Inverse Problems **29** (2013) 075006 (19pp) doi:10.1088/0266-5611/29/7/075006

A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation

T Saratoon[1], T Tarvainen[2,3], B T Cox[1] and S R Arridge[3]

[1] Department of Medical Physics and Bioengineering, University College London, London WC1E 6BT, UK
[2] Department of Applied Physics, University of Eastern Finland, PO Box 1627, FI-70211 Kuopio, Finland
[3] Centre for Medical Image Computing, University College London, London WC1E 6BT, UK

E-mail: teedah.saratoon.09@ucl.ac.uk, tanja.tarvainen@uef.fi, bcox@ucl.ac.uk and S.R.Arridge@cs.ucl.ac.uk

Received 18 September 2012, in final form 2 May 2013
Published 12 June 2013
Online at stacks.iop.org/IP/29/075006

Abstract

Quantitative photoacoustic tomography (QPAT) offers the possibility of high-resolution molecular imaging by quantifying molecular concentrations in biological tissue. QPAT comprises two inverse problems: (1) the construction of a photoacoustic image from surface measurements of photoacoustic wave pulses over time, and (2) the determination of the optical properties of the imaged region. The first is a well-studied area for which a number of solution methods are available, while the second is, in general, a nonlinear, ill-posed inverse problem. Model-based inversion techniques to solve (2) are usually based on the diffusion approximation to the radiative transfer equation (RTE) and typically assume the acoustic inversion step has been solved exactly. Here, neither simplification is made: the full RTE is used to model the light propagation, and the acoustic propagation and image reconstruction are included in the simulations of measured data. Since Hessian- and Jacobian-based minimizations are computationally expensive for the large data sets typically encountered in QPAT, gradient-based minimization schemes provide a practical alternative. The acoustic pressure time series were simulated using a k-space, pseudo-spectral time domain model, and a time-reversal reconstruction algorithm was used to form a set of photoacoustic images corresponding to four illumination positions. A regularized, adjoint-assisted gradient inversion using a finite element model of the RTE was then used to determine the optical absorption and scattering coefficients.

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

1. Introduction

Photoacoustic tomography (PAT) uses laser-generated ultrasound to produce high-contrast, high-resolution images of biological tissue *in vivo* [17, 19, 21, 29, 34]. PAT images represent the initial distribution of acoustic pressure generated from the absorption of short (ns) pulses of laser light, delivered to a region of soft tissue. When light is absorbed by chromophores within the tissue, there quickly follows a small, localized increase in temperature and pressure which, due to the elasticity of soft tissue, results in the propagation of acoustic waves that are detected by an ultrasound transducer array placed at the tissue surface [30]. Temporal measurements of these photoacoustic waves and knowledge of the speed of sound in biological tissue enable the reconstruction of the initial pressure distribution following the absorption of the light pulse.

Quantitative photoacoustic tomography (QPAT) comprises two separable inverse problems: the first involves the reconstruction of the distribution of initial pressure from time-series measurements of the propagated acoustic wave pulses, and is a well-studied area for which a number of solution techniques are available [18]. This work is mostly concerned with the second inverse problem of QPAT, often referred to as the optical inverse problem, which aims to determine the chromophore concentration distributions from a reconstructed PAT image. Since the contrast in a PAT image is provided by the absorption of light within the tissue, PAT images are dependent on the concentrations of chromophores present within the tissue region. The absorption-based contrast also means that PAT images may be selectively enhanced by tuning the excitation wavelength to a known peak in the absorption spectra of a particular chromophore. This means that a technique to determine quantitatively accurate spatial maps of specific chromophore concentrations could be combined with a spectroscopic approach to obtain quantitative estimates of important physiological parameters, e.g. blood oxygen saturation. Moreover, the ability to quantify the concentrations of externally administered contrast agents would enable photoacoustic molecular imaging. Since chromophore concentrations are linearly related to the optical absorption coefficient, an equivalent solution to the problem is to develop a method to determine the absorption coefficient from a measured PAT image. However, whether recovering chromophore concentrations or optical absorption, the optical inverse problem of QPAT is, in general, a nonlinear, ill-posed, large-scale inverse problem.

Several inversion methods have been proposed for QPAT. These typically assume that the propagation of light throughout the tissue is near-isotropic, and so can be modelled using the diffusion approximation (DA) to the radiative transfer equation (RTE). However, light propagation in biological tissue is highly anisotropic in regions close to light sources, and does not behave diffusely until it has travelled a few transport mean free paths. The DA does not therefore hold in regions of great interest in QPAT, e.g. close to the tissue surface. For this reason we have used the full RTE to model light transport within the tissue. The acoustic propagation and image reconstruction were also included in the simulation of PAT images, so that the inversion technique tackles both the acoustic and optical inverse problems, as would be required in practice.

In this paper we present a model-based inversion technique which utilizes a transport-based light model and a quasi-Newton minimization scheme. Minimization schemes iteratively reduce the difference between a measured and modelled image, and often use curvature information provided by the Hessian matrix of second derivatives, or an approximation to it, to direct this minimization. Since PAT images are three dimensional and high resolution, the data sets involved can be very large. The calculation, inversion and even the storage of a full Hessian matrix may therefore be unfeasible, and in this case even Jacobian-based methods, which approximate the full Hessian using the Jacobian matrix of first derivatives, can become

impractical. Quasi-Newton methods require only gradient information to approximate the Hessian matrix and are therefore memory-efficient even for large data sets, making them a practical alternative for the full QPAT inversion.

The structure of the paper is as follows. Section 2 describes the formulation of the photoacoustic model used to simulate PAT images and perform the quasi-Newton minimization. In section 3 we motivate the choice of light transport model and minimization algorithm used in the presented method, and include a brief discussion on existing minimization-based methods for QPAT. The presented adjoint-assisted, gradient-based method is detailed in section 4, and the results of the method using two-dimensional, simulated PAT images are presented and discussed in section 5.

2. Formulation of the problem

In PAT, a spatially dependent ultrasound signal is generated by illuminating an optically absorbing and scattering medium with short pulses of visible or near-infrared light. As the time-scale for the acoustic wave propagation is typically around three orders of magnitude longer than the optical illumination and absorption, the optical and acoustic inverse problems may be decoupled and tackled separately [6].

In practice, a PAT image is reconstructed from temporal surface measurements of propagated photoacoustic waves. By modelling the physical processes involved in the generation of a photoacoustic signal and performing the image reconstruction, PAT images can be simulated and used in a model-based inversion scheme. Consider a tissue region of interest $\Omega \in \mathbb{R}^n$ with $n = 2, 3$. Light propagation in a turbid medium such as biological tissue is characterized by its absorption and scattering properties, namely the absorption coefficient $\mu_a(\mathrm{r})$ and scattering coefficient $\mu_s(\mathrm{r})$. These represent the probability per unit length of an absorption and scattering event, respectively, at a point $\mathrm{r} \in \Omega$. Light transport in biological tissue may be modelled analytically using the RTE, an integro-differential equation which represents the conservation of energy in a particular control volume. The RTE (given below in equation (1)) states that, as light travels throughout a region in a particular direction, energy can be lost through the absorption of a photon, the scattering of a photon out of the direction of interest or the net outflow of the region due to the gradient, and can be gained via the scattering of a photon into the direction of interest or from any light sources in the medium. Since the optical propagation and absorption can be treated as instantaneous on an acoustic timescale, the quantity of interest is the time-integrated radiance $\phi(\mathrm{r}, \hat{\mathbf{s}})$, which represents the energy per unit area at a point $\mathrm{r} \in \Omega$ and in a direction of interest $\hat{\mathbf{s}} \in S^{n-1}$. $\phi(\mathrm{r}, \hat{\mathbf{s}})$ is the solution of the time-integrated RTE, which takes the form

$$(\hat{\mathbf{s}} \cdot \nabla + \mu_a(\mathrm{r}) + \mu_s(\mathrm{r}))\phi(\mathrm{r}, \hat{\mathbf{s}}) - \mu_s \int_{S^{n-1}} \Theta(\hat{\mathbf{s}}, \hat{\mathbf{s}}')\phi(\hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}' = q(\mathrm{r}, \hat{\mathbf{s}}), \tag{1}$$

where q is a light source and $\Theta(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$ is the scattering phase function, a probability density function which describes the probability that a photon travelling in a direction $\hat{\mathbf{s}}$ will be scattered into a direction $\hat{\mathbf{s}}'$. Here, Θ is represented by the Henyey–Greenstein phase function [15], a commonly used phase function for biological tissue [5] which is characterized by the mean cosine of the angle between the incident and scattered light, usually called the anisotropy factor and denoted g. It is assumed that no photons travel in an inward direction at the boundary $\partial\Omega$ except at the source position $\mathrm{r}_s \subset \partial\Omega$, which gives the boundary condition

$$\phi(\mathrm{r}, \hat{\mathbf{s}}) = \begin{cases} \phi_0(\mathrm{r}, \hat{\mathbf{s}}), & \mathrm{r} \in \cup \mathrm{r}_s, \quad \hat{\mathbf{s}} \cdot \hat{\mathbf{n}} < 0 \\ 0, & \mathrm{r} \in \partial\Omega \setminus \cup \mathrm{r}_s, \quad \hat{\mathbf{s}} \cdot \hat{\mathbf{n}} < 0, \end{cases} \tag{2}$$

where ϕ_0 is the boundary source and $\hat{\mathbf{n}}$ is a unit vector normal to $\partial\Omega$. Since the radiance $\phi(\mathrm{r}, \hat{\mathbf{s}})$ is the energy per unit area at a point r in a direction $\hat{\mathbf{s}}$, the total energy at position r, often called the fluence and denoted here by $\Phi(\mathrm{r})$, is equal to its integral over all directions:

$$\Phi(\mathrm{r}) = \int_{S^{n-1}} \phi(\mathrm{r}, \hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}}. \tag{3}$$

