A PDE-constrained SQP algorithm for optical tomography based on the frequency-domain equation of radiative transfer

Hyun Keol Kim[1] **and Andreas H Hielscher**[1,2]

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

E-mail: hkk2107@columbia.edu and ahh2004@columbia.edu

Abstract

It is well acknowledged that transport-theory-based reconstruction algorithm can provide the most accurate reconstruction results especially when small tissue volumes or high absorbing media are considered. However, these codes have a high computational burden and are often only slowly converging. Therefore, methods that accelerate the computation are highly desirable. To this end, we introduce in this work a partial-differential-equation (PDE) constrained approach to optical tomography that makes use of an *all-at-once* reduced Hessian sequential quadratic programming (rSQP) scheme. The proposed scheme treats the forward and inverse variables independently, which makes it possible to update the radiation intensities and the optical coefficients simultaneously by solving the forward and inverse problems, all at once. We evaluate the performance of the proposed scheme with numerical and experimental data, and find that the rSQP scheme can reduce the computation time by a factor of 10–25, as compared to the commonly employed limited memory BFGS method. At the same time accuracy and robustness even in the presence of noise are not compromised.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The past decades have seen considerable developments in the theory and application of diffuse optical tomography (DOT). This emerging biomedical imaging modality has been mainly applied to brain imaging [1–5], breast imaging [6–8], finger-joint imaging [9–11] and small animal studies [12–15]. This method estimates the spatial distribution of optical properties in

tissues by analyzing intensity signatures measured at boundary surfaces. State-of-the-art image reconstruction codes employ a forward model of light propagation that leads to predictions of measured values on the boundary, assuming a certain distribution of optical properties inside the medium. An objective function is defined that quantifies the differences between predicted and actually measured values. The so-called model-based iterative image reconstruction algorithms (MOBIIR) [16–26] are employed to find the minimum of this objective function, by updating the parameters of the forward model.

Light propagation in tissue is typically modeled either by the equation of radiative transfer (ERT) or by its diffusion approximation (DA) depending on the physical character of the medium. It is well known that the validity of the DA becomes less accurate when applied to imaging of small tissue volumes and is further compromised if highly absorbing objects or fluid-filled regions, which contain, for example, cerebrospinal or synovial fluids, are considered [27]. Employing the ERT alleviates these problems and provides accurate prediction of intensity measurements for imaging of small-tissue geometries. However, simulating light propagation using the ERT requires much longer computation times as compared to the DA. Thus, it remains a challenging problem to develop computationally efficient transport-theory-based image-reconstruction schemes, since the reconstruction process requires a large number of repeated forward simulations that often lead to prohibitively long computing times. In this work we present a novel approach that is based on PDE-constrained optimization. Combining this with a reduced Hessian sequential quadratic programming (rSQP) scheme this method promises to accelerate the ERT-based image reconstruction process while maintaining the reconstruction accuracy.

To understand the differences between PDE-constrained and more traditional image reconstruction methods employed in DOT, we post the optical tomographic problem in the most general terms as

$$
\begin{aligned}
\text{minimize} \quad & f(x; u) = \tfrac{1}{2}|Qu - z^{\text{obs}}|^2 \\
\text{subject to} \quad & c(x; u) = A(x)u - b = 0,
\end{aligned}
\tag{1}
$$

where $x = (\mu_a, \mu_s) \in \mathbb{R}^n$ is the model parameter vector, $u \in \mathbb{Z}^m$ is the radiance vector and Q is the measurement operator; $f(x; u)$ is an objective function that quantifies the difference between measured and predicted intensities and $A(x)u = b$ (or $c(x; u) = 0$) is a discretized version of the forward transport equation. The problem given by (1) is often referred to as '*equality-constrained*' since the optimal solution at minimum of f has to satisfy the equality condition represented by $A(x)u - b = 0$.

Methods for solving (1) can be categorized into two groups depending on how the forward variable u and the inverse variable x are treated. The most common approach in DOT is to treat the forward variable u as a dependent variable of the inverse variable x, which makes it possible to replace the prediction vector u in f of (1) by its forward solution vector $A^{-1}b$. As a result, the problem (1) can be reformulated as

$$
\text{minimize} \quad \tilde{f}(x) = \tfrac{1}{2}|QA(x)^{-1}b - z^{\text{obs}}|^2
\tag{2}
$$

which is often referred to as '*unconstrained*' because equality $A(x)u - b = 0$ no longer appears in (2), i.e. $\tilde{f}$ is now a function of x only. Thus the forward solution vector $A^{-1}b$ has to be obtained for evaluation of the objective function (2), which explains why the unconstrained optimization scheme requires the complete solution of the forward problem at each iteration of optimization. As a consequence, the associated optimization procedure is a computationally very demanding process, both with respect to time and memory. Nonetheless, this approach has been widely used for the solution of optical tomographic problems mainly because of easiness of implementation. The existing optimization schemes [16–26, 28–31] belong to

this category and include the conjugate gradient (CG) approach [17, 25, 30], the quasi-Newton (QN) approach [21, 22, 24, 29] and the so-called Levenberg–Marquart (LM) method [11, 13, 26, 31]. It should be noted that equation (2) is often modified by adding simple-bound constraint [22, 29] (e.g. μ_a, $\mu_s > 0$) and regularization terms, such as the Tikhonov regularization [30, 31]. However, these modifications just constrain the model parameters and should not be confused with equality constraints, in which the constrained is the forward equation itself!

In another approach to solve equation (1), often referred to as 'PDE-constrained' optimization, the forward variable u and the inverse variable x are treated independently. This enables solving the equality-constrained problem (1) directly by updating the forward and inverse variables simultaneously at each iteration of optimization. Typically an extended objective function called '*Lagrangian*' is introduced as follows:

$$L(x, u; \lambda) \stackrel{\Delta}{=} \tfrac{1}{2}|Qu - z^{\mathrm{obs}}|^2 + \lambda^T (Au - b). \tag{3}$$

Here $\lambda \in \mathbb{Z}^m$ is called the vector of Lagrange multipliers. The simultaneous solutions of forward and inverse problems can be achieved at points satisfying the so-called *first-order Karush–Khun–Tucker (KKT) conditions* [32] where the gradient of L in (3) vanishes with respect to λ, u and x, respectively. Optimization problems of this kind are referred to as '*PDE-constrained*' because the forward and inverse solutions satisfy the first-order PDEs of L (i.e. constrained to the KKT conditions).

One major advantage of this PDE-constrained approach is that the complete solution of the forward problem is not required until the convergence is reached. So the solution accuracy of the forward problem can be appropriately controlled depending on the solution accuracy of the inverse problem, which leads to a significant saving in the total reconstruction time. In recent years, the field of PDE-constrained optimization has seen rapid developments mainly in applications related to airfoil design, flow variable optimization and electromagnetic inverse problems [33–39]. Recently our group (Abdoulaev *et al* [40]) has introduced this approach to optical tomographic imaging.

In particular Abdoulaev *et al* implemented an augmented Lagrangian method (ALM) for simultaneous solutions of the forward and inverse problems. However, the implemented code is required to preset a large number of convergence parameters, which differs from application to application, and makes the use of the code cumbersome. Furthermore, the use of non-reentry boundary condition for the frequency-domain transport-theory-based forward solver does not allow using the code with experimental data. In this case partially reflective boundary conditions are necessary to accurately model the light propagation in tissue. Finally, only a first-order discretization scheme was used for the forward model, which limits the accuracy of the reconstruction especially when applied to experimental data.

The approach suggested in this study goes beyond Abdoulaev's work in several respects. First we replace the ALM scheme with an all-at-once reduced Hessian SQP scheme (rSQP). The SQP scheme, which has been extensively studied in other fields [41–46], has the advantage that a much smaller number of convergence parameters have to be preset. In addition, many instability issues inherent to the ALM approach can be avoided that relate to the choice of a penalty parameter, the accuracy of a Lagrange multiplier and the solution unboundness [39]. Furthermore, we implement a second-order discretizing scheme for the forward model based on the frequency-domain transport equation and applied partially reflective boundary condition. These features allow us for the first time to apply the code to experimental data. We evaluate the performance of the rSQP scheme, with an emphasis on computational efficiency, by comparing the algorithm with the limited-memory Broyden–Fletcher–Goldfarb–Shanno (lm-BFGS) method [32] that is known to be the most efficient of the existing gradient-based

unconstrained optimization methods [21]. The remainder of the paper is organized as follows. We first describe a general framework of a reduced Hessian sequential quadratic programming scheme in section 2. The application of this approach to frequency-domain optical tomographic is presented in section 3. This is followed by numerical and experimental results addressing the performance evaluation of the rSQP scheme in sections 4 and 5. Finally we draw some conclusions in section 6.

