Inverse Problems **19** (2003) 387–409 PII: S0266-5611(03)38037-2

Quasi-Newton methods in optical tomographic image reconstruction

Alexander D Klose and Andreas H Hielscher

Departments of Biomedical Engineering and Radiology, Columbia University, 500 West 120th Street, MC8904, NY 10027, USA

E-mail: ak2083@columbia.edu

Received 6 June 2002, in final form 13 January 2003
Published 7 March 2003
Online at stacks.iop.org/IP/19/387

Abstract

Optical tomography (OT) recovers the cross-sectional distribution of optical parameters inside a highly scattering medium from information contained in measurements that are performed on the boundaries of the medium. The image reconstruction problem in OT can be considered as a large-scale optimization problem, in which an appropriately defined objective function needs to be minimized. In the simplest case, the objective function is the least-square error norm between the measured and the predicted data. In biomedical applications that apply near-infrared light as the probing tool the predictions are obtained from a model of light propagation in tissue. Gradient techniques are commonly used as optimization methods, which employ the gradient of the objective function with respect to the optical parameters to find the minimum. Conjugate gradient (CG) techniques that use information about the first derivative of the objective function have shown some good results in the past. However, this approach is frequently characterized by low convergence rates. To alleviate this problem we have implemented and studied so-called quasi-Newton (QN) methods, which use approximations to the second derivative. The performance of the QN and CG methods are compared by utilizing both synthetic and experimental data.

1. Introduction

Over recent years optical tomography (OT) has made considerable advances and promises to become a novel biomedical imaging modality. For example, initial studies have begun that explore the clinical usefulness of this emerging technique for imaging breast cancer, brain function or rheumatoid arthritis in finger joints [Colak99, Benaron00, Netz01]. In these and similar studies near-infrared light (λ = 650–900 nm) is delivered through optical fibres to multiple sites on the surface of the body part that is under investigation. Another set of optical fibres is used to collect the transmitted and reflected light intensities. From these measurements the distribution of optical properties inside the medium is sought. The optical properties can be further used to derive physiologically important parameters such as blood oxygenation or blood volume. In recent years, considerable advances have been made with respect to instrument design, which allows for more accurate as well as faster data acquisition. A major challenge remains the reconstruction of two-dimensional cross sections of the optical parameters μ.

The image reconstruction problem in OT is distinctly different from the imaging problem that arises in other imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), position emission tomography (PET) or single photon emission computed tomography (SPECT) [Herman80, Natterer99]. For example, in x-ray tomography the probing photons traverse the medium on a straight line and well-known backprojection algorithms can be applied that are based on the inverse radon transform [Radon17]. In OT, multiple scattering of photons is significant and backprojection algorithms have only been of limited use [Benaron94, Colak97, Matson97]. Therefore, most of the currently available reconstruction algorithms for OT use some form of a model-based iterative image reconstruction (MOBIIR) scheme [Hanson98, Arridge99, Hielscher99] for recovering the distribution of μ. The most prevalent MOBIIR schemes in OT are the perturbation approach and other similar techniques [Schottland93, OLeary95, Arridge95, Paulsen95, Chang96, Jiang96, Arridge97, Yao97, Dorn98, Ye99]. The image reconstruction problem, however, can also be formulated within the MOBIIR scheme as a numerical optimization problem [Saquib97, Arridge98b, Hielscher99, Klose99, Roy01] consisting of three major parts.

First, a forward model for light transportation is used to predict the measured data. This model is a function of the distribution of optical properties inside the medium and the position and strength of the light source. Initially, a estimation of the distribution of optical parameters μ_0 is used to calculate the first prediction of the measurement data. Secondly, a scalar objective function, Φ, is evaluated to obtain a measure of difference between the predicted and measured data. In a third step the initial estimation of the optical properties is updated in a way that reduces the difference between predicted and measured data as defined by the objective function. These steps are repeated until a distribution μ^* is found for which the objective function is minimal.