The amount of optical energy absorbed per unit volume in Ω, denoted by $h(\mathrm{r})$, is then

$$h(\mathrm{r}) = \mu_{\mathrm{a}}(\mathrm{r})\Phi(\mathrm{r}). \tag{4}$$

Thermodynamic considerations enable the initial pressure p_0 arising from this optical absorption to be written as

$$p_0(\mathrm{r}) = \Gamma(\mathrm{r})h(\mathrm{r}), \tag{5}$$

where Γ is the photoacoustic efficiency, sometimes equated to the Grüneisen parameter, a dimensionless, tissue-specific property representing the conversion efficiency of heat energy to pressure, which we will assume is known throughout the tissue. When the sound speed and density are uniform and the optical excitation is regarded as instantaneous, the acoustic propagation may be described by an initial value problem for the homogenous wave equation

$$\left(\frac{\partial^2}{\partial t^2} - c_s^2\nabla^2\right) p = 0, \tag{6}$$

whose initial conditions are given by

$$p(\mathrm{r}, t)|_{t=0} = p_0(\mathrm{r}), \qquad \left.\frac{\partial p}{\partial t}\right|_{t=0} = 0, \tag{7}$$

where c_s is the sound speed and $p(\mathrm{r}, t)$ represents the acoustic pressure at a point $\mathrm{r} \in \Omega$ and time t [11]. The initial pressure distribution $p_0(\mathrm{r})$, which constitutes the PAT image, is reconstructed from measurements of the acoustic pressure over some arbitrary measurement surface. Numerous image reconstruction algorithms for reconstructing the initial pressure distribution exist [18]; here, (6) was solved using a k-space, pseudo-spectral time domain model and a time-reversal reconstruction algorithm was used to simulate a 'measured' photoacoustic image p_0^m [27, 28]. As Γ is known throughout the domain it is straightforward to obtain a measured absorbed energy map

$$h^{\mathrm{m}}(\mathrm{r}) = \frac{p_0^m(\mathrm{r})}{\Gamma(\mathrm{r})}, \tag{8}$$

from which the chromophore concentration distributions are to be recovered. Since chromophore concentrations are linearly related to the optical absorption coefficient via the chromophores' molar absorption coefficients, they can be obtained straightforwardly from μ_{a} provided that all contributing chromophore types are known. We therefore seek to determine a quantitatively accurate estimate of μ_{a} from the measured absorbed energy map h_m.

2.1. Challenges

The propagation of light in turbid media is strongly characterized by the wavelength-dependent optical absorption and scattering, so that (4) can be written as

$$h(\mathrm{r}, \lambda) = \mu_{\mathrm{a}}(\mathrm{r}, \lambda)\Phi(\mathrm{r}, \lambda; \mu_{\mathrm{a}}(\mathrm{r}, \lambda), \mu_{\mathrm{s}}(\mathrm{r}, \lambda)), \tag{9}$$

where we have noted explicitly any dependence on the optical wavelength λ. Since Φ and μ_{a} vary both spatially and with wavelength, the resulting image suffers from both structural and spectral distortion. Furthermore, the dependence of Φ on μ_{a} means that h (and hence p_0) is nonlinearly related to the absorption coefficient, making it much harder to recover without an

estimate of the fluence. The spectral distortion means that the light is coloured as it passes through the medium, and the spectrum of the absorbed energy at any point may be dependent on a combination of the spectra of the absorbing components at other points in the illuminated region. Moreover, the diffuse nature of light propagation in a turbid medium such as biological tissue means that information is quickly lost as it travels further away from the source.

The dependence on both optical absorption and scattering, and the fact that neither are likely to be known or easily measured, means that it will be necessary to recover quantitative estimates of both coefficients simultaneously. However, recovering these two parameters from a single PAT image may not have a unique solution, since two different sets of absorption and scattering coefficients may give rise to the same absorbed energy image. It is therefore necessary to include a sufficient amount of additional information to avoid this nonuniqueness. This has been highlighted recently in [2, 3, 7, 33], where it has been demonstrated that a set of PAT images obtained using multiple illumination positions can ensure a unique solution. The spectral dependence of a PAT image can also be exploited; in this case, a set of PAT images obtained using multiple wavelengths are used to avoid the nonuniqueness [2, 9, 22]. In this work we have used a multiple-illumination approach, which does not require the prior knowledge of the wavelength-dependence of the scattering coefficient needed when using a multiple-wavelength approach.

A further, more practical challenge when dealing with this inversion is the large-scale nature of the problem. PAT images are three dimensional and high resolution, and hence contain a significantly large amount of data. Any inversion technique must therefore be able to perform successfully with images of this size in order to provide a practical method of recovering chromophore concentrations from real PAT images. A discussion of the effect of the size and resolution of PAT images on minimization-based techniques can be found in section 3.2.

Though memory-efficient methods have been successfully used for QPAT [2, 11, 13, 14], existing methods typically use the DA to the RTE to model the light distribution. The DA assumes that the probability of an absorption event is much smaller than that of a scattering event ($\mu_a \ll \mu_s'$), and that the light is near-isotropic throughout the whole domain. It is therefore only valid in regions more than a few transport mean free paths away from a light source, where the light has undergone multiple scattering events and has therefore lost its directionality and become diffuse [1]. In PAT, however, regions close to the light source and domain boundaries can constitute a significant portion of the image and often contain information of great interest. The DA therefore breaks down in regions of interest to PAT, and so does not provide a suitably accurate model.

3. Minimization-based approaches in QPAT

Early attempts at developing optical inversion techniques have simplified the problem by assuming μ_s is known, in which case a number of methods can be used to recover μ_a [4, 10, 12, 31, 32]. However, since this assumption is unlikely to be true in most cases, it is desirable to develop a more general method. In this case, error minimization approaches have been proposed [2, 7, 8, 13, 14, 22, 26], whereby the sum of squared differences between the measured absorbed optical energy h^m and the modelled absorbed optical energy $h(\mu_a, \mu_s)$ is minimized with respect to the optical coefficients. This nonlinear least-squares formulation therefore seeks to find (μ_a, μ_s) which minimizes the error functional

$$\mathcal{E} = \frac{1}{2}\int_\Omega (h^m(\mathrm{r}) - h(\mathrm{r}; \mu_a, \mu_s))^2 \,\mathrm{dr}. \tag{10}$$

This type of error minimization scheme is effective at dealing with the nonlinearity of the problem, as it aims to iteratively adjust (μ_a, μ_s) until the difference between the measured and modelled data is satisfactorily reduced. The update to (μ_a, μ_s) can be calculated using the inverse of the Hessian matrix of second partial derivatives, however, since calculating and inverting this large, dense matrix is a costly procedure, alternative algorithms have been developed which make approximations to the inverse Hessian. Whichever method is chosen, the success of the inversion will be largely determined by the accuracy of the numerical model. In the following sections we discuss which class of minimization algorithm and which models of light propagation are most suitable for QPAT.

3.1. Jacobian-based methods

Minimization-based approaches for QPAT were proposed in [7] and [26], where it was demonstrated that both optical coefficients may be simultaneously determined from a set of PAT images by using a Gauss–Newton minimization scheme and a finite element (FE) light model based on the full RTE [25]. The main drawback of the Gauss–Newton method is the computational expense required for its execution; the technique looks to solve a regularized version of (10) by constructing an approximation to the Hessian matrix using the absorption and scattering Jacobian matrices, which is then used in a Gauss–Newton step to update (μ_a, μ_s). Given that there are N_e elements in the FE mesh and the optical coefficients are defined as piecewise constant over the mesh elements, the Jacobian matrix could have dimensions as large as $N_e \times N_e$. For two-dimensional examples the storage of these large, dense matrices may be computationally feasible, but real PAT images are inherently three-dimensional, and as the data sets become larger this Jacobian-based technique becomes much less practical. Furthermore, calculating the functional gradients using an adjoint-assisted Jacobian-based technique will require significantly more evaluations of the forward and adjoint models than the adjoint-assisted gradient-based method we propose here. (If the number of unknowns equals the number of data points equals the number of elements in the mesh N_e, then a Gauss–Newton iteration will require N_e runs of each of the forward and adjoint models, compared to one run each for a gradient-based method.) Matrix-free methods, which evaluate only matrix-vector products when calculating the Hessian approximation [24], could be used to avoid the memory problem, however, the operation count will remain high. Section 3.2 illustrates the computational expense when using Jacobian-based methods for QPAT.