2. Methods

2.1. Sequential quadratic programming

2.1.1. Background. To proceed with the SQP scheme, we begin with the PDE-constrained algebraic equations of (3) given as

$$
\begin{aligned}
L_\lambda &\triangleq \partial_\lambda L = Au - b = 0 \\
L_u &\triangleq \partial_u L = A^T \lambda + Q^T (Qu - z^{\text{obs}}) = 0. \\
L_x &\triangleq \partial_x L = \partial_x (Au) = 0
\end{aligned}
\tag{4}
$$

The first PDE equation in (4) is equivalent to the discretized version of the forward transport equation, and the second PDE equation can be viewed as the adjoint equation

$$
A^T \lambda = -Q^T (Qu - z^{\text{obs}}).
\tag{5}
$$

If vector $g(x, u)$ represents the gradient of $f(x, u)$ with respect to x and u, and the matrix C represent the Jacobian of constraints $c(x)$ with respect to x and u, we may define the following:

$$
\begin{aligned}
g &\triangleq \nabla_{x(u)} f = [f_x, f_u], \\
C &\triangleq \nabla_{x(u)} c = [C_x, C_u].
\end{aligned}
\tag{6}
$$

Using (6) the first-order KKT system given by equations (4) can be rewritten as

$$
\begin{aligned}
f_x + C_x^T \lambda &= 0 \\
f_u + C_u^T \lambda &= 0 \\
c &= 0
\end{aligned}
\tag{7}
$$

which can be solved by using the Newton method as

$$
\begin{bmatrix} W_{xx} & W_{ux} & C_x^T \\ W_{xu} & W_{uu} & C_u^T \\ C_x & C_u & 0 \end{bmatrix}
\begin{bmatrix} \Delta x \\ \Delta u \\ \Delta \lambda \end{bmatrix}
= -\begin{bmatrix} f_x + C_x^T \lambda \\ f_u + C_u^T \lambda \\ c \end{bmatrix}.
\tag{8}
$$

Here the matrix W denotes the Hessian matrix of the Lagrangian function with respect to x and u. The algebraic system given by (8) can be solved efficiently through the reduced Hessian SQP scheme as will be shown next.

2.1.2. Reduced Hessian SQP scheme. The reduced Hessian SQP scheme has been used outside the field of optical tomography to solve nonlinear constrained optimization problems with relatively low computational cost and fast convergence [33]. Employing the SQP method for solving the above system (8) iteratively is equivalent to minimizing the quadratic approximation of the Lagrangian function given by (3) subject to the linearization of the forward equation. We can formulate the following quadratic problem:

$$
\begin{aligned}
\text{minimize} \quad & \Delta p^{kT} g^k + \tfrac{1}{2} \Delta p^{kT} W^k \Delta p^k \\
\text{subject to} \quad & C^k \Delta p^k + c^k = 0.
\end{aligned}
\tag{9}
$$

4

Here $\Delta p = (\Delta x, \Delta u)$ at (x^k, u^k) and W^k is the full Hessian (or approximations) of the Lagrangian function. The linearized QP problems given by (9) only lead to a locally convergent algorithm. However, this can be overcome by incorporating a line search to ensure the global convergence as will be shown later. Equation (9) has a unique solution that satisfies

$$
\begin{aligned}
W^k \Delta p^k + C^{kT} \lambda^{k+1} &= -g^k, \\
C_x^k \Delta x^k + C_u^k \Delta u^k &= -c^k.
\end{aligned}
\tag{10}
$$

The full Hessian of the Lagrangian function is often difficult to obtain and its approximation by the updating schemes tends to create large dense matrices of size $(n + m) \times (n + m)$. These problems can be overcome by dropping certain non-critical second-order terms of the full Hessian matrix. In what follows, we describe the standard reduced Hessian SQP method based on separation of variables.

Since A_u^k is invertible, the vector Δp^k can be decomposed into two parts as follows:

$$
\Delta p^k = Z^k + Y^k \Delta x^k.
\tag{11}
$$

The choices of Z^k and Y^k are a challenging problem arising in practical implementation of the reduced SQP scheme. In this study, we have followed the popular choice for Z^k and Y^k given in (12) [33, 37, 42, 43, 46]:

$$
Z^k = \begin{bmatrix} 0 \\ -\left(C_u^k\right)^{-1} c^k \end{bmatrix} \qquad \text{and} \qquad Y^k = \begin{bmatrix} I \\ -\left(C_u^k\right)^{-1} C_x^k \end{bmatrix}.
\tag{12}
$$

Substituting (11) and (12) into the system (10) and using the equality $Y^{kT} C^{kT} = 0$, we can rewrite the full-space system (10) in the following reduced space form:

$$
H_r^k \Delta x^k = -g_r^k - d_r^k
\tag{13}
$$

where $H_r^k = Y^{kT} W^k Y^k$ denotes the reduced Hessian, $g_r^k = Y^{kT} g^k$ denotes the reduced gradient and $d_r^k = Y^{kT} W^k Z^k$ is called the cross-term. Thus the reduced SQP method requires much less memory than the full SQP method, i.e. only a small $(n \times n)$ matrix needs to be maintained and updated at each optimization iteration.

From (13), the inverse solution Δx^k and the forward solution Δu^k can be obtained respectively as

$$
\Delta x^k = -\left(H_r^k\right)^{-1} \left(g_r^k + d_r^k\right),
\tag{14}
$$

$$
\Delta u^k = Z^k - Y^k \left(H_r^k\right)^{-1} \left(g_r^k + d_r^k\right).
\tag{15}
$$

At the new iterate, the Lagrangian multiplier vector is updated from the second block of (7)

$$
\lambda^{k+1} = -\left[C_u^{(k+1)T}\right]^{-1} f_u^{k+1}.
\tag{16}
$$

Using (16) the reduced gradient can be reformulated as

$$
g_r^k = \left[I - C_x^{kT}\left(C_u^{kT}\right)^{-1}\right] \begin{bmatrix} f_x^k \\ f_u^k \end{bmatrix} = C_x^{kT} \lambda^k + f_x^k.
\tag{17}
$$

For large-scale applications, it is desirable to avoid the direct computation of the reduced Hessian H_r^k and its matrix inversion $\left(H_r^k\right)^{-1}$. Accordingly we approximate the matrix-vector product of $\left(H_r^k\right)^{-1} g_r^k$ directly by using the limited-memory updating formula, which is another important feature of our study that enables the proposed rSQP scheme to be applied to large-scale optimization. Also note that, similar to other works in this area [33, 42, 43], we ignore the cross-term vector $d_r^k = Y^{kT} W^k Z^k$. The effect of neglecting the cross-term vector may be

different from application to application. One major disadvantage is that we lose superlinear convergence. However we can still ensure a two-step superlinear convergence rate, which is good enough for our applications. Especially in this study the effect of the cross-term is not significant because the cross-term goes to zero as the iteration proceeds toward the convergence [33].

Theoretically, the rSQP method is not as effective as the full-SQP method of quadratic convergence. However it has been shown that the rSQP approach has a two-step superlinear convergence [33, 43, 46]. Therefore, one can expect an acceleration of optical tomographic imaging codes.

2.1.3. Merit function and line search.   The global convergence of the rSQP scheme is ensured by a line search on the following l_1 merit function defined as

$$\varphi_{\eta_k}(u, x) = f(u, x) + \eta \|Au - b\|_1, \tag{18}$$

which is chosen here for its simplicity and low computational cost. The directional derivative of $\varphi_\eta(u, x)$ along the descent direction $\Delta p = (\Delta x, \Delta u)$ is given by

$$\nabla \varphi_\eta(u, x; \Delta p) = g^{kT} \Delta p - \eta_k \|Au - b\|_1. \tag{19}$$

Thus, the descent property $\nabla \varphi_{\eta_k}(u, x; \Delta p) < 0$ can be maintained by choosing [43]

$$\eta_k > \lambda_k^T (Au - b) / \|Au - b\|_1. \tag{20}$$

At the new iterate given by $x^{k+1} = x^k + \alpha^k \Delta x$ and $u^{k+1} = u^k + \alpha^k \Delta u^k$, the merit function (18) is successively monitored to ensure the global progress toward the solution, while a line search is performed to find a step length that can provide a sufficient decrease in the merit function as [46]

$$\varphi_{\eta_k}(x^k + \alpha^k \Delta x^k; u^k + \alpha^k \Delta u^k) \leqslant \varphi_{\eta_k}(x^k, u^k) + 0.1 \alpha_k \nabla_{\eta_k}(x^k, u^k; \Delta p^k). \tag{21}$$

