INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **21** (2005) 1507–1530 doi:10.1088/0266-5611/21/5/002

Optical tomography as a PDE-constrained optimization problem

Gassan S Abdoulaev[1], Kui Ren[2] and Andreas H Hielscher[3]

[1] Department of Biomedical Engineering, Columbia University, ET351 Mudd Building, MC 8904, 500 West 120th Street, New York, NY 10027, USA
[2] Department of Applied Physics and Applied Mathematics, Columbia University, 200 Mudd Building, 500 West 120th Street, New York, NY 10027, USA
[3] Departments of Biomedical Engineering and Radiology, Columbia University, ET351 Mudd Building, MC 8904, 500 West 120th Street, New York, NY 10027, USA

E-mail: gsa2001@columbia.edu, kr2002@columbia.edu and ahh2004@columbia.edu

Received 6 December 2004, in final form 13 May 2005
Published 23 August 2005
Online at stacks.iop.org/IP/21/1507

Abstract

We report on the implementation of an augmented Lagrangian approach for solving the inverse problems in diffuse optical tomography (DOT). The forward model of light propagation is the radiative transport equation (RTE). The inverse problem is formulated as a minimization problem with the RTE being considered as an equality constraint on the set of 'optical properties—radiance' pairs. This approach allows the incorporation of the recently developed technique of PDE-constrained optimization, which has shown great promise in many applications that can be formulated as infinite-dimensional optimization problems. Compared to the traditional unconstrained optimization approaches for optical tomographic imaging where one solves several forward and adjoint problems at each optimization iteration, the method proposed in this work solves the forward and inverse problems simultaneously. We found in initial studies, using synthetic data, that the image reconstruction time can typically be reduced by a factor of 10 to 30, which depends on a combination of noise level, regularization parameter, mesh size, initial guess, optical properties and system geometry.

1. Introduction

Optical tomography (OT) is a fast developing area of medical imaging that provides a number of challenging problems to researchers from the engineering, physics and mathematics communities [3, 4, 8, 12, 21, 27, 34, 65]. In OT near-infrared light (wavelength from 700 to 900 nm) is used to illuminated the tissue and measurements of transmitted and reflected light are taken. From these surface measurements one attempts to determine the distribution of absorption and scattering coefficients (μ_a and μ_s, respectively) inside the tissue [4]. Tissues

such as fat, muscle, white matter, grey matter, etc, have different values of μ_a and μ_s, which gives rise to the contrast of this imaging modality. Moreover, the absorption coefficients of oxygenated and deoxygenated blood are different, which allows performing spatially dependent blood oximetry. Clinical applications of this technology include, for instance, brain imaging [1, 8, 11, 13, 32, 62], breast imaging [37, 53, 66], and imaging of finger joints [35, 60, 70].

One of the major challenges of OT is the development of efficient numerical algorithms for solving both the forward and inverse light propagation problems. In general, the propagation of the near-infrared light in tissue can be modelled by a set of equations and boundary conditions that can be written in the following abstract form,

$$\mathcal{G}(\boldsymbol{\mu}, \psi) = 0, \qquad (\boldsymbol{\mu}, \psi) \in \mathcal{U} \times \mathcal{W}, \tag{1}$$

where $\boldsymbol{\mu} \equiv (\mu_a, \mu_s)$ denotes the spatially dependent optical properties of the tissue, and ψ denotes the intensity (or fluence) distribution of light in the tissue. The operator $\mathcal{G}(\boldsymbol{\mu}, \psi)$ should be understood as an operator from $\mathcal{U} \times \mathcal{W}$ to $\mathcal{V}$, where $\mathcal{U}$ and $\mathcal{W}$ denote the Banach spaces of all admissible optical properties $\boldsymbol{\mu}$ and solutions ψ of the forward problem (1), respectively, and $\mathcal{V}$ is another Banach space.

Models most often used are either the radiative transport equation (RTE) [14, 18], which is believed to be a very accurate model of light propagation; or the diffusion equation [4], which is an approximation of the RTE. In general, the diffusion equation is much easier to solve than the RTE. However, diffusion equations are only valid in scattering-dominated ($\mu_s \gg 1 \gg \mu_a$) media. For tissues with non-scattering regions present or tissues of small volume, the diffusion approximation fails to describe light propagation accurately. Detailed discussion on the RTE and its diffusion approximation can be found in classical references such as [14, 18] and more recent papers concerning problems in optical tomography [20, 25, 31].

The measured data (in data space $\mathcal{Z}$) in OT is usually a bounded linear functional of ψ, $\mathcal{M}\psi$, where $\mathcal{M} : \mathcal{W} \mapsto \mathcal{Z}$ is called a measurement operator. Note that $\mathcal{M}$ must be introduced because in practice only partial information about the forward solution, ψ, can be measured [4]. We will provide in section 2 a specific form of such an operator, but remark here that when ψ can be measured directly, $\mathcal{M}$ is nothing but an identity operator. The objective of OT is to transform the measured data into accurate approximation of the spatial distribution of optical properties $\boldsymbol{\mu}$ inside the tissue. Due to the lack of analytical inversion formulae, this transformation is usually done through numerical optimization [4, 5]. The guiding principle of the optimization approach to the inverse problems in OT is to iteratively update the model parameter $\boldsymbol{\mu}$ in such a way that the forward model (1) generates sets of data (predictions) that match measurements with higher and higher accuracy. To be more specific, let us first denote by $M \in \mathcal{Z}$ given experimental measurements. We then introduce a real-valued non-negative function, called the objective function, $\phi : \mathcal{U} \times \mathcal{W} \longmapsto \mathbb{R}^+$,

$$\phi(\boldsymbol{\mu}, \psi) = \frac{\alpha}{2}\|\mathcal{M}\psi - M\|^2_{\mathcal{Z}} + \frac{\beta}{2}\mathcal{R}(\boldsymbol{\mu}), \tag{2}$$

which has to be minimized to generate the inverse solution, for some positive α and β. The first term in (2) is used to quantify the difference between the measurements and model predictions, while the second term, $\mathcal{R}(\boldsymbol{\mu})$, which is called a regularization term, is used to impose additional constraints on the parameter $\boldsymbol{\mu}$ [22]. The optical tomographic imaging can now be formulated as the following equation-constrained optimization problem:

$$\begin{aligned} &\min_{(\boldsymbol{\mu},\psi)} \phi(\boldsymbol{\mu}, \psi) \\ &\text{subject to (1).} \end{aligned} \tag{3}$$

Note that we do not consider here further constraints on the optical property $\boldsymbol{\mu}$, such as simple-bound constraints specifying upper and lower bounds for the optical properties, which

might be added to (3). We refer interested readers to references [33, 54, 58, 59] for detailed discussions on that.

There are several techniques to solve the problem (3) [49]. All existing algorithms in the optical tomography community convert (3) into an unconstrained optimization problem. Employing this approach, one first solves equation (1) to obtain ψ as a function of $\boldsymbol{\mu}$, $\psi = \psi(\boldsymbol{\mu})$, and then uses this expression to eliminate the explicit dependence of the objective function (2) on function ψ, giving rise to another function that should be minimized only with respect to the optical property $\boldsymbol{\mu}$, $\widetilde{\phi} : \mathcal{U} \longmapsto \mathbb{R}^+$:

$$\widetilde{\phi}(\boldsymbol{\mu}) \equiv \phi(\boldsymbol{\mu}, \psi(\boldsymbol{\mu})). \tag{4}$$

To minimize $\widetilde{\phi}(\boldsymbol{\mu})$, one can use methods such as conjugate gradient and quasi-Newton methods; see [1, 4, 12, 23, 39, 40, 57–59, 61, 63] for details of the implementation of algorithms based on this unconstrained approach, as applied to optical tomography.

One drawback of the above-mentioned method is that for each evaluation of the function $\widetilde{\phi}(\boldsymbol{\mu})$ during the minimization process, one has to solve the forward problem (1) to obtain $\psi(\boldsymbol{\mu})$. Therefore each inversion procedure will require solving the forward problem hundreds of times, which requires extensive computing times.

In this work, we adopt a numerical method that, instead of solving the unconstrained optimization problem (4), solves the constrained optimization problem (3) directly. To do that, we introduce the following Lagrangian functional for problem (3), $\mathcal{L} : \mathcal{U} \times \mathcal{W} \times \mathcal{V}^* \mapsto \mathbb{R}$,

$$\mathcal{L}(\boldsymbol{\mu}, \psi; \lambda) = \phi(\boldsymbol{\mu}, \psi) - \langle \lambda, \mathcal{G}(\boldsymbol{\mu}, \psi) \rangle, \tag{5}$$

where $\lambda \in \mathcal{V}^*$ is the Lagrange multiplier, $\mathcal{V}^*$ is the dual space of $\mathcal{V}$, and $\langle \cdot, \cdot \rangle$ denotes the duality pair between $\mathcal{V}$ and $\mathcal{V}^*$. From the theory of constrained optimization it is known that the solution to (3) satisfies the following optimality condition for $\mathcal{L}$ [9, 49]:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}^*, \psi^*; \lambda^*) = 0, \qquad \frac{\partial \mathcal{L}}{\partial \psi}(\boldsymbol{\mu}^*, \psi^*; \lambda^*) = 0, \qquad \mathcal{G}(\boldsymbol{\mu}^*, \psi^*) = 0. \tag{6}$$

The system (6) is also known as the Karush–Kuhn–Tacker (KKT) condition [49]. Now the optimum point of the problem (3) can be found by solving this KKT system. This approach is often referred to as the *all-at-once* method because it solves the forward and minimization problems simultaneously as compared to the unconstrained optimization approach.

This kind of constrained optimization technique promises a much faster solution of the inverse problem and has already proved to be very useful in other applications such as shape design in aerodynamics [51, 64] and optimal control of incompressible flows [26]. (See also [10] for a general overview). However, implementation of this PDE-constrained optimization approach for DOT has never been reported to the best of our knowledge.

The remainder of the paper is structured as follows. After briefly introducing numerical schemes to discretize the radiative transport equation (7) in section 2, we describe in detail the implementation of an augmented Lagrangian method to solve the optimality condition (6) in section 3. This method was introduced by Hestenes [30] and Powell [55] and proved to be efficient in other applications. We then present some numerical reconstruction results from synthetic data and compare the performance of our method to that of an quasi-Newton unconstrained minimization method in section 4. Conclusions are drawn in section 5.

2. The forward problem

We consider the frequency-domain version of the radiative transport equation as the forward model of light propagation in tissues. In the case where time harmonic sources are used, this

model can be written for a fixed modulation frequency ω as [56],

$$\left(-\frac{\mathrm{i}\omega}{v} + \boldsymbol{\Omega}\cdot\nabla + \mu_t(\mathbf{x})\right)\psi(\mathbf{x},\boldsymbol{\Omega}) = \mu_s(\mathbf{x})\int_{\mathbf{S}^2}\mathbf{k}(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}')\psi(\mathbf{x},\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}', \quad \text{in} \quad X = \mathcal{D}\times S^2, \tag{7}$$

where $\psi(\mathbf{x},\boldsymbol{\Omega})$ is the radiance at location $\mathbf{x}\in\mathcal{D}\subset\mathbb{R}^3$ and in direction $\Omega\in S^2$ (unit sphere in $\mathbb{R}^3$). Here we drop the explicit dependence of ψ on the modulation frequency ω. The function $\mu_t(\mathbf{x}) = \mu_{\mathbf{a}}(\mathbf{x}) + \mu_{\mathbf{s}}(\mathbf{x})$ is a total attenuation, with $\mu_a(\mathbf{x})$ and $\mu_s(\mathbf{x})$ the absorption and scattering coefficients, respectively. The function $k(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}')$ is a scattering kernel that describes a probability that photons propagating in direction Ω' are scattered into direction Ω. The angular element $\mathrm{d}\Omega'$ is normalized, $\int_{\mathbf{S}^2}\mathrm{d}\boldsymbol{\Omega}' = 1$.

In this work, we use the Henyey–Greenstein phase function [29]

$$k(\cos\theta) = \frac{1-g^2}{(1+g^2-2g\cos\theta)^{3/2}},$$

where $\cos\theta = \boldsymbol{\Omega}\cdot\boldsymbol{\Omega}'$. This phase function has been used in many cases to describe the scattering process in tissues.