3.2. Gradient-based methods

A gradient-based minimization approach was presented in [8], which demonstrated that a quasi-Newton minimization scheme can be used to recover one optical coefficient given that the other is known *a priori*. This technique has recently been extended to recover both absorption and scattering coefficients simultaneously [2, 13, 14, 22]. Quasi-Newton methods use the gradient of the error functional with respect to the optical coefficients to obtain the update to (μ_a, μ_s), avoiding the calculation and storage of the Jacobian and Hessian matrices entirely. Their use therefore provides a computationally efficient alternative which may be more practical for three-dimensional reconstructions.

To illustrate the advantage of using a gradient-based method, consider recent PAT images with typical dimensions 8 mm $\times$ 8 mm $\times$ 3.5 mm and spatial resolution 70 μm [20]. The spatial resolution of the image will dictate the required discretization of the computational mesh and hence the number of elements required. For this image size, varying the spatial resolution between 1 mm and 50 μm results in a range of 484 to $>$2.8 million elements. Given

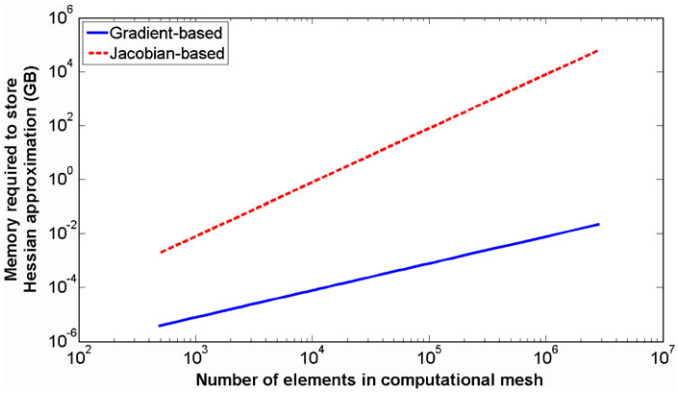

Figure 1. Memory required to store the Hessian approximation using gradient- and Jacobian-based minimization methods. The number of elements in the computational domain were based on a typical domain size of 8 mm × 8 mm × 3.5 mm and varying spatial resolution of 1 mm–50 μm.

that we require single precision data and that we are recovering both absorption and scattering coefficients, the total memory required to store the Hessian approximation for the gradient- and Jacobian-based methods is shown in figure 1. It is clear that the memory requirement for a Jacobian-based method becomes significantly large as the resolution improves. For example, if using a computational mesh with the same size and resolution as the PAT images obtained in [20], a gradient-based method would require 8.34 MB of memory to store the Hessian approximation, while a Jacobian-based method would require 8.7 TB.

4. An adjoint-assisted gradient-based method using the RTE

In this section we propose an inversion technique for QPAT, which utilizes an RTE-based model of light transport and an efficient quasi-Newton minimization scheme. The quasi-Newton minimization is based on the limited-memory BFGS method, which stores the gradient information from a user-defined number of iterations to build an iteratively improved approximation to the Hessian matrix in order to form an update to (μ_a, μ_s). For more information on quasi-Newton methods, including the limited-memory BFGS, see e.g. [23]. The light distribution is calculated using a numerical model of light transport based on the time-independent RTE [25], ensuring accuracy in all regions of interest for PAT, and the adjoint RTE operator is used to assist the gradient calculation. Recall that the idea of the minimization-based approach for QPAT is to minimize the error functional (10). The functional gradient calculation in the case of a single PAT image is considered first, and an extension to multiple images using different illumination positions will be included later in section 4.4.

4.1. Definition of operators

The natural space for the radiance is $L^1(\Omega \times S^{n-1})$, for the fluence is $L^1(\Omega)$ and for the optical coefficients is $L^\infty(\Omega)$. The RTE operator may therefore be written as a mapping $\mathcal{L}(\mu_a, \mu_s) : L^1(\Omega \times S^{n-1}) \to L^1(\Omega \times S^{n-1})$ such that

$$\mathcal{L}(\mu_a, \mu_s)\phi(\mathrm{r}, \hat{\mathbf{s}}) = (\hat{\mathbf{s}} \cdot \nabla + \mu_a + \mu_s)\phi(\mathrm{r}, \hat{\mathbf{s}}) - \mu_s \mathcal{S}\phi(\mathrm{r}, \hat{\mathbf{s}}), \tag{11}$$

where $\mathcal{S}$ is the scattering operator, defined by $\mathcal{S} : L^1(\Omega \times S^{n-1}) \to L^1(\Omega \times S^{n-1})$, with

$$\mathcal{S}\phi(\mathrm{r}, \hat{\mathbf{s}}) := \int_{S^{n-1}} \Theta(\hat{\mathbf{s}}, \hat{\mathbf{s}}')\phi(\mathrm{r}, \hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}'. \tag{12}$$

To assist in the functional gradient calculations we will make use of an adjoint equation for the RTE. The adjoint RTE operator is defined as the mapping $\mathcal{L}^*(\mu_\mathrm{a}, \mu_\mathrm{s}) : L^1(\Omega \times S^{n-1}) \to L^1(\Omega \times S^{n-1})$, with

$$\mathcal{L}^*(\mu_\mathrm{a}, \mu_\mathrm{s})\phi(\mathrm{r}, \hat{\mathbf{s}}) = (-\hat{\mathbf{s}} \cdot \nabla + \mu_\mathrm{a} + \mu_\mathrm{s})\phi(\mathrm{r}, \hat{\mathbf{s}}) - \mu_\mathrm{s}\mathcal{S}\phi(\mathrm{r}, \hat{\mathbf{s}}). \tag{13}$$

In PAT, the absorbed energy (4) can be written as a mapping $h : L^\infty(\Omega) \times L^\infty(\Omega) \to L^1(\Omega)$ with

$$h(\mu_\mathrm{a}, \mu_\mathrm{s}) = \mu_\mathrm{a} \int_{S^{n-1}} \phi(\hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}} = \mu_\mathrm{a}\Phi. \tag{14}$$

4.2. Error functional gradients

In order to develop a computational scheme we need to use inner product notation, and therefore modify our choice of spaces such that $\phi \in L^2(\Omega \times S^{n-1})$, $\Phi \in L^2(\Omega)$ and $\mu_\mathrm{a}, \mu_\mathrm{s} \in L^2(\Omega)$. From this, the error functional (10) can be written as a mapping $\mathcal{E} : L^2(\Omega) \times L^2(\Omega) \to \mathbb{R}$ such that

$$\begin{aligned} \mathcal{E}(\mu_\mathrm{a}, \mu_\mathrm{s}) &= \frac{1}{2} \int_\Omega (h^\mathrm{m} - h)^2\, \mathrm{dr} \\ &= \frac{1}{2} \langle h^\mathrm{m} - h, h^\mathrm{m} - h \rangle_{L^2(\Omega)}. \end{aligned} \tag{15}$$

Using the definition of the Fréchet derivative $D\mathcal{E}$ of $\mathcal{E}$, we can write

$$D\mathcal{E}\mu^\delta = -\langle h^\mathrm{m} - h, Dh\mu^\delta \rangle_{L^2(\Omega)}, \tag{16}$$

where $\mu^\delta = \begin{bmatrix} \mu_\mathrm{a}^\delta & \mu_\mathrm{s}^\delta \end{bmatrix}^\mathrm{T}$ represents a small change in the absorption and scattering. Since the product rule applies to Fréchet derivatives, and since the Fréchet derivative of μ_a is simply the identity function, the Fréchet derivative of h is given by

$$Dh = \Phi + \mu_\mathrm{a}(D\Phi). \tag{17}$$

Substituting this into equation (16) gives

$$\begin{aligned} D\mathcal{E}\mu^\delta &= -\left\langle h^\mathrm{m} - h, \mu_\mathrm{a}^\delta \Phi + \mu_\mathrm{a}\Phi^\delta \right\rangle_{L^2(\Omega)} \\ &= -\left\langle \Phi(h^\mathrm{m} - h), \mu_\mathrm{a}^\delta \right\rangle_{L^2(\Omega)} - \langle \mu_\mathrm{a}(h^\mathrm{m} - h), \Phi^\delta \rangle_{L^2(\Omega)}, \end{aligned} \tag{18}$$

where Φ^δ represents the change in the fluence due to a small change in the parameters, and is defined by

$$\Phi^\delta = \Phi\left(\mu_\mathrm{a} + \mu_\mathrm{a}^\delta, \mu_\mathrm{s} + \mu_\mathrm{s}^\delta\right) - \Phi(\mu_\mathrm{a}, \mu_\mathrm{s}). \tag{19}$$