2.2. Applications to frequency-domain optical tomography

2.2.1. The light propagation model.   In this work we employ frequency-domain data. In this case the light source is typically amplitude modulated by frequencies in the range of 50–1000 MHz, and the demodulation and phase shift of the resulting photon-density waves that propagate inside the tissue are measured on the tissue boundary. It has been shown that the quality of frequency-domain reconstructions is superior to the steady-state approach [24, 25]. The frequency domain forward problem for light propagation in absorbing, scattering media can be accurately modeled by the frequency-dependent equation of radiative transfer [24, 25, 40], given by

$$(\nabla \cdot \Omega) \psi(\mathbf{r}, \Omega, \omega) + \left(\mu_a + \mu_s + \frac{i\omega}{c}\right) \psi(\mathbf{r}, \Omega, \omega) = \frac{\mu_s}{4\pi} \int_{4\pi} \psi(\mathbf{r}, \Omega^+, \omega) \Phi(\Omega^+, \Omega) \, d\Omega^+, \tag{22}$$

where $\psi(\mathbf{r}, \Omega, \omega)$ is complex radiation intensity in unit (W cm^{-2} sr^{-1}), μ_a and μ_s are the absorption and scattering coefficients, respectively, in units of (cm^{-1}), ω is the source modulation frequency, c is the speed of light inside the medium and $\Phi(\Omega^+, \Omega)$ is the scattering phase function that describes scattering from the incoming direction Ω^+ into the scattering direction Ω. Here we use the widely employed Henyey–Greestein phase function [47], given by

$$\Phi = \frac{1 - g^2}{(1 + g^2 - 2g \cos\theta)^{3/2}}. \tag{23}$$

Furthermore, we implemented a partially reflective boundary condition that allows us to consider the refractive index mismatch between the tissue and air. In particular this boundary condition is given by [48]

$$\psi_b(\mathbf{r}_b, \Omega, \omega)|_{\vec{n}_b \cdot \Omega < 0} = \psi^0(\mathbf{r}_b, \Omega, \omega) + R(\Omega', \Omega) \cdot \psi(\mathbf{r}, \Omega', \omega)|_{\vec{n}_b \cdot \Omega' > 0}, \quad (24)$$

where $R(\Omega', \Omega)$ is the reflectivity at Fresnel interface from direction Ω' to direction Ω, $\psi_b^0(\mathbf{r}, \Omega, \omega)$ is the radiation intensity due to the external source function and subscript b denotes the boundary surface of the medium, while $\vec{n}_b$ is the unit normal vector pointing outward at the boundary surface.

The spatial domain of medium under consideration is discretized using a node-centered finite-volume approach in combination with a discrete-ordinate formulation for the angular domain.   The node-centered finite-volume method combines the energy conservation properties of the finite-volume formulation and the geometric flexibility of the finite element approach.   When using an unstructured finite-volume discrete-ordinates method [48], the discretized form of radiative transfer equation is obtained by integrating equation (22) over the control volume with a divergence theorem as

$$\sum_{j=1}^{N_{\text{surf}}} (\vec{n}_j \cdot \Omega^m) \psi_j^m \, dA_j + \left(\mu_a + \mu_s + \frac{i\omega}{c} \right) \psi_N^m = \frac{\mu_s}{4\pi} \Delta V_N \sum_{m'=1}^{N_\Omega} \psi_N^{m'} \Phi^{m'm} w^{m'}, \quad (25)$$

where N_{surf} and N_Ω are the number of surfaces surrounding the node N and the number of discrete ordinates based on the level symmetric scheme, respectively, $\vec{n}_j$ and ψ_j^m denote the surface normal vector and the radiation intensity defined on the jth surface. Also the surface intensity ψ_j^m is related to the nodal intensity ψ_N^m by the second-order spatial differencing scheme that can be applied to unstructured meshes [49]. In this work we chose a node-centered scheme for constructing unstructured control volumes since it calculates more accurately the flux and requires much less memory, as compared to the cell-centered schemes [50]. For example, in three-dimensional cases, the number of tetrahedrons (i.e. cells) is usually about six times the number of nodes, which causes the cell-centered scheme to require six times as much memory as the node-centered scheme. Therefore the node-centered scheme offers an efficient way of saving memory and further accelerating the convergence.

After discretizations for all nodes, the resulting algebraic equation can be written as follows:

$$\sum_{j=1}^{N_{\text{nei}}} \gamma_{N_j}^m \psi_{N_j}^m + \sum_{m'=1}^{N_\Omega} \beta_N^{m'm} \psi_N^{m'} = b_N^m, \quad (26)$$

where $\gamma_{N_j}^m$ and $\beta_N^{m'm}$ depend only on medium optical properties or geometric properties. The boundary condition comes into the flux term after discretization on the boundary node as

$$b_N^m = -\sum_{j \in \Gamma} [1 - \max(n_j \cdot \Omega^m / |n_j \cdot \Omega^m|, 0)](n_j \cdot \Omega^m) \, dA_j \psi_{N_b \in \Gamma}^{0,m}, \quad (27)$$

where $\psi_{N_b \in \Gamma}^{0,m}$ is the external source function on boundary node N_b in direction m. It can be easily seen that equation (26) involves N_{nei} spatial unknown intensities at the neighboring nodes N_j in direction m, and N_Ω unknown intensities at node N. Equation (26) can be solved by any iterative solvers where all the radiation intensities ψ_N^m are updated simultaneously after each iteration. We use here a matrix-based iterative solver for fast convergence. To formulate the matrix system of the discretized equations (26), the radiation intensity ψ_N^m at node N and direction m will be represented by ψ_l, where l is expressed in terms of N and m as $l = (N-1) \cdot N_\Omega + m$. We define here the radiation intensity vector as $u = (\psi_1, \ldots, \psi_{N_t N_\Omega})$,

where N_t denotes the total number of nodes and finally can obtain the resulting linear algebraic equation as

$$Au = b. \tag{28}$$

Each line denoted by $l(l = 1, \ldots, N_t N_\Omega)$ of the matrix A contains the coefficients of the discretized form (26) established at node number N and direction m. Thus the coefficients of the matrix A_{lk} and the vector b_l can be given explicitly as

$$A_{lk} = \begin{cases} \beta_N^{m'm} & \text{for } m' = 1, \ldots, N_\Omega; k = (N-1)N_\Omega + m'; l = (N-1)N_\Omega + m \\ \gamma_{N_j}^m & \text{for } N_j = N_1, \ldots, N_{N_{nei}}; k = (N_j - 1)N_\Omega + m; l = (N-1)N_\Omega + m \\ 0 & \text{for other } l \text{ and } k \end{cases} \tag{29}$$

$b_l = b_N^m$.

In this study, the sparse matrix given by (29) contains complex-valued elements since we treat the frequency-domain equation of radiative transfer (22) directly, instead of separating it into two real-valued equations as found in other works [24, 40]. As a result, the complex-valued algebraic linear equation given by (28) is solved with a complex version of the GMRES linear solver [51, 52].

2.2.2. Optical tomographic inverse problems. In an all-at-once rSQP approach, the inverse problem associated with optical tomography is to find the radiation intensity vector $u = (\psi_1, \ldots, \psi_{N_t N_\Omega})$ and the optical property vector $x = (\mu_a^1, \ldots, \mu_a^{N_t}, \mu_s^1, \ldots, \mu_s^{N_t})$ so that

$$\min \quad f(x; u) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{d=1}^{N_d} (Q_d u_s - z_{s,d})(Q_d u_s - z_{s,d})^*$$
$$\text{s.t.} \quad A(\mu_a, \mu_s)u_s = b_s; \qquad s = 1, \ldots, N_s \tag{30}$$

where N_s and N_d are the numbers of sources and detectors used for measurements and predictions, $z_{s,d}$ and $Q_d u_s$ are the measurements and the predictions for source–detector pairs (s, d), and the operator $(\cdot)^*$ denotes the complex conjugate of the complex vector. Thus the optimization problem (30) is characterized by $N_t N_\Omega$ forward variables and N_t (or $2N_t$) inverse variables.

Applying the KKT condition to equation (30) gives the frequency-domain adjoint equation as

$$A^T \lambda_s = -\sum_{d=1}^{N_d} Q_d^T (Q_d u_s - z_{s,d})^*, \tag{31}$$

which is solved by the iterative GMRES method. The gradients of the objective function are obtained by differentiating $f(x; u)$ given by (30) with respect to $x^i = (\mu_a^i, \mu_s^i)$ at the ith control volume δV^i as

$$\nabla_{\mu_a^i} f = \sum_{s=1}^{N_s} (\lambda_s^T \delta V^i u_s)_{\text{Re}}, \tag{32}$$

$$\nabla_{\mu_s^i} f = \sum_{s=1}^{N_s} \left(\lambda_s^T \delta V^i \left(u_s - \frac{1}{4\pi} \sum_{m'} u_s \Phi w \right) \right)_{\text{Re}}, \tag{33}$$

which represents the reduced gradient vector $g_r = (\nabla_{\mu_a} f, \nabla_{\mu_s} f)^T$ as described in (17).