We supply equation (7) with a boundary condition

$$\psi(\mathbf{x},\boldsymbol{\Omega}) = f(\mathbf{x},\boldsymbol{\Omega}), \quad \text{on } \Gamma_- = \{(\mathbf{x},\boldsymbol{\Omega})\in\partial\mathcal{D}\times\mathbf{S}^2 \text{ s.t. } \boldsymbol{\Omega}\cdot\mathbf{n}(\mathbf{x}) < 0\}, \tag{8}$$

where $\mathbf{n}(\mathbf{x})$ is the unit outer normal vector at the boundary point $\mathbf{x}\in\partial\mathcal{D}$, and $f(\mathbf{x},\boldsymbol{\Omega})$ represents a modulated source. The forward problem of optical tomography thus consists in computing radiance distribution $\psi(\mathbf{x},\boldsymbol{\Omega})$ for given optical properties and a source.

The reason for using the frequency domain version of the RTE, as in the form of (7), is that frequency domain measurements contain more information than measurements in the stationary case. To be more specific, let us define the measurement operator in DOT. It is a linear operator that maps the forward solution of (7) to the boundary photon current measurement,

$$(\mathcal{M}\psi)(\mathbf{x}) = \int_{\mathbf{S}^2_+}\boldsymbol{\Omega}\cdot\mathbf{n}(\mathbf{x})\psi(\mathbf{x},\boldsymbol{\Omega})\,\mathrm{d}\boldsymbol{\Omega}, \tag{9}$$

where $S^2_+ = \{\boldsymbol{\Omega} : \boldsymbol{\Omega}\in S^2$ such that $\boldsymbol{\Omega}\cdot\mathbf{n}(\mathbf{x}) > 0\}$. Because $\psi(\mathbf{x},\boldsymbol{\Omega})$ is a complex function, $\mathcal{M}\psi$ is thus a complex functional. The measurement thus contains both intensity and phase information.

The numerical solution of the RTE (7) is extremely expensive because the equation is posed in phase space. One thus needs to discretize in both angular and space variables. Let us first introduce a quadrature formula for the integration of a function $g(\boldsymbol{\Omega})$ on the unit sphere S^2,

$$\int_{S^2} g(\boldsymbol{\Omega})\,\mathrm{d}\boldsymbol{\Omega} \approx \sum_{l=1}^{L} w_l g(\Omega_l), \tag{10}$$

where w_l is the quadrature weight in direction Ω_l, and satisfies the following normalization condition:

$$\sum_{l=1}^{L} w_l = 1.$$

We can then approximate the RTE (7) by the following coupled system of partial differential equations,

$$-\frac{\mathrm{i}\omega}{v}\psi^l(\mathbf{x}) + \Omega_l \cdot \nabla\psi^l(\mathbf{x}) + \mu_t(\mathbf{x})\psi^l(\mathbf{x}) = \mu_s(\mathbf{x})\sum_{l'=1}^{L} w_{l'}k_{l,l'}\psi^{l'}(\mathbf{x}), \qquad l = 1,\ldots,L \tag{11}$$

where $\psi^l(\mathbf{x}) \equiv \psi(\mathbf{x}, \Omega_l)$, $k_{l,l'} \equiv k(\Omega_l \cdot \Omega_{l'})$. System (11) is called the discrete ordinate formulation of the RTE, which has been studied extensively in the past decades [2, 46].

We can separate the real and imaginary parts of the system (11) to obtain

$$\begin{aligned}
\Omega_l \cdot \nabla\psi^l_{\mathrm{Re}} + \mu_t\psi^l_{\mathrm{Re}} + \frac{\omega}{v}\psi^l_{\mathrm{Im}} &= \mu_s\sum_{l'=1}^{L} w_{l'}k_{l,l'}\psi^{l'}_{\mathrm{Re}} \\
\Omega_l \cdot \nabla\psi^l_{\mathrm{Im}} + \mu_t\psi^l_{\mathrm{Im}} - \frac{\omega}{v}\psi^l_{\mathrm{Re}} &= \mu_s\sum_{l'=1}^{L} w_{l'}k_{l,l'}\psi^{l'}_{\mathrm{Im}}
\end{aligned} \qquad l = 1,\ldots,L \tag{12}$$

where we use the representation $\psi^l = \psi^l_{\mathrm{Re}} + \mathrm{i}\psi^l_{\mathrm{Im}}$.

To discretize the system (12) spatially we use the finite-volume method [24, 56], which is a conservative scheme on a discrete level, and allows approximation of problems on domains of complex and irregular geometries. We cover the computational domain by a mesh $\mathcal{D}_h$ of small polyhedrons (finite volumes), for instance, tetrahedrons (figures 2(b) and 10(b)).

Integrating (12) over a tetrahedron E and using the divergence theorem we obtain

$$\begin{aligned}
\int_{\partial E} \Omega_l \cdot \mathbf{n}_{\partial E}\psi^l_{\mathrm{Re}}\,\mathrm{d}\sigma + \int_E \left(\mu_{\mathbf{t}}\psi^l_{\mathrm{Re}} + \frac{\omega}{v}\psi^l_{\mathrm{Im}}\right)\mathrm{d}\mathbf{x} &= \int_E \mu_{\mathbf{s}}\sum_{l'=1}^{L} w_{l'}k_{l,l'}\psi^{l'}_{\mathrm{Re}}\,\mathrm{d}\mathbf{x}, \\
\int_{\partial E} \Omega_l \cdot \mathbf{n}_{\partial E}\psi^l_{\mathrm{Im}}\,\mathrm{d}\sigma + \int_E \left(\mu_{\mathbf{t}}\psi^l_{\mathrm{Im}} - \frac{\omega}{v}\psi^l_{\mathrm{Re}}\right)\mathrm{d}\mathbf{x} &= \int_E \mu_{\mathbf{s}}\sum_{l'=1}^{L} w_{l'}k_{l,l'}\psi^{l'}_{\mathrm{Im}}\,\mathrm{d}\mathbf{x},
\end{aligned} \tag{13}$$

for $l = 1,\ldots,L$, where $\mathbf{n}_{\partial E}$ is an outward normal vector at point $\mathbf{x} \in \partial E$, $\mathrm{d}\sigma(\mathbf{x})$ is the Lebesgue surface measure of ∂E.

To get the finite-volume discretization equations, we make the hypothesis that each of the finite volumes is small enough, so that we can assume that the values of all space-dependent functions are constant in a volume. We also have to approximate the flux terms, i.e., the terms involving boundary integrals. In fact, this approximation will determine the order of accuracy of the finite-volume scheme. Here we adopt a first-order upwind scheme [56],

$$\int_{\partial E} \Omega_l \cdot \mathbf{n}_{\partial E}\psi^l_{\mathrm{Re}}\,\mathrm{d}\sigma = \sum_m \Omega_l \cdot \mathbf{n}_{\tau_m}|\tau_m|\left[\psi^l_{\mathrm{Re}}\right]_{\tau_m}, \tag{14}$$

with

$$\left[\psi^l_{\mathrm{Re}}\right]_{\tau_m} = \begin{cases} \left(\psi^l_{\mathrm{Re}}\right)_E, & \Omega_l \cdot \mathbf{n}_{\tau_m} \geqslant 0 \\ \left(\psi^l_{\mathrm{Re}}\right)_{E'}, & \Omega_l \cdot \mathbf{n}_{\tau_m} < 0, \end{cases} \tag{15}$$

where $\tau_m \subset \partial E$ is the mth triangle that forms ∂E with measure $|\tau_m|$. The vector $\mathbf{n}_{\tau_m}$ is the unit outer normal vector on τ_m, and E' is the neighbour volume that shares a common face τ_m with E. If the face τ_m happens to be part of the boundary of the computational domain, $\tau_m \subset \partial\mathcal{D}_h$, and $\Omega_l \cdot \mathbf{n}_{\tau_m} < 0$, then the boundary condition (8) requires that $\left[\psi^l_{\mathrm{Re}}\right]_{\tau_m} = f_{\mathrm{Re}}(\mathbf{x}, \boldsymbol{\Omega}_l)|_{\tau_m}$.

The reason for choosing such a first-order scheme is that it preserves the positivity of the solution of the transport equation in the case when $\omega = 0$, which is important in optical tomographic imaging.

After both space and angular discretization, we arrive at an algebraic system for the equation of radiative transfer:

$$G(\Sigma, \Psi) \equiv \mathbf{T}(\Sigma)\Psi - \mathbf{S} = 0. \tag{16}$$

Here $\mathbf{T}(\Sigma) \in \mathbb{R}^{2LN} \times \mathbb{R}^{2LN}$ is an $2L \times 2L$-block matrix that denotes the discretized transport operator, $\Psi \in \mathbb{R}^{2LN}$ is a vector of the discretized radiance arranged as $\left(\left(\psi^1_{\mathrm{Re}}\right)^T, \ldots, \left(\psi^L_{\mathrm{Re}}\right)^T, \left(\psi^1_{\mathrm{Im}}\right)^T, \ldots, \left(\psi^L_{\mathrm{Im}}\right)^T\right)^T$. $\Sigma = \left(\Sigma_a^T, \Sigma_s^T\right)^T \in \mathbb{R}^{2N}$ represents the values of optical properties μ_a and μ_s on mesh volumes, where N is the total number of volumes of the mesh. Matrix $\mathbf{T}$ has the following block structure:

$$\mathbf{T} = \begin{pmatrix} \mathbf{B} & \mathbf{D} \\ -\mathbf{D} & \mathbf{B} \end{pmatrix} + \begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}. \tag{17}$$

Matrices $\mathbf{B}$ and $\mathbf{D}$ are block-diagonal,

$$\mathbf{B} = \mathrm{diag}\{\mathbf{B}_1, \ldots, \mathbf{B}_L\}, \qquad \mathbf{D} = \mathrm{diag}\{\mathbf{D}_0, \ldots, \mathbf{D}_0\},$$

where matrix B_l approximates the streaming operator $\mathcal{B}_l\psi = \Omega_l \cdot \nabla\psi + \mu_t\psi$, matrix D_0 is diagonal, and $[D_0]_{ii} = V_i\omega/v$, V_i denotes the volume of the ith tetrahedron (finite volume) of the mesh. Matrix $\mathbf{C}$ is an $L \times L$-block matrix,

$$\mathbf{C} = \begin{pmatrix} C_{11} & \cdots & C_{1L} \\ \vdots & \ddots & \vdots \\ C_{L1} & \cdots & C_{LL} \end{pmatrix},$$

with blocks C_{lm} that are diagonal matrices, and $[C_{lm}]_{ii} = -V_i[\Sigma_s]_i\omega_m k_{lm}$. Throughout the paper $[X]_i$ (respectively $[X]_{ij}$) denotes the ith (ijth) entry of the vector (respectively matrix) X.

3. The inverse problem

As we have mentioned before, the objective of optical tomographic imaging is to reconstruct parameter $\boldsymbol{\mu}$ through boundary measurements of outgoing photon current. This is very similar to many other inverse problems such as optimal control [26], shape optimization [64], and parameter identification for partial differential systems [28]. In all those problems, there are forward models that can be written in the form of (1), with a model parameter $\boldsymbol{\mu}$ that is sought in inverse problems. Usually, the inversion can only be done by using numerical optimization techniques. This kind of problem is often called simulation-based optimization in the literature. In the case when the forward model is a partial differential equation, the term *PDE-constrained optimization* is used instead [10]. For the OT problem, the integro-differential radiative transport equation can be written as a system of coupled first-order partial differential equations after passing to the discrete ordinate formulation (see (11)), we can thus put our inverse problem in a general PDE-constrained optimization framework.

Usually in optical tomographic imaging, a set of measurements with different source and detector pairs are performed to collect as much information as possible for the reconstruction. We thus denote by N_{src} the total number of light sources and denote by N_j^{det} the number of detector readings corresponding to source S_j. The solution of the forward problem (1) corresponding to S_j is denoted by Ψ_j and the set of solutions for all sources is denoted by $\widehat{\Psi} \equiv \left(\Psi_1, \ldots, \Psi_{N_{\mathrm{src}}}\right)$. The set of measurements is denoted by $\mathbf{M} \equiv \left\{M_{jd}, d = 1, \ldots, N_j^{\mathrm{det}}, j = 1, \ldots, N_{\mathrm{src}}\right\}$. Finally, for the detector located at a mesh node $\mathbf{x}_d \in \partial D$, we denote by $P_d : \mathbb{R}^{2LN} \mapsto \mathbb{C}$ the discretized version of the measurement operator (9) which measures the outgoing photon flux at $\mathbf{x_d}$.

