

# Gauss–Newton method for image reconstruction in diffuse optical tomography

Martin Schweiger<sup>1</sup>, Simon R Arridge<sup>1</sup> and Ilkka Nissilä<sup>2</sup>

<sup>1</sup> Department of Computer Science, University College London, Gower Street London, WC1E 6BT, UK

<sup>2</sup> Laboratory of Biomedical Engineering, Helsinki University of Technology, Finland

Received 21 January 2005, in final form 30 March 2005

Published 5 May 2005

Online at [stacks.iop.org/PMB/50/2365](http://stacks.iop.org/PMB/50/2365)

## Abstract

We present a regularized Gauss–Newton method for solving the inverse problem of parameter reconstruction from boundary data in frequency-domain diffuse optical tomography. To avoid the explicit formation and inversion of the Hessian which is often prohibitively expensive in terms of memory resources and runtime for large-scale problems, we propose to solve the normal equation at each Newton step by means of an iterative Krylov method, which accesses the Hessian only in the form of matrix–vector products. This allows us to represent the Hessian implicitly by the Jacobian and regularization term. Further we introduce transformation strategies for data and parameter space to improve the reconstruction performance. We present simultaneous reconstructions of absorption and scattering distributions using this method for a simulated test case and experimental phantom data.

## 1. Introduction

### 1.1. Optical tomography

Diffuse optical tomography is a new medical imaging modality with potential applications in functional imaging of the brain and in breast cancer detection, amongst other applications. This method seeks to recover optical parameters of blood and tissue from boundary measurements of light transmission in the visible and near-infrared range. The reconstructed images of the spatial distribution of tissue parameters can be related directly to physiologically important properties such as blood and tissue oxygenation state. Instrumentation for optical tomography is portable and relatively inexpensive, and can provide a viable alternative to currently available systems such as functional magnetic resonance imaging.

Data acquisition systems consist of a light source such as an infrared laser, illuminating the body surface at different source locations in succession. The light which has propagated through the tissue is then measured at multiple detector locations on the surface. Biological

tissue is strongly scattering at the wavelengths used in optical tomography, which generally makes the recovery of tissue parameters from the boundary data a highly nonlinear problem.

The experimental systems in use today utilize either ultrashort input pulses (time domain systems) or continuous intensity-modulated input (frequency-domain systems). In the former case, measurements consist of the temporal dispersion of the transmitted pulse, measured at a resolution in the order of pico-seconds. In the latter case, the measurements consist of the complex intensity of the transmitted photon density wave, most commonly measured in terms of the phase shift and modulation amplitude. The frequency domain version of the problem is the one considered in this paper.

### *1.2. Image reconstruction in optical tomography*

Image reconstruction methods in optical tomography differ in the type of data being considered, the type of solutions being sought, the physical model assumed for light propagation, and in the algorithmic details; for reviews see Arridge (1993, 1999), Yodh and Chance (1995), Hebden *et al* (1997), Arridge and Hebden (1997), Hawysz and Sevick-Muraca (2000), Boas *et al* (2001) and Gibson *et al* (2005). Two main distinctions are between linear methods based on inverse scattering theory, and non-linear methods based on model fitting. In the former category, a perturbation model is postulated which corresponds to the first term in a Born or Rytov expansion of the Lippman–Schwinger equation (Arridge *et al* 1991, O’Leary *et al* 1995, O’Leary 1996, Schotland *et al* 1993, Schotland 1997, Schotland and Markel 2001, Markel and Schotland 2001, 2002). This model can be constructed using Green’s functions for the diffusion equation, either in infinite space or in the presence of boundaries, or by a numerical technique such as Monte Carlo, finite difference or finite elements. The resultant linear system is usually solved by an iterative method such as the Kaczmarz method, conjugate gradients or projection onto convex sets (Barbour *et al* 1990, Wang *et al* 1992, Walker *et al* 1997, Ripoll *et al* 2001). The second approach considers the model in terms of explicit parameters and adjusts these parameters in order to optimize an objective function combining a data fitting and regularization term (Arridge *et al* 1992, Schweiger *et al* 1993, Paulsen and Jiang 1995, Klose and Hielscher 1999, Ye *et al* 2001). These methods, based either on the diffusion equation or on the radiative transfer equation, use a variety of algorithms, including Levenberg–Marquardt (Schweiger *et al* 1993), quasi-Newton (Klose and Hielscher 2003), truncated Newton (Roy and Sevick-Muraca 1999) and non-linear conjugate gradients (Arridge and Schweiger 1998) to minimize the objective function over the search space of model parameters.

### *1.3. Contribution of this paper*

In this paper we compare some implementation strategies for a Gauss–Newton approach to the inverse solver in optical tomography. In this approach the non-linear forward model is linearized to produce a Jacobian, and a system of normal equations is developed wherein the Hessian of the forward model is approximated by the Jacobian transposed with itself, plus a regularization term. This approach has to be combined with a *globalization* strategy to ensure convergence. When the dimension of the problem is large the representation of the Hessian can be infeasible in terms of available memory, whereas the representation of the Jacobian is possible. However, if the solution of the normal equations is carried out with an iterative procedure, the Hessian is only utilized in terms of matrix–vector products. Therefore we propose a method to implicitly represent the Hessian in a functional form that proves to be computationally efficient.

We use this representation to compare both a Levenberg–Marquardt, and damped Gauss–Newton globalization strategy and show that the latter has better convergence. In addition, performance of the reconstruction technique is strongly dependent on transformation of the data and conditioning of the Jacobian.

## 2. Formulation of the problem

Let  $\Omega \subset \mathbb{R}^n$  be a simply connected domain containing two real positive scalar functions  $\mu_a(\mathbf{r}), \kappa(\mathbf{r})$ . Let  $f(\mathbf{m}; \omega) e^{i\omega t}$  be a time-harmonic incoming radiation boundary condition. Then the *photon density*  $\Phi(\mathbf{r}; \omega)$  inside the domain satisfies

$$-\nabla \cdot \kappa(\mathbf{r}) \nabla \Phi(\mathbf{r}; \omega) + \left( \mu_a(\mathbf{r}) + \frac{i\omega}{c} \right) \Phi(\mathbf{r}; \omega) = 0 \quad \mathbf{r} \in \Omega \setminus \partial\Omega \quad (1)$$

$$\Phi(\mathbf{m}; \omega) + 2\zeta \kappa(\mathbf{m}) \frac{\partial \Phi(\mathbf{m}; \omega)}{\partial \nu} = f(\mathbf{m}; \omega) \quad \mathbf{m} \in \partial\Omega \quad (2)$$

where  $\zeta$  is a boundary term which incorporates the refractive index mismatch at the tissue-air boundary (Moulton 1990, Schweiger *et al* 1995), and  $\nu$  is the outward normal of the boundary  $\partial\Omega$  at  $\mathbf{m}$ . In the literature, the optical coefficients are often expressed in terms of absorption and scattering. The scattering coefficient,  $\mu_s$ , is related to  $\mu_a$  and  $\kappa$  by

$$\kappa = \frac{1}{3(\mu_a + \mu'_s)}. \quad (3)$$

The distribution of scattering angles for individual scattering events in biological tissues is generally not isotropic, but strongly biased towards small angles. If we assume that all light contributing to measurements has undergone a large number of scattering events, then we can introduce an equivalent isotropic scattering coefficient (often called *transport* or *reduced* scattering coefficient)  $\mu'_s = (1 - g)\mu_s$ , where  $g$  is the average cosine of the single scattering phase function, with a typical range of  $g = 0.9, \dots, 0.99$ .

The measurable *exitance* is the (diffuse) outgoing radiation and corresponds to the Neumann data

$$y(\mathbf{m}; \omega) = -\kappa(\mathbf{m}) \frac{\partial \Phi(\mathbf{m}; \omega)}{\partial \nu}, \quad \mathbf{m} \in \partial\Omega. \quad (4)$$

We use  $X$  to denote the function space of the parameters,  $Q$  the function space of the sources and  $Y$  the function space of the data. Some discussion of the form of these spaces can be found in Arridge (1999) and Dorn (1998).

The *Robin-to-Neumann* (RtN) map

$$\Lambda(\mu_a, \kappa; \omega) : Q \rightarrow Y \quad (5)$$

is a linear mapping from the incoming radiation  $f$  to the outgoing radiation  $y$ , for any given pair of functions in the solution space  $\{\mu_a, \kappa\}$ :

$$y(\mathbf{m}; \omega) = \Lambda(\mu_a, \kappa; \omega) f(\mathbf{m}; \omega), \quad \mathbf{m} \in \partial\Omega. \quad (6)$$

The *forward operator*

$$F_s : X \rightarrow Y \quad (7)$$

is a non-linear mapping from pairs of functions in the solution space to the data on the boundary, for given incoming radiation  $f_s$ , where index  $s$  is introduced to allow specific indexing of different sources. The *direct Fréchet derivative*

$$F'_s(\mu_a, \kappa; \omega) : X \rightarrow Y \quad (8)$$

is a linear mapping from pairs of functions in the solution space to the data on the boundary, for given incoming radiation  $f_s$ .  $F'_s$  is the linearization of  $F_s$  around the solution point  $\{\mu_a, \kappa\}$  that maps changes  $\{\mu_a^\delta, \kappa^\delta\} \equiv \{\alpha, \beta\}$  in solution space functions to changes  $y_s^\delta$  in data. The value of the mapping

$$y_s^\delta(\mathbf{m}; \omega) = F'_s(\mu_a, \kappa; \omega) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad (9)$$

is given by

$$\left\{ -\nabla \cdot \kappa(\mathbf{r})\nabla + \left( \mu_a(\mathbf{r}) + \frac{i\omega}{c} \right) \right\} \Phi_s^\delta(\mathbf{r}; \omega) = \left\{ \nabla \cdot \beta(\mathbf{r})\nabla - \left( \alpha(\mathbf{r}) + \frac{i\omega}{c} \right) \right\} \Phi_s(\mathbf{r}; \omega) \quad \mathbf{r} \in \Omega \setminus \partial\Omega \quad (10)$$

$$\Phi_s^\delta(\mathbf{m}; \omega) + 2\zeta\kappa(\mathbf{m}) \frac{\partial \Phi_s^\delta(\mathbf{m}; \omega)}{\partial \nu} = 0 \quad \mathbf{m} \in \partial\Omega \quad (11)$$

$$y_s^\delta(\mathbf{m}; \omega) = -\kappa(\mathbf{m}) \frac{\partial \Phi_s^\delta(\mathbf{m}; \omega)}{\partial \nu}, \quad \mathbf{m} \in \partial\Omega. \quad (12)$$

The adjoint Fréchet derivative

$$F_s'^* (\mu_a, \kappa; \omega) : Y \rightarrow X \quad (13)$$

is a linear mapping from functions on the boundary to pairs of functions in the solution space, for given incoming radiation  $f_s$ . The value of the mapping