2.2.3. Computational implementation of the reduced SQP scheme. The algorithm that makes use of the reduced Hessian SQP method as described in sections 3.1 and 3.2, has the following structure:

(1) Set $k = 0$. Pick initial guess x^0 and obtain the starting point u^0 from solving $A(x^0)u^0 = b$, and initialize the merit function parameter as $\eta_0 = 1$. Assume the initial reduced Hessian as $H^0 = I$.

(2) Evaluate c^0, g^0, $C_{x(u)}^0$ at (x^0, u^0), and compute Y^0 and Z^0 given by (12).

(3) Solve for λ^0 from $C_u^{0T}\lambda^0 = -f_u^0$ by using a GMRES iterative solver.

(4) Evaluate the reduced gradient $g_r^0 = C_x^{oT}\lambda^0 + f_x^0$.

(5) Check the convergence: if the stopping criterion is satisfied, stop.

(6) Calculate the search direction Δx^k from $\Delta x^k = (H_r^k)^{-1}g_r^k$ via the limited memory BFGS updating formula.

(7) Solve for the forward variables Δu^k from $C_u^k \Delta u^k = -C_x^k \Delta x^k - c^k$ by a GMRES linear solver.

(8) Set $\alpha_k = 1$ and check if it ensures the sufficient decrease in the merit function

$$\varphi_{\eta_k}(x^k + \alpha^k \Delta x^k; u^k + \alpha^k \Delta u^k) \leqslant \varphi_{\eta_k}(x^k, u^k) + 0.1\alpha_k \nabla_{\eta_k}(x^k, u^k; \Delta p^k)$$

where $\varphi_\eta(u, x) = f(u, x) + \eta \|Au - b\|_1$.

(9) If the sufficient decreasing condition is satisfied by the searched step length, then set

$$x^{k+1} = x^k + \alpha^k \Delta x^k \qquad \text{and} \qquad u^{k+1} = u^k + \alpha^k \Delta u^k;$$

otherwise,

$$\alpha^k = \max\left\{\frac{-0.5\nabla\varphi_{\eta_k}(x^k, u^k; \Delta p^k)(\alpha^k)^2}{\varphi_{\eta_k}(x^k + \alpha^k \Delta x^k; u^k + \alpha^k \Delta u^k) - \varphi_{\eta_k}(x^k) - \alpha^k \nabla\varphi_{\eta_k}(x^k, u^k; \Delta p^k)}, 0.1\right\}$$

when line search fails, the algorithm solves the linearized forward solution (step 7) more accurately and performs line search again.

(10) Evaluate c^{k+1}, g^{k+1} and C^{k+1}, and compute Y^{k+1} and Z^{k+1}.

(11) Solve for λ^{k+1} from $\lambda^{k+1} = -\left[C_u^{(k+1)T}\right]^{-1}f_u^{k+1}$ with a GMRES solver and update the merit function parameter η_k by

$$\eta_{k+1} = \max(1.001 + \|\lambda_{k+1}\|_\infty, (3\eta_k + \|\lambda_{k+1}\|_\infty)/4, 10^{-6})$$

(12) Evaluate $g_r^{k+1} = C_x^{(k+1)T}\lambda^{k+1} + f_x^{k+1}$.

(13) Evaluate $y^k = g_r^{k+1} - g_r^k$ and $s^k = x^{k+1} - x^k$.

(14) Return to step 5 with $k = k + 1$ to check the convergence.

Note that our algorithm stated above is similar to the reduced Hessian SQP algorithm proposed by Biegler and Nocedal [46]. However, we employ here the frequency-domain ERT as constraint, while Biegler and Nocedal used a much simpler nonlinear quadratic function as constraint.

As mentioned earlier, the rSQP algorithm does not require the exact solution of the radiative transfer equation during the reconstruction process. Instead the rSQP scheme solves the linearized forward equation as described in step 7, which allows us to utilize the incomplete solution obtained with the loose tolerance (10^{-2}–10^{-3}). On the other hand, the unconstrained (lm-BFGS) method requires the accurate solution of a forward problem with the tolerance of 10^{-10}.
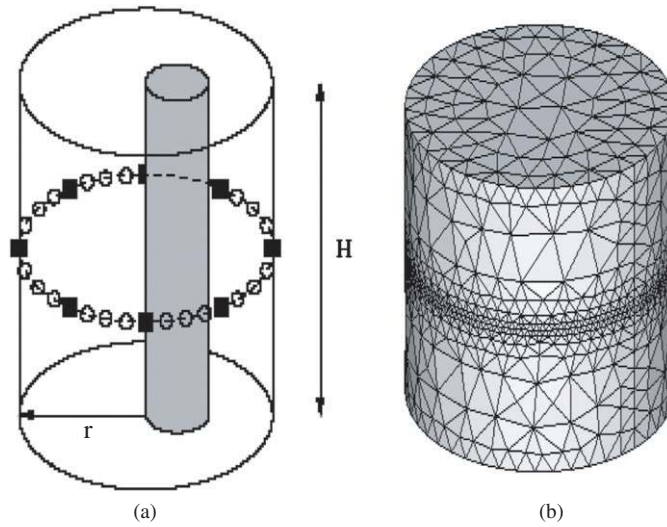
Figure 1. The schematic of the test problems 1 and 2: cylinder height $H = 2$ cm and radius $r = 1$ cm. (a) Source–detector configuration: 8 sources (■) and 64 detectors (○), and (b) computation domain with 6764 tetrahedrons.

3. Results

3.1. Numerical experiments

In this section, we show numerical results for reconstructions of the spatial distributions of optical properties inside the medium by using the rSQP method and the lm-BFGS method. Note that the lm-BFGS approach implemented in this study generates the quasi-Newton matrices using information from the last six iterations in a very efficient way by employing the recursive formula given in [32]. To illustrate the code performance, we compare the results of both schemes on three types of test problems, all with cylindrical geometries.

3.1.1. Setup of the test problems. The geometry for the first two cases in this study consists of a cylinder with diameter of 2 cm and a height of 2 cm in which a smaller cylinder with a diameter of 0.5 cm is embedded (figure 1). This geometry mimics small animal experiments [12–15] or measurements on finger joints [9–11]. In the first example, the smaller cylinder differs in μ_a from the larger cylinder, while μ_s is the same (see table 1). The absorption coefficient of the smaller cylinder is set to be twice as high ($\mu_a = 0.2$ cm^{-1}) as that of the background medium ($\mu_a = 0.1$ cm^{-1}). Note that all sources and detectors are located on a circle defined by $\Gamma = \{(x, y, z) | x^2 + y^2 = 1, z = 1\}$. The second problem involves the reconstruction of μ_s. Therefore instead of varying μ_a inside the smaller cylinder, we now increase μ_s inside the smaller cylinder to 15 cm^{-1}, while the background medium stays at $\mu_s = 10$ cm^{-1}.

In the third example, we consider the case of reconstructing μ_a and μ_s simultaneously. In this case, two spherical objects are embedded in a background medium that varies in both the absorption coefficient μ_a and the scattering coefficient μ_s. The sources and detectors are located on two circles, defined by $\Gamma = \{(x, y, z) | x^2 + y^2 = 1.5^2, z = 2.2\}$
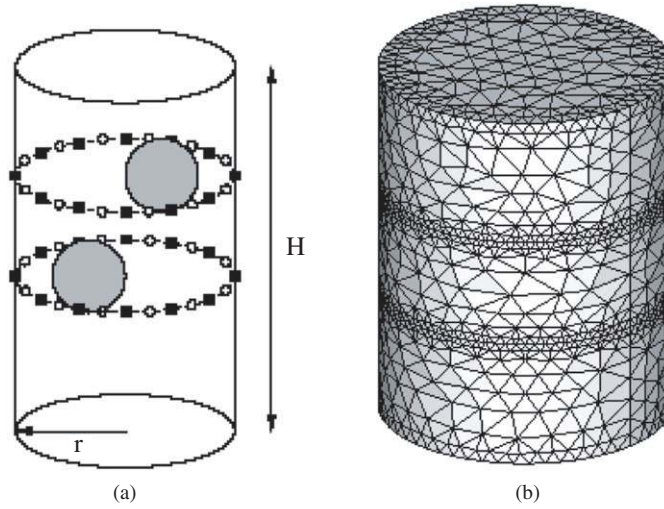
Figure 2. The schematic of the test problem 3: cylinder height $H = 5$ cm, radius $r = 1.5$ cm and two embedded spherical objects with diameter of 0.5 cm. (a) Source–detector configuration at planes $z = 2.2$ cm and $z = 3.2$ cm: 24 sources (■) and 24 detectors (○), and (b) computation domain with 14 114 tetrahedrons.

Table 1. Parameters used in three different examples.

	Problem 1	Problem 2	Problem 3
Anisotropic 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	600
Number of finite volumes	6764	6764	14 114
Number of discrete ordinates	24	48	80

and $\Gamma = \{(x, y, z)|x^2 + y^2 = 1.5^2, z = 3.2\}$, respectively (see figure 2). The corresponding values of optical properties for all three cases are given in table 1.