We now define the adjoint field ϕ^* as the solution to

$$\mathcal{L}^*\phi^* = \mu_\mathrm{a}(h^\mathrm{m} - h), \tag{20}$$

which can be substituted into equation (18) to give

$$D\mathcal{E}\mu^\delta = -\left\langle \Phi(h^\mathrm{m} - h), \mu_\mathrm{a}^\delta \right\rangle_{L^2(\Omega)} - \langle \mathcal{L}^*\phi^*, \Phi^\delta \rangle_{L^2(\Omega)}. \tag{21}$$

Since neither $\mathcal{L}^*\phi^*$ nor Φ^δ depends on the direction $\hat{\mathbf{s}}$, equation (21) can be written equivalently as

$$D\mathcal{E}\mu^\delta = -\left\langle \Phi(h^\mathrm{m} - h), \mu_\mathrm{a}^\delta \right\rangle_{L^2(\Omega)} - \langle \mathcal{L}^*\phi^*, \phi^\delta \rangle_{L^2(\Omega \times S^{n-1})}. \tag{22}$$

To obtain an expression for the second term on the right-hand side, we will require a form of the divergence theorem, which states that

$$\begin{aligned}\int_{\partial\Omega}(\hat{\mathbf{s}}\cdot\hat{\mathbf{n}})\phi^*(\mathrm{r},\hat{\mathbf{s}})\phi^\delta(\mathrm{r},\hat{\mathbf{s}})\,\mathrm{dr} &= \int_\Omega(\hat{\mathbf{s}}\cdot\nabla)\phi^*(\mathrm{r},\hat{\mathbf{s}})\phi^\delta(\mathrm{r},\hat{\mathbf{s}})\,\mathrm{dr}\\ &= \int_\Omega \phi^*(\mathrm{r},\hat{\mathbf{s}})(\hat{\mathbf{s}}\cdot\nabla)\phi^\delta(\mathrm{r},\hat{\mathbf{s}}) + \phi^\delta(\mathrm{r},\hat{\mathbf{s}})(\hat{\mathbf{s}}\cdot\nabla)\phi^*(\mathrm{r},\hat{\mathbf{s}})\,\mathrm{dr}\\ &= \langle\phi^*,\mathcal{L}\phi^\delta\rangle_{L^2(\Omega\times S^{n-1})} - \langle\mathcal{L}^*\phi^*,\phi^\delta\rangle_{L^2(\Omega\times S^{n-1})}. \end{aligned}\tag{23}$$

Assuming that $\phi^\delta = 0$ on $\partial\Omega$, the left-hand side equates to zero, leaving

$$\langle\mathcal{L}^*\phi^*,\phi^\delta\rangle_{L^2(\Omega\times S^{n-1})} = \langle\phi^*,\mathcal{L}\phi^\delta\rangle_{L^2(\Omega\times S^{n-1})}.\tag{24}$$

Now consider the solution to the RTE (1) for a given source q:

$$\mathcal{L}(\mu_\mathrm{a},\mu_\mathrm{s})\phi = q,\tag{25}$$

and the change in the solution under a small change in the parameters:

$$\mathcal{L}\big(\mu_\mathrm{a}+\mu_\mathrm{a}^\delta,\mu_\mathrm{s}+\mu_\mathrm{s}^\delta\big)(\phi+\phi^\delta) = q.\tag{26}$$

By subtraction, we are left with

$$\mathcal{L}(\mu_\mathrm{a},\mu_\mathrm{s})\phi^\delta = -\mathcal{L}^\delta\big(\mu_\mathrm{a}^\delta,\mu_\mathrm{s}^\delta\big)\phi + \mathcal{O}\left(\left|\big(\mu_\mathrm{a}^\delta+\mu_\mathrm{s}^\delta\big)\phi^\delta\right|\right),\tag{27}$$

where the perturbation operator is defined as

$$\mathcal{L}^\delta\big(\mu_\mathrm{a}^\delta,\mu_\mathrm{s}^\delta\big)\phi(\mathrm{r},\hat{\mathbf{s}}) := \mu_\mathrm{a}^\delta\phi(\mathrm{r},\hat{\mathbf{s}}) + \mu_\mathrm{s}^\delta\phi(\mathrm{r},\hat{\mathbf{s}}) - \mu_\mathrm{s}^\delta\mathcal{S}\phi(\mathrm{r},\hat{\mathbf{s}}).\tag{28}$$

This allows equation (24) to be written as

$$\langle\mathcal{L}^*\phi^*,\phi^\delta\rangle_{L^2(\Omega\times S^{n-1})} = -\langle\phi^*,\mathcal{L}^\delta\phi\rangle_{L^2(\Omega\times S^{n-1})},\tag{29}$$

which can be substituted into equation (22) to give

$$D\mathcal{E}\mu^\delta = -\big\langle\Phi(h^\mathrm{m}-h),\mu_\mathrm{a}^\delta\big\rangle_{L^2(\Omega)} + \langle\phi^*,\mathcal{L}^\delta\phi\rangle_{L^2(\Omega\times S^{n-1})}.\tag{30}$$

Substituting the definition of the pertubation operator (28) and rearranging gives

$$\begin{aligned}D\mathcal{E}\mu^\delta = -\big\langle\Phi(h^\mathrm{m}-h),\mu_\mathrm{a}^\delta\big\rangle_{L^2(\Omega)} &+ \big\langle\phi\phi^*,\mu_\mathrm{a}^\delta\big\rangle_{L^2(\Omega\times S^{n-1})} + \big\langle\phi\phi^*,\mu_\mathrm{s}^\delta\big\rangle_{L^2(\Omega\times S^{n-1})}\\ &- \big\langle\phi\mathcal{S}\phi^*,\mu_\mathrm{s}^\delta\big\rangle_{L^2(\Omega\times S^{n-1})}.\end{aligned}\tag{31}$$

The gradient of $\mathcal{E}$ with respect to the absorption and scattering coefficients, which we denote by $\mathcal{E}'_{\mu_\mathrm{a}}$ and $\mathcal{E}'_{\mu_\mathrm{s}}$, respectively, are related to the Fréchet derivative by

$$D\mathcal{E} = \langle\mathcal{E}'_{\mu_\mathrm{a},\mu_\mathrm{s}},\mu^\delta\rangle_{L^2(\Omega)\times L^2(\Omega)},\tag{32}$$

which leads to the expressions for the error functional gradients

$$\mathcal{E}'_{\mu_\mathrm{a}} = -\Phi(h^\mathrm{m}-h) + \int_{S^{n-1}}\phi(\hat{\mathbf{s}})\phi^*(\hat{\mathbf{s}})\,\mathrm{d}\hat{\mathbf{s}}\tag{33}$$

$$\mathcal{E}'_{\mu_\mathrm{s}} = \int_{S^{n-1}}\phi(\hat{\mathbf{s}})\phi^*(\hat{\mathbf{s}})\,\mathrm{d}\hat{\mathbf{s}} - \int_{S^{n-1}}\int_{S^{n-1}}\phi(\hat{\mathbf{s}})\Theta(\hat{\mathbf{s}},\hat{\mathbf{s}}')\phi^*(\hat{\mathbf{s}})\,\mathrm{d}\hat{\mathbf{s}}\,\mathrm{d}\hat{\mathbf{s}}'.\tag{34}$$

4.3. Implementation

When implementing the minimization, the formulation of the numerical light model must be accounted for, and the continuous gradient calculations should be transferred to their discrete counterparts. In this work, the absorption and scattering coefficients are considered to be piecewise constant on the elements of the computational mesh, and can hence be written as

$$\mu_{\mathrm{a}} \approx \mu_{\mathrm{a}}^{e} = \sum_{k=1}^{N_e} \mu_{\mathrm{a}}^{k} \chi_k(\mathrm{r}), \quad \mu_{\mathrm{s}} \approx \mu_{\mathrm{s}}^{e} = \sum_{k=1}^{N_e} \mu_{\mathrm{s}}^{k} \chi_k(\mathrm{r}), \tag{35}$$

where N_e is the number of elements in the computational mesh and χ_k is the characteristic function of element k, which is equal to 1 in the support of element k and 0 elsewhere. In the FE formulation of the time-integrated RTE, its solution ϕ is approximated by

$$\phi(\mathrm{r}, \hat{\mathbf{s}}) \approx \phi^h(\mathrm{r}, \hat{\mathbf{s}}) = \sum_{j=1}^{N_n} \sum_{l=1}^{N_a} \phi_{jl} \psi_j(\mathrm{r}) \psi_l(\hat{\mathbf{s}}), \tag{36}$$

where N_n is the number of spatial nodes in the computational mesh and N_a is the number of angular directions. ϕ_{jl} represents the radiance and $\psi_j(\mathrm{r})$ and $\psi_l(\hat{\mathbf{s}})$ the nodal basis functions of the spatial and angular FE meshes, respectively, in the nodal point j and direction l. The FE RTE model calculates the radiance ϕ by solving

$$A\phi^h = b, \tag{37}$$

where A is the FE system matrix representing the terms in the RTE and b is the source term. The matrix A can be written as

$$A = A_1 + A_2 + A_3 + A_4, \tag{38}$$

where $A_{1,\ldots,4}$ have been obtained by determining a variational formulation of the RTE and making a finite-dimensional approximation to this variational formulation, constructed using piecewise linear bases for both the spatial and angular discretizations. The components of A are given by