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = F_s'^* (\mu_a, \kappa; \omega) y_s^\delta(\mathbf{m}; \omega) \quad (14)$$

is given by

$$\left\{ -\nabla \cdot \kappa(\mathbf{r})\nabla + \left( \mu_a(\mathbf{r}) - \frac{i\omega}{c} \right) \right\} \Phi_s^+(\mathbf{r}; \omega) = 0 \quad \mathbf{r} \in \Omega \setminus \partial\Omega \quad (15)$$

$$\Phi_s^+(\mathbf{m}; \omega) + 2\zeta\kappa(\mathbf{m}) \frac{\partial \Phi_s^+(\mathbf{m}; \omega)}{\partial \nu} = y_s^\delta(\mathbf{m}; \omega) \quad \mathbf{m} \in \partial\Omega \quad (16)$$

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \text{Re} \left( \begin{pmatrix} -\overline{\Phi_s^+} \Phi_s \\ -\nabla \overline{\Phi_s^+} \cdot \nabla \Phi_s \end{pmatrix} \right) \quad \mathbf{r} \in \Omega \setminus \partial\Omega. \quad (17)$$

We assume a finite number of incoming radiation sources  $\mathbb{S} = \{f_s(\mathbf{m}); s = 1, \dots, S\}$ , where  $f_s(\mathbf{m})$  is a function of local support on  $\partial\Omega$  representing the finite width and profile of the  $s$ th source. The forward operator is now considered as a stacked set of operators

$$\mathbf{y}(\mathbf{m}; \omega) = \begin{bmatrix} y_1(\mathbf{m}; \omega) \\ y_2(\mathbf{m}; \omega) \\ \vdots \\ y_S(\mathbf{m}; \omega) \end{bmatrix} = \mathbf{F}(\mu_a, \kappa; \omega) = \begin{bmatrix} F_1(\mu_a, \kappa; \omega) \\ F_2(\mu_a, \kappa; \omega) \\ \vdots \\ F_S(\mu_a, \kappa; \omega) \end{bmatrix}. \quad (18)$$

We also consider a finite sampling of the outgoing distributions  $y_s(\mathbf{m}; \omega)$ , leading to a measurement model:

$$y_{s,i}(\omega) = \mathcal{M}_i[y_s(\mathbf{m}; \omega)] = \int_{\partial\Omega} w_i(\mathbf{m}) y_s(\mathbf{m}; \omega) \quad (19)$$

where  $w_i(\mathbf{m})$  represents the finite aperture of a detector. In the following we will use  $\mathbb{W} = \{w_i(\mathbf{m}); i = 1, \dots, D\}$  to assume a finite number  $D$  of detectors, with  $\mathbb{W}_s \subset \mathbb{W}$  the

subset of size  $D_s \leq D$  of detectors that see source  $s$ . An ideal choice for the functions in sets  $\mathbb{S}, \mathbb{W}$  would form a basis for the set of complete functions on  $\partial\Omega$  that could be used to represent the Robin-to-Neumann operator  $\Lambda$ . In this paper we use Hanning functions with finite width chosen to span  $\partial\Omega$ . These functions can be analytically integrated with the FEM basis functions used for the discrete forward model (section 4.1).

The discrete data in (18) is complex, and inversion would lead to a complex parameter update. For this reason the data vector is usually split into real and imaginary parts with a commensurate splitting of the linearized derivative operators. In the following we assume  $\mathbf{y} \in \mathbb{R}^{2M}$  where  $M = \sum_{s=1}^S D_s$ . In addition, when considering log of the data the splitting associates the real part with logarithmic amplitude, and the imaginary part with phase. This is discussed further below in section 4.4.

### 3. Approaches to the inverse problem

#### 3.1. Output least-squares approach

In the following we combine the two scalar functions into one symbol

$$\mathbf{x} := \begin{pmatrix} \mu_a \\ \kappa \end{pmatrix}, \quad \mathbf{x} \in \mathbb{R}^N, \quad (20)$$

with  $N = 2P_b$ , where  $P_b$  is the dimension of the basis expansion used for representing the optical parameter distributions in the context of the inverse solver. We consider the regularized output least-squares approach to the inverse problem. We seek the solution

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \Psi(\mathbf{x}), \quad (21)$$

where

$$\Psi(\mathbf{x}) = \frac{1}{2} \sum_{i,s} |y_{s,i}(\omega) - F_{s,i}(\mathbf{x}; \omega)|^2 + \tau \mathcal{R}(\mathbf{x}) \quad (22)$$

is the *objective function* and  $\mathcal{R}(\mathbf{x})$  is a regularizing functional whose relative contribution is governed by hyperparameter  $\tau$ . The Newton scheme for solving equation (21) is given by the iteration

$$\begin{aligned} \mathbf{x}^{(k+1)} = & \mathbf{x}^{(k)} + \zeta_k (\mathbf{F}'^*(\mathbf{x}^{(k)}; \omega) \mathbf{F}'(\mathbf{x}^{(k)}; \omega) + \mathcal{H}(\mathbf{x}^{(k)}; \omega) + \tau \mathcal{R}''(\mathbf{x}^{(k)}))^{-1} \\ & \times [\mathbf{F}'^*(\mathbf{x}^{(k)}; \omega) (\mathbf{y}(\omega) - \mathbf{F}(\mathbf{x}^{(k)}; \omega)) - \tau \mathcal{R}'(\mathbf{x}^{(k)})] \end{aligned} \quad (23)$$

starting from an initial estimate  $\mathbf{x}^{(0)}$ , where  $\mathcal{H} = \langle \mathbf{y} - \mathbf{F}, \mathbf{F}'' \rangle$  is a term depending on both the data and the second derivative of the forward model. Because this term is difficult to compute, and can also lead to a loss of positive-definiteness of the Hessian, it is left out in the Gauss–Newton scheme at the cost of potential loss of the local quadratic convergence of the standard Newton iteration.

#### 3.2. Globalization

There are two common approaches to restore local convergence (Dennis and Schnabel 1983): (i) a trust region approach which leads to the Levenberg–Marquardt (LM) method

$$\begin{aligned} \mathbf{x}_{\text{LM}}^{(k+1)} = & \mathbf{x}^{(k)} + (\mathbf{F}'^*(\mathbf{x}^{(k)}; \omega) \mathbf{F}'(\mathbf{x}^{(k)}; \omega) + \tau \mathcal{R}''(\mathbf{x}^{(k)}) + \lambda_k \mathbf{I})^{-1} \\ & \times [\mathbf{F}'^*(\mathbf{x}^{(k)}; \omega) (\mathbf{y}(\omega) - \mathbf{F}(\mathbf{x}^{(k)}; \omega)) - \tau \mathcal{R}'(\mathbf{x}^{(k)})], \end{aligned} \quad (24)$$

where the control parameter  $\lambda_k > 0$  is adjusted at each iteration, and (ii) a damped Gauss–Newton (DGN) approach given by

$$\mathbf{x}_{\text{DGN}}^{(k+1)} = \mathbf{x}^{(k)} + \varsigma_k (\mathbf{F}'^{/*}(\mathbf{x}^{(k)}; \omega) \mathbf{F}'(\mathbf{x}^{(k)}; \omega) + \tau \mathcal{R}''(\mathbf{x}^{(k)}))^{-1} \times [\mathbf{F}'^{/*}(\mathbf{x}^{(k)}; \omega)(\mathbf{y}(\omega) - \mathbf{F}(\mathbf{x}^{(k)}; \omega)) - \tau \mathcal{R}'(\mathbf{x}^{(k)})], \quad (25)$$

where  $\varsigma_k$  is a step length parameter obtained by an inexact line search at each iteration. For simplification we introduce the following notation for the Hessian and gradient terms:

$$\mathbf{H}_k^{(\lambda_k)} = \mathbf{F}'^{/*}(\mathbf{x}^{(k)}; \omega) \mathbf{F}'(\mathbf{x}^{(k)}; \omega) + \tau \mathcal{R}''(\mathbf{x}^{(k)}) + \lambda_k \mathbf{I} \quad (26)$$

and

$$\mathbf{g}_k = \mathbf{F}'^{/*}(\mathbf{x}^{(k)}; \omega)(\mathbf{F}(\mathbf{x}^{(k)}; \omega) - \mathbf{y}(\omega)) + \tau \mathcal{R}'(\mathbf{x}^{(k)}). \quad (27)$$

A conventional way to implement these schemes is to build the explicit Jacobian  $\mathbf{J} \in \mathbb{R}^{2M \times N}$  as the discrete representation of  $\mathbf{F}'$ , and solve the linear normal equations

$$(\mathbf{J}^T \mathbf{J} + \tau \mathcal{R}''(\mathbf{x}^{(k)}) + \lambda_k \mathbf{I}) \mathbf{x}^\delta = \mathbf{g}_k. \quad (28)$$

### 3.3. Solution of normal equations

Solution of a linear system such as (28) can be approached in a number of ways. For a relatively small solution space, the system is considered overdetermined and solved directly using Gaussian substitution or indirectly using an iterative solution scheme.

If the regularization functional is quadratic we have the relations

$$\begin{aligned} \mathcal{R}(\mathbf{x}^{(k)}) &= \|\mathbf{L}(\mathbf{x}^{(k)} - \mathbf{x}_g)\|_2^2 \\ \mathcal{R}'(\mathbf{x}^{(k)}) &= \mathbf{L}^T \mathbf{L}(\mathbf{x}^{(k)} - \mathbf{x}_g) \\ \mathcal{R}''(\mathbf{x}^{(k)}) &= \mathbf{L}^T \mathbf{L} \end{aligned}$$

where  $\mathbf{x}_g$  is a given reference state. In the Bayesian framework and for quadratic regularization functionals  $\mathbf{x}_g$  is the mean of a Gaussian prior distribution. Note that  $\mathbf{x}_g$  is not necessarily identical to the initial estimate  $\mathbf{x}^{(0)}$ . The normal equations can then be expressed as

$$\begin{pmatrix} \mathbf{J} \\ \tau^{1/2} \mathbf{L} \end{pmatrix} \mathbf{x}^\delta = \begin{pmatrix} \mathbf{y} - \mathbf{F}(\mathbf{x}^{(k)}) \\ \tau^{1/2} \mathbf{L}(\mathbf{x}^{(k)} - \mathbf{x}_g) \end{pmatrix} \quad (29)$$

which are usually solved by QR decomposition. The matrix in (29) is of dimension  $(2M + N) \times N$ . A standard QR algorithm decomposes into a rectangular  $(2M + N) \times N$  matrix  $\mathbf{Q}$  and a square  $N \times N$  matrix  $\mathbf{R}$ , thus requiring storage for  $N(2M + 2N)$  floating point values.

Alternatively, an underdetermined form can be constructed using the matrix inversion lemma (Eppstein *et al* 2001, Dehghani *et al* 2003):

$$(\mathbf{J}^T \mathbf{J} + \tau \mathbf{L}^T \mathbf{L})^{-1} \mathbf{J}^T = \mathbf{J}^T (\mathbf{J} (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{J}^T + \tau \mathbf{I})^{-1}.$$

Typically the regularization functional is in terms of a PDE and its inverse is non-sparse and equivalent to a smoothing operation. For this reason the regularization is sometimes replaced by an image filtering operation.