For the numerical experiments, we use synthetic data that are obtained from the solution of the frequency-domain forward problem given by (22) at specified detector locations for the exact optical properties. All synthetic data are generated on a mesh that is two times finer than the mesh used for the reconstructions. The solution of the forward problem at detector locations provides the exact 'measurements' z^{ex}. Measurements containing noise are simulated by adding an error term to z^{ex} in the form $z^{\text{obs}} = z^{\text{ex}} + \varpi\sigma$, where σ is the standard deviation of measurement errors and ϖ is the random variable with normal distribution, zero mean and unitary standard deviation. With the use of such simulated measurements as the input data for the inverse analysis, we examine the stability of the inverse solution with respect to various noise levels by generating data with different standard deviations obtained from the signal-to-noise ratio, SNR $= 10\log(z/\sigma)$.

11

We stop the optimization process when the following stopping criteria are satisfied:

$$|f^{k+1}(u; x) - f^k(u; x)|/f^k(u; x) \leqslant \varepsilon_1, \tag{34a}$$

$$f^k(u; x) \leqslant \varepsilon_2. \tag{34b}$$

The tolerance ε_1 is set to 10^{-5}, and ε_2 is chosen to have the same order of magnitude of measurement errors, which leads to sufficiently stable results in the *principle of discrepancy* [53]. When noise-free data is considered, the tolerance ε_2 is assigned a sufficiently small number (typically 10^{-6}). The stopping criteria given by (34a), (34b) are applied in the same way to the reconstruction codes that make use of the rSQP method and the lm-BFGS method.

To quantify the quality of reconstruction, we use the correlation factor $\rho(\mu^e, \mu^r)$ and the deviation factor $\delta(\mu^e, \mu^r)$ as introduced in [21]:

$$\rho(\mu^e, \mu^r) = \frac{\sum_{i=1}^{N_t} (\mu_i^e - \bar{\mu}_i^e)(\mu_i^r - \bar{\mu}_i^r)}{(N_t - 1)\sigma(\mu^e)\sigma(\mu^r)}, \qquad \delta(\mu^e, \mu^r) = \frac{\sqrt{\sum_{i=1}^{N_t} (\mu_i^e - \mu_i^r)^2 N_t}}{\sigma(\mu^e)}. \tag{35}$$

Here $\bar{\mu}$ and $\sigma(\mu)$ are the mean value and the standard deviation for the spatial function of the optical property than can be either μ_a and μ_s. Similarly, μ^e and μ^r are the exact and reconstructed distributions of optical properties, respectively. The correlation coefficient indicates the degree of correlation between exact and estimated quantities while the deviation factor describes the discrepancy in absolute values of exact and estimated quantities. Accordingly, the closer $\rho(\mu^e, \mu^r)$ gets to unity, and the closer $\delta(\mu^e, \mu^r)$ gets to zero, the better is the quality of the reconstruction.

In the following sections, the rSQP method and the lm-BFGS method are applied to functional estimations of unknown optical properties for the three test problems as given in table 1. All the simulations are carried out on a Pentium IV 3.0 GHz CPU processor.

3.1.2. CPU time and influence of noise. To compare the CPU time and the influence of noise in rSQP-based and lm-BFGS-based algorithms, we consider the reconstruction of μ_a inside the target medium as shown in figure 1. We test the effects of noise on the algorithm by varying the SNR from infinity (no noise) to 20 dB and 15 dB, with the later two values representing typical noise levels encountered in optical tomography [21].

As shown in table 2, the usage of the rSQP method results in a significant reduction of computation time in all cases considered here. For the case of the noise-free data, the rSQP method converges in 0.44 h while the lm-BFGS method takes about 5 h to reach the same convergence criterion. Therefore, the rSQP method reduces the computation time by a factor of about 11. We observe a similar reduction in the two other cases of different noise levels. The 20 dB data requires 0.2 h using the rSQP method and 2.18 h using the lm-BFGS method, respectively. With the 15 dB data the rSQP method converges in 0.15 h while the lm-BFGS method requires 3.72 h. Therefore, in this case, using the rSQP method reduces computation time by a factor of about 25.

The main reason for this significant reduction in CPU time can be explained by the fact that the rSQP method does not require the exact solution of the forward problem at each iteration of optimization until it converges to the optimal solution, as mentioned earlier. Indeed we obtain the incomplete solution of the linearized forward equation by applying the loose tolerance of 10^{-2} that is empirically chosen from our extensive study. We found that even a tolerance of 10^{-2} can generate a sequence of feasible solutions satisfying the first-order necessary conditions (KKT conditions), while advancing to the true solution through the optimization process. For example, we usually stop the GMRES iteration for the forward
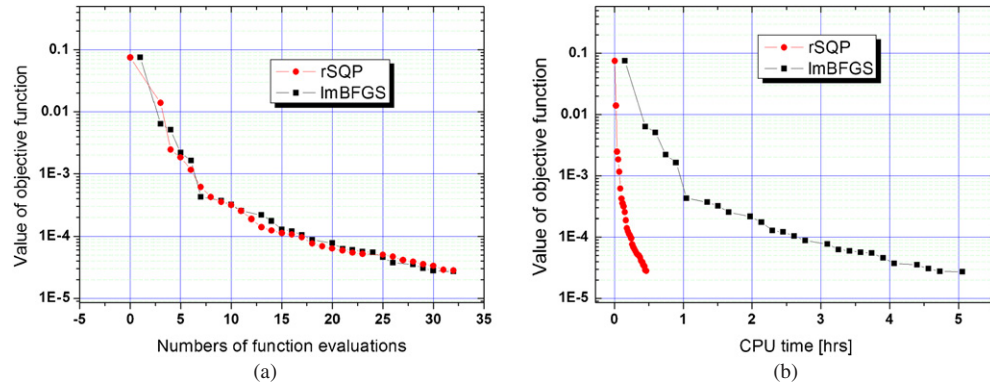
Figure 3. Variations of the value of the objective function (in $\log_{10}$ scale) obtained with the rSQP and lm-BFGS methods for the reconstruction μ_a with noise-free data. A convergence history for each of the two methods is depicted with respect to (a) the number of function evaluations and (b) the CPU time.

Table 2. Comparison of the CPU time and the reconstruction quality for codes that make use of rSQP and lm-BFGS methods. CPU time, rho and delta are given for varying noise levels in the input data of problem 1 (see table 1).

	Schemes	CPU time (*)	Correlation $\rho(\mu^e, \mu^r)$	Deviation $\delta(\mu^e, \mu^r)$
Noise free	rSQP	**0.44 h (11.5)**	**0.79**	**0.65**
	lm-BFGS	5.05 h	0.78	0.65
20 dB	rSQP	**0.20 h (10.9)**	**0.63**	**0.93**
	lm-BFGS	2.18 h	0.64	0.85
15 dB	rSQP	**0.15 h (24.8)**	**0.47**	**1.03**
	lm-BFGS	3.72 h	0.42	0.99

* denotes the acceleration factor by the rSQP method.

solution when its relative residual becomes smaller than 10^{-10}. Thus the GMRES iteration process essentially has to perform a sufficient number of matrix-vector multiplications to reach the desired accuracy, which is the case using the unconstrained lm-BFGS method. However, the rSQP scheme stops the GMRES iteration at a much earlier stage of the iterations based on the loose tolerance. As a consequence, the rSQP scheme only requires a relatively smaller number of matrix-vector multiplications, which leads to a significant time savings through the overall optimization process. Figures 3(a) and (b) show the convergence history of the two methods versus the number of function evaluations and the CPU times. As expected, it is observed that the rSQP scheme produces a similar convergence history with respect to the iteration number (figure 3(a)) but it shows much faster convergence with respect to the computation time (figure 3(b)).

