 124002

[31] Pulkkinen A, Cox B T, Arridge S R, Goh H, Kaipio J P and Tarvainen T 2016 Direct estimation of optical parameters from photoacoustic time series in quantitative photoacoustic tomography *IEEE Trans. Med. Imaging* **35** 2497–508

[32] Pulkkinen A, Cox B T, Arridge S R, Kaipio J P and Tarvainen T 2014 A bayesian approach to spectral quantitative photoacoustic tomography *Inverse Problems* **30** 065012

[33] Qian J, Stefanov P, Uhlmann G and Zhao H 2011 An efficient neumann series-based algorithm for thermoacoustic and photoacoustic tomography with variable sound speed *SIAM J. Imaging Sci.* **4** 850–83

[34] Rosenthal A, Ntziachristos V and Razansky D 2013 Acoustic inversion in optoacoustic tomography: a review *Curr. Med. Imaging Rev.* **9** 318–36

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

[36] Schweiger M, Arridge S R and Nissilä I 2005 Gauss–Newton method for image reconstruction in diffuse optical tomography *Phys. Med. Biol.* **50** 2365

[37] Shao P, Harrison T and Zemp R J 2012 Iterative algorithm for multiple illumination photoacoustic tomography (mipat) using ultrasound channel data *Biomed. Opt. Express* **3** 3240–9

[38] Song N, Deumié C and Silva A D 2014 Considering sources and detectors distributions for quantitative photoacoustic tomography *Biomed. Opt. Express* **5** 3960–74

[39] Stefanov P and Uhlmann G 2009 Thermoacoustic tomography with variable sound speed *Inverse Problems* **25** 075011

[40] Tabei M, Mast T D and Waag R C 2002 A k-space method for coupled first-order acoustic propagation equations *J. Acoust. Soc. Am.* **111** 53–63

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

[42] Tarvainen T, Pulkkinen A, Cox B T, Kaipio J P and Arridge S R 2013 Bayesian image reconstruction in quantitative photoacoustic tomography *IEEE Trans. Med. Imaging* **32** 2287–98

[43] Tarvainen T, Vauhkonen M, Kolehmainen V and Kaipio J P 2006 Finite element model for the coupled radiative transfer equation, diffusion approximation *Int. J. Numer. Methods Eng.* **65** 383–405

[44] Treeby B E and Cox B T 2010 k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields *J. Biomed. Opt.* **15** 021314

[45] Treeby B E and Cox B T 2010 Modeling power law absorption and dispersion for acoustic propagation using the fractional Laplacian *J. Acoust. Soc. Am.* **127** 2741–8

[46] Treeby B E, Zhang E Z and Cox B T 2010 Photoacoustic tomography in absorbing acoustic media using time reversal *Inverse Problems* **26** 115003

[47] Vogel C and Oman M 1996 Iterative methods for total variation denoising *SIAM J. Sci. Comput.* **17** 227–38

[48] Wang K and Anastasio M A 2011 *Photoacoustic and Thermoacoustic Tomography: Image Formation Principles* (New York: Springer) pp 781–815

[49] Wang K, Su R, Oraevsky A A and Anastasio M A 2012 Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography *Phys. Med. Biol.* **57** 5399–423

[50] Wang L V and Yao J 2016 A practical guide to photoacoustic tomography in the life sciences *Nat. Methods* **13** 627–38

[51] Xu M and Wang L V 2002 Time-domain reconstruction for thermoacoustic tomography in a spherical geometry *IEEE Trans. Med. Imaging* **21** 814–22

[52] Xu M and Wang L V 2005 Universal back-projection algorithm for photoacoustic computed tomography *Phys. Rev.* E **71** 016706

[53] Xu Y, Feng D and Wang L V 2002 Exact frequency-domain reconstruction for thermoacoustic tomography. I. Planar geometry *IEEE Trans. Med. Imaging* **21** 823–8

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

[55] Xu Y, Xu M and Wang L V 2002 Exact frequency-domain reconstruction for thermoacoustic tomography. II. Cylindrical geometry *IEEE Trans. Med. Imaging* **21** 829–33