With all this notation, we are ready to introduce the objective function to be minimized,

$$\Phi(\Sigma, \widehat{\Psi}) = \frac{1}{2}\sum_{j=1}^{N_{\mathrm{src}}}\sum_{d=1}^{N_j^d}\frac{|P_d\Psi_j - M_{jd}|^2}{|M_{jd}|^2} + \frac{\beta}{2}R(\Sigma), \tag{18}$$

where β is the regularization parameter that will be used later. $R(\Sigma)$ is the discretized version of the following Tikhonov regularization functional [22, 67],

$$\mathcal{R}(\boldsymbol{\mu}) = \|\mu_a\|^2_{\mathcal{H}^1(\mathcal{D})} + \|\mu_s\|^2_{\mathcal{H}^1(\mathcal{D})}, \tag{19}$$

with

$$\|p\|^2_{\mathcal{H}^1(\mathcal{D})} = \int_{\mathcal{D}} (p^2 + |\nabla p|^2)\, \mathrm{d}\mathbf{x}, \qquad \forall\, \mathrm{p} \in \mathcal{H}^1(\mathcal{D}). \tag{20}$$

The choice of a strong $\mathcal{H}^1$ (rather than L^2) norm in the regularization term is necessary since the OT problem is in general severely ill-posed [4, 6]. One usually does not expect to recover high-frequency components of the optical properties. $\mathcal{H}^1$ norms help to suppress those high-frequency modes. See [33, 52, 54, 62, 68] for various types of regularization used in optical tomographic imaging.

It is important to note that the objective function Φ takes into account solutions Ψ_j of the forward problem for all N_{src} sources simultaneously. Also one has to pay special attention to the discretization of the regularizing term $\mathcal{R}(\boldsymbol{\mu})$ in the case when an unstructured mesh has been used to partition the computational domain because the gradient of μ_a and μ_s has to be calculated properly. Here we approximate the gradient of μ_a (the same for μ_s) on control volume E based on the Green–Gauss formulation [7]

$$(\nabla \mu_a)_E = \frac{1}{V_E} \sum_m \mu_a^m |\tau_m| \mathbf{n}_{\tau_m}, \qquad \mu_a^m = \frac{(\mu_a)_E + (\mu_a)_{E'}}{2}, \tag{21}$$

where, as before, V_E is the volume of E, $\tau_m \subset \partial E$ is the mth triangle that forms ∂E. $\mathbf{n}_{\tau_m}$ is the unit outer normal of τ_m, and E' is the neighbour volume that shares a common face τ_m with E.

On the discretized level, OT can be formulated as a minimization problem subject to the constraints (16) for N_{src} different sources:

$$\min_{(\Sigma, \widehat{\Psi}) \in \mathbb{R}^{2N} \times \mathbb{R}^{2LN}} \Phi(\Sigma, \widehat{\Psi}), \quad \text{subject to} \quad G_j(\Sigma, \Psi_j) = 0, \quad j = 1, \ldots, N_{\mathrm{src}}. \tag{22}$$

Note that each discrete constraint $G_j(\Sigma, \Psi_j) = 0$ binds the vector of optical properties Σ and the vector of fluences Ψ_j for a particular source S_j. The overall number of discrete constraints is (number of sources) $\times$ (number of ordinates) $\times$ (number of finite volumes) $\times$ 2.

3.1. An augmented Lagrangian approach

We introduce a Lagrangian function $\mathcal{L} : \mathbb{R}^{2N} \times \mathbb{R}^{2LNN_{\mathrm{src}}} \times \mathbb{R}^{2LNN_{\mathrm{src}}} \mapsto \mathbb{R}$ defined by

$$\mathcal{L}(\Sigma, \widehat{\Psi}; \widehat{\lambda}) = \Phi(\Sigma, \widehat{\Psi}) - \sum_{j=1}^{N_{\mathrm{src}}} \lambda_j^T G_j(\Sigma, \Psi_j), \qquad \lambda_j \in \mathbb{R}^{2LN} \tag{23}$$

where $\widehat{\lambda} \equiv (\lambda_1, \ldots, \lambda_{N_{\mathrm{src}}})$. The solution to the optimization problem (22) satisfies the KKT optimality condition with (23), which is given by the following system:

$$\begin{aligned}
\nabla_\Sigma \mathcal{L}(\Sigma, \widehat{\Psi}; \widehat{\lambda}) &= \nabla_\Sigma \Phi(\Sigma, \widehat{\Psi}) - \sum_{j=1}^{N_{\mathrm{src}}} \lambda_j^T \nabla_\Sigma G_j(\Sigma, \Psi_j) = 0, \\
\nabla_{\widehat{\Psi}} \mathcal{L}(\Sigma, \widehat{\Psi}; \widehat{\lambda}) &= \nabla_{\widehat{\Psi}} \Phi(\Sigma, \widehat{\Psi}) - \sum_{j=1}^{N_{\mathrm{src}}} \lambda_j^T \nabla_\Psi G_j(\Sigma, \Psi_j) = 0, \\
\nabla_{\lambda_j} \mathcal{L}(\Sigma, \widehat{\Psi}; \widehat{\lambda}) &= G_j(\Sigma, \Psi_j) = 0, \qquad j = 1, \ldots, N_{\mathrm{src}}.
\end{aligned} \tag{24}$$

In other words, if $(\Sigma^*, \widehat{\Psi}^*)$ provides an optimal solution of (22), then there exist $\lambda_j^* \in \mathbb{R}^{2LN}$, $j = 1, \ldots, N_{\mathrm{src}}$, such that $(\Sigma^*, \widehat{\Psi}^*, \widehat{\lambda}^*)$ is a stationary point of the Lagrangian function (23).

There exist several methods to find the stationary point of the Lagrangian function. For example, one can solve system (24) directly if an efficient nonlinear algebraic solver is available [10]. Here, we propose to use an augmented Lagrangian method (ALM), which is an iterative method easy to implement. The method defines an augmented Lagrangian function by

$$\mathcal{L}_K(\Sigma, \widehat{\Psi}; \widehat{\lambda}) = \mathcal{L}(\Sigma, \widehat{\Psi}; \widehat{\lambda}) + \frac{1}{2K} \sum_{j=1}^{N_{\text{src}}} \|G_j(\Sigma, \Psi_j)\|_{l^2}^2, \tag{25}$$

where the term $(1/2K) \sum_{j=1}^{N_{\text{src}}} \|G_j(\Sigma, \Psi_j)\|_{l^2}^2$ is a penalty for violating the constraints $G_j(\Sigma, \Psi_j) = 0$, $j = 1, \ldots, N_{\text{src}}$. In the augmented Lagrangian method, we look for a stationary point of the function $\mathcal{L}_K$ instead of $\mathcal{L}$. Note that in the limit, when the constraints are satisfied exactly, $\mathcal{L}_K = \mathcal{L}$.

Assume that at the kth iteration of the algorithm we have an approximation $(\Sigma^k, \widehat{\Psi}^k, \widehat{\lambda}^k)$ to the stationary point $(\Sigma^*, \widehat{\Psi}^*, \widehat{\lambda}^*)$ of the Lagrangian function. Let us fix the current estimates of the Lagrangian multipliers $\widehat{\lambda}^k$ and a penalty parameter K_k. Minimization of $\mathcal{L}_{K_k}(\Sigma, \widehat{\Psi}; \widehat{\lambda}^k)$ with respect to Σ and $\widehat{\Psi}$ yields the following system:

$$\begin{aligned}
&\nabla_\Sigma \Phi(\Sigma, \widehat{\Psi}) - \sum_{j=1}^{N_{\text{src}}} \left[\left(\lambda_j^k\right)^T - \frac{1}{K_k} G_j^T(\Sigma, \Psi_j) \right] \nabla_\Sigma G_j(\Sigma, \Psi_j) = 0, \\
&\nabla_{\widehat{\Psi}} \Phi(\Sigma, \widehat{\Psi}) - \sum_{j=1}^{N_{\text{src}}} \left[\left(\lambda_j^k\right)^T - \frac{1}{K_k} G_j^T(\Sigma, \Psi_j) \right] \nabla_{\widehat{\Psi}} G_j(\Sigma, \Psi_j) = 0.
\end{aligned} \tag{26}$$

Suppose that $(\Sigma^k, \widehat{\Psi}^k)$ is an approximate minimizer of $\mathcal{L}_{K_k}(\Sigma, \widehat{\Psi}; \widehat{\lambda}^k)$, i.e. an approximate solution of system (26). Comparing this system with the optimality condition (24), we conclude that $\lambda_j^k - (1/K_k) G_j\left(\Sigma^k, \Psi_j^k\right)$ approximates λ_j^*:

$$\lambda_j^* \approx \lambda_j^k - (1/K_k) G_j\left(\Sigma^k, \Psi_j^k\right), \qquad j = 1, \ldots, N_{\text{src}}. \tag{27}$$

This formula can be rearranged to produce an estimate of $G_j\left(\Sigma, \Psi_j^k\right)$:

$$G_j\left(\Sigma^k, \Psi_j^k\right) \approx K_k\left(\lambda_j^k - \lambda_j^*\right). \tag{28}$$

Hence, we deduce that if λ_j^k is close to the optimal Lagrangian multiplier λ_j^*, and K_k is small enough, then the pair $(\Sigma^k, \widehat{\Psi}^k)$ satisfies the corresponding constraint with a high accuracy. Formula (27) prompts a rule for iterative updating of the Lagrangian multipliers:

$$\lambda_j^{k+1} = \lambda_j^k - (1/K_k) G_j\left(\Sigma^k, \Psi_j^k\right), \qquad j = 1, \ldots, N_{\text{src}}. \tag{29}$$

We thus arrive at the following augmented Lagrangian algorithm [17, 49].

Augmented Lagrangian method

- Choose positive constants $\overline{\eta}, \overline{\varepsilon}, \overline{K} \leqslant 1, \tau < 1, \overline{\gamma} \leqslant 1, \alpha_\varepsilon, \beta_\varepsilon, \alpha_\eta, \beta_\eta, \alpha_*, \beta_*$, such that $\alpha_\eta < \min(1, \alpha_\varepsilon)$, $\beta_\eta < \min(1, \beta_\varepsilon)$;
- Choose initial Lagrange multiplier λ^0;
- Choose maximum number of iterations $k_{\max}$;
- Set $K_0 = \overline{K}$, $\alpha_0 = \min(K_0, \overline{\gamma})$, $\varepsilon_0 = \overline{\varepsilon}(\alpha_0)^{\alpha_\varepsilon}$, $\eta_0 = \overline{\eta}(\alpha_0)^{\alpha_\eta}$;
- **for** $k = 0, 1, 2, \ldots, k_{\max}$
  1. Find the minimizer $(\Sigma^k, \widehat{\Psi}^k)$ of $\mathcal{L}_{K_k}(\Sigma, \widehat{\Psi}; \widehat{\lambda}^k)$ by an iterative method that
     - Starts from initial guess $(\widetilde{\Sigma}^k, \widetilde{\Psi}^k)$;
     - Terminates when $\left\| \nabla_{(\Sigma, \widehat{\Psi})} \mathcal{L}_{K_k}(\Sigma^k, \widehat{\Psi}^k; \widehat{\lambda}^k) \right\|_{l^2} \leqslant \varepsilon_k$ is satisfied;
  2. **if** $\sum \left\| G_j\left(\Sigma^k, \Psi_j^k\right) \right\|_{l^2} \leqslant \eta_k$

– Test convergence:
if $\sum \left\| G_j\left(\Sigma^k, \Psi_j^k\right)\right\|_{l^2} \leqslant \eta_*$ and $\left\|\nabla_{(\Sigma,\widehat{\Psi})}\mathcal{L}_{K_k}(\Sigma^k, \widehat{\Psi}^k; \widehat{\lambda}^k)\right\|_{l^2} \leqslant \varepsilon_*$
stop, $(\Sigma^k, \widehat{\Psi}^k)$ is the final solution;
end if
– Update Lagrangian multipliers:
$\lambda_j^{k+1} = \lambda_j^k - (1/K_k)G_j\left(\Sigma^k, \Psi_j^k\right), \quad j = 1, \ldots, N_{\text{src}};$
$K_{k+1} = K_k;$
$\alpha_{k+1} = K_{k+1};$
$\eta_{k+1} = \eta_k \alpha_{k+1}^{\beta_\eta};$
$\varepsilon_{k+1} = \varepsilon_k \alpha_{k+1}^{\beta_\varepsilon};$
else
– Decrease penalty parameters:
$\lambda_j^{k+1} = \lambda_j^k;$
$K_{k+1} = \tau K_k;$
$\alpha_{k+1} = K_{k+1}\overline{\gamma};$
$\eta_{k+1} = \overline{\eta}\alpha_{k+1}^{\beta_\eta};$
$\varepsilon_{k+1} = \overline{\varepsilon}\alpha_{k+1}^{\beta_\varepsilon};$
end if
3. Set starting point for the next subproblem in Step 1:
$(\widetilde{\Sigma}^{k+1}, \widetilde{\Psi}^{k+1}) = (\Sigma^k, \widehat{\Psi}^k);$

- **end for**

We refer interested readers to [9, 36, 49] for more detailed discussion of the ALM.