In addition to the CPU time, we measured the accuracy of reconstruction as a function of the SNR. The correlation factor $\rho(\mu^e, \mu^r)$ and the deviation factor $\delta(\mu^e, \mu^r)$ as defined in (35) are computed for the plane $z = 1$ where the sources and the detectors are located. The corresponding values of $\rho(\mu^e, \mu^r)$ and $\delta(\mu^e, \mu^r)$ are given in table 2. Figure 4 shows the cross-section maps of the reconstructed μ_a for the xy-plane (first row in figure 4) at $z = 1$ and the xz-plane (second row in figure 4) at $y = 0$. The upper and lower parts of the xz-plane images contain the optical properties of the unchanged initial guess due to the lack of sources
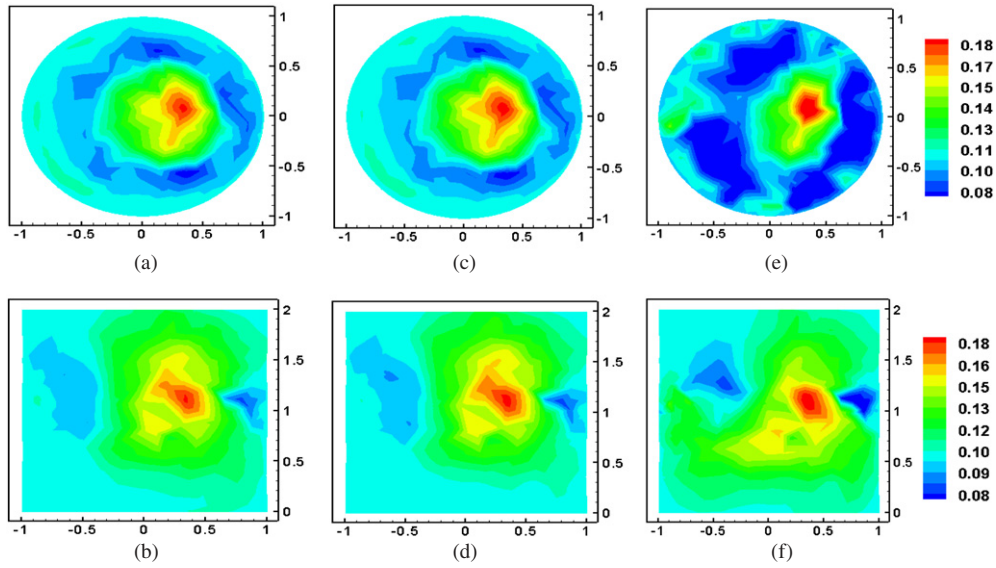
Figure 4. Cross-section maps of the reconstructed μ_a-value obtained for problem 1 (see table 1). The maps are shown for the *xy*-plane at $z = 1$ and *xz*-plane at $y = 0$. (a) and (b) lm-BFGS results with noise-free data; (c) and (d) rSQP result with noise-free data; (e) and (f) rSQP reconstructions with 15 dB noise-added data.

and detectors at these regions. As a result, one can observe the reconstructed object only around the region where all sources and detectors are placed.

At noise levels of 15–20 dB, the difference between the rSQP and lm-BFGS methods are not significant; both schemes show a decrease in the correlation factor and an increase in the deviation factor as compared to the corresponding values of the noise-free data (∞ dB, see table 2). Therefore, it can be stated that the rSQP and lm-BFGS methods are similar to each other in terms of response to noise in the data.

3.1.3. Effect of initial guess. Typically the exact optical properties of the background medium are not known *a priori*, and the optimization scheme starts with certain reasonable choice for these properties. It can be assumed that there is always some mismatch between the true background medium and the guessed background medium, which can affect the reconstruction accuracy. For this reason, we evaluate the robustness of the rSQP code to this initial guess of optical properties.

For this study we consider the same geometry as the first example but this time we change μ_s inside the smaller cylinder as shown in figure 1. The corresponding optical properties are given in table 1. We looked at two cases. First we assumed that the initial guess of the background optical properties is 10% higher than the true value; in the second example the initial guess is 20% higher. The simulations are performed with noise-free data and CPU times are measured. In addition, we calculated the correlation factor $\rho(\mu^e, \mu^r)$ and the deviation factor $\delta(\mu^e, \mu^r)$ to discuss the robustness of the rSQP and lm-BFGS codes to initial guess. The results are given in figure 5 and table 3.

It can be seen from table 3 that the rSQP method leads to much more accurate results than the lm-BFGS method when the initial guesses of the optical properties of the background
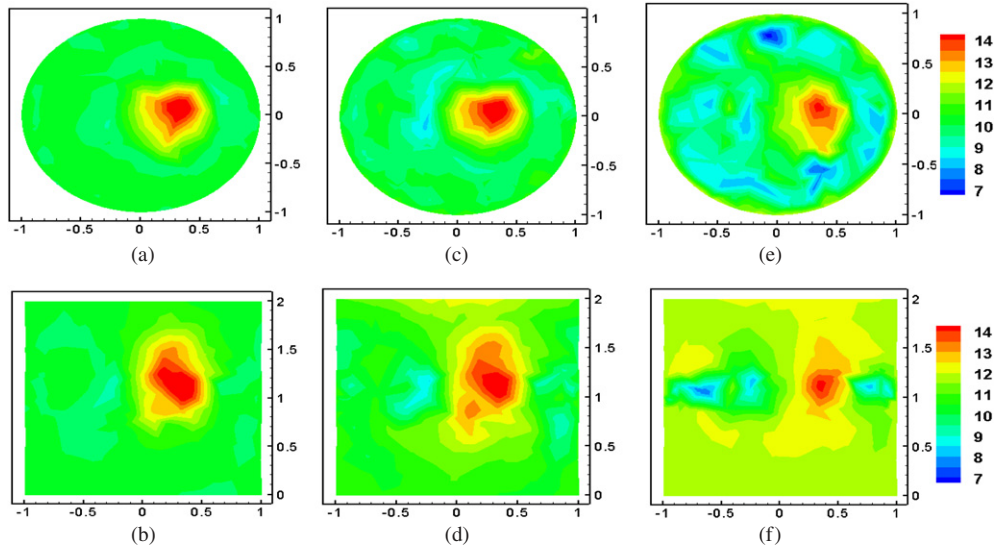
Figure 5. Cross-section maps of the reconstructed μ_s value for problem 2 (see table 1) obtained by using the rSQP method. Maps shown represent result for the xy-plane at $z = 1.0$ and xz-plane at $y = 0$, respectively. (a) and (b) Reconstruction results based on noise-free data; (c) and (d) reconstruction results obtained by starting with initial guess of optical properties that is 10% higher than true optical properties; (e) and (f) reconstruction results obtained by starting with initial guess of optical properties that is 20% higher than true optical properties.

Table 3. The reconstruction quality and the computation time obtained with the two methods for different initial guesses for the second example of reconstructing μ_s.

	Schemes	CPU time (*)	Correlation $\rho(\mu^e, \mu^r)$	Deviation $\delta(\mu^e, \mu^r)$
Background	RSQP	**4.3 h (12.2)**	**0.89**	**0.57**
	lm-BFGS	52.4 h	0.88	0.53
10% higher	RSQP	**3.7 h (16)**	**0.78**	**0.73**
	lm-BFGS	59.2 h	0.62	0.87
20% higher	RSQP	**3.6 h (16.9)**	**0.53**	**0.98**
	lm-BFGS	60.8 h	0.41	1.69

* denotes the acceleration factor by the rSQP method.

medium are 10% or 20% higher. The correlation factor $\rho(\mu^e, \mu^r)$ using the rSQP method is 0.78 and 0.53 for the 10% and 20% cases, respectively. This is almost 20% better than when the lm-BFGS method is used ($\rho(\mu^e, \mu^r) = 0.62$ and 0.41, respectively). Also the rSQP method shows 20–50% smaller deviation factors ($\delta(\mu^e, \mu^r) = 0.73$ and 0.98 for 10% and 20% cases, respectively), as compared to the lm-BFGS method ($\delta(\mu^e, \mu^r) = 0.87$ and 1.69).

Furthermore, the CPU times are 16 times shorter (3.7 h as compared to 59.2 h in 10% case, and 3.6 h versus 60.8 h in 20% case). We also observed that the rSQP method still converges to reasonable solutions even when initial guess is made 50% larger than the true value, whereas the lm-BFGS method leads to a premature convergence for this same case. The reason for this observation is not quite clear but may be understood by looking at how the variables are treated in these two different schemes. The unconstrained scheme treats the intensity vector u as a dependent variable of the model parameter vector x, while the PDE-constrained scheme
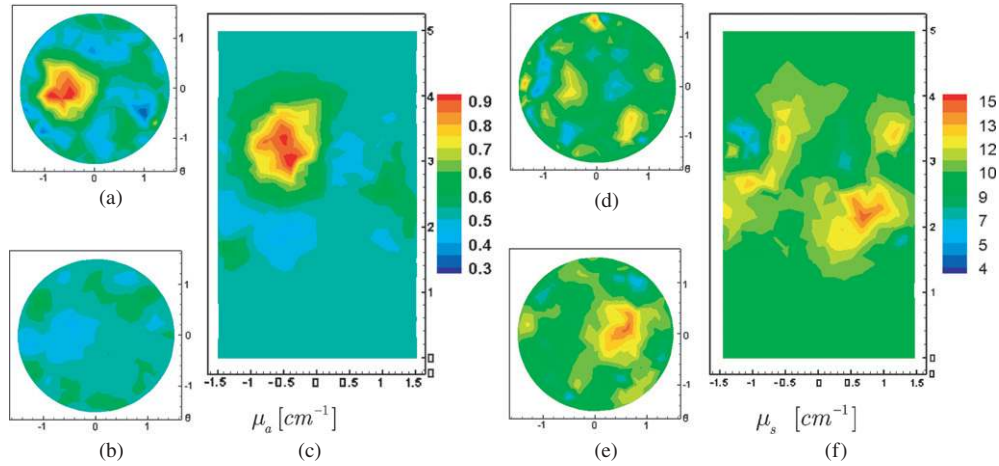
15

Figure 6. Cross-section maps of the reconstructed μ_a(a) –(c) and μ_s (d)–(f) values for problem 3 (see table 1) obtained by using the rSQP method. The maps show results for the *xy*-planes at $z = 2.2$, $z = 3.2$ and *xz*-plane at $y = 0$.