None of these methods take account of the particular structure of the Hessian. It contains a part that is a dense representation of a compact operator and a part that is sparse, representing a local operator which may have a non-trivial null space. In general the regularizer may not be quadratic, or even convex. In this paper we consider an inexact Gauss–Newton approach to solve equation (28). We use a Krylov method for the Hessian inversion that computes a small number of matrix–vector products. These products are efficiently calculated by applying the

forward and adjoint Fréchet derivatives plus the explicit Hessian of the regularizer. In this paper we consider that the Fréchet derivative is explicitly represented by a Jacobian matrix, although if this is infeasible, it could be replaced by an operator involving additional solutions of the original PDE.

#### 4. Implementation

##### 4.1. Finite element method forward solver

We employ a finite element discretization for numerically computing the forward problem defined in equations (1) and (2). We divide the domain  $\Omega$  into an unstructured mesh of nonoverlapping elements of simple shape, and define a basis in  $\Omega$  which is polynomial within each element, and continuous across elements. Given  $P_h$  nodal points with associated basis functions  $u_i(\mathbf{r}), i = 1, \dots, P_h$ , of limited support, a function  $\Phi$  defined in  $\Omega$  is then approximated in this basis by the piecewise polynomial interpolation  $\Phi^h(\mathbf{r}) = \sum_{i=1}^{P_h} \Phi_i u_i(\mathbf{r})$ , where  $\Phi = \{\Phi_i\}$  is a  $P_h$ -dimensional array of basis coefficients. We use superscript  $h$  to denote the finite element mesh basis expansion. The coefficients  $\kappa(\mathbf{r})$  and  $\mu_a(\mathbf{r})$  are expanded in the same manner into  $\kappa^h(\mathbf{r})$  and  $\mu_a^h(\mathbf{r})$ . As developed previously (Arridge *et al* 1993, Schweiger *et al* 1995), the basis representation transforms the continuous problem (1), (2) into a linear system which is expressed in terms of the nodal values of the coefficients and field:

$$(\mathbf{K}(\kappa^h) + \mathbf{C}(\mu_a^h) + \gamma \mathbf{A} + i\omega \mathbf{B})\Phi(\omega) = \mathbf{Q}(\omega) \quad (30)$$

with  $\Phi, \mathbf{Q} \in \mathbb{R}^{P_h}$  and system matrices  $\mathbf{K}, \mathbf{C}, \mathbf{A}, \mathbf{B} \in \mathbb{R}^{P_h \times P_h}$  given by

$$\begin{aligned} K_{ij} &= \int_{\Omega} \kappa^h(\mathbf{r}) \nabla u_i(\mathbf{r}) \cdot \nabla u_j(\mathbf{r}) \, d\mathbf{r}, & C_{ij} &= \int_{\Omega} \mu_a^h(\mathbf{r}) u_i(\mathbf{r}) u_j(\mathbf{r}) \, d\mathbf{r}, \\ A_{ij} &= \int_{\partial\Omega} u_i(\mathbf{r}) u_j(\mathbf{r}) \, d\mathbf{r}, & B_{ij} &= \frac{1}{c} \int_{\Omega} u_i(\mathbf{r}) u_j(\mathbf{r}) \, d\mathbf{r}. \end{aligned} \quad (31)$$

Analytic solutions for the integrals of polynomial basis functions over simple elements such as triangles or tetrahedra are readily available (Zienkiewicz and Taylor 1987). In the general case, e.g. for elements with curved boundaries, numerical quadrature rules must be applied.

##### 4.2. Krylov solution of the Gauss–Newton step

When solving the LM problem (24) or DGN problem (25), the explicit computation and storage of the approximate Hessian  $\mathbf{J}^T \mathbf{J}$  is often intractable in large-scale problems. Instead, we use a Krylov method as an iterative solver for the linearized problem (28) at each Newton step  $k$ . The computational advantage of using a Krylov solver lies in the fact that  $\mathbf{H}$  is only accessed for matrix–vector products. This approach avoids the explicit formation of  $\mathbf{H}$ . Instead we can evaluate matrix–vector operations directly by applying the right-hand side of (26). We can therefore represent  $\mathbf{H}$  formally by its components as an operator  $\mathbf{H} = \mathbf{H}[\mathbf{J}, \mathcal{R}'', \tau, \lambda]$  and replace all instances of operations  $\mathbf{H}\mathbf{x}$  by the subroutine

```
proc Hx (J, R'', tau, lambda, x)
  z = Jx
  return J^T z + tau R'' x + lambda x
end
```

To estimate the computational efficiency of the implicit formulation of  $\mathbf{H}$  we note that the product  $\mathbf{H}\mathbf{x}$  requires  $N^2$  floating point multiplications when  $\mathbf{H}$  is given explicitly, while in

the implicit representation the calculation of  $\mathbf{J}^T \mathbf{J}$  requires  $4NM$  floating point multiplications. Assuming that the computational cost of the product  $\mathcal{R}'' \mathbf{x}$  is small due to its sparsity we find that the implicit formulation is more efficient when  $2M < \frac{N}{2}$ .

#### 4.3. GMRES linear solver

Equation (28) defines a linear system of the form  $\mathbf{H}\mathbf{x} = \mathbf{g}$ , where GN iteration subscript  $k$  and superscript  $\delta$  have been omitted for simplicity. In this paper we use the generalized minimal residual (GMRES) Krylov method (Saad and Schultz 1986) to solve this system.

At iteration  $j$  of the GMRES solver an approximation  $\mathbf{x}^{(j)}$  of  $\mathbf{x}$  is produced from a Krylov space generated by a vector  $\mathbf{c}$ :

$$\mathcal{K}_j(\mathbf{H}, \mathbf{c}) = \text{span}\{\mathbf{c}, \mathbf{H}\mathbf{c}, \dots, \mathbf{H}^{j-1}\mathbf{c}\} \quad (32)$$

where  $\mathbf{c} = \mathbf{g}$  is commonly used, and the matrix power is defined by  $H^1 = H$ ,  $H^{i+1} = HH^i$  ( $i \geq 1$ ).

GMRES finds in iteration  $j$  a solution  $\mathbf{x}^{(j)}$  which solves the least-squares problem

$$\mathbf{x}^{(j)} = \arg \min_{\mathbf{x} \in \mathcal{K}_j(\mathbf{H}, \mathbf{g})} \|\mathbf{g} - \mathbf{H}\mathbf{x}\| \quad (33)$$

by constructing an orthonormal basis  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_j\}$  for  $\mathcal{K}_j(\mathbf{H}, \mathbf{g})$  using Arnoldi's method (Arnoldi 1951). Starting from  $\mathbf{v}_1 = \mathbf{g}/\|\mathbf{g}\|$ , the basis for  $\mathcal{K}_{j+1}(\mathbf{H}, \mathbf{g})$  is built recursively by

$$\mathbf{v}_{j+1} = \hat{\mathbf{v}}_{j+1}/\|\hat{\mathbf{v}}_{j+1}\| \quad (34)$$

where

$$\hat{\mathbf{v}}_{j+1} = \mathbf{H}\mathbf{v}_j - (h_{1j}\mathbf{v}_1 + \dots + h_{jj}\mathbf{v}_j) \quad (35)$$

and  $h_{ij} = \mathbf{v}_i^* \mathbf{H}\mathbf{v}_j$ . The basis vectors for  $\mathcal{K}_j(\mathbf{H}, \mathbf{g})$  are collected in a matrix  $\mathbf{V}_j = \{\mathbf{v}_1, \dots, \mathbf{v}_j\}$  which leads to a decomposition

$$\mathbf{H}\mathbf{V}_j = \mathbf{V}_{j+1}\mathbf{U}_j \quad (36)$$

where  $\mathbf{U}_j$  is an upper Hessenberg matrix of size  $(j+1) \times j$  storing coefficients  $h_{ij}$  and norms  $\|\hat{\mathbf{v}}_j\|$ . We can now represent  $\mathbf{x}$  as  $\mathbf{x} = \mathbf{V}_j\mathbf{w}$  for some  $\mathbf{w}$ , and thus

$$\mathbf{H}\mathbf{x} = \mathbf{V}_{j+1}\mathbf{U}_j\mathbf{w}, \quad \mathbf{g} = \|\mathbf{g}\|\mathbf{V}_{j+1}\mathbf{e}_1 \quad (37)$$

where  $\mathbf{e}_1$  is the first column of the identity matrix. The least-squares problem in (33) at iteration  $j$  reduces therefore to

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \|(\|\mathbf{g}\|\mathbf{e}_1) - \mathbf{U}_j\mathbf{w}\|, \quad \mathbf{x}^{(j)} = \mathbf{V}_j\hat{\mathbf{w}}. \quad (38)$$

A standard method to reduce memory requirements for GMRES is to employ a restarting strategy, where after  $m$  iterations the accumulated data in  $\mathbf{U}$  and  $\mathbf{V}$  are cleared and the intermediate results are used as the starting values for the next  $m$  iterations, until convergence is achieved. For the calculations in this paper we used  $m = 10$ , and a stopping criterion defined by

$$\|\mathbf{g} - \mathbf{H}\mathbf{x}^{(j)}\|_{L^2} \leq \eta\|\mathbf{g}\|_{L^2} \quad (39)$$

for a prescribed tolerance  $\eta$ . We find that  $\mathbf{H}$  is referenced only in (35) and (39) for vector multiplication operations, and can therefore be represented implicitly as discussed in section 4.2.

#### 4.4. Rescaling and conditioning

The linearized problem in optical tomography given by (28) is very badly conditioned, and one strategy to compensate for this is rescaling (or functional transformation) of the data space or the solution space or both (Dennis and Schnabel 1983). Transformation of the data spaces constitutes left preconditioning of the normal equations and leads to a change in the definition of the data-space norm in (22). Commensurately, transformation of the solution space constitutes right preconditioning and leads to a change in the regularization functional in (22). In this paper we use the following transformations:

$$\tilde{\mathbf{y}} = \begin{pmatrix} \tilde{\mathbf{y}}^A \\ \tilde{\mathbf{y}}^\varphi \end{pmatrix} = \begin{pmatrix} c_{\text{re}} \text{Re} \\ c_{\text{im}} \text{Im} \end{pmatrix} \ln \mathbf{y} \quad (40)$$

$$\tilde{\mathbf{x}} = \ln \frac{\mathbf{x}}{\bar{\mathbf{x}}}. \quad (41)$$

The logarithmic transformation of data corresponds to the choice of logarithmic modulation amplitude  $A$  and phase shift  $\varphi$  as the measurables with an additional scaling of the log amplitude and phase parts of the data vector. This choice of measurable is widely used in optical tomography, and is very similar to the choice of logarithmic DC intensity and mean time as used in Schweiger *et al* (1993). The purpose of the constants  $c_{\text{re}}, c_{\text{im}}$  is to normalize between the logarithmic amplitude (which typically has values in the range  $-1 \rightarrow -20$ ) and phase (which is by definition in the range  $0 \rightarrow 2\pi$ ). In terms of the solution space, the purpose of scaling  $\mathbf{x}$  by the average value  $\bar{\mathbf{x}}$  is to render the units of the solution space ‘dimensionless’, and the purpose of the logarithmic transformation is to ensure a positive solution. For more details on the choice of transformations and on the choice of the constants  $c_{\text{re}}, c_{\text{im}}$  see Schweiger and Arridge (1999).