To solve the sub-optimization problem in step 1 of the above algorithm, we use a limited-memory version of the Broyden–Fletcher–Goldfarb–Shanno method (lm-BFGS) [40, 49]. The lm-BFGS algorithm requires computing of the gradient of the objective function with respect to Σ and $\widehat{\Psi}$. Those gradients can be analytically computed through

$$\nabla_\Sigma \mathcal{L}_K(\Sigma, \widehat{\Psi}; \widehat{\lambda}) = \nabla_\Sigma R(\Sigma) - \sum_{j=1}^{N_{\text{src}}} \lambda_j^T \nabla_\Sigma G_j(\Sigma, \Psi_j) + \frac{1}{K}\sum_{j=1}^{N_{\text{src}}} (G_j(\Sigma, \Psi_j))^T \nabla_\Sigma G_j(\Sigma, \Psi_j), \tag{30}$$

where $\nabla_\Sigma G_j = \left(\nabla_{\Sigma_a} G_j, \nabla_{\Sigma_s} G_j\right)$, and partial derivatives

$$\frac{\partial G_j}{\partial [\Sigma_a]_i} \equiv \begin{pmatrix} \frac{\partial [G_j]_1}{\partial [\Sigma_a]_i} \\ \frac{\partial [G_j]_2}{\partial [\Sigma_a]_i} \\ \vdots \end{pmatrix}$$

can be easily computed explicitly using the matrix representation (16) and (17). Gradient $\nabla_{\widehat{\Psi}} \mathcal{L}_K(\Sigma, \widehat{\Psi}; \widehat{\lambda})$ can similarly be found in a closed form.

3.2. Interpretation and discussion

As follows from the estimate (28), the constraints $G_j(\Sigma, \Psi_j) = 0$, $j = 1, \ldots, N_{\text{src}}$ are in general not satisfied with $(\Sigma^k, \widehat{\Psi}^k)$ on every iteration of the ALM. This means that Ψ_j^k is not the exact solution of the RTE with the optical properties $\Sigma^k = \left(\Sigma_a^k, \Sigma_s^k\right)$ and source S_j, which is quite different from the unconstrained optimization approach (4), in which Ψ_j^k is the solution of the RTE with the optical properties $\left(\Sigma_a^k, \Sigma_s^k\right)$ on every step k of an iterative minimization algorithm such as the quasi-Newton method. To outline the difference between the two

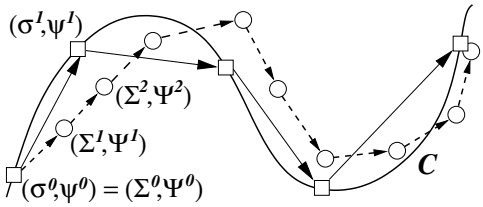

Figure 1. A simple illustration of the iteration process of unconstrained (□) and constrained (○) optimization approaches to optical tomographic imaging. Iterates (σ^k, ψ^k) are generated by an unconstrained optimization approach; iterates (Σ^k, Ψ^k) are generated by a constrained optimization algorithm.

methods we give a simple geometrical interpretation of the unconstrained and constrained optimization approaches. The equations $G_j(\Sigma, \Psi_j) = 0$, $j = 1, \ldots, N_{\text{src}}$ define a set C of the space $\mathbb{R}^N \times \mathbb{R}^N \times (\mathbb{R}^{2LN})^{N_{\text{src}}}$. On every iteration of the unconstrained optimization method vectors $(\Sigma^k, \widehat{\Psi}^k)$ are forced to belong to the set C by construction. Satisfying this condition exactly requires solving forward problems for different right-hand sides on every iteration of the algorithm. In the constrained optimization approach, however, conditions $G_j(\Sigma, \Psi_j) = 0$, $j = 1, \ldots, N_{\text{src}}$ will be satisfied only in the limit when $k \to \infty$. At a specific iteration $k < \infty$, $(\Sigma^k, \widehat{\Psi}^k)$ does not have to belong to the set C, see figure 1. This property provides a higher degree of freedom in choosing a new iteration, and thus can lead to computationally less intensive algorithms.

An advantageous property of the augmented Lagrangian method for optical tomographic imaging is that it does not have significantly larger memory requirements than unconstrained optimization methods. Since the discretized transport operator T is the same for all sources S_j, $j = 1, \ldots, N_{\text{src}}$, one needs to store only one T as in the unconstrained optimization cases. Actually, the only extra storage needed are the several source, radiance vectors and Lagrangian multipliers, S_j, Ψ_j and λ_j. This storage is much smaller than the storage of $\mathbf{T}$.

We also remark that the augmented Lagrangian method proposed above can be easily parallelized. First, in the sub-optimization problem, step 1, the gradient of the augmented Lagrangian function has an analytical form (30), which involves only the summation of local matrix–vector and vector–vector products. The computation of this gradient can thus be done on separate processors and then collected. A trivial example is to send those terms with different subscript j to different processors. Second, the update of Lagrangian multipliers in step 2 of the method can also be sent to parallel processors.

4. Numerical studies

4.1. The test problem set-up

To illustrate the performance of the ALM we consider three types of media and measurement geometries. In the first problem, we reconstruct the absorption coefficient in a cylinder with a smaller cylindrical inhomogeneity, in which the absorption coefficient is twice as high as in the background (see figure 2 for the geometrical set-up). In the second problem, which uses the same overall geometry, we reconstruct the spatial distribution of the scattering coefficient in a moderately anisotropic medium. In these two problems all sources and detectors are located on the circle defined by $\Gamma = \{(x, y, z) : x^2 + y^2 = 1, z = 1\}$. Finally, we show an example in which we simultaneously reconstruct the absorption and scattering coefficients in a highly anisotropic medium (figure 10, table 1). In this case, all sources and detectors

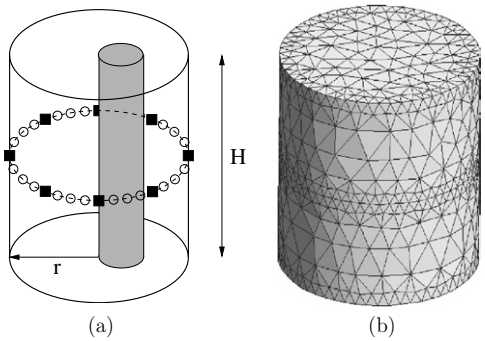

Figure 2. The test problems set-up. Cylinder height: $H = 2$ cm, radius $r = 1$ cm; the radius of the embedded small cylinder $r = 0.25$ cm. (a) Source–detector layout with 8 sources (■), 64 detectors (○). (b) Finite-volume mesh with 6727 tetrahedrons.

Table 1. Parameters used in three different problems.

	Problem 1	Problem 2	Problem 3
Anisotropy factor g	0.0	0.5	0.9
Background μ_a (cm^{-1})	0.1	0.1	0.5
Inhomogeneity μ_a (cm^{-1})	0.2	0.1	1.0
Background μ_s (cm^{-1})	10.0	10.0	10.0
Inhomogeneity μ_s (cm^{-1})	10.0	15.0	15.0
Number of sources	8	8	24
Number of detectors	64	64	24
Modulation frequency ω (MHz)	400	400	400
Number of finite volumes	6747	6747	13 867
Number of ordinates	8	48	80

are located on two circles defined by $\Gamma_1 = \{(x, y, z) : x^2 + y^2 = 1.5^2, z_1 = 2.2\}$ and $\Gamma_2 = \{(x, y, z) : x^2 + y^2 = 1.5^2, z_2 = 3.5\}$. Similar measurement configurations are commonly used in our and other laboratories [38, 44, 47, 69], for example, for optical measurements in finger joints, or small animal studies. For the discrete ordinate method, we use the level symmetric discrete ordinate arrangement and the corresponding weight set from [46]. All reconstructions were performed on a Linux workstation with a 700 MHz Pentium XEON processor.

4.1.1. Synthetic measurements. As measurements we use synthetic data, which are obtained by implementing a forward solver to the problem with exact optical properties. An algorithm for the forward problem solution in frequency domain is presented in [56]. To avoid a cancellation of numerical errors when the same model is used for the inverse problem and for generation of the synthetic data, which is often referred to as an 'inverse crime' [16], these data were generated using a much finer mesh (64 280 finite volumes) and 48 discrete ordinates.

Along with the 'exact' synthetic measurements we consider synthetic measurements with added interval Gaussian noise. If $M_{s,d}$ is a synthetic measurements corresponding to the source s and detector d, then instead of $M_{s,d}$ we use $\tilde{M}_{s,d} = M_{s,d} + \nu_{s,d}\mathcal{N}(0, 1)I_{[-1,1]}$, where $\mathcal{N}(0, 1)$ a standard normal distribution, and $I_{[-1,1]}$ is an interval indicator function that is equal

to 1 on the interval $[-1, 1]$ and zero otherwise. Parameter $v_{s,d}$ is the standard deviation of the added noise. The signal-to-noise ratio χ is then defined as

$$\chi = 10 \log_{10} \frac{M_{sd}}{v_{sd}}. \tag{31}$$

In our experiments, the signal-to-noise ratio χ is the same for all source–detector pairs and is equal to 20 dB or 15 dB, which are typical values for OT systems [40].

4.1.2. The initial guess. In all three problems considered in the paper we use homogeneous distributions as initial guesses μ_a^0 and μ_s^0 for the absorption and scattering coefficients. Solution of the forward problem with these optical properties provides us with the initial guess for the radiance ψ. We solve the forward problems with different sources only once to ensure that the ALM starts from the point in the space of 'optical properties—radiance' pairs that belongs to the set C (figure 1).

4.1.3. The stopping criterion. The stopping criterion for the reconstruction algorithm is given by

$$|E(\widehat{\Psi}^{k+1}) - E(\widehat{\Psi}^k)| < \varepsilon,$$

with $\varepsilon = 10^{-6}$, and the error function $E(\widehat{\Psi})$ being the first term in (18), in other words,

$$E(\widehat{\Psi}) = \frac{1}{2} \sum_{j=1}^{N_{\text{src}}} \sum_{d=1}^{N_j^d} \frac{|P_d \Psi_j - M_{jd}|^2}{|M_{jd}|^2}.$$

Therefore the algorithm is stopped when the error function $E(\widehat{\Psi^k})$ does not decrease anymore.

4.1.4. Reconstruction error measurement. To evaluate the quality of reconstruction we use the correlation coefficient $\rho(\mu^e, \mu^r)$ and the deviation factor $\delta(\mu^e, \mu^r)$ [40] that are defined as

$$\rho(\mu^e, \mu^r) = \frac{\sum_{i=1}^{N} \left(\mu_i^e - \overline{\mu^e}\right)\left(\mu_i^r - \overline{\mu^r}\right)}{(N-1)\sigma(\mu^e)\sigma(\mu^r)}, \qquad \delta(\mu^e, \mu^r) = \frac{\sqrt{\sum_{i=1}^{N} \left(\mu_i^e - \mu_i^r\right)^2 / N}}{\sigma(\mu^e)}, \tag{32}$$

where $\overline{\mu}$ and $\sigma(\mu)$ are the mean value and the standard deviation of distribution μ, respectively, μ_i^e is the value of the exact optical property in ith volume as given by the problem set-up, and μ_i^r is the corresponding value reconstructed by the algorithm. The larger the correlation coefficient the closer the shape of μ_r resembles the shape of μ_e. If the images are identical then $\rho(\mu^e, \mu^r) = 1$. The smaller $\delta(\mu^e, \mu^r)$ the smaller is the absolute difference between μ^e and μ^r. In the ideal case $\delta(\mu^e, \mu^r) = 0$.

Note that, since it is possible to obtain reasonably good reconstructions only in the vicinity of the planes where sources and detectors are located (figure 2(a)), we compute $\rho\left(\mu_a^e, \mu_a^r\right)$ and $\delta\left(\mu_a^e, \mu_a^r\right)$ only in this plane.

4.1.5. Unconstrained optimization code. To compare our constrained optimization algorithm with the unconstrained optimization approach we use a recently presented algorithm by Ren *et al* [57]. Like the algorithm presented in this paper, this code solves the inverse problem for the frequency-domain radiative transport equation using the discrete ordinate method and the finite-volume method for angular and spatial approximation, respectively. However, the unconstrained optimization code is based on minimization of the objective function (4) only

Table 2. Quality of reconstruction of the absorption coefficient for different reconstruction methods, different noise levels, and different regularization parameters. The parameter β is given in units of $[10^{-10}]$.