Table 4. The reconstruction quality and the computation time obtained with the rSQP method for the third example in which the spatial distributions of μ_a and μ_s are reconstructed.

CPU time	Plane	Correlation $\rho(\mu^e, \mu^r)$	Deviation $\delta(\mu^e, \mu^r)$
51 h	$z = 2.25$ cm	**0.79**	**0.60**
	$z = 3.25$ cm	**0.53**	**0.92**

treats these two variables as independent variables. Accordingly the lm-BFGS method can be stronger influenced by a wrong starting value as compared to the rSQP method.

3.1.4. Reconstruction of absorption and scattering coefficients. Finally we reconstruct μ_a and μ_s simultaneously. The test medium contains two objects: one object that has an increased μ_a value and another object that has an increased μ_s value as compared to the background medium (see table 1—third column, and figure 2). Finer spatial (14 114 tetrahedrons) and angular (S_8) meshes are used in order to capture the effects of the highly anisotropic scattering. Since we already discuss the code performances of the two methods in the previous two sections, we limit our attention here to the rSQP scheme.

The 20 dB noise-added synthetic data are used as the input. The optimization process is started with an initial guess that assumes a homogeneous medium with the optical properties being the same as the optical properties of the true background medium. The reconstructed cross-section maps of the μ_a and μ_s are shown in figure 6 for *xy*-planes at $z = 2.2$ and $z = 3.2$, and for the *xz*-plane at $y = 0$. As expected (see for example, Kui *et al* [24]), we observe that the absorbing and scattering objects are well separated, which is mainly due to the use of a higher modulation frequency. Also it can be seen from table 4 that the rSQP method offers reasonable accuracy with respect to the two metrics $\rho(\mu^e, \mu^r)$. As a result, we demonstrate that the rSQP scheme can reconstruct μ_a and μ_s simultaneously within reasonable accuracy by mitigating the cross-talk problem between the objects.
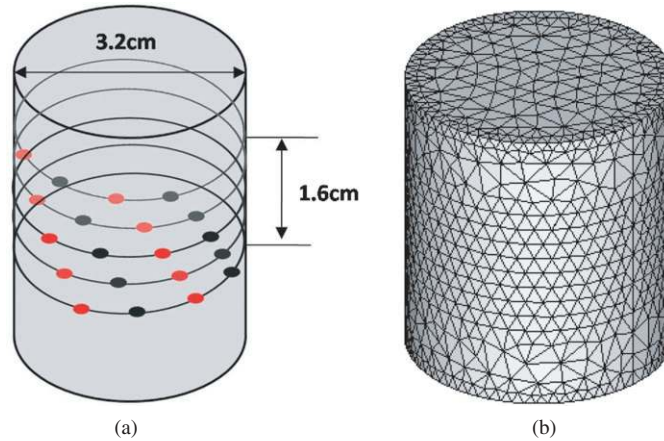
16

Figure 7. Schematic of the phantom used for experimental studies; (a) locations of 25 sources (●) and 25 detectors (●), (b) the computation domain with 28 852 tetrahedrons. The diameter of a cylinder is 3.2 cm and the height 3.6 cm.
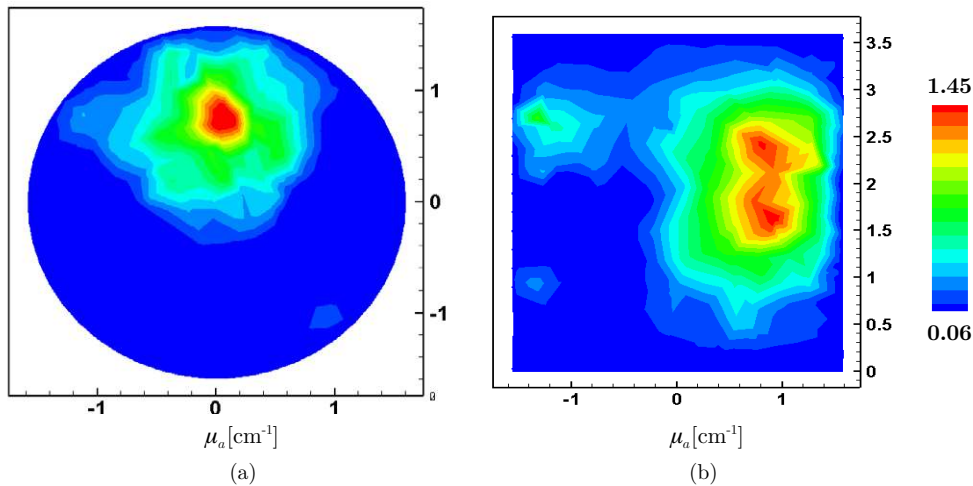

Figure 8. Cross-section maps of the reconstructed-μ_a value using experimental data obtained at 800 MHz with the tissue phantom shown in figure 7. Reconstruction results were generated with the rSQP code and results are shown for the xy-plane at $z = 1.8$ (a) and the yz-plane at $x = 0$ (b).

3.2. Application to experimental data

In addition to the numerical studies, we also apply our code to experimental data. For the experimental studies we use a tissue phantom with optical properties similar to those used in the numerical studies. A cylinder with a 3.2 cm diameter (figure 7) is filled with 0.55% intralipid. The anisotropy of intralipid g is 0.71. The reduced scattering coefficient of intralipid is $\mu'_s = 3.5$ cm^{-1}. The absorption coefficient of the background is known as $\mu_a = 0.027$ cm^{-1}. The perturbation consists of a simple absorption inclusion rod (with the same scattering properties as the background) placed vertically about 1 cm away from the surface.

The absorbing cylinder rod is filled with India ink whose μ_a is about 3.5 cm^{-1}. The 25 sources and 25 detectors are placed around the phantom surface. The source–detector configuration is shown in figure 7. The measurement data were obtained at modulation frequency of 800 MHz, which yields reasonable amplitude SNR using our instrument and provides for large enough phase shifts that can be measured with sufficient SNR.

The reconstruction of μ_a is carried out with the rSQP method and the lm-BFGS method. The reconstruction results using the rSQP methods can be seen in figure 8. As found in the numerical study, we observe that both the location of the smaller cylinder and the spatial distribution of μ_a can be recovered using the rSQP method. As for the total image reconstruction time, the rSQP method only took 9 min to converge while the lm-BFGS method converged after 3.15 h. Therefore, the convergence speed was increased by a factor of 21, while the quality of the reconstructed images was maintained.

4. Conclusions

To increase the convergence speed of transport-theory-based reconstruction algorithm for optical tomography, we considered in this work a PDE-constrained approach that makes use of a reduced Hessian all-at-once sequential quadratic programming (rSQP) method. The proposed algorithm solves the forward and inverse problems *all at once* by updating the radiances and the optical coefficients simultaneously in one iteration process. The frequency-domain equation of radiative transfer was employed as a light transport model. To evaluate the performance of the rSQP scheme, we performed numerical experiments varying the geometry and the optical coefficients of a test problem and compare the results of the rSQP method and the lm-BFGS method (the fastest unconstrained scheme in OT) in terms of computation time, accuracy and robustness.

We found that the rSQP method can lead to a significant reduction in CPU time. Depending on the measurement noise and the optical properties of the medium considered, we observed 10- to 25-fold faster convergence compared to lm-BFGS-based codes. We also studied the impact of noise on the image quality in the reconstruction process. While the image quality depends on the signal-to-noise level of the data, there appear to be little difference between the cases when rSQP or lm-BFGS methods were used. Investigating the influence of the initial guess of the background medium on the accuracy on the reconstruction results, we found that the rSQP method is more robust than the lm-BFGS method. Therefore, in the examples considered, the rSQP method would converge to a reasonable solution, even when the lm-BFGS code would not. Furthermore, we showed that the rSQP scheme can be successfully applied to simultaneous estimation of μ_a and μ_s. Finally, we employed the rSQP and lm-BFGS codes to reconstruct the optical property distribution in a tissue phantom using real measurement data. Increase in convergence speed and reconstruction accuracy were comparable to the results observed in the numerical studies.