$$A_1 = -\int_{\Omega} \int_{S^{n-1}} \hat{\mathbf{s}} \cdot \nabla \psi_{j'}(\mathrm{r}) \psi_{l'}(\hat{\mathbf{s}}) \psi_l(\hat{\mathbf{s}}) \, \mathrm{d}\hat{\mathbf{s}} \psi_j(\mathrm{r}) \, \mathrm{dr}, \tag{39}$$

$$A_2 = \int_{\partial\Omega} \psi_j(\mathrm{r}) \psi_{j'}(\mathrm{r}) \, \mathrm{d}S \int_{S^{n-1}} (\hat{\mathbf{s}} \cdot \hat{\mathbf{n}})_+ \psi_l(\hat{\mathbf{s}}) \psi_{l'}(\hat{\mathbf{s}}) \, \mathrm{d}\hat{\mathbf{s}}, \tag{40}$$

$$A_3 = \int_{\Omega} \left(\mu_{\mathrm{s}}^{e} + \mu_{\mathrm{a}}^{e}\right) \psi_j(\mathrm{r}) \psi_{j'}(\mathrm{r}) \, \mathrm{dr} \int_{S^{n-1}} \psi_l(\hat{\mathbf{s}}) \psi_{l'}(\hat{\mathbf{s}}) \, \mathrm{d}\hat{\mathbf{s}}, \tag{41}$$

$$A_4 = -\int_{\Omega} \mu_{\mathrm{s}}^{e} \psi_j(\mathrm{r}) \psi_{j'}(\mathrm{r}) \, \mathrm{dr} \int_{S^{n-1}} \int_{S^{n-1}} \Theta(\hat{\mathbf{s}}, \hat{\mathbf{s}}') \psi_l(\hat{\mathbf{s}}') \, \mathrm{d}\hat{\mathbf{s}}' \psi_{l'}(\hat{\mathbf{s}}) \, \mathrm{d}\hat{\mathbf{s}}. \tag{42}$$

The fluence Φ^e is equal to the integral of the radiance over all directions, and can therefore be calculated from

$$\Phi^e = M\phi^h, \tag{43}$$

where M is a measurement matrix which performs the summation over the angular directions. The measurement matrix also has the effect of expressing the fluence as piecewise constant over the mesh elements; its value over a single element is the average of its value at each node attached to that element. This average is taken because the absorbed energy is the product of the absorption coefficient and the fluence, with

$$h^e = \mu_{\mathrm{a}}^{e} \Phi^e, \tag{44}$$

and since the absorption coefficient is piecewise constant over the mesh elements it is not defined at element boundaries and hence not at the mesh nodes. Another way to overcome this problem would be to express the absorption coefficient using nodal basis functions. For notational convenience we will now write $\mu_a = \mu_a^e$, $\mu_s = \mu_s^e$, $\phi = \phi^h$, $\Phi = \Phi^e$ and $h = h^e$. Using this implementation, the gradients of the least-squares error functional (10) with respect to μ_a and μ_s, corresponding to equations (33) and (34), are given by

$$\mathcal{E}'_{\mu_a} = -\Phi^T(h^m - h) + \phi^T\left(A'_{\mu_a}\right)^T\phi^*, \tag{45}$$

$$\mathcal{E}'_{\mu_s} = \phi^T\left(A'_{\mu_s}\right)^T\phi^*, \tag{46}$$

where ϕ^* can be found by solving the adjoint model

$$A^T\phi^* = (\mu_a M)^T(h^m - h). \tag{47}$$

The terms A'_{μ_a} and A'_{μ_s} are the gradients of the FE system matrix (38) with respect to the absorption and scattering coefficients, respectively. In the absorption case for each element $k = 1, \ldots, N_e$ this is the single term

$$A'_{\mu_a^k} = \int_\Omega \chi_k(\mathrm{r})\psi_j(\mathrm{r})\psi_{j'}(\mathrm{r})\,\mathrm{dr} \int_{S^{n-1}} \psi_l(\hat{\mathbf{s}})\psi_{l'}(\hat{\mathbf{s}})\,\mathrm{d}\hat{\mathbf{s}}, \tag{48}$$

while in the scattering case there are the two terms

$$A'_{\mu_s^k} = \int_\Omega \chi_k(\mathrm{r})\psi_j(\mathrm{r})\psi_{j'}(\mathrm{r})\,\mathrm{dr} \int_{S^{n-1}} \psi_l(\hat{\mathbf{s}})\psi_{l'}(\hat{\mathbf{s}})\,\mathrm{d}\hat{\mathbf{s}} \tag{49}$$

$$- \int_\Omega \chi_k(\mathrm{r})\psi_j(x)\psi_{j'}(x)\,\mathrm{dr} \int_{S^{n-1}}\int_{S^{n-1}} \Theta(\hat{\mathbf{s}}, \hat{\mathbf{s}}')\psi_l(\hat{\mathbf{s}}')\,\mathrm{d}\hat{\mathbf{s}}'\psi_{l'}(\hat{\mathbf{s}})\,\mathrm{d}\hat{\mathbf{s}}. \tag{50}$$

4.4. Extension to multiple illumination positions

To avoid the nonuniqueness encountered when recovering both absorption and scattering from a single PAT image, we must include additional information into the problem [3]. In this work we have chosen to use a multiple-illumination approach, whereby a set of images are obtained using sources placed at different positions around the image domain. This technique therefore requires that we sum over the number of sources used. If N_p is the number of source positions, then the error functional becomes

$$\mathcal{E} = \frac{1}{2}\sum_{p=1}^{N_p}\int_\Omega \left(h_p^m - h_p\right)^2 \mathrm{dr} \tag{51}$$

and the functional gradients are given by

$$\mathcal{E}'_{\mu_a} = -\sum_{p=1}^{N_p}\Phi_p^T\left(h_p^m - h_p\right) + \sum_{p=1}^{N_p}\phi_p^T\left(A'_{\mu_a}\right)^T\phi_p^*, \tag{52}$$

$$\mathcal{E}'_{\mu_s} = \sum_{p=1}^{N_p}\phi_p^T\left(A'_{\mu_s}\right)^T\phi_p^*, \tag{53}$$

where the subscript $(\cdot)_p$ denotes the quantity at the pth source position.

4.5. Regularization

To dampen the effects of noise in the inversions, a first order Tikhonov regularization penalty term is added to the error functional whenever noise is present in the simulated data. With this inclusion, the error functional becomes

$$\mathcal{E} = \sum_{p=1}^{N_p} \int_\Omega \frac{1}{2}\left(h_p^{\mathrm{m}} - h_p\right)^2 \mathrm{d}\mathrm{r} + \frac{\alpha}{2} L\mu_{\mathrm{a}} + \frac{\beta}{2} L\mu_{\mathrm{s}}, \tag{54}$$

where α and β are two constant scaling factors and L is a discrete implementation of the Laplacian matrix. The corresponding functional gradient expressions are then given by

$$\mathcal{E}'_{\mu_{\mathrm{a}}} = -\sum_{p=1}^{N_p} \Phi_p^{\mathrm{T}}\left(h_p^{\mathrm{m}} - h_p\right) + \sum_{p=1}^{N_p} \phi_p^{\mathrm{T}}\left(A'_{\mu_{\mathrm{a}}}\right)^{\mathrm{T}} \phi_p^* - \alpha \mu_{\mathrm{a}}^{\mathrm{T}} L\mu_{\mathrm{a}}, \tag{55}$$

$$\mathcal{E}'_{\mu_{\mathrm{s}}} = \sum_{p=1}^{N_p} \phi_p^{\mathrm{T}}\left(A'_{\mu_{\mathrm{s}}}\right)^{\mathrm{T}} \phi_p^* - \beta \mu_{\mathrm{a}}^{\mathrm{T}} L\mu_{\mathrm{a}}. \tag{56}$$

5. Results and discussion

5.1. Simulation of PAT images

The gradient-based method described in section 4 was tested using two-dimensional simulated PAT images. To simulate a set of PAT images using multiple illumination positions, the optical coefficients $(\mu_{\mathrm{a}}, \mu_{\mathrm{s}})$ were defined on the square $\Omega = [0, 8] \times [0, 8]$ mm^2, represented by a discretized mesh containing 3720 triangular elements. Using a source term $q = (q_1, q_2, q_3, q_4)$, representing four different planar illuminations along each edge of the domain, a numerical model based on the time-integrated RTE (1) with boundary condition (2) was solved for the radiance ϕ. The streamline diffusion modification was included to overcome the ray effect which may disturb the FE solution to the RTE (see the appendix for more details). Equations (3) and (4) were used to obtain four absorbed energy maps h, from which a corresponding set of initial pressure distributions p_0 were found using (5). This provides the initial condition (7) required to solve the photoacoustic wave equation (6) and obtain four sets of acoustic pressure time series. The data was interpolated onto a rectangular grid and equation (6) was solved using a collocation pseudo-spectral time-domain acoustic model [27], with acoustic detectors placed along each edge of the domain. Gaussian noise, at a level of 1% of the maximum value of the acoustic pressure, was added to these time series to replicate the type of measurement noise encountered in practice, e.g. thermal noise in the detector. The initial pressure distribution was then found using a time-reversal image reconstruction technique [27, 28]. The Grüneisen parameter was assumed to be known and constant throughout the tissue, with $\Gamma(\mathrm{r}) = 1\ \forall r \in \Omega$, so that the measured absorbed energy maps h^{m} were found trivially from equation (8). Due to the presence of noise in the acoustic data it is possible that the resulting images may contain negative values of absorbed optical energy; since these do not make physical sense, any negative values of h^{m} were set equal to zero. This set of simulated, noisy images then provide the measured data from which the optical coefficients are to be recovered.