With these transformations, the objective function (22) becomes

$$\tilde{\Psi}(\tilde{\mathbf{x}}) = \frac{c_{\text{re}}}{2} \sum_{i,s} \left| y_{s,i}^A(\omega) - F_{s,i}^A(\tilde{\mathbf{x}}e^{\tilde{\mathbf{x}}}; \omega) \right|^2 + \frac{c_{\text{im}}}{2} \sum_{i,s} \left| y_{s,i}^\varphi(\omega) - F_{s,i}^\varphi(\tilde{\mathbf{x}}e^{\tilde{\mathbf{x}}}; \omega) \right|^2 + \tau \mathcal{R}(\tilde{\mathbf{x}}e^{\tilde{\mathbf{x}}}). \quad (42)$$

Making use of the chain rule for Fréchet differentiation we can write the Jacobian and Hessian terms

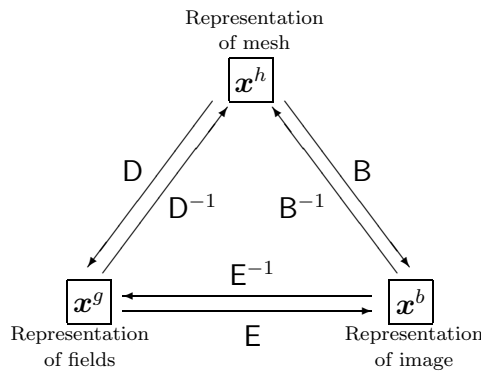
$$\begin{aligned} \tilde{\mathbf{J}} &= \begin{pmatrix} c_{\text{re}} \mathbf{J}^{(A)} \\ c_{\text{im}} \mathbf{J}^{(\varphi)} \end{pmatrix} \text{diag}(\mathbf{x}) = \begin{pmatrix} c_{\text{re}} \text{Re} \\ c_{\text{im}} \text{Im} \end{pmatrix} \text{diag}\left(\frac{1}{\mathbf{F}}\right) \mathbf{J} \text{diag}(\mathbf{x}) \\ \tilde{\mathbf{H}} &= \text{diag}(\mathbf{x}) (c_{\text{re}}^2 \mathbf{J}^{(A)\text{T}} \mathbf{J}^{(A)} + c_{\text{im}}^2 \mathbf{J}^{(\varphi)\text{T}} \mathbf{J}^{(\varphi)} + \tau \mathcal{R}'') \text{diag}(\mathbf{x}) \end{aligned} \quad (43)$$

where the separation of amplitude and phase components leads to real-valued matrices  $\tilde{\mathbf{J}}$  and  $\tilde{\mathbf{H}}$ . Finally we note that in Marquardt’s paper (Marquardt 1963), the trust region method requires ‘sphereing’ of the solution space, in order to give a Hessian with 1’s on its diagonal. This is achieved by constructing another diagonal scaling matrix  $\mathbf{T}$

$$T_{ii} = \frac{1}{\sqrt{\sum_j J_{ji}^2 + \tau \mathcal{R}''_{ii}}} \quad (44)$$

which is applied in conjunction with the above rescalings:

$$\begin{aligned} \tilde{\mathbf{J}} &\rightarrow \tilde{\mathbf{J}}\mathbf{T} \\ \tilde{\mathbf{H}} &\rightarrow \mathbf{T}\tilde{\mathbf{H}}\mathbf{T}. \end{aligned}$$



**Figure 1.** Illustration of the mapping between mesh and image bases, via an intermediate high-resolution regular grid.

#### 4.5. Gauss–Newton update strategies

We need to define the update criteria for the control parameters  $\lambda_k$  in (24) and  $\zeta_k$  in (25). For the adjustment of  $\lambda$  in the LM updates we adopt the following strategy:

- Set  $\lambda$  to an initial value  $\lambda_0 > 0$ , for example  $\lambda_0 = 0.01$ .
- Calculate the objective function  $\tilde{\Psi}_k$  with equation (42).
- If  $\tilde{\Psi}_k < \tilde{\Psi}_{k-1}$ , accept the update and set  $\lambda_{k+1} = \lambda_k/4$ .
- If  $\tilde{\Psi}_k \geq \tilde{\Psi}_{k-1}$ , set  $\lambda_k \leftarrow 4\lambda_k$  and repeat until a reduction in the objective function is achieved.

For the DGN method we determine the step  $\zeta_k$  at each iteration by an inexact line search:

- Define an initial estimate of the solution  $\mathbf{x}_0$  and initial step length  $\zeta_0 > 0$ .
- At GN iteration  $k \geq 1$ , obtain a search direction  $\mathbf{x}^\delta$  by solving equation (28).
- Set  $\zeta_a = 0$  and  $\zeta_b = \zeta_{k-1}$ . Calculate the corresponding objective functions  $\tilde{\Psi}_a = \tilde{\Psi}(\mathbf{x}^{(k-1)} + \zeta_a \mathbf{x}^\delta)$  and  $\tilde{\Psi}_b = \tilde{\Psi}(\mathbf{x}^{(k-1)} + \zeta_b \mathbf{x}^\delta)$  from equation (42).
- Bracket the minimum:
  - (a) if  $\tilde{\Psi}_b > \tilde{\Psi}_a$ : set  $\zeta_m \leftarrow \zeta_b/2$ . While if  $\tilde{\Psi}_m > \tilde{\Psi}_a$ : set  $\zeta_m \leftarrow \zeta_m/2$ ,  $\zeta_b \leftarrow \zeta_b/2$ .
  - (b) if  $\tilde{\Psi}_b \leq \tilde{\Psi}_a$ : set  $\zeta_m \leftarrow \zeta_b$ ,  $\zeta_b \leftarrow 2\zeta_b$ . While if  $\tilde{\Psi}_m > \tilde{\Psi}_b$ : set  $\zeta_a \leftarrow \zeta_m$ ,  $\zeta_m \leftarrow \zeta_b$ ,  $\zeta_b \leftarrow 2\zeta_b$ .
- Obtain step length  $\zeta_k$  as the minimum of the quadratic interpolation of  $\zeta_a, \zeta_m, \zeta_b$ :  $\zeta_k = -b/(2a)$ , where  $a = \left( \frac{\tilde{\Psi}_a - \tilde{\Psi}_b}{\zeta_a - \zeta_b} - \frac{\tilde{\Psi}_a - \tilde{\Psi}_m}{\zeta_a - \zeta_m} \right) (\zeta_b - \zeta_m)^{-1}$ , and  $b = \frac{\tilde{\Psi}_a - \tilde{\Psi}_b}{\zeta_a - \zeta_b} - a(\zeta_a + \zeta_b)$ .

#### 4.6. Basis expansion

Although the forward FEM model typically defines the coefficients  $\mu_a$  and  $\kappa$  on an unstructured mesh, it is convenient to represent the images in terms of a regular grid using well-defined local basis functions (Schweiger and Arridge 2003). This also has the effect of making the definition of image-based regularization terms much simpler. Mapping between the representations can be described in terms of sparse prolongation and restriction operators which are conveniently obtained by sampling the unstructured mesh solution into an intermediate high-resolution regular grid. This high-resolution grid is also used to represent the fields for the purpose of calculating the Fréchet derivatives and Jacobians. A summary of this scheme is shown in figure 1. When the mapping between bases is linear, the prolongation operators ( $D, B, E$ ) and associated restriction operators ( $D^{-1}, B^{-1}, E^{-1}$ ) can be represented in matrix form.

#### 4.7. Regularization

We use a first-order Tikhonov prior of the form

$$\mathcal{R}(\mathbf{x}_k) = \|\mathbf{L}\Delta\mathbf{x}_k\|_2^2 \quad (45)$$

where  $\mathbf{L}^T\mathbf{L} \in \mathbb{R}^{N \times N}$  denotes the Laplacian, and  $\Delta\mathbf{x}_k = \mathbf{x}_k - \mathbf{x}_0$ . We define  $\mathbf{L}^T\mathbf{L}$  as

$$(\mathbf{L}^T\mathbf{L})_{ij} = \begin{cases} nn & \text{if } i = j \\ -1 & \text{if } j \text{ is neighbour of } i \\ 0 & \text{otherwise} \end{cases} \quad (46)$$

where  $nn$  is the number of neighbours of basis component  $i$ . In this paper we use a regular pixel (in 2D) or voxel (in 3D) grid basis, and use the four-connected neighbourhood in 2D or six-connected neighbourhood in 3D. An alternative, not presented here, is to define a neighbourhood radius  $r$ , such that all pixels (voxels) inside  $r$  are considered neighbours of  $i$ . To obtain a value for the regularization parameter  $\tau$  we use the L-curve method (Hansen and O’Leary 1993, Vogel 2002).

#### 4.8. Calculation of the Jacobian

For the  $s$ th source and  $i$ th detector we calculate the forward field  $\Phi_s(\mathbf{r}, \omega)$  from equation (1), with boundary condition  $f_s(\mathbf{m})$  in (2) and the adjoint field  $\Phi_i^+$  from (15) with boundary condition  $w_i(\mathbf{m})$  in equation (16). These fields are mapped from the nodal basis of the forward solver into the high-resolution grid basis expansion  $g$ :

$$\Phi_s^g = \mathbf{D}\Phi_s^h, \quad \Phi_i^{g+} = \mathbf{D}\Phi_i^{h+} \quad (47)$$

and used to define the complex *photon measurement density functions* (Arridge 1995)

$$\rho_{is}^{(\mu)} = -\frac{\Phi_s^g \Phi_i^{g+}}{y_{s,i}}, \quad \rho_{is}^{(\kappa)} = -\frac{\nabla\Phi_s^g \cdot \nabla\Phi_i^{g+}}{y_{s,i}}. \quad (48)$$

These vectors are mapped back into the image basis and split into real and imaginary parts to construct the Jacobian