Acknowledgments

References

[1] Xu Y, Graber H L and Barbour R L 2007 Image correction algorithm for functional three-dimensional diffuse optical tomography brain imaging *Appl. Opt.* **46** 1693–704

[2] Joseph D K, Huppert T J, Franceschini M A and Boas D A 2006 Diffuse optical tomography system to image brain activation with improved spatial resolution and validation with functional magnetic resonance imaging *Appl. Opt.* **45** 8142–51

[3] 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

[4] Hebden J C, Gibson A P, Austin T, Yusof R M, Everdell N, Delpy D T, Arridge S R, Meek J H and Wyatt J S 2004 Imaging changes in blood volume and oxygenation in the newborn infant brain using three-dimensional optical tomography *Phys. Med. Biol.* **49** 1117–30

[5] Blueston 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

[6] Davis S C, Dehghani H, Wang J, Jiang S, Pogue B W and Paulsen K D 2007 Image-guided diffuse optical fluorescence tomography implemented with Laplacian-type regularization *Opt. Express* **15** 4066–82

[7] Corlu A, Choe R, Durduran T, Rosen M A, Schweiger M, Arridge S A, Schnall M D and Yodh A G 2007 Three-dimensional in vivo fluorescence diffuse optical tomography of breast cancer in humans *Opt. Express* **15** 6696

[8] Choe R *et al* 2005 Diffuse optical tomography of breast cancer during neoadjuvant chemotherapy: a case study with comparison to MRI *Med. Phys.* **32** 1128–39

[9] 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

[10] Scheel A K, Backhaus M, Klose A D, Moa-Anderson B, Netz U J, Hermann K G, Beuthan J, Muller 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.* **64** 239–45

[11] Zhang Q and Jiang H 2005 Three-dimensional diffuse optical tomography of simulated hand joints with a 64 × 64-channel photodiodes-based optical system *J. Opt. A Pure Appl. Opt.* **7** 224–31

[12] Hielscher A H 2005 Optical tomographic imaging of small animals *Curr. Opin. Biotechnol.* **16** 79–88

[13] Xu H, Springett R, Dehghani H, Pogue B W and Paulsen K D 2005 Magnetic-resonance-imaging-coupled broadband near-infrared tomography system for small animal brain studies *Appl. Opt.* **44** 2177–88

[14] Bluestone A Y, Stewart M, Lei B, Kass I S, Lasker J, Abdoulaev G S and Hielscher A H 2004 Three-dimensional optical tomographic brain imaging in small animals: Part I. Hypercapnia *J. Biomed. Opt.* **9** 1046–62

[15] Bluestone A Y, Stewart M, Lasker J, Abdoulaev G S and Hielscher A H 2004 Three-dimensional optical tomographic brain imaging in small animals: Part II. Unilateral carotid occlusion *J. Biomed. Opt.* **9** 1063–73

[16] Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** R41–93

[17] Hielscher A H, Klose A D and Hanson K M 1999 Gradient-based iterative image reconstruction scheme for time-resolved optical tomography *IEEE Trans. Med. Imaging* **18** 262–71

[18] Yao Y Q, Wang Y, Pei Y L, Zhu W W and Barbour R L 1997 Frequency-domain optical imaging of absorption and scattering distributions by Born iterative method *J. Opt. Soc. Am. A* **14** 325–42

[19] Ye J C, Webb K J, Bouman C A and Millane R P 1999 Optical diffusion tomography by iterative-coordinate-descent optimization in a Bayesian framework *J. Opt. Soc. Am. A* **16** 2400–12

[20] Ripoll J and Ntziachristos V 2003 Iterative boundary method for diffuse optical tomography *J. Opt. Soc. Am. A* **20** 1103–10

[21] Klose A D and Hielscher A H 2003 Quasi-newton methods in optical tomographic image reconstruction *Inverse Problems* **19** 309–87

[22] Roy R and Sevick-Muraca E M 2000 Active constrained truncated Newton method for simple-bound optical tomography *J. Opt. Soc. Am. A* **17** 1627–41

[23] Intes X, Ntziachristos V, Culver J P, Yodh A and Chance B 2002 Projection access order in algebraic reconstruction technique for diffuse optical tomography *Phys. Med. Biol.* **47** N1–10

[24] Ren R, Bal G and Hielscher A H 2006 Frequency domain optical tomography with the equation of radiative transfer *SIAM J. Sci. Comput.* **28** 1463–89

[25] Kim H K and Charette A 2007 A sensitivity function-based conjugate gradient method for optical tomography with the frequency-domain equation of radiative transfer *J. Quantum Spectrosc. RadIat. Transfer* **104** 24–39

[26] Schweiger M, Arridge S and Nassila I 2005 Gauss–Newton method for image reconstruction in diffuse optical tomography *Phys. Med. Biol.* **50** 2365–86

[27] Hielscher A H, Alcouffe A E and Barbour R L 1998 Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues *Phys. Med. Biol.* **43** 1285–302

[28] Roy R, Godavarty A and Sevick-Muraca E M 2003 Fluorescence-enhanced optical tomography using referenced measurements of heterogeneous media *IEEE Trans. Med. Imaing* **22** 824–36

[29] 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

[30] Hielscher A and Bartel S 2001 Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography *J. Biomed. Opt.* **6** 183

[31] Yalavarthy P K, Pogue B W, Dehghani H, Carpenter C M, Jiang S and Paulsen K D 2007 Structural information within regularization matrices improves near infrared diffuse optical tomography *Opt. Express* **15** 8043–58

[32] Nocedal J and Wright S J 2006 *Numerical Optimization* (New York: Springer)

[33] Heinkenschloss M 1996 Projected sequential quadratic programming methods *SIAM J. Optim.* **6** 373–417

[34] Byrd R, Curtis F and Nocedal J 2008 An inexact SQP method for equality constrained optimization *SIAM J. Optim.* **19** 351–69

[35] Gill P, Murray W and Saunders M 2005 SNOPT: an SQP algorithm for large-scale constrained optimization *SIAM Rev.* **47** 99–131

[36] Biros G and Ghattas O 2003 Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization: Part I. The Krylov–Schur solver *SIAM J. Sci. Comput.* **27** 687–713

[37] Biegler L, Schmid C and Ternet D 1997 *A Multiplier-Free, Reduced Hessian Method For Process Optimization, Large-Scale Optimization with Applications: Part II. Optimal Design and Control* (New York: Springer) p 101

[38] Haber E and Ascher U 2001 Preconditioned all-at-once methods for large, sparse parameter estimation problems *Inverse Problems* **17** 1847

[39] Bonnas J, Gilbert J, Lemaréchal C and Sagastizábal C 2003 *Numerical optimization: theoretical and practical aspects* (New York: Springer)

[40] Abddoulaev G S, Ren K and Hielscher A H 2005 Optical tomography as a PDE-constrained optimization problem *Inverse Problems* **21** 1507–30

[41] Burger M and Muhlhuber W 2002 Iterative regularization of parameter identification problems by sequential quadratic programming methods *Inverse Problems* **18** 943–69
Boggs P and Tolle J 2000 Sequential quadratic programming for large-scale nonlinear optimization *J. Comput. Appl. Math.* **124** 123–37

[42] Hu JL, Wu Z, McCann H, Davis L E and Xie CG 2005 Sequential quadratic programming method for solution of electromagnetic inverse problems *IEEE Trans. Antennas Propag.* **53** 2680

[43] Feng D and Pulliam T 1995 *All at-at-once reduced Hessian SQP scheme for aerodynamics design optimization* Technical Report NASA Ames Research Center

[44] Coleman T, Liu J and Yuan W 2000 A quasi-newton quadratic penalty method for minimization subject to nonlinear equality constraints *Comput. Optim. Appl.* **15** 103–23

[45] Lalee M, Nocedal J and Plantenga T 2003 On the implementation of an algorithm for large-scale equality constrained optimization *SIAM J. Optim.* **8** 682–706

[46] Biegler L, Nocedal J, Schmid C and Ternet D 2000 Numerical experience with a reduced Hessian method for large scale constrained optimization *Comput. Optim. Appl.* **15** 45–67

[47] Henyey L G and Greenstein L J 1941 Diffuse radiation in the galaxy *Astrophysics* **90** 70

[48] Modest M 2003 *Radiative Heat Transfer* (New York: McGraw-Hill)

[49] Minkowycz S, Sparrow E and Murthy J 2006 *Handbook of Numerical Heat Transfer* (Hoboken, NJ: Wiley)

[50] Meese E A 1998 Finite volume methods for the incompressible Navier–Stokes equations on unstructured grids Ph.D. Thesis Norwegian University of Science and Technology, Trondheim, Norway

[51] Saad Y 2003 *Iterative Methods for Sparse Linear Systems* (Philadelphia: SIAM)

[52] Saad Y and Schultz M H 1986 GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems *SIAM J. Sci. Stat. Comput.* **3** 856–69

[53] Alifanov O M 1994 *Inverse Heat Transfer Problems* (New York: Springer)