5.2. Inversion example: smooth optical coefficients

Using the method described above, a set of PAT images were simulated using smoothly-varying optical coefficients such that $0 \leqslant \mu_{\mathrm{a}} \leqslant 0.3$ mm^{-1}, $1 \leqslant \mu_{\mathrm{s}} \leqslant 3$ mm^{-1} and $g = 0.6$.

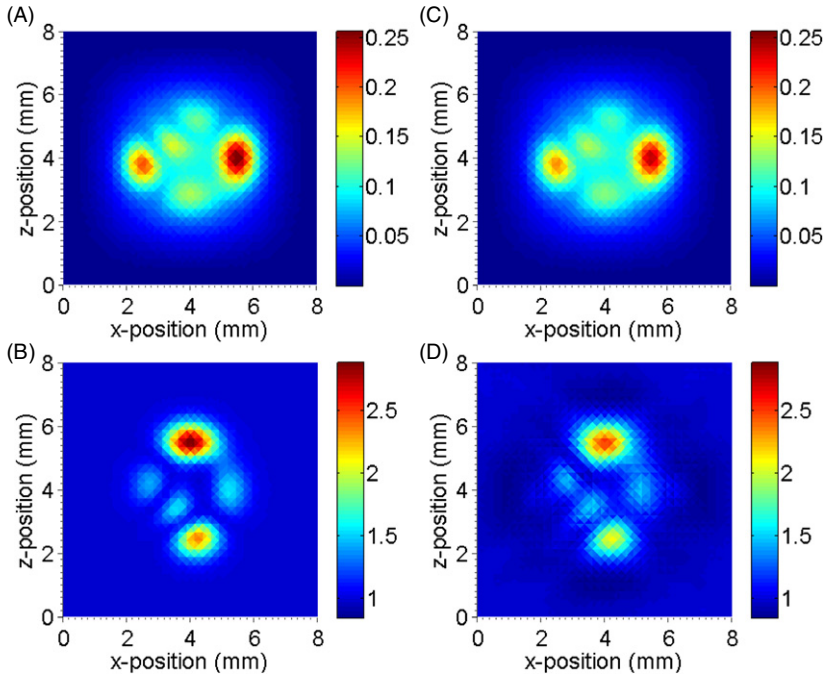

Figure 2. Results of the gradient-based inversion using smooth coefficients after 400 iterations. The figures show the true and recovered absorption coefficient μ_a ((A) and (C), respectively) and the true and recovered scattering coefficient μ_s ((B) and (D), respectively). The simulated PAT image included the acoustic inversion step and additive 1% Gaussian noise. A first order Tikhonov scheme was included to dampen the effect of noise in the inversion. The colour scale represents mm^{-1}.

In biological tissue the ansitropy factor g is typically closer to 0.9, however, using a higher value of g in the RTE model requires a finer angular discretization, and subsequently increases the computational burden. When $g = 0.6$, the RTE model is accurate if $N_a \geqslant 16$. The value of g directly affects the reduced scattering coefficient $\mu_s' = \mu_s(1 - g)$, which describes the scattering of photons in the diffusive regime. The values of μ_a and μ_s were therefore chosen to ensure that μ_a and μ_s' were typical of problems in biomedical applications. The results of the gradient-based inversion using noisy simulated PAT images can be seen in figure 2. These show the true and recovered optical coefficients after 400 iterations of the l-BFGS method. The percent relative errors in the absorption and scattering coefficients, denoted ϵ_a and ϵ_s, respectively, were calculated using

$$\epsilon_{a,s} = \frac{\|\mu_{a,s} - \tilde{\mu}_{a,s}\|}{\|\mu_{a,s}\|} \times 100, \tag{57}$$

where $\mu_{a,s}$ represents the true value of the absorption or scattering coefficient and $\tilde{\mu}_{a,s}$ represents their estimated values. In these reconstructions, values of $\epsilon_a = 4.4\%$ and $\epsilon_s = 6.7\%$ were achieved. A cross-section in the x-direction of the inversion results can be seen in figure 3.

The directionally randomizing nature of light propagation in scattering media and the relative insensitivity of the objective function on μ_s mean that recovering the scattering coefficient is typically more difficult than recovering the absorption coefficient. A study of the sensitivity to changes in the optical parameters can be found in [26]. Since it is the absorption coefficient that will be required to obtain a map of the chromophore concentrations the strong

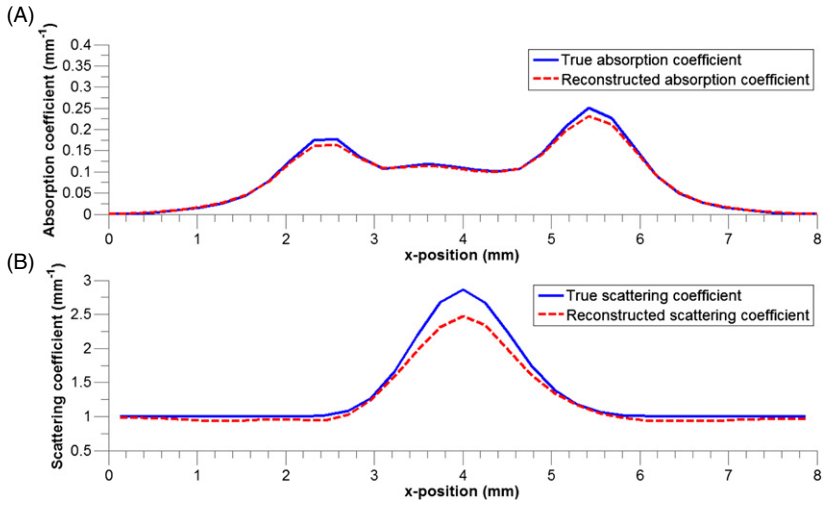

Figure 3. Results of the gradient-based inversion using smooth coefficients after 400 iterations. (A) and (B) are profiles, taken in the x-direction, corresponding to figure 2, and show the true (solid) and recovered (dashed) absorption coefficient μ_a and scattering coefficient μ_s, respectively.

dependence on μ_a is favourable, however, since its accuracy depends on the accuracy of the scattering estimate, recovering μ_s is still necessary.

Gradient-based methods typically take a larger number of iterations than Jacobian-based methods; this is certainly the case here, as it has been shown that Jacobian-based methods have successfully recovered both optical coefficients in as few as five iterations [26]. A brief discussion of the relative advantages of gradient- and Jacobian-based methods can be found later in section 5.4.

5.3. Inversion example: non-smooth coefficients

Since PAT is particularly well-suited to imaging the blood vasculature, it is common to see non-smooth structures such as blood vessels in an image. For this reason, the method was also tested using non-smooth optical coefficients. The PAT images were simulated in the same way as described above, though now the absorption and scattering coefficients consisted of a constant background value with absorbing and scattering rectangular structures, respectively, which were designed to partially overlap. The optical coefficients were defined by

$$\begin{aligned} &\mu_a(\mathrm{r}) = 0.3, \quad \mathrm{r} \in \Omega_1, && \Omega_1 = [1.5, 4.5] \times [3.5, 6.5]\,\mathrm{mm}^2 \\ &\mu_a(\mathrm{r}) = 4/3, \quad \mathrm{r} \in \Omega_2, && \Omega_2 = [5, 7] \times [1, 3]\,\mathrm{mm}^2 \\ &\mu_a(\mathrm{r}) = 0.1, \quad \mathrm{r} \in (\Omega \backslash (\Omega_1 \cup \Omega_2)) \end{aligned} \tag{58}$$

$$\begin{aligned} &\mu_s(\mathrm{r}) = 4/3, \quad \mathrm{r} \in \Omega_1, && \Omega_1 = [1, 3] \times [5, 7]\,\mathrm{mm}^2 \\ &\mu_s(\mathrm{r}) = 3, \quad \mathrm{r} \in \Omega_2, && \Omega_2 = [3.5, 6.5] \times [1.5, 4.5]\,\mathrm{mm}^2 \\ &\mu_s(\mathrm{r}) = 1, \quad \mathrm{r} \in (\Omega \backslash (\Omega_1 \cup \Omega_2)). \end{aligned} \tag{59}$$

The simulated PAT images were obtained using the same method as above, and also included Gaussian noise at a level of 1% of the maximum value of the acoustic pressure and a first-order Tikhonov regularization scheme. The results of the inversion after 400 iterations using these non-smooth optical coefficients can be seen in figure 4, and a profile in the x-direction is shown in figure 5. In these reconstructions, values of $\epsilon_a = 11.7\%$ and $\epsilon_s = 21.3\%$ were achieved. As