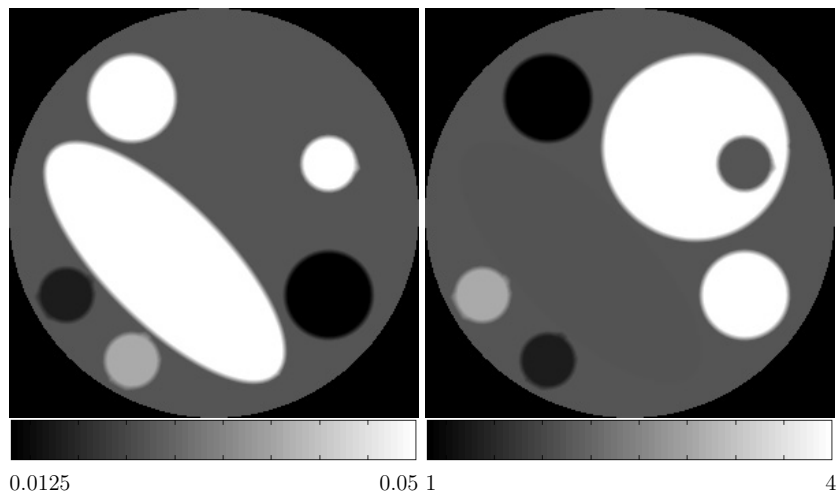
$$\begin{aligned} \mathbf{J}_{is}^{(A,\mu)} &= \text{Re}[\mathbf{E}\rho_{is}^{(\mu)}] & \mathbf{J}_{is}^{(A,\kappa)} &= \text{Re}[\mathbf{E}\rho_{is}^{(\kappa)}] \\ \mathbf{J}_{is}^{(\varphi,\mu)} &= \text{Im}[\mathbf{E}\rho_{is}^{(\mu)}] & \mathbf{J}_{is}^{(\varphi,\kappa)} &= \text{Im}[\mathbf{E}\rho_{is}^{(\kappa)}]. \end{aligned}$$

## 5. Numerical results

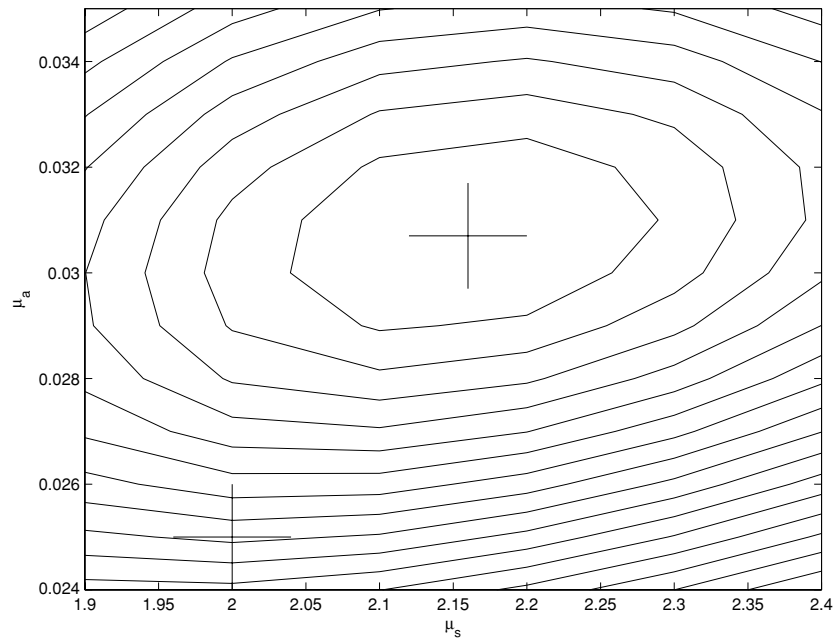
### 5.1. Simulated test case

We first consider a 2D test problem of recovering the parameter distributions of  $\mu_a$  and  $\mu_s$  in a circular object of radius 25 mm, shown in figure 2. A total of 32 source sites are placed at equidistant angular spacing along the surface, and 32 detector sites are placed so that each detector is located between two source sites. For each source, measurements are produced at all detector sites except the two sites closest to the source, leading to a total of 960 measurements. We consider sources to be intensity modulated with a frequency of 50 MHz, and each measurement consisting of the logarithmic modulation amplitude  $\tilde{y}_i^A$  and phase shift  $\tilde{y}_i^\varphi$ . The target data  $\tilde{\mathbf{y}} = (\tilde{\mathbf{y}}^A, \tilde{\mathbf{y}}^\varphi)$  are calculated with the FEM diffusion forward model using a mesh consisting of 7261 ten-noded triangles and 32 971 nodes, defining a piecewise quadratic unstructured basis expansion. Both  $\tilde{\mathbf{y}}^A$  and  $\tilde{\mathbf{y}}^\varphi$  are then contaminated with 1% additive Gaussian random noise.

For the reconstruction we use homogeneous distributions for the initial estimates of optical coefficients,  $\mu_a^{(0)}(\mathbf{r}) = \bar{\mu}_a$  and  $\mu_s^{(0)}(\mathbf{r}) = \bar{\mu}_s$ , which are obtained from a global fit of the two

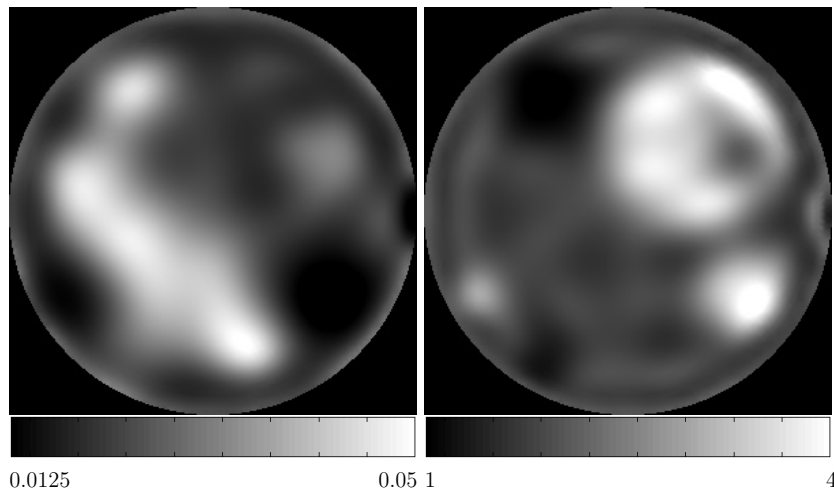


**Figure 2.** Target absorption (left) and scattering distribution (right) of the circular 2D test problem. Background parameters are  $\mu_a = 0.025 \text{ mm}^{-1}$  and  $\mu_s = 2 \text{ mm}^{-1}$ .



**Figure 3.** Data residual  $\tilde{\Psi}$  of initial estimate as a function of homogeneous parameters  $\bar{\mu}_a$  and  $\bar{\mu}_s$ . Indicated are the background values of the target at  $\bar{\mu}_a = 0.025$ ,  $\bar{\mu}_s = 2$ , and the minimum of  $\tilde{\Psi}$  at  $\bar{\mu}_a = 0.0307$  and  $\bar{\mu}_s = 2.16$ .

parameter values to the data. Figure 3 shows the norm of the data residual,  $\tilde{\Psi}(\bar{\mu}_a, \bar{\mu}_s)$ , as defined in equation (42), as a function of the homogeneous parameters  $\bar{\mu}_a$  and  $\bar{\mu}_s$ . The minimum occurs at  $\bar{\mu}_a = 0.0307 \text{ mm}^{-1}$  and  $\bar{\mu}_s = 2.16 \text{ mm}^{-1}$ , which is offset from the true background values of  $\mu_a = 0.025 \text{ mm}^{-1}$  and  $\mu_s = 2 \text{ mm}^{-1}$  due to the influence of the inclusions.



**Figure 4.** Reconstructed absorption (left) and scattering (right) distributions with DGN algorithm. Updates generated using a line search for parameter  $\zeta$ .

Starting from this estimate, we perform a reconstruction into a linear  $20 \times 20$  pixel grid  $b$ . The FEM basis  $h$  consists of a mesh with 6840 three-noded triangles and 3511 nodes, using linear shape functions. The intermediate high-resolution pixel grid  $g$  is of dimension  $80 \times 80$ .

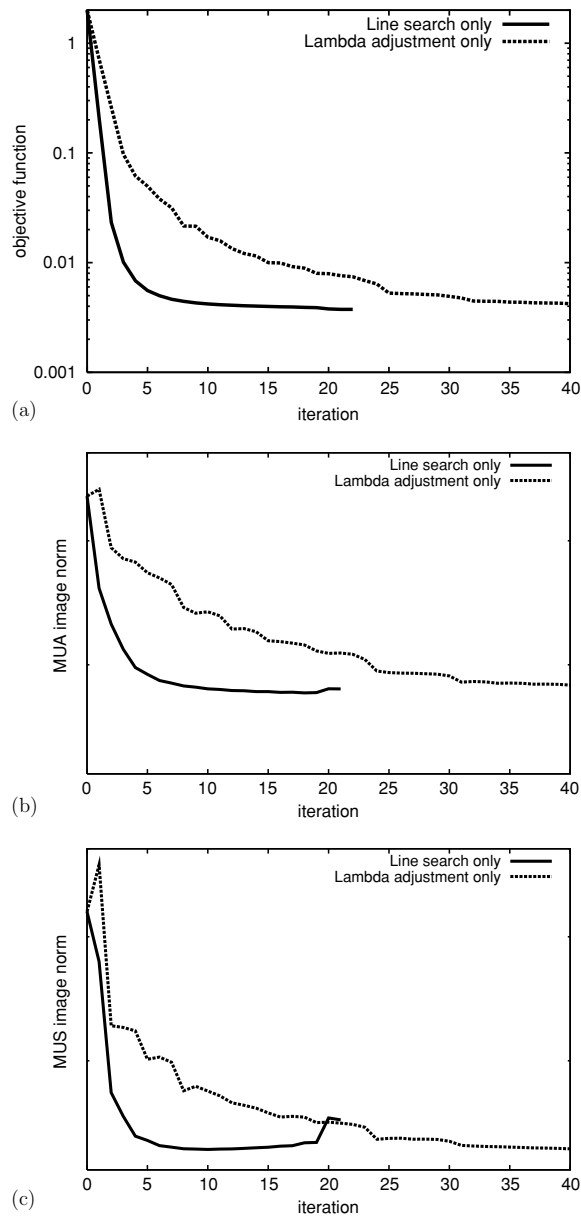
The results of the reconstruction are shown in figure 4. We used the DGN version of the Gauss–Newton solver, performing a line search for  $\zeta$  at each iteration. The Tikhonov prior (45) was used, and a regularization parameter  $\tau = 10^{-5}$ , obtained from an L-curve calculation. The linear step is solved by GMRES, using a tolerance value of  $\eta = 10^{-3}$  for the stopping criterion (39).

A reconstruction with the Levenberg–Marquardt version of the Gauss–Newton solver was also performed. While both solvers produce similar final images, the convergence behaviour is quite different. Figure 5(a) shows the objective functions  $\tilde{\Psi}(\mathbf{x}^{(k)})$  as a function of Newton iteration  $k$  for both LM and DGN. We find that the DGN solver consistently converges significantly faster than the LM solver, for all the reconstructions presented in this paper, both in terms of the number of iterations required, and the total run time. The superior performance of the DGN solver is also reflected in the behaviour of the normalized  $L_1$  residuals

$$\varepsilon_{\mu_a}^{(k)} = \frac{\int_{\Omega} |\mu_a^{(k)} - \mu_a^{(\text{tgt})}| \, d\mathbf{r}}{\int_{\Omega} \mu_a^{(\text{tgt})} \, d\mathbf{r}}, \quad \varepsilon_{\mu_s}^{(k)} = \frac{\int_{\Omega} |\mu_s^{(k)} - \mu_s^{(\text{tgt})}| \, d\mathbf{r}}{\int_{\Omega} \mu_s^{(\text{tgt})} \, d\mathbf{r}}$$

of the reconstructed images as a function of iteration  $k$ , shown in figures 5(b) and (c). We conclude that the DGN solver provides better performance than the LM solver for the type of problems considered here, despite the additional computational cost per iteration incurred by the line search.