Method	ALM	ALM	ALM	ALM	ALM	lm-BFGS
Signal-to-noise ratio χ (dB)	15	20	20	20	∞	∞
Regularization parameter β	200	10	200	500	200	–
Iterations to convergence	512	422	367	315	248	60
CPU time	9.1 h	7.9 h	6.4 h	5.5 h	3.9 h	103.2 h
Correlation $\rho(\mu_a^e, \mu_a^r)$	0.63	0.53	0.68	0.71	0.76	0.79
Deviation $\delta(\mu_a^e, \mu_a^r)$	0.85	0.97	0.79	0.81	0.69	0.64

with respect to optical properties μ_a and μ_s using the limited-memory quasi-Newton (lm-BFGS) method, and the adjoint differentiation approach to compute a gradient of the objective function with respect to the optical properties. This code utilizes a GMRES iterative method with the ILU pre-conditioner to solve the forward and adjoint algebraic systems arising after discretization of the RTE [56].

4.2. Reconstruction of the absorption coefficient

The constrained and unconstrained algorithms were used to perform reconstructions for various combinations of different noise levels and regularization parameters. The results with respect to the correlation coefficient, deviation factor, CPU time, number of iterations to convergence are summarized in table 2, and the corresponding convergence history (the error function $E(\Psi^k)$ versus the iteration number k) of the iterative methods is shown in figure 3.

4.2.1. Comparison of constrained and unconstrained optimization. We first compare the performance of our newly developed constrained optimization code with the unconstrained optimization code presented in section 4.1.5. Figures 4(a) and (b) show the reconstruction of the absorber obtained with that unconstrained optimization code using noise-free data ($\chi = \infty$). It took approximately 103.2 h of CPU time to complete 60 lm-BFGS iterations (table 2, figure 3(a)). The correlation coefficient is $\rho(\mu_a^e, \mu_a^r) = 0.79$ and the deviation factor is $\delta(\mu_a^e, \mu_a^r) = 0.64$. Figures 4(c) and (d) display the images obtained with the augmented Lagrangian method, using the same spatial and angular discretization and the same finite-volume mesh. A comparable image quality is achieved in only 3.9 h (248 ALM iterations), which is almost 27 times faster. This acceleration factor is essentially determined by the number of matrix–vector multiplications with the matrix $\mathbf{T}$ defined in (16) and (17), required to complete the reconstruction. For example, for the absorption reconstruction problem considered in this section, on every ALM iteration one has to make on average 1.5×10^3 multiplications with the matrix $\mathbf{T}$ for each source. The unconstrained optimization code makes approximately 1.5×10^5 matrix–vector multiplications for each source on every iteration of the lm-BFGS method. Since 60 lm-BFGS iterations are needed to get figures 4(a) and (b), while 248 ALM iterations are required for figures 4(c) and (d), the overall number of matrix–vector multiplications used by the ALM code is approximately 24 times less than in the unconstrained minimization code, which accords well with the observed CPU time differences.

We also explored if relaxing the accuracy requirements for the forward and adjoint solutions in the unconstrained optimization code can yield similar acceleration. We observed a moderate (up to 25%) acceleration of the reconstruction in the cases where appropriate accuracy requirements were chosen. Here an appropriate accuracy requirement means a

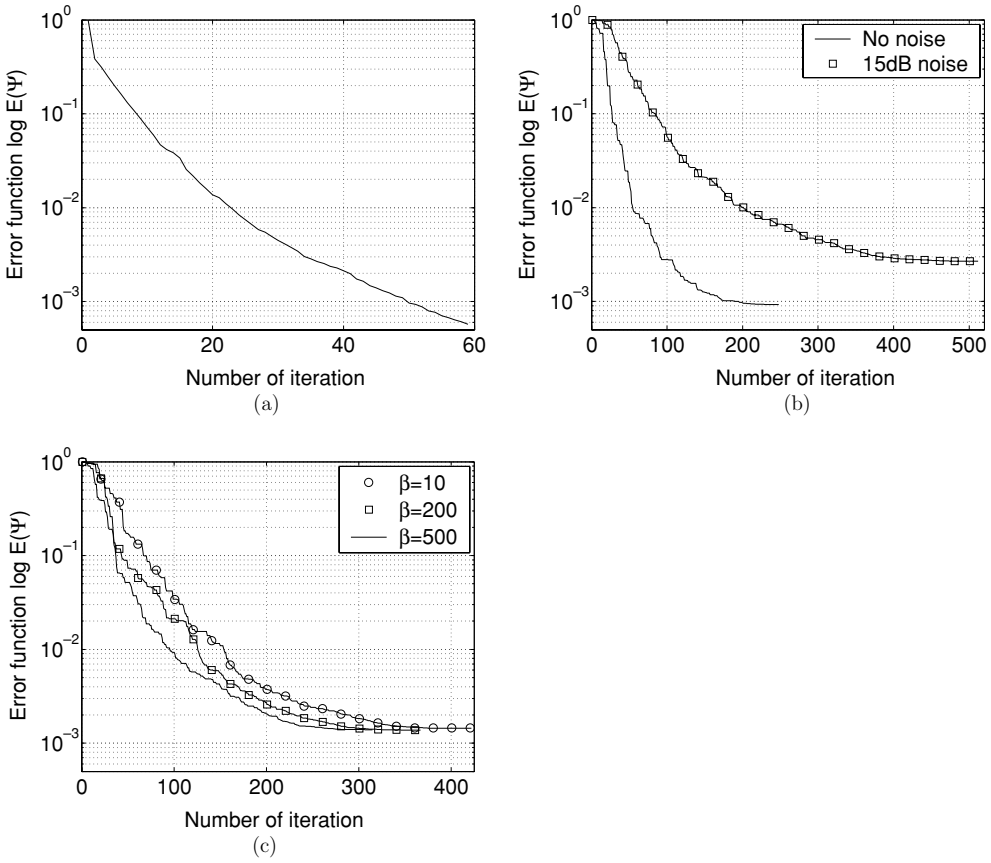

Figure 3. Convergence history of $E(\Psi^k)/E(\Psi^0)$ for μ_a reconstruction (in $\log_{10}$ scale). (a) The lm-BFGS unconstrained optimization method [57], with no noise. (b) The augmented Lagrangian method, $\chi = \infty$ (no noise), and $\chi = 15$ dB; (c) The augmented Lagrangian method, $\chi = 20$ dB and different regularization parameters. All the values of β are given in units of $[10^{-10}]$.

stopping criterion for the forward GMRES solver such that it still ensures convergence of the reconstruction algorithm. For example, we usually stop the GMRES solver when the relative residual is smaller than 10^{-13}. If we stop the GMRES when the relative residual is smaller than 10^{-7} in the first several quasi-Newton steps, the reconstruction still converges to very similar results, and the computational time is reduced by approximately 25%. However, if we further relax the accuracy requirements our minimization algorithm no longer converges to the right solution. The reason for this, in our opinion, is because we use the adjoint differentiation method to compute the gradient of the objective function with respect to optical parameters. This gradient is calculated using the solution of the forward and adjoint problems. In the cases where the forward and adjoint solutions are not accurate enough, the approximate gradient can be computed only approximately which results in poor convergence of the minimization algorithm, or no convergence at all.

Note that the reconstructions of the absorption coefficient obtained by different methods (ALM versus lm-BFGS) do not coincide completely. But this fact is expected and can be easily explained. Indeed, ALM and lm-BFGS generate iterates $\left(\Sigma_a^k, \Psi_j^k\right)$ and $\left(\sigma_a^k, \psi_j^k\right)$, respectively, ($k$ is the number of iteration) that do not have to be the same (figure 1), but converge to each other only in the limit as $k \to \infty$. In the lm-BFGS method for the unconstrained optimization

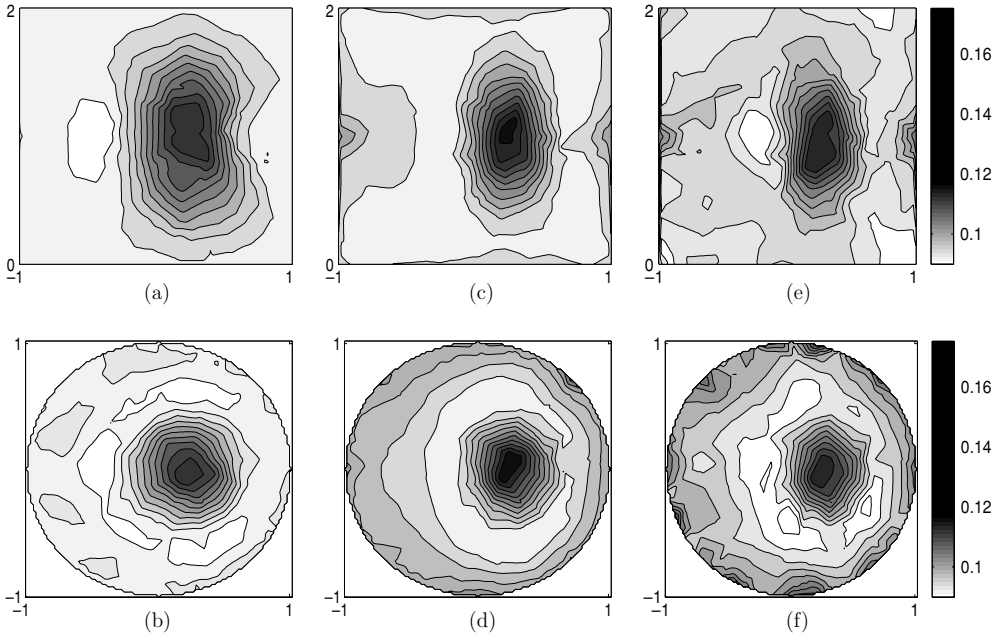

Figure 4. Cross sections of the reconstructed absorption coefficient in the planes $y = 0$ (a), (c), (e) and $z = 1$ (b), (d), (f) with the quasi-Newton lm-BFGS method for the unconstrained optimization [57] and the ALM for problem 1 with different noise levels. The target optical properties are $\mu_a = 0.2$ cm^{-1} in the inclusion, $\mu_a = 0.1$ cm^{-1} in the background. (a), (b) correspond to the reconstruction with the unconstrained minimization approach; (c), (d) correspond to the ALM reconstruction with noise-free data; (e), (f) correspond to the ALM reconstruction with 15 dB added noise.

the intensity ψ_j^k is a solution of the forward problem (RTE) with $\mu_a = \sigma_a^k$, on each iteration k, by construction, whereas this property does not have to hold for the augmented Lagrangian method.

4.2.2. Impact of noise level. Next we test the impact of noise in the synthetic measurement data on the reconstruction results. An example of a reconstruction performed with data with a signal-to-noise ratio $\chi = 15$ dB is shown in figures 4(e) and (f) next to the images obtained with noise-free data (figures 4(c) and (d)). As expected the quality of reconstruction decreases with a decreased signal-to-noise ratio. The correlation coefficient drops from 0.76 in the case of noise-free data to 0.63, while the deviation factor increases from 0.69 to 0.85 (table 2). Note that as the noise level increases (and χ decreases), the number of necessary ALM iterations and CPU time increases. The results for $\chi = 20$ dB (the fourth column in table 2) are in between the values for 15 dB and noise-free data ($\chi = \infty$). However, even for $\chi = 25$ dB which is a very reasonable noise level for currently available instrumentation, the 9.1 h CPU time is still over 11 times faster than the time it takes to complete unconstrained optimization with noise-free data. Only if the signal-to-noise ratio drops below 10 dB, does the augmented Lagrangian method fail to converge to any solution.