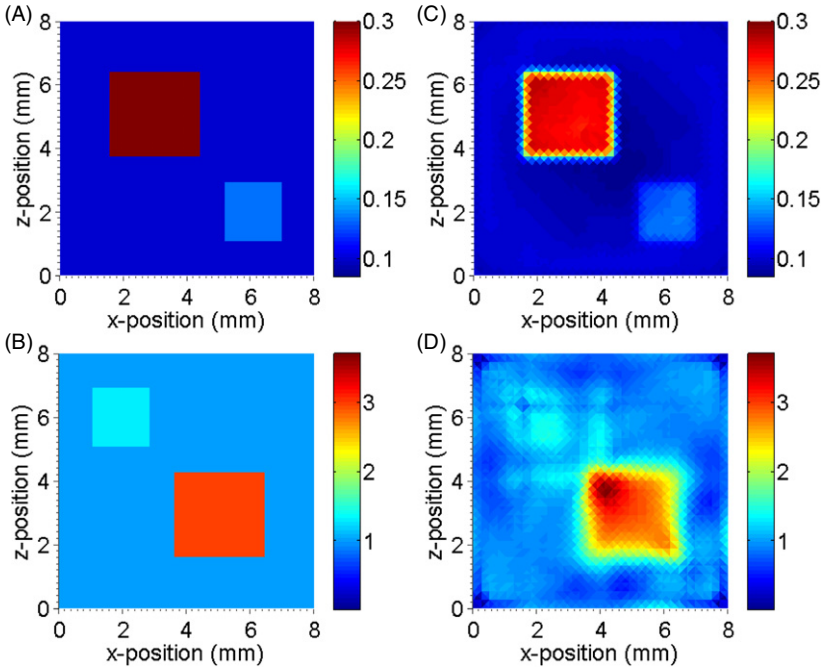

Figure 4. Results of the gradient-based inversion using non-smooth coefficients after 400 iterations. The figures show the true and recovered absorption coefficient μ_a ((A) and (C), respectively) and the true and recovered scattering coefficient μ_s ((B) and (D), respectively). The simulated PAT image included the acoustic inversion step and additive 1% Gaussian noise. A first-order Tikhonov scheme was included to dampen the effect of noise in the inversion. The colour scale represents mm^{-1}.

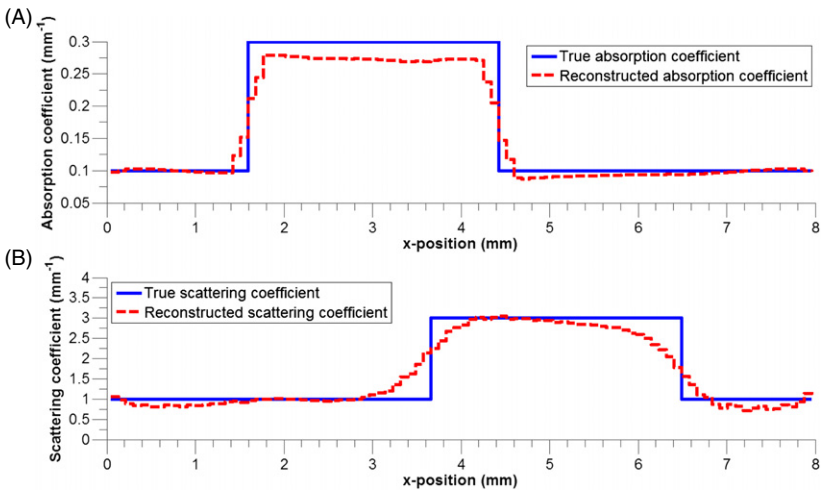

Figure 5. Results of the gradient-based inversion using non-smooth coefficients after 400 iterations. (A) and (B) are profiles, taken in the x-direction, corresponding to figure 4, and show the true (solid) and recovered (dashed) absorption coefficient μ_a (A) and scattering coefficient μ_s (B), respectively.

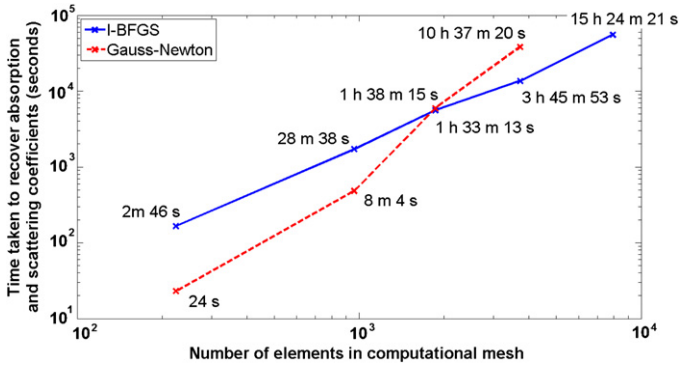

Figure 6. Relationship between the number of elements in the computational mesh and the total time taken to recover both optical coefficients when using the l-BFGS method (solid) and the Gauss–Newton method (dashed). Results are based on noise-free data which does not include the acoustic inversion step.

in the case of smooth optical coefficients, the absorption coefficient is recovered quicker and more accurately than the scattering coefficient. A Tikhonov regularization scheme was used for consistency with section 5.2, but in the case of non-smooth coefficients it is likely that the results would be improved by the inclusion of an edge-preserving regularization method, such as total variation [14].

5.4. Comparison with Jacobian-based methods

The presented gradient-based inversion method has been shown to successfully recover both absorption and scattering coefficients from a simulated PAT image. Although it is more efficient from a computational storage perspective, gradient-based methods typically require a larger number of iterations to perform the inversion than a Jacobian-based method. However, since Jacobian-based methods require N_e times more evaluations of the forward and adjoint model to calculate the functional gradients, one iteration of the adjoint-assisted gradient-based inversion is significantly quicker than one iteration of a Jacobian-based method. This means that the increased number of iterations required of a gradient-based method may not have a negative effect on the total computational time taken to recover (μ_a, μ_s). To investigate how the completion time of the inversion scales as the data set becomes larger, the presented gradient-based method was performed on increasingly fine computational meshes and compared with the performance of a corresponding Jacobian-based minimization scheme, namely the Gauss–Newton method. In this case $0.01 \leqslant \mu_a \leqslant 0.3$, $1 \leqslant \mu_s \leqslant 3$, $g = 0.6$, no additive Gaussian noise was added to the data and the acoustic inversion step was not included.

Minimizations using the gradient- and Jacobian-based methods were performed on computational meshes with 224, 960, 1860, 3720 and 7938 elements, and the time taken to recover both optical coefficients was recorded. To compare the two methods fairly, the time taken to recover (μ_a, μ_s) was taken to be the time required to achieve an estimate of the scattering coefficient to within 5% of its true value (i.e. $\epsilon_s < 5$). The time taken to perform the inversions for each computational mesh using both the l-BFGS and Gauss–Newton methods can be seen in figure 6. In all cases, the gradient-based l-BFGS method took a larger number of iterations than the Jacobian-based Gauss–Newton method to recover (μ_a, μ_s). In the case of the 224- and 960-element meshes the Gauss–Newton method took less time than the l-BFGS

method. The methods took roughly the same amount of time to recover both coefficients when using the 1860-element mesh, with the l-BFGS taking just over 1 h 33 min and the Gauss–Newton method taking just over 1 h 38 min. For meshes finer than this the total time to recover (μ_a, μ_s) when using the Gauss–Newton method was significantly longer than the l-BFGS method. Though neither the l-BFGS or Gauss–Newton codes used to achieve these results are optimized, so the absolute time carry no significance, the trend suggests that a gradient-based minimization scheme is more efficient not only in terms of memory required, but in realistically sized cases may also perform the full inversion more quickly than a corresponding Jacobian-based method.

6. Conclusions

An adjoint-assisted, gradient-based minimization method has been shown to succesfully determine the absorption and scattering coefficients from a set of two-dimensional, simulated PAT images. The model-based inversion utilizes the full radiative transfer equation and includes the acoustic propagation and reconstruction step in the simulation of the images, so that the full quantitative inversion has been tackled. The gradient-based minimization seeks to minimize a nonlinear least-squares error functional quantifying the difference between the set of measured PAT images and the modelled data, and was performed using the l-BFGS quasi-Newton method. The advantage of using a gradient-based method comes from the fact that the storage of large, dense matrices is avoided, so that from a storage perspective they are more computationally efficient that a corresponding Jacobian- or Hessian-based method. Gradient-based methods typically require a greater number of iterations to perform the inversion, however, a brief comparison of the performance of the gradient-based l-BFGS method and the Jacobian-based Gauss–Newton method has indicated that it scales better for large problems, despite the larger number of iterations required.

There are a number of reasons why one might wish to obtain quantitative estimates of chromophore concentrations, and if a minimization-based method is to be used it may be worth considering to what degree of accuracy these concentrations are required before performing the inversion. Since the absorption estimate is likely to improve only fractionally after some point and at significant computational expense, depending on the application it may not be worthwhile to continue with the minimization to further improve the reconstructions and determine the scattering coefficient.