The influence of the stopping criterion  $\eta$  (39) of the DGN-GMRES solver on the convergence behaviour of the reconstruction is shown in figure 6. The graphs show the data residual as a function of iteration count and runtime, for  $\eta = 0.1, 0.01$  and  $0.001$ . For comparison, the results for the Cholesky solution of the explicit Hessian are also plotted. We find that at  $\eta = 0.001$  the convergence rate of the DGN-GMRES solver is nearly identical to the explicit DGN solver. All GMRES solutions compare favourably with the explicit DGN solver in terms of runtime. Figure 7 shows the corresponding  $L_1$  image errors for  $\mu_a$  and  $\mu_s$ .

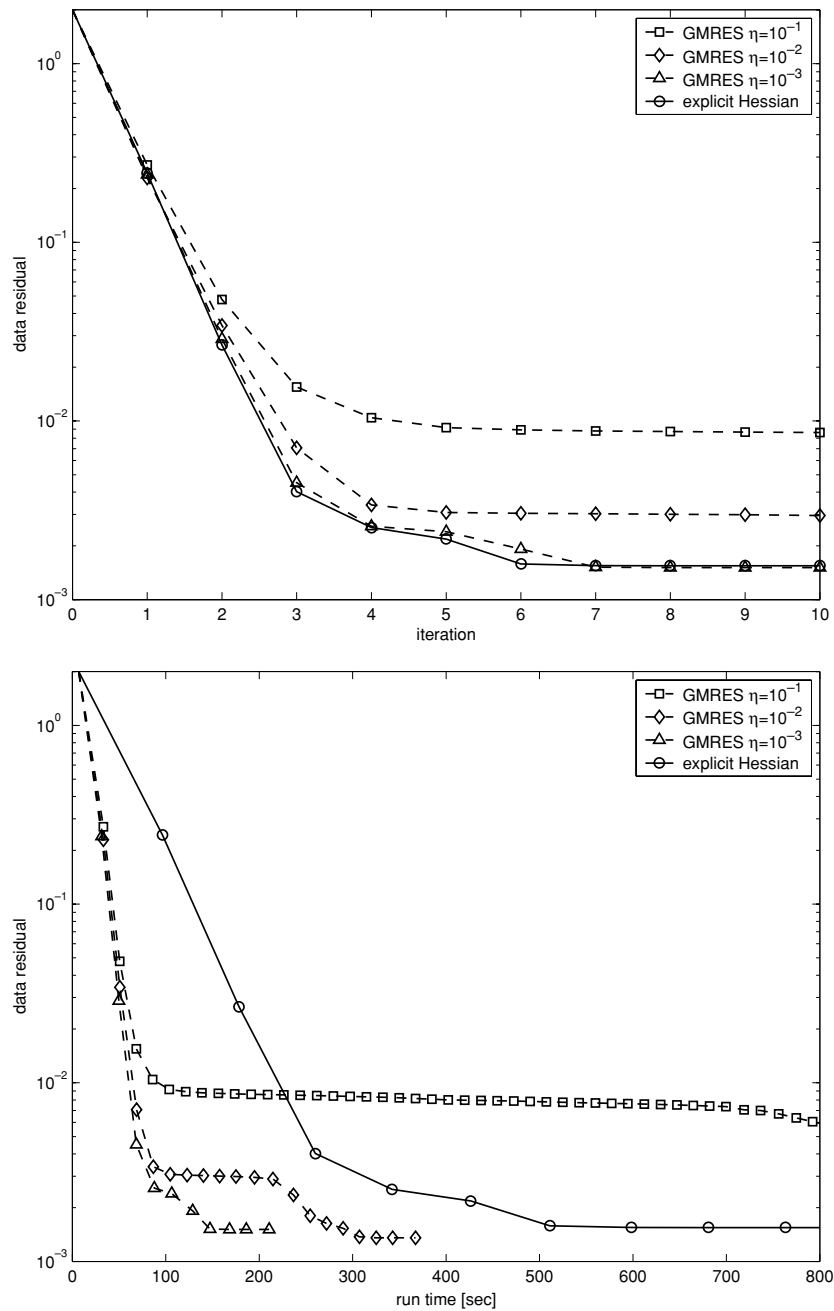


**Figure 5.** Comparison of convergence of objective function  $\Psi$  (a), and  $L_1$  image residuals of  $\mu_a$  (b) and  $\mu_s$  images (c), for reconstruction with line search (DGN), and with control parameter adjustment (LM).

Again, the convergence of the GMRES solver approaches that of the solution of the explicit Hessian, with  $\eta = 0.001$  producing similar results to the explicit DGN solver.

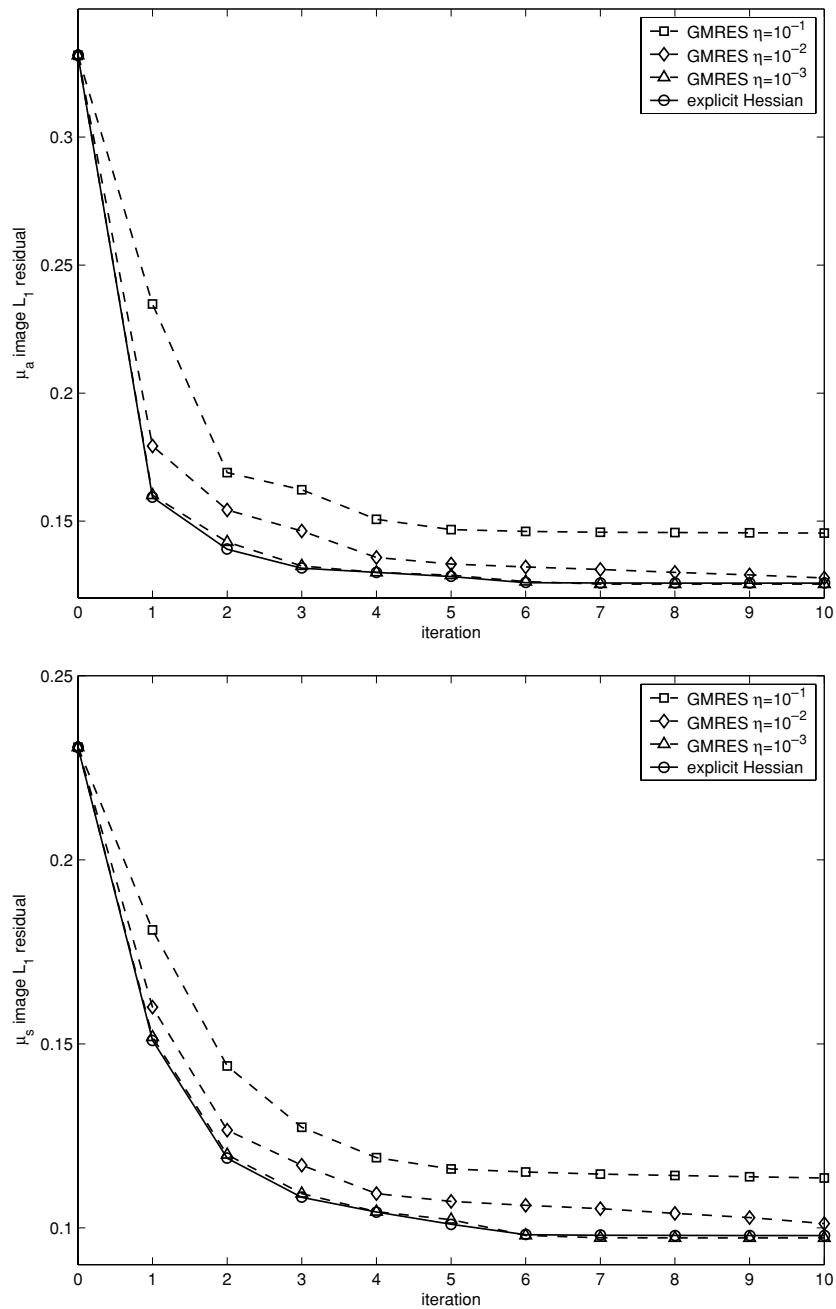
To determine the robustness of the solution with respect to the initial starting parameter distribution we compare the reconstruction results for different combinations of homogeneous initial  $\mu_a^{(0)}$  and  $\mu_s^{(0)}$  parameters. Figure 8 shows the combined  $L_1$  image residuals

$$\varepsilon^{(k)} = \varepsilon_{\mu_a}^{(k)} + \varepsilon_{\mu_s}^{(k)} \quad (49)$$



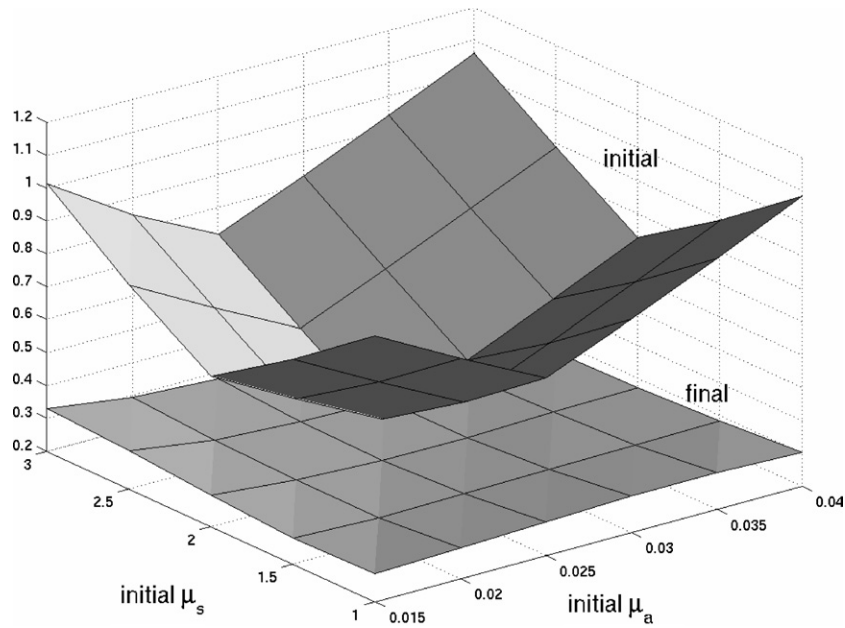
**Figure 6.** Influence of GMRES stopping criterion  $\eta$  on LM convergence. The graphs show the data residual as a function of iteration count (top) and as a function of runtime (bottom). For comparison, the convergence rate of the direct solution of the explicit Hessian is also plotted.

for both the initial estimates ( $k = 0$ ) and the final reconstructions, as a function of the initial homogeneous  $\mu_a$  and  $\mu_s$  parameters. Again the simulated data were contaminated with 1% Gaussian random noise. The top surface shows the image residual of the initial homogeneous



**Figure 7.**  $\mu_a$  (top) and  $\mu_s$  image error (bottom) as a function of iteration count for different choices of GMRES stopping criterion  $\eta$  and direct solution of the explicit Hessian.

distribution, while the bottom surface shows the image residual of the final reconstruction. We find that over the range of initial parameters considered here ( $0.015 \leq \mu_a \leq 0.04$  and  $1 \leq \mu_s \leq 3$ ) the reconstructed images show a very similar residual, indicating good robustness of the reconstruction with respect to the starting values.