4.2.3. Impact of the regularization parameter. In the next series of numerical experiments we test the performance of the ALM for different values of the regularization parameter β (see equation (18), section 3). In these experiments 20 dB Gaussian noise is added and the

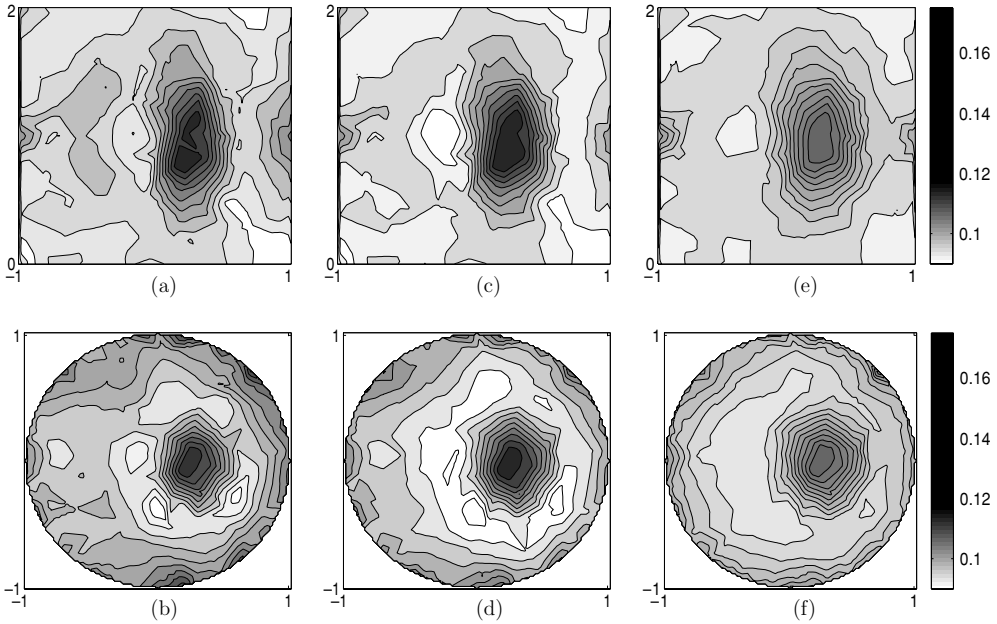

Figure 5. Cross sections of the reconstructed absorption coefficient in the planes $y = 0$ (a), (c), (e) and $z = 1$ (b), (d), (f) with the ALM for problem 1 with different regularization parameters. The target optical properties are $\mu_a = 0.2$ cm^{-1} in the inclusion, $\mu_a = 0.1$ cm^{-1} in the background. (a), (b) correspond to the ALM reconstruction with $\beta = 10 \times 10^{-10}$; (c), (d) correspond to the ALM reconstruction with $\beta = 200 \times 10^{-10}$; (e), (f) correspond to the ALM reconstruction with $\beta = 500 \times 10^{-10}$.

parameter β is equal to 10, 200 and 500×10^{-10}. As we can see, when the regularization parameter is small (figure 5(a) and (b) and table 2), the impact of noise is more significant, whereas for a large regularization parameter the absolute value of the absorption coefficient in the centre of the inhomogeneity is found less accurately (figures 5(e) and (f)). This is reflected in the deviation factor $\delta(\mu_a^e, \mu_a^r)$ which first drops from 0.97 (in the case of a small value of the regularization parameter ($\beta = 10 \times 10^{-10}$) to 0.79 ($\beta = 200 \times 10^{-10}$), but then increases slightly to 0.81 for $\beta = 500 \times 10^{-10}$ (see columns 3–5 in table 2). This suggests that there is an optimal value of β between 10 and 500×10^{-10}, for which $\delta(\mu_a^e, \mu_a^r)$ becomes smallest. On the other hand, the correlation factor $\rho(\mu_a^e, \mu_a^r)$ keeps increasing (improving) as β increases. It is notable that for a rather large range of values of β 'reasonable' reconstruction results can be obtained. In the cases considered in this study, values of β up to 2000×10^{-10} allowed perturbation to be located, even though with incorrect absolute values. It appears that for each particular application (e.g., breast, brain or joint imaging), some preliminary studies should be performed to find optimal regularization parameters.

4.3. Reconstruction of the scattering coefficient

In the second example, we show the reconstruction of the scattering coefficient with an anisotropic factor $g = 0.5$. We use synthetic measurements with $\chi = 20$ dB and $\beta = 500 \times 10^{-10}$. A large number of discrete ordinates are necessary to capture anisotropy effects as compared to the case of problem 1 where isotropic scattering is assumed. This leads to a total number of 5181 696 constraints in the constrained optimization formulation. The results of the reconstruction are presented in figure 6 and the convergence history is given in figure 7(a)

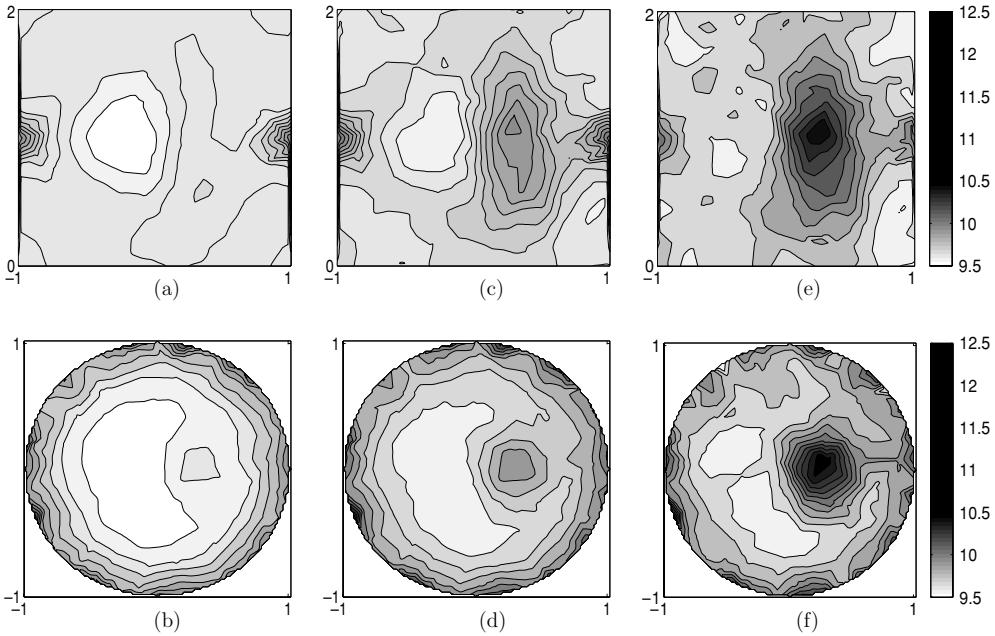

Figure 6. Cross sections of the reconstructed scattering coefficient in the planes $y = 0$ (a), (c), (e) and $z = 1$ (b), (d), (f) with the augmented Lagrangian method for problem 2. The target optical properties are $\mu_s = 15\,\mathrm{cm}^{-1}$ in the inclusion, $\mu_s = 10\,\mathrm{cm}^{-1}$ in the background. (a), (b) correspond to the reconstruction after 50 iterations of the ALM; (c), (d) correspond to the reconstruction after 200 iterations of the ALM; (e), (f) correspond to the reconstruction at convergence (498 iterations).

Table 3. Quality of reconstruction of the scattering coefficient as a function of the ALM iteration step.

Iteration number	50	200	498
Signal-to-noise ratio χ (dB)	20	20	20
Correlation $\rho(\mu_s^e, \mu_s^r)$	−0.12	0.42	0.67
Deviation $\delta(\mu_s^e, \mu_s^r)$	1.12	0.97	0.81

and table 3. We observe that the reconstruction algorithm converges slower than for μ_a, which is consistent with the results presented in [57]. Note that the error function $E(\Psi)$ can actually increase on some iterations. This can be best explained by the fact that the ALM tries to minimize the augmented Lagrangian function $\mathcal{L}_K(\Sigma, \widehat{\Psi}; \widehat{\lambda})$, not only the error function $E(\Psi)$. The acceleration rate we obtained as compared to the unconstrained optimization method in [57] is approximately 18.

4.3.1. Impact of the initial guess. We use reconstructions of the scattering coefficient to illustrate convergence of the ALM from different initial guesses μ_s^0 ($\chi = \infty$). In the first example, the initial guess for the scattering coefficient is equal to the background value $\left(\mu_s^0 = 10\ \mathrm{cm}^{-1}\right)$, in the second it is 10% higher $\left(\mu_s^0 = 11\ \mathrm{cm}^{-1}\right)$, and in the third it is 20% higher than the background value $\left(\mu_s^0 = 12\ \mathrm{cm}^{-1}\right)$. The results of reconstruction are presented in figure 8 and table 4. It can be seen that the correlation factor depends weakly on the initial guess. It only drops by approximately 10% from 0.69 to 0.62, when the initial guess is increased from 10 to 12 cm^{-1}. This just expresses the fact that the location and

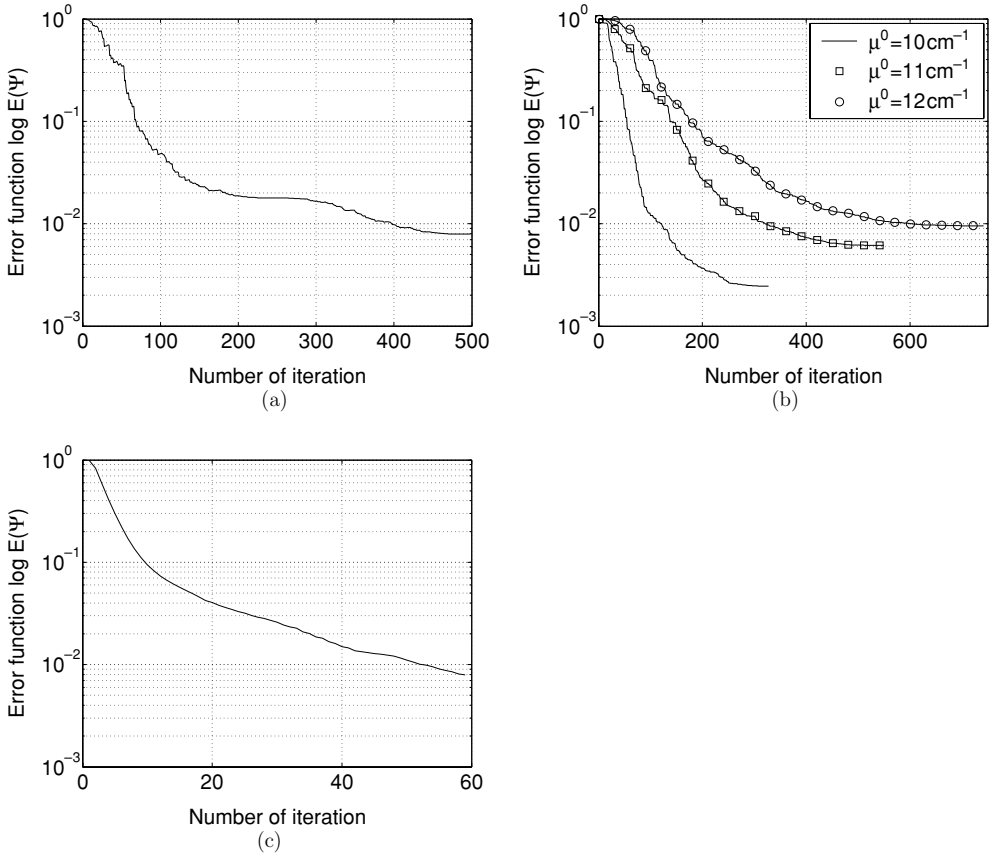

Figure 7. Convergence history of $E(\Psi^k)/E(\Psi^0)$ for μ_s reconstruction (in $\log_{10}$ scale). (a) The augmented Lagrangian method, $\chi = 20$ dB and $\beta = 500 \times 10^{-10}$; (b) the augmented Lagrangian method with different initial guesses, $\chi = \infty$ dB and $\beta = 300 \times 10^{-10}$; (c) the lm-BFGS unconstrained optimization method [57], with no noise.

Table 4. Quality of reconstruction of the scattering coefficient as a function of the initial guess.

Initial guess μ_s^0 (cm^{-1})	10	11	12
Number of ALM iterations to convergence	327	549	741
Correlation $\rho(\mu_s^e, \mu_s^r)$	0.69	0.66	0.62
Deviation $\delta(\mu_s^e, \mu_s^r)$	0.76	1.20	1.34

shape of the inhomogeneity is still rather accurately reconstructed. The absolute values of the optical properties in the image are strongly affected by the initial guess, which is obvious in the images as well as in the values of the deviation factor $\delta(\mu_s^e, \mu_s^r)$ which increases from 0.74 to 1.34 when the initial guess is $\mu_s^0 = 12$ cm^{-1} instead of $\mu_s^0 = 10$ cm^{-1}. Similar results were observed for the case of absorption reconstruction. Overall we found that the ALM will produce images with $\rho(\mu^e, \mu^r) > 0.5$ as long as the initial guess is not more than 40% higher than the actual background value.