The success of the inversion when including the acoustic inversion step in the PAT image simulation is encouraging, since these simulations mimic more closely the types of artifacts and noise present in a real PAT image. Extending the method to three-dimensional data in order to observe exactly how practical the presented gradient-based method will be the subject of future work.

Acknowledgments

This work was supported by the Department of Medical Physics and Bioengineering at University College London, the Engineering and Physical Sciences Research Council, UK, the Academy of Finland (projects 136220, 140984, and 250215 Finnish Centre of Excellence in Inverse Problems Research) and by the strategic funding of the University of Eastern Finland.

Appendix. Streamline diffusion modification

The streamline diffusion modification was included in the numerical light transport model to overcome the ray effect which may disturb the finite element solution to the RTE [16, 25]. Including this modification means that the terms of (38) become

$$A_1 = -\int_\Omega \int_{S^{n-1}} \hat{\mathbf{s}} \cdot \nabla \psi_{j'}(\mathrm{r}) \psi_{l'}(\hat{\mathbf{s}}) \psi_l(\hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}} \psi_j(\mathrm{r})\, \mathrm{dr} + \int_\Omega \delta \int_{S^{n-1}} (\hat{\mathbf{s}} \cdot \nabla \psi_{j'}(\mathrm{r}) \psi_{l'}(\hat{\mathbf{s}}))(\hat{\mathbf{s}} \cdot \nabla \psi_j(\mathrm{r}) \psi_l(\hat{\mathbf{s}}))\, \mathrm{d}\hat{\mathbf{s}}\, \mathrm{dr}, \tag{A.1}$$

$$A_2 = \int_{\partial\Omega} \psi_j(\mathrm{r}) \psi_{j'}(\mathrm{r})\, \mathrm{d}S \int_{S^{n-1}} (\hat{\mathbf{s}} \cdot \hat{\mathbf{n}})_+ \psi_l(\hat{\mathbf{s}}) \psi_{l'}(\hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}}, \tag{A.2}$$

$$A_3 = \int_\Omega \left(\mu_\mathrm{s}^e + \mu_\mathrm{a}^e\right) \psi_j(\mathrm{r}) \psi_{j'}(\mathrm{r})\, \mathrm{dr} \int_{S^{n-1}} \psi_l(\hat{\mathbf{s}}) \psi_{l'}(\hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}} + \int_\Omega \delta\left(\mu_\mathrm{a}^e + \mu_\mathrm{s}^e\right) \int_{S^{n-1}} \hat{\mathbf{s}} \cdot \nabla \psi_{j'}(\mathrm{r}) \psi_{l'}(\hat{\mathbf{s}}) \psi_l(\hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}} \psi_j(\mathrm{r})\, \mathrm{dr}, \tag{A.3}$$

$$A_4 = -\int_\Omega \mu_\mathrm{s}^e \psi_j(\mathrm{r}) \psi_{j'}(\mathrm{r})\, \mathrm{dr} \int_{S^{n-1}} \int_{S^{n-1}} \Theta(\hat{\mathbf{s}}, \hat{\mathbf{s}}') \psi_l(\hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}' \psi_{l'}(\hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}} - \int_\Omega \delta \mu_\mathrm{s}^e \int_{S^{n-1}} \hat{\mathbf{s}} \cdot \nabla \psi_{j'}(\mathrm{r}) \psi_{l'}(\hat{\mathbf{s}}) \int_{S^{n-1}} \Theta(\hat{\mathbf{s}}, \hat{\mathbf{s}}') \psi_l(\hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}'\, \mathrm{d}\hat{\mathbf{s}} \psi_j(\mathrm{r})\, \mathrm{dr}, \tag{A.4}$$

where δ is a smoothing parameter which depends on the local absorption and scattering.

References

[1] Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** 41–93

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

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

[4] Banerjee B, Bagchi S, Vasu R M and Roy D 2008 Quantitative photoacoustic tomography from boundary pressure measurements: noniterative recovery of optical absorption coefficient from the reconstructed absorbed energy map *J. Opt. Soc. Am.* A **25** 2347–56

[5] Binzoni T, Leung T S, Gandjbakhche A H, Rüfenacht D and Delpy D T 2006 The use of the Henyey–Greenstein phase function in Monte Carlo simulations in biomedical optics *Phys. Med. Biol.* **51** N313–22

[6] Cox B, Laufer J G, Arridge S R and Beard P C 2012 Quantitative spectroscopic photoacoustic imaging: a review *J. Biomed. Opt.* **17** 061202–1 061202

[7] Cox B, Tarvainen T and Arridge S 2011 *Multiple Illumination Quantitative Photoacoustic Tomography using Transport and Diffusion Models* (*American Mathematical Society Contemporary Mathematics Series* vol 559) (Providence, RI: American Mathematical Society)

[8] Cox B T, Arridge S R and Beard P C 2007 Gradient-based quantitative photoacoustic image reconstruction for molecular imaging *Proc. SPIE* **6437** 64371T

[9] Cox B T, Arridge S R and Beard P C 2009 Estimating chromophore distributions from multiwavelength photoacoustic images *J. Opt. Soc. Am.* A **26** 443–55

[10] 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

[11] 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

[12] Cox B T, Laufer J G, Köstli K P and Beard P C 2005 Quantitative photoacoustic imaging: fitting a model of light transport to the initial pressure distribution *Proc. SPIE* **5697** 49–55

[13] Gao H, Zhao H and Osher S 2010 Bregman methods in quantitative photoacoustic tomography *UCLA CAM Report* 10–42

[14] Gao H, Osher S and Zhao H 2012 Quantitative photoacoustic tomography *Mathematical Modeling in Biomedical Imaging II* (Berlin/Heidelberg: Springer) pp 131–58
[15] Henyey L G and Greenstein J L 1941 Diffuse radiation in the galaxy *Astrophys. J.* **93** 70–83
[16] Kanschat G 1998 A robust finite element discretization for radiative transfer problems with scattering *East West J. Numer. Math.* **6** 265–72
[17] Kruger R A, Lam R B, Reinecke D R, Del Rio S P and Doyle R P 2010 Photoacoustic angiography of the breast *Med. Phys.* **37** 6096
[18] Kuchment P and Kunyansky L 2011 Mathematics of photoacoustic and thermoacousic tomography *Handbook of Mathematical Methods in Imaging* (Berlin: Springer) pp 817–65 chapter 19
[19] Laufer J, Johnson P, Zhang E, Treeby B, Cox B, Pedley B and Beard P 2011 *In vivo* longitudinal photoacoustic imaging of subcutaneous tumours in mice *Proc. SPIE* **7899** 789915
[20] Laufer J, Johnson P, Zhang E, Treeby B, Cox B, Pedley B and Beard P 2012 *In vivo* preclinical photoacoustic imaging of tumor vasculature development and therapy *J. Biomed. Opt.* **17** 056016
[21] Laufer J, Zhang E, Raivich G and Beard P 2009 Three-dimensional noninvasive imaging of the vasculature in the mouse brain using a high resolution photoacoustic scanner *Appl. Opt.* **48** D299–306
[22] Mamonov A V and Ren K 2012 Quantitative photoacoustic imaging in radiative transport regime arXiv:1207.4664
[23] Nocedal J and Wright S 2006 *Numerical Optimization Springer Series in Operations Research* 2nd edn (New York, NY: Springer)
[24] Schweiger M, Arridge S R and Nissilä I 2005 Gauss–Newton method for image reconstruction in diffuse optical tomography *Phys. Med. Biol.* **50** 2365
[25] Tarvainen T 2006 Computational methods for light transport in optical tomography *PhD Thesis* University of Kuopio
[26] 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
[27] 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
[28] Treeby B E, Zhang E Z and Cox B T 2010 Photoacoustic tomography in absorbing acoustic media using time reversal *Inverse Problems* **26** 115003
[29] Wang X, Pang Y, Ku G, Xie X, Stoica G and Wang L V 2003 Noninvasive laser-induced photoacoustic tomography for structural and functional *in vivo* imaging of the brain *Nature Biotechnol.* **21** 803–6
[30] Yao J and Wang L V 2011 Photoacoustic tomography: fundamentals, advances and prospects *Contrast Media Mol. Imaging* **6** 332–45
[31] Yao L, Sun Y and Jiang H 2009 Quantitative photoacoustic tomography based on the radiative transfer equation *Opt. Lett.* **34** 1765–7
[32] Yuan Z and Jiang H 2006 Quantitative photoacoustic tomography: recovery of optical absorption coefficient maps of heterogeneous media *Appl. Phys. Lett.* **88** 231101
[33] Zemp R J 2010 Quantitative photoacoustic tomography with multiple optical sources *Appl. Opt.* **49** 3566–72
[34] Zhang E Z, Laufer J G, Pedley R B and Beard P C 2009 *In vivo* high-resolution 3D photoacoustic imaging of superficial vascular anatomy *Phys. Med. Biol.* **54** 1035–46