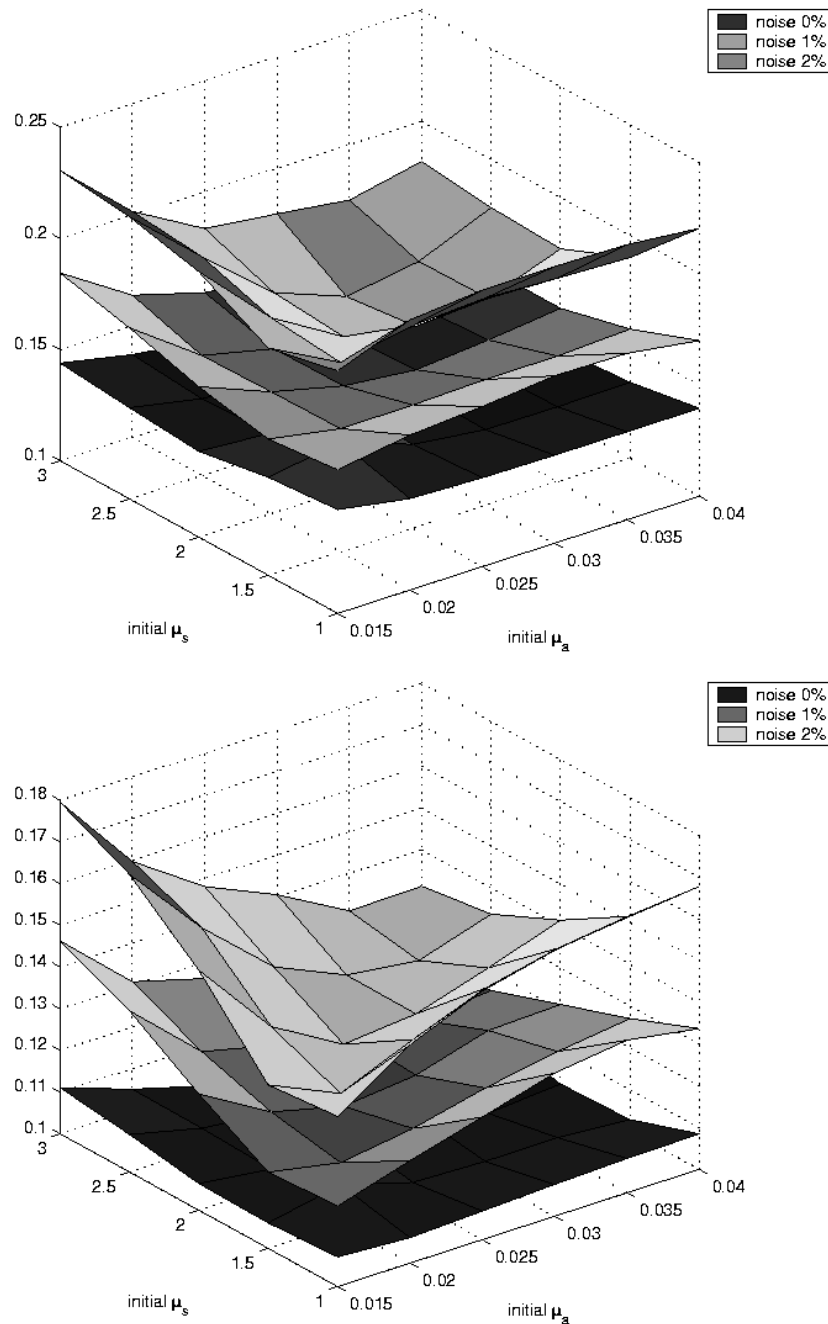
**Figure 8.**  $L_1$  image error  $\varepsilon$  for initial estimate and final reconstruction as a function of homogeneous initial parameters.

The choice of starting values is expected to be of particular importance in the presence of noise. To test the susceptibility of the convergence to data noise, reconstructions from the same range of starting parameters as above were performed, with three different levels of additive random Gaussian noise, at 0%, 1% and 2%. The  $L_1$  error norms of the images of the resulting reconstructions are shown in figure 9, separated for  $\mu_a$  and  $\mu_s$ . As expected, higher noise levels lead to a degradation of the image quality. It is interesting to note that the distribution of residuals shows a characteristic saddle shape, in particular for  $\mu_s$ , indicating that an over- or underestimation of both parameters is more tolerable for the reconstruction than an overestimation of one parameter combined with an underestimation of the other.

### 5.2. Experimental results

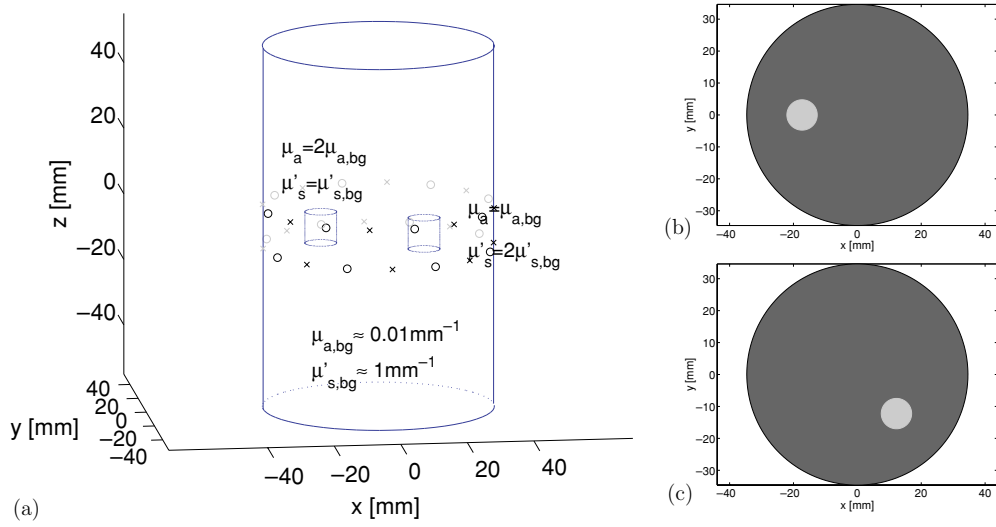
A cylindrical phantom of diameter 69.25 mm and height 110 mm made from epoxy resin was used to acquire data using a frequency-domain instrument developed at Helsinki University of Technology (Nissilä *et al* 2002, Tarvainen *et al* 2005). The phantom uses  $\text{TiO}_2$  particles and an infrared dye to provide scattering and absorption properties similar to that of biological tissue (Firbank and Delpy 1993). The homogeneous optical coefficients of the background material were approximately  $\mu'_s = 1 \pm 0.1 \text{ mm}^{-1}$  and  $\mu_a = 0.01 \pm 0.001 \text{ mm}^{-1}$  at a wavelength 800 nm. The speed of light for the phantom material was  $c = 0.19 \text{ mm ps}^{-1}$ . Two cylindrical inhomogeneities of diameter 9.5 mm and height 9.5 mm were located in the central plane  $z = 0$  of the cylinder. The optical properties of the two targets relative to the background were set to  $(\mu_a, 2\mu'_s)$  and  $(2\mu_a, \mu'_s)$ , respectively. Cross sections of the central plane of the phantom are shown in figure 10.

Optical fibres were used to transmit light from the source to the detectors. In this experiment, 16 source and 16 detector sites arranged in two rings at a spacing of 12 mm were used. Calibrated phase and amplitude data were used to make the reconstructions (Nissilä *et al*

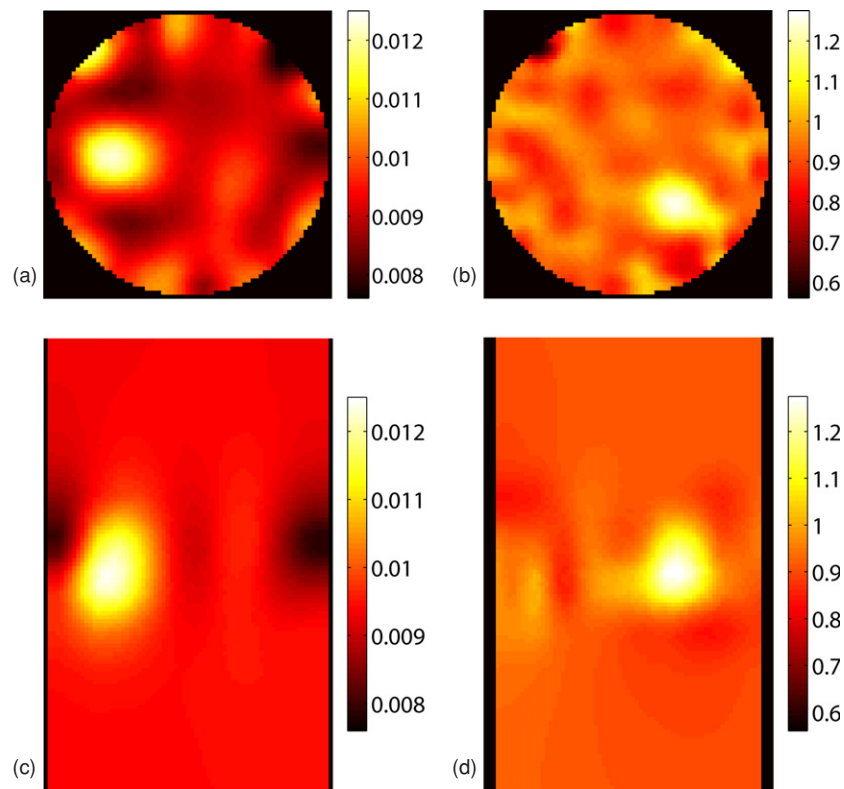


**Figure 9.** Comparison of  $L_1$  image residuals  $\varepsilon_{\mu_a}$  (top) and  $\varepsilon_{\mu_s}$  (bottom) for different data noise levels as a function of initial homogeneous  $\mu_a$  and  $\mu_s$  values.

2005). The homogeneous initial property values in the model and the global amplitude and phase coupling coefficients which link the model and the measurement to each other were determined by a four-parameter search using the measured data. Although there is a small error which is made by the fact that we are fitting a homogeneous model with inhomogeneous data,



**Figure 10.** Geometry of cylindrical phantom with embedded targets, and source and detector locations on the surface (a). Cross sections in the central  $xy$  plane show the position of the absorption (b) and scattering target (c).