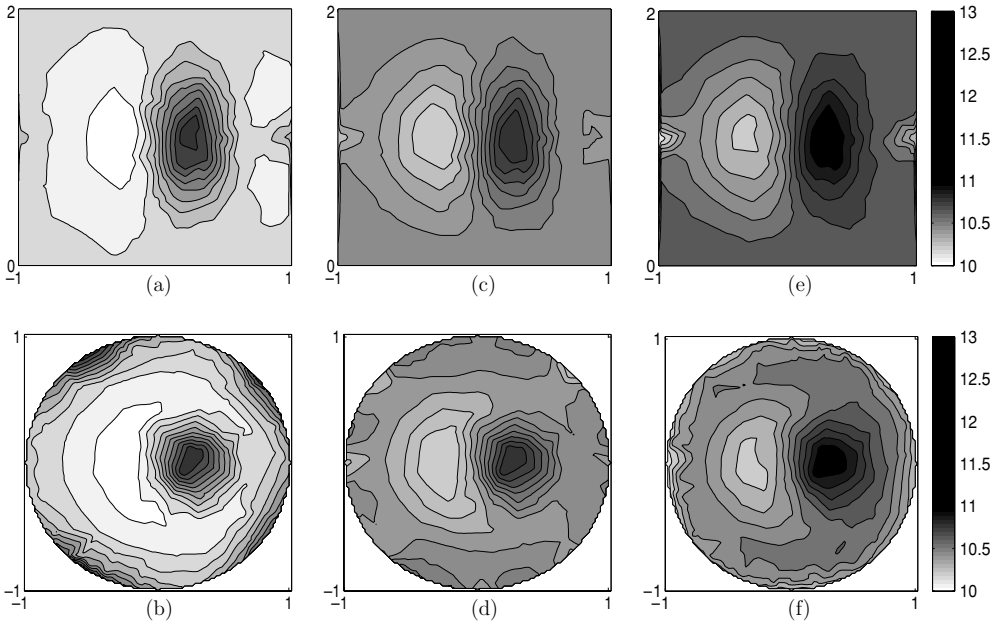

Figure 8. Cross sections of the reconstructed scattering coefficient in the planes $y = 0$ (a), (c), (e) and $z = 1$ (b), (d), (f) with the augmented Lagrangian method for problem 2 with different initial guesses. The target optical properties are $\mu_s = 15\ \text{cm}^{-1}$ in the inclusion, $\mu_s = 10\ \text{cm}^{-1}$ in the background. (a), (b) correspond to the initial guess $\mu_s^0 = 10\ \text{cm}^{-1}$; (c), (d) correspond to the initial guess $\mu_s^0 = 11\ \text{cm}^{-1}$; (e), (f) correspond to the initial guess $\mu_s^0 = 12\ \text{cm}^{-1}$.

Table 5. Quality of reconstruction of the scattering coefficient as a function of the mesh size.

Number of finite volumes	6747	10 062	15 612	19 489
Correlation $\rho(\mu_s^e, \mu_s^r)$	0.69	0.72	0.75	0.75
Deviation $\delta(\mu_s^e, \mu_s^r)$	0.76	0.75	0.73	0.74

4.3.2. Impact of mesh size. Finally, we compare the reconstructions obtained with different finite-volume meshes, using noise-free synthetic measurement data. Images are displayed in figure 9, while the corresponding $\rho(\mu_s^e, \mu_s^r)$ and $\delta(\mu_s^e, \mu_s^r)$ are shown in table 5. The images show almost no visible differences. Indeed, refining the mesh from 6747 tetrahedrons to 15 615 leads only to slight improvements in the correlation coefficient $\rho(\mu_s^e, \mu_s^r)$ (from 0.69 to 0.75) and the deviation factor $\delta(\mu_s^e, \mu_s^r)$ (from 0.76 to 0.73). Further mesh refinement does not improve these image quality parameters. This shows that when the mesh is fine enough so that the average distance between the mesh points is of the order of the scattering mean free path $(1/\mu_s)$, the discretization is sufficient to capture most pertinent effects.

4.4. Simultaneous reconstruction of absorption and scattering coefficients

As a last example we present a simultaneous reconstruction of both the absorption and scattering coefficients in highly anisotropic scattering media, with $g = 0.9$. In this case, we set the regularization parameter β to 700×10^{-10}. The synthetic measurements for this problem were generated using 80 ordinates and a finite-volume mesh with 23 793 tetrahedrons and 20 dB noise was added to the measurements.

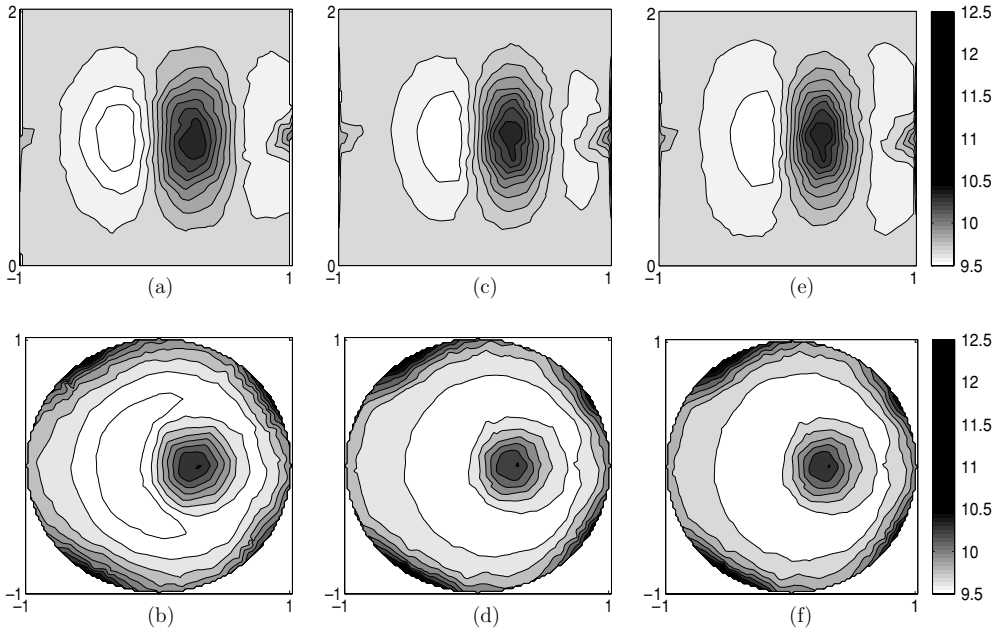

Figure 9. Cross sections of the reconstructed scattering coefficient in the planes $y = 0$ (a), (c), (e) and $z = 1$ (b), (d), (f) with the augmented Lagrangian method for problem 2 with different meshes. The target optical properties are $\mu_s = 15$ cm^{-1} in the inclusion, $\mu_s = 10$ cm^{-1} in the background. (a), (b) correspond to the mesh with 10 062 tetrahedrons; (c), (d) correspond to the mesh with 15 612 tetrahedrons; (e), (f) correspond to the mesh with 19 489 tetrahedrons.

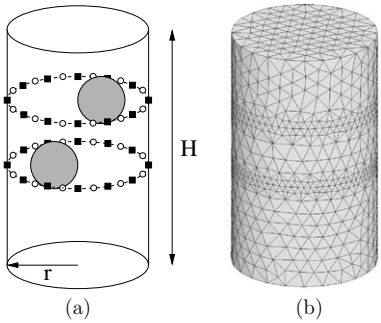

Figure 10. Test problem 3 set-up. Cylinder height: $H = 5$ cm, radius $r = 1.5$ cm; the radius of the embedded small cylinder $r = 0.5$ cm. (a) Source–detector layout with 24 sources (■), 24 detectors (○). (b) Finite-volume mesh with 13 867 tetrahedrons.

Target optical properties represent media with a relatively high absorption and a low scattering, for which the diffusion theory fails to describe the light propagation accurately. The cross sections of the reconstructed absorption and scattering coefficients are presented in figure 11, and the corresponding values of $\rho\left(\mu_a^e, \mu_a^r\right)$, $\rho\left(\mu_s^e, \mu_s^r\right)$ and $\delta\left(\mu_a^e, \mu_a^r\right)$, $\delta\left(\mu_s^e, \mu_s^r\right)$ are given in table 6. As for problems 1 and 2, the parameters $\rho(\mu^e, \mu^r)$ and $\delta(\mu^e, \mu^r)$ are computed only in the planes $z = 2.2$ and $z = 3.5$, in which sources and detectors are located.

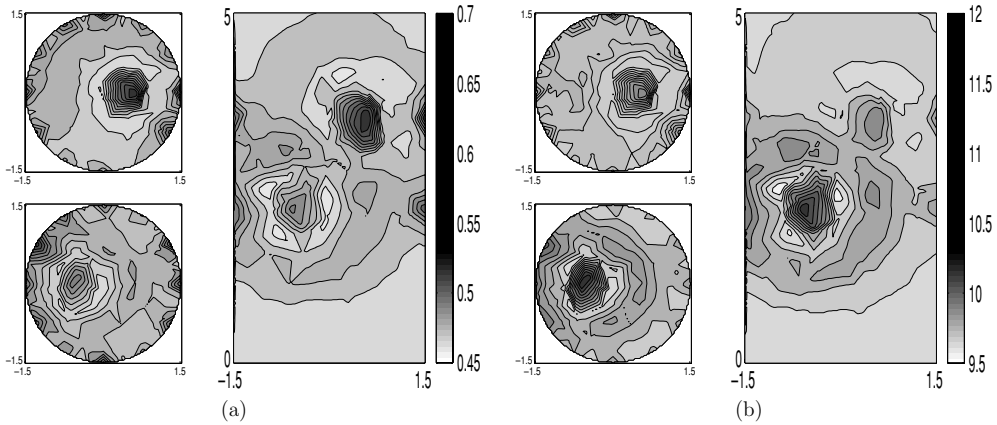

Figure 11. Cross sections of the reconstructed absorption and scattering coefficient in the planes $y = 0, z = 2.2$ and $z = 3.5$ with the augmented Lagrangian method for problem 3. The target optical properties are $\mu_a = 1.0\ \text{cm}^{-1}$, $\mu_s = 15\ \text{cm}^{-1}$ in the inclusion, $\mu_a = 0.5\ \text{cm}^{-1}$, $\mu_s = 10\ \text{cm}^{-1}$ in the background. (a) Reconstruction of μ_a at convergence (712 iterations), left-top: cross section $z = 3.5$, left-bottom: cross section $z = 2.2$, right: cross section $y = 0$. (b) Reconstruction of μ_s at convergence (712 iterations).

Table 6. Quality of reconstruction of the absorption and scattering coefficients as a function of the ALM iteration step.

Iteration number	50	300	712
Correlation $\rho(\mu_a^e, \mu_a^r)$	−0.33	0.20	0.29
Deviation $\delta(\mu_a^e, \mu_a^r)$	1.14	0.99	0.96
Correlation $\rho(\mu_s^e, \mu_s^r)$	−0.32	0.07	0.39
Deviation $\delta(\mu_s^e, \mu_s^r)$	1.20	1.01	0.95

5. Summary and conclusion

In this work, we present a novel approach to solving the inverse problems encountered in optical tomography. We have implemented a PDE-constrained optimization method that uses a finite-volume method for the discretization of the frequency-domain radiative transport equation (RTE). The finite-volume discretization gives rise to an algebraic nonlinear programming problem that is solved using the iterative augmented Lagrangian method. By simultaneously updating both radiance and optical properties, the method solves the forward and inverse problems in optical tomography all at once. In this way, the computing time is greatly reduced as compared to traditional unconstrained optimization methods, during which one has to repeatedly solve the forward problem many times. We tested and quantified the performance of the algorithm for various combinations of mesh sizes, noise, regularization parameters, initial guesses, optical properties and measurement geometries. Besides the speed of the code, we compared image qualities by defining a correlation coefficient ρ as well as a deviation factor δ.

In the cases that involve image reconstruction from synthetic measurement data we observe 10–30-fold decrease in computing time for the constrained optimization code compared to the unconstrained optimization code. The regularization parameter β has some influence on the computing time, but with reasonable values of β which in our case are of order $10–10^3 \times 10^{-10}$, the computational time changes less than 20%. In general, reconstruction of both

absorption and scattering together took longer than reconstructions of only the scattering coefficient or only the absorption coefficient. As expected the correlation coefficients (ρ) and deviation factors (δ) worsen as the signal-to-noise ratio decreases. Similarly δ decrease substantially as the (homogeneous) initial guess is not chosen close to the optical properties of the actual background medium. Interestingly ρ is only weakly affected by the initial guess. As long as the optical properties are chosen within 50% of the actual background medium ρ changes by only 10–20%. Finally ρ and δ do not change once the mesh is fine enough so that the average size of finite volumes becomes less than the average scattering mean free path ($1/\mu_s$).

Another positive aspect of the augmented Lagrangian method is that it maintains storage requirements that are comparable to requirements encountered in unconstrained optimization methods. The augmented Lagrangian also provides the flexibility of being easily implemented on parallel processors.

Finally, it should be noted that the constrained optimization method introduced in this work does not require that the forward model is the frequency-domain RTE. For example, one can also implement similar codes with the diffusion equation as the forward model of light propagation in tissues or with a system of two RTEs as the forward model as in the case of fluorescence tomography [42].

Acknowledgments

We would like to thank Guillaume Bal and Alexander Klose for fruitful discussions on the inverse problems in optical tomography. We also acknowledge anonymous referees for their constructive comments that helped us to improve both the quality and clarity of this paper. This work was supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant no 5R01 AR046255-05) and the National Institute of Biomedical Imaging and Bioengineering (grant no R01 EB001900-01), which are both part of the National Institutes of Health.

References

[1] Abdoulaev G S and Hielscher A H 2003 Three-dimensional optical tomography with the equation of radiative transfer *J. Electron. Imaging* **9** 594

[2] Adams M and Larsen E W 2002 Fast iterative methods for discrete-ordinates particle transport calculations *Prog. Nucl. Energy* **40** 3

[3] Aronson R, Barbour R L, Lubowsky J and Graber H 1991 Application of transport theory to infra-red medical imaging *Modern Mathematical Methods in Transport Theory* ed W Greenberg and J Polewczak (Basle: Birkhäuser)

[4] Arridge S 1999 Optical tomography in medical imaging *Inverse Problems* **15** 41

[5] Arridge S R and Hebden J 1997 Optical imaging in medicine: II. Modeling and reconstruction *Phys. Med. Biol.* **42** 841

[6] Arridge S R and Lionheart W R B 1998 Non-uniqueness in diffusion-based optical tomography *Opt. Lett.* **23** 882

[7] Barth T J and Jespersen D C 1989 The design and application of upwind schemes on unstructured meshes *AIAA Paper 89-0366*

[8] Benaron D A *et al* 2000 Noninvasive functional imaging of human brain using light *J. Cereb. Blood Flow Metab.* **20** 469

[9] Bertsekas D 1982 *Constrained Optimization and Lagrange Multiplier Methods* (New York: Academic)

[10] Biegler L, Ghattas O, Heinkenschloss M and van Bloemen Waanders B 2003 *Large-Scale PDE-Constrained Optimization (Lecture Notes in Computational Science and Engineering* vol 30*)* (Berlin: Springer)

[11] Bluestone A, Abdoulaev G S, Schmitz C, Barbour R and Hielscher A H 2001 Three-dimensional optical-tomography of hemodynamics in the human head *Opt. Express* **9** 272

[12] 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.* **18** 57

[13] Boas D A, Chen K, Grebert D and Franceschini M A 2004 Improving the diffuse optical imaging spatial resolution of the cerebral hemodynamic response to brain activation in humans *Opt. Lett.* **29** 1506
[14] Case K M and Zweifel P F 1967 *Linear Transport Theory* (Reading, MA: Addison-Wesley)
[15] Chang J, Barbour R, Graber H and Aronson R 1995 Fluorescence optical tomography *Proc. SPIE, Experimental and Numerical Methods for Solving Ill-Posed Inverse Problems: Medical and Non-medical Applications* vol 2570 ed R Barbour *et al* (Bellingham, WA: SPIE)
[16] Colton D and Kress R 1998 *Inverse Acoustic and Electromagnetic Scattering Theory* (New York: Springer)
[17] Conn A, Gould N and Toint P 1992 *LANCELOT: a FORTRAN Package for Large-scale Nonlinear Optimization (Release A) (Springer Series in Computational Mathematics* vol 17*)* (New York: Springer)
[18] Dautray R and Lions J-L 1993 *Mathematical Analysis and Numerical Methods for Science and Technology* vol 6 (Berlin: Springer)
[19] Dehghani H, Brooksby B, Vishwanath K, Pogue B W and Paulsen K D 2003 The effects of internal refractive index variation in near infrared optical tomography: a finite element modeling approach *Phys. Med. Biol.* **48** 2713
[20] Dehghani H, Delpy D T and Arridge S R 1999 Photon migration in non-scattering tissue and the effects on image reconstruction *Phys. Med. Biol.* **44** 2897
[21] Dorn O 1998 A transport–backtransport method for optical tomography *Inverse Problems* **14** 1107
[22] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[23] Eppstein M J, Hawrysz D J, Godavarty A and Sevick-Muraca E M 2002 Three dimensional near infrared fluorescence tomography with Bayesian methodologies for image reconstruction from sparse and noisy data sets *Proc. Natl Acad. Sci.* **99** 9619
[24] Eymard R, Gallouet T and Herbin R 2000 *Handbook of Numerical Analysis VII* ed P Ciarlet and J L Lions (Amsterdam: North-Holland) p 715
[25] Firbank M, Arridge S, Schweiger M and Delpy D 1996 An investigation of light transport through scattering bodies with non-scattering regions *Phys. Med. Biol.* **41** 767
[26] Ghattas O and Bark J-H 1997 Optimal control of two- and three-dimensional incompressible Navier–Stokes flows *J. Comput. Phys.* **136** 231
[27] Gurfinkel M, Ke S, Wen X X, Li C and Sevick-Muraca E M 2003 Near-infrared fluorescence optical imaging and tomography *Disease Markers* **19** 107
[28] Haber E and Ascher U 2001 Preconditioned all-at-once methods for large, sparse parameter estimation problems *Inverse Problems* **17** 1847
[29] Henyey L G and Greenstein L J 1941 Diffuse radiation in the galaxy *Astrophys.* **90** 70
[30] Hestenes M R 1969 Multiplier and gradient methods *J. Optim. Theory Appl.* **4** 303
[31] Hielscher A, Alcouffe R and Barbour R 1998 Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues *Phys. Med. Biol.* **43** 1285
[32] Hielscher A, Bluestone A, Abdoulaev G S, Klose A D, Lasker J, Stewart M, Netz U and Beuthan J 2002 Near-infrared diffuse optical tomography *Dis. Markers* **18** 313
[33] Hielscher A and Bartel S 2001 Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography *J. Biomed. Opt.* **6** 183
[34] Hielscher A, Klose A D and Hanson K 1999 Gradient-based iterative image reconstruction scheme for time-resolved optical tomography *IEEE Trans. Med. Imaging* **18** 262
[35] Hielscher A H, Klose A D, Scheel A, Moa-Anderson B, Backhaus M, Netz U and Beuthan J 2004 Sagittal laser optical tomography for imaging of rheumatoid finger joints *Phys. Med. Biol.* **49** 1147
[36] Ito K and Kunish K 1990 The augmented Lagrangian method for parameter estimation in elliptic systems *SIAM J. Control Optim.* **28** 137
[37] Jiang H, Xu Y, Iftimia N, Baron L and Eggert J 2001 Three-dimensional optical tomographic imaging of breast in a human subject *IEEE Trans. Med. Imaging* **20** 1334
[38] Jiang S, Pogue B W, McBride T O and Paulsen K D 2003 Quantitative analysis of near-infrared tomography: sensitivity to the tissue-simulated precalibration phantom *J. Biomed. Opt.* **8** 308
[39] 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
[40] Klose A D and Hielscher A H 2003 Quasi-Newton methods in optical tomographic imaging *Inverse Problems* **19** 387
[41] Klose A D and Hielscher A H 2003 Fluorescence tomography with simulated data based on the equation of radiative transfer *Opt. Lett.* **28** 1019
[42] Klose A D, Ntziachristos V and Hielscher A H 2005 The inverse source problem based on the radiative transfer equation in molecular optical imaging *J. Comput. Phys.* **202** 323
[43] Lakowicz J 1983 *Principles of Fluorescence Spectroscopy* (New York: Plenum)

[44] Lasker J M, Dwyer E and Hielscher A H Optical tomographic imaging of vascular and metabolic reactivity in rheumatoid joints *Optical Tomography and Spectroscopy of Tissue VII (Proc. SPIE—International Society for Optical Engineering* vol 5693*)* ed B Chance, R R Alfano, B J Tromberg, M Tamura and E M Sevick-Muraca (at press)

[45] Lee J and Sevick-Muraca E M 2002 Three-dimensional fluorescence enhanced optical tomography using referenced frequency-domain photon migration measurements at emission and excitation wavelengths *J. Opt. Soc. Am.* A **19** 759

[46] Lewis E E and Miller W F 1993 *Computational Methods in Neutron Transport* (Le Grange Park, IL: American Nuclear Society)

[47] Masciotti J, Abdoulaev G A, Hur J, Papa J, Bae J, Huang J, Yamashiro D, Kandel J and Hielscher A H Combined optical tomographic and magnetic resonance imaging of tumor bearing mice *Optical Tomography and Spectroscopy of Tissue VII (Proc. SPIE—International Society for Optical Engineering* vol 5693*)* ed B Chance, R R Alfano, B J Tromberg, M Tamura and E M Sevick-Muraca (at press)

[48] Milstein A B, Oh S, Webb K J, Bouman C A, Zhang Q, Boas D A and Millane R P 2003 Fluorescence optical diffusion tomography *Appl. Opt.* **42** 3081

[49] Nocedal J and Wright S J 1999 *Numerical Optimization* (New York: Springer)

[50] Ntziachristos V and Weissleder R 2002 Charge-coupled-device based scanner for tomography of fluorescent near-infrared probes in turbid media ed probes in turbid media *Med. Phys.* **29** 803

[51] Orozco C and Ghattas G 1992 Massively parallel aerodynamic shape optimization *Comput. Syst. Eng.* **1–4** 311

[52] Paulsen K and Jiang H 1996 Enhanced frequency-domain optical image reconstruction in tissue through total variation minimization *Appl. Opt.* **35** 3447

[53] Pogue B W, Jiang S D, Dehghani H, Kogel C, Soho S, Srinivasan S, Song X M, Tosteson T D, Poplack S P and Paulsen K D 2004 Characterization of hemoglobin, water, and NIR scattering in breast tissue: analysis of inter-subject variability and menstrual cycle changes *J. Biomed. Opt.* **9** 541

[54] Pogue B, McBride T, Prewitt J, Österberg U L and Paulsen K 1999 Spatially variant regularization improves diffuse optical tomography *Appl. Opt.* **38** 2950

[55] Powell M J D 1969 A method for nonlinear constraints in minimization problems *Optimization* ed R Fletcher (New York: Academic) p 283

[56] Ren K, Abdoulaev G S, Bal G and Hielscher A H 2004 Algorithm for solving the equation of radiative transfer in the frequency domain *Opt. Lett.* **29** 578

[57] Ren R, Bal G and Hielscher A H 2004 Frequency domain optical tomography with the equation of radiative transfer *SIAM J. Sci. Comput.* at press

[58] Roy R and Sevick-Muraca E M 2000 Active constrained truncated Newton method for simple-bound optical tomography *J. Opt. Soc. Am.* A **17** 1627

[59] Roy R and Sevick-Muraca E M 2001 Three-dimensional unconstrained and constrained image-reconstruction techniques applied to fluorescence, frequency-domain photon migration *Appl. Opt.* **40** 2206

[60] Scheel A K, Backhaus M, Klose A D, Moa-Anderson B, Netz U, Hermann K G, Beuthan J, Müller G A, Burmester G R and Hielscher A H 2004 First clinical evaluation of sagittal laser optical tomography for detection of synovitis in arthritic finger joints *Ann. Rheum. Dis.* at press

[61] Schotland J and Markel V 2001 Inverse scattering with diffusing waves *J. Opt. Soc. Am.* A **18** 2767

[62] Schweiger M and Arridge S 1999 Optical tomographic reconstruction in a complex head model using *a priori* region boundary information *Phys. Med. Biol.* **44** 2703

[63] Schweiger M, Arridge S and Delpy D 1993 Application of the finite element method for the forward and inverse models in optical tomography *J. Math. Imaging Vis.* **3** 263

[64] Shenoy A, Heinkenschloss M and Cliff E M 1998 Airfoil design by an all-at-once method *Int. J. Comput. Fluid Dyn.* **11** 3

[65] Stefanov P and Uhlmann G 2003 Optical tomography in two dimensions *Methods Appl. Anal.* **10** 1

[66] Taroni P, Danesini G, Torricelli A, Pifferi A, Spinelli L and Cubeddu R 2004 Clinical trial of time-resolved scanning optical mammography at 4 wavelengths between 683 and 975 nm *J. Biomed. Opt.* **9** 464

[67] Tikhonov A and Arsenin V 1977 *Methods for Solving Ill-posed Problems* (New York: Wiley)

[68] Ye J, Webb K, Bouman C and Millane R P 1999 Optical diffusion tomography by iterative coordinate-descent optimization in a Bayesian framework *J. Opt. Soc. Am.* A **16** 2400

[69] Xu Y, Iftimia N, Jiang H, Key L L and Bolster M B 2001 Imaging of *in vitro* and *in vivo* bones and joints with continuous-wave diffuse optical tomography *Opt. Express* **8** 447

[70] Xu Y, Iftimia N, Jiang H, Key L and Bolster M 2002 Three-dimensional diffuse optical tomography of bones and joints *J. Biomed. Opt.* **7** 88