**Figure 11.** Reconstruction results: horizontal and vertical cross sections of absorption and scattering images from experimental phantom data. (a)  $\mu_a$  at  $z = 0$ , (b)  $\mu_s$  at  $z = 0$ , (c)  $\mu_a$  at  $y = 0$ , (d)  $\mu_s$  at  $y = -12.24$  mm.

in practice the procedure works reasonably well. Hopefully, this step will be made unnecessary by future developments in the absolute accuracy of the forward model in particular near the source and detector positions. The finite element model used in the reconstruction consisted of an unstructured cylindrical mesh with 63 443 nodes and 336 030 linear tetrahedral elements.

The reconstruction results are shown in figure 11, which represent the fourth iteration of the Gauss–Newton method, taking 1 h on a 3.2 GHz pentium processor with 2 GB RAM. Both the absorbing and scattering object are recovered well in terms of size as well as localization, although the contrast of the reconstructed images is lower than the target. There is very little cross-talk between the absorption and scattering solutions, but some artefacts are evident, in particular close to the boundaries.

## 6. Conclusion

We have presented a Gauss–Newton method for solving the nonlinear inverse problem of image reconstruction in diffuse optical tomography. At each Newton step, an iterative Krylov GMRES method is used to solve the normal equations. This approach allows us to represent the Hessian matrix implicitly by its components, the Jacobian of the forward model and the second derivative of the regularization term. Avoiding an explicit formulation of the Hessian reduces memory requirements, in particular where the dimension of the solution space is large, which is commonly the case in three-dimensional reconstruction problems.

We have compared two strategies for calculating updates at each GN iteration. One strategy holds the step size  $\zeta$  constant, but adjusts the control parameter  $\lambda$  at each step. The other strategy performs a line search for  $\zeta$  at each step, but keeps  $\lambda$  fixed. We have found that both strategies converge to similar final images, but that the line search strategy with  $\lambda = 0$  converges significantly faster.

We have further shown that the convergence rate of the iterative GMRES solver using a moderate stopping criterion of  $\eta = 10^{-3}$  approaches that of a direct Cholesky solution of the explicit Hessian, but without incurring the same runtime and storage penalties as the latter.

We have applied the GN-GMRES algorithm to image reconstruction in the frequency domain, using both a simulated two-dimensional circular test case, and experimental three-dimensional measurements on a cylindrical phantom. We have shown that a simultaneous reconstruction of the spatial distribution of absorption and scattering parameters can be obtained from frequency domain boundary data in both cases, with very good localization and good separation of absorption and scattering features.

## Acknowledgments

This work was supported by the EPSRC-MRC IRC ‘From medical images and signals to clinical information’, EPSRC GR/N14248/01, UK Medical Research Council Grant No D2025/31, EPSRC grants GR/N14248101 and GR/R86201, and the Academy of Finland. A collaborative travel grant was provided by the Wellcome Trust.

## References

- Arnoldi W E 1951 The principle of minimized iterations in the solution of the matrix eigenvalue problem *Q. J. Appl. Math.* **9** 17–29
- Arridge S R 1993 The forward and inverse problems in time-resolved infrared imaging *Medical Optical Tomography: Functional Imaging and Monitoring* ed G Muller, B Chance, R Alfano, S Arridge, J Beuthan, E Gratton, M Kaschke, B Masters, S Svanberg and P van der Zee (Bellingham, WA: SPIE) pp 35–64

- Arridge S R 1995 Photon measurement density functions: Part 1. Analytical forms *Appl. Opt.* **34** 7395–409
- Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** R41–92
- Arridge S R and Hebden J C 1997 Optical imaging in medicine: II. Modelling and reconstruction *Phys. Med. Biol.* **42** 841–53
- Arridge S R and Schweiger M 1998 A gradient-based optimisation scheme for optical tomography *Opt. Express* **2** 213–26 (<http://www.epubs.osa.org/oearchive/source/4014htm>)
- Arridge S R, Schweiger M and Delpy D T 1992 Iterative reconstruction of near-infrared absorption images *Inverse Problems in Scattering and Imaging (Proc. SPIE vol 1767)* ed M A Fiddy pp 372–83
- Arridge S R, Schweiger M, Hiraoka M and Delpy D T 1993 A finite element approach for modeling photon transport in tissue *Med. Phys.* **20** 299–309
- Arridge S R, van der Zee P, Delpy D T and Cope M 1991 Reconstruction methods for infra-red absorption imaging *Time-Resolved Spectroscopy and Imaging of Tissues (Proc. SPIE vol 1431)* ed B Chance and A Katzir pp 204–15
- Barbour R L, Graber H L, Lubowsky J and Aronson R 1990 Model for 3-D optical imaging of tissue *10th Annual IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* vol 2 ed J Ormsby pp 1395–9
- Boas D A, Brooks D H, Miller E L, DiMarzio C A, Kilmer M, Gaudette R J and Zhang Q 2001 Imaging the body with diffuse optical tomography *IEEE Signal Process. Mag.* 57–75
- Dehghani H, Pogue B W, Poplack S P and Paulsen K D 2003 Multiwavelength three-dimensional near-infrared tomography of the breast: initial simulation, phantom, and clinical results *Appl. Opt.* **42**
- Dennis J E and Schnabel R B 1983 *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Englewood Cliffs, NJ: Prentice-Hall)
- Dorn O 1998 A transport-backtransport method for optical tomography *Inverse Problems* **14** 1107–30
- Eppstein M J, Dougherty D E, Hawysz D J and Sevick-Muraca E M 2001 Three-dimensional Bayesian optical image reconstruction with domain decomposition *IEEE Trans. Med. Imaging* **20** 147–63
- Firbank M and Delpy D T 1993 A design for a stable and reproducible phantom for use in near infrared imaging and spectroscopy *Phys. Med. Biol.* **38** 847–53
- Gibson A, Hebden J C and Arridge S R 2005 Recent advances in diffuse optical tomography *Phys. Med. Biol.* **50** R1–43
- Hansen P C and O’Leary D P 1993 The use of the L-curve in the regularization of discrete ill-posed problems *SIAM J. Sci. Comput.* **14** 1487–503
- Hawysz D and Sevick-Muraca E M 2000 Developments towards diagnostic breast cancer imaging using near-infrared optical measurements and fluorescent contrast agents *Neoplasia* **2** 388–417
- Hebden J C, Arridge S R and Delpy D T 1997 Optical imaging in medicine: I. Experimental techniques *Phys. Med. Biol.* **42** 825–40
- Klose A D and Hielscher A H 1999 Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer *Med. Phys.* **26** 1698–707
- Klose A D and Hielscher A H 2003 Quasi-Newton methods in optical tomographic image reconstruction *Inverse Problems* **19** 387–409
- Markel V A and Schotland J C 2001 Inverse problem in optical diffusion tomography: I. Fourier Laplace inversion formulas *J. Opt. Soc. Am. A* **18** 1336–47
- Markel V and Schotland J C 2002 Inverse problem in optical diffusion tomography: II. Role of boundary conditions *J. Opt. Soc. Am. A* **19** 558–66
- Marquardt D W 1963 An algorithm for least-squares estimation of nonlinear parameters *J. SIAM* **11** 431–41
- Moulton J D 1990 Diffusion modelling of picosecond laser pulse propagation in turbid media *M. Eng. Thesis* McMaster University, Hamilton, Ontario
- Nissilä I, Kotilahti K, Fallström K and Katila T 2002 Instrumentation for the accurate measurement of phase and amplitude in optical tomography *Rev. Sci. Instrum.* **73** 3306–12
- Nissilä I, Noponen T, Kotilahti K, Tarvainen T, Schweiger M, Lipiäinen L, Arridge S R and Katila T 2005 Instrumentation and calibration methods for the multichannel measurement of phase and amplitude in optical tomography *Rev. Sci. Instrum.* **76** 044302
- O’Leary M A 1996 Imaging with diffuse photon density waves *PhD Thesis* University of Pennsylvania
- O’Leary M A, Boas D A, Chance B and Yodh A G 1995 Experimental images of heterogeneous turbid media by frequency-domain diffusing-photon tomography *Opt. Lett.* **20** 426–8
- Paulsen K D and Jiang H 1995 Spatially-varying optical property reconstruction using a finite element diffusion equation approximation *Med. Phys.* **22** 691–701
- Ripoll J, Ntziachristos V and Nieto-Vesperinas M 2001 The Kirchoff approximation for diffusive waves *Phys. Rev. E* **64** 1–8
- Roy R and Sevick-Muraca E M 1999 Truncated Newton’s optimization scheme for absorption and fluorescence optical tomography: Part I. Theory and formulation *Opt. Express* **4** 353–71

- Saad Y and Schultz M H 1986 GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems *SIAM J. Sci. Stat. Comput.* **7** 856–69
- Schotland J C 1997 Continuous-wave diffusion imaging *J. Opt. Soc. Am. A* **14** 275–9
- Schotland J C, Haselgrove J C and Leigh J S 1993 Photon hitting density *Appl. Opt.* **32** 448–53
- Schotland J C and Markel V 2001 Inverse scattering with diffusing waves *J. Opt. Soc. Am. A* **18** 2767–77
- Schweiger M and Arridge S R 1999 Application of temporal filters to time resolved data in optical tomography *Phys. Med. Biol.* **44** 1699–717
- Schweiger M and Arridge S R 2003 Image reconstruction in optical tomography using local basis functions *J. Electron. Imaging* **12** 583–93
- Schweiger M, Arridge S R and Delpy D T 1993 Application of the finite-element method for the forward and inverse models in optical tomography *J. Math. Imag. Vision* **3** 263–83
- Schweiger M, Arridge S R, Hiraoka M and Delpy D T 1995 The finite element model for the propagation of light in scattering media: boundary and source conditions *Med. Phys.* **22** 1779–92
- Tarvainen T, Kolehmainen V, Vauhkonen M, Vanne A, Gibson A P, Schweiger M, Arridge S R and Kaipio J P 2005 Computational calibration method for optical tomography *Appl. Opt.* **44** 1879–88
- Vogel C 2002 *Computational Methods for Inverse Problems* (Society for Industrial and Applied Mathematics)
- Walker S A, Fantini S and Gratton E 1997 Image reconstruction by backprojection from frequency-domain optical measurements in highly scattering media *Appl. Opt.* **36** 170–9
- Wang Y, Chang J-H, Aronson R, Barbour R L, Graber H L and Lubowsky J 1992 Imaging of scattering media by diffusion tomography: an iterative perturbation approach *Physiological Monitoring and Early Detection Diagnostic Methods (Proc. SPIE vol 1641)* ed T S Mang pp 58–71
- Yodh A and Chance B 1995 Spectroscopy and imaging with diffusing light *Phys. Today* **48** 34–40
- Ye J C, Bouman C A, Webb K J and Millane R P 2001 Nonlinear multigrid algorithms for bayesian optical diffusion tomography *IEEE Trans. Image Proc.* **10** 909–22
- Zienkiewicz O C and Taylor R L 1987 *The Finite Element Method* 4th edn (London: McGraw-Hill)