In addition to the type of forward model being used (e.g. diffusion-equation-based models or radiative-transfer-equation-based models), the algorithms that are currently available for OT mainly differ in what type of updating scheme is employed. In general, the updating procedure can be formulated as [Nocedal99]

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k + \alpha_k \boldsymbol{u}_k, \tag{1}$$

where $\boldsymbol{\mu}_k$ is a vector containing a set of optical properties from which the new set $\boldsymbol{\mu}_{k+1}$ is obtained. The vector $\boldsymbol{u}_k$ is a search direction in N-dimensional space, given a problem with N unknowns. The parameter α_k is the step length in the direction $\boldsymbol{u}_k$. In the most general form the search direction can be written as [Nocedal99]

$$\boldsymbol{u}_k = \boldsymbol{A}_k \nabla_\mu \Phi(\boldsymbol{\mu}_k) + \beta_k \boldsymbol{u}_{k-1}. \tag{2}$$

For example, using the *steepest descent* method, one chooses $\beta_k = 0$ and $\boldsymbol{A}_k = -\boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. In OT the most common method to determine the search direction has been the *nonlinear conjugate gradient* (CG) technique with $\boldsymbol{A}_k = -\boldsymbol{I}$ and $\beta_k \neq 0$ to ensure that $\boldsymbol{u}_k$ and $\boldsymbol{u}_{k-1}$ are conjugate. A commonly applied formula for β_k is given by the *Polak–Ribiere* formula [Luenberger84, Fletcher87]

$$\beta_k = \frac{\boldsymbol{y}_{k-1}^T \nabla_\mu \Phi(\boldsymbol{\mu}_k)}{\nabla_\mu \Phi(\boldsymbol{\mu}_{k-1})^T \nabla_\mu \Phi(\boldsymbol{\mu}_{k-1})}. \tag{3}$$

In this work we explore the performance of so-called *quasi-Newton* (QN) methods, in which $\beta_k = 0$ and $\boldsymbol{A}_k$ is chosen as an approximation to the inverse of the second derivative of the objective function [Davidon91, Martinez00]. In general, QN methods are often found to be more reliable and converge faster than CG methods [Luenberger84, Nash96]. However, they have not yet been applied to the image reconstruction problems that are encountered in OT. In particular, we will focus on two implementations known as the *Broyden–Fletcher–Goldfarb–Shanno* (BFGS) method [Broyden65, Broyden65, Fletcher70, Goldfarb70, Shanno70, Dennis77] and the *limited-memory Broyden–Fletcher–Goldfarb–Shanno* (lm-BFGS) method [Nocedal80, Liu89]. We compare reconstruction results obtained with the CG, lm-BFGS and BFGS methods and discuss the impact of measurement noise and different initial estimations on the performance of the updating scheme. Noise corrupted measurement data and the appropriate choice of an initial estimation still constitute major difficulties in OT.

2. Numerical methods

2.1. Forward model

As a forward model for light propagation in biological tissue we use the equation of radiative transfer (ERT) [Chandrasekhar60, Case67] for the domain Ω, which is given by

$$\boldsymbol{\omega} \cdot \nabla\psi(\boldsymbol{r}, \boldsymbol{\omega}) + (\mu_a(\boldsymbol{r}) + \mu_s(\boldsymbol{r}))\psi(\boldsymbol{r}, \boldsymbol{\omega}) = S(\boldsymbol{r}, \boldsymbol{\omega}) + \mu_s(\boldsymbol{r}) \int_{\partial\Omega} p(\boldsymbol{\omega}, \boldsymbol{\omega}')\psi(\boldsymbol{r}, \boldsymbol{\omega}')\,\mathrm{d}\boldsymbol{\omega}'. \quad (4)$$

The fundamental quantity in radiative transport theory is the radiance $\psi(\boldsymbol{r}, \boldsymbol{\omega})$, with units of W cm^{-2} sr^{-1}, at the spatial position $\boldsymbol{r}$ and unit direction $\boldsymbol{\omega}$ in the three-dimensional domain Ω. Other quantities in addition to the radiance ψ that are included in the ERT are the source term $S(\boldsymbol{r}, \boldsymbol{\omega})$ with the unit W cm^{-3} sr^{-1}, the scattering coefficient, $\mu_s(\boldsymbol{r})$, the absorption coefficient, $\mu_a(\boldsymbol{r})$, both given in units of cm^{-1}, and the scattering phase function $p(\boldsymbol{\omega}, \boldsymbol{\omega}')$ with units of sr^{-1} [Patterson91]. The scattering phase function gives the probability that a single photon is deflected by an angle θ. The angle θ encloses the two directions formed by $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ in the interval $\theta \in [0, \pi]$ with $\boldsymbol{\omega} \cdot \boldsymbol{\omega}' = \cos\theta$. A commonly applied scattering phase function in tissue optics is the Henyey–Greenstein function with g depicting the anisotropy factor

$$p(\cos\theta) = \frac{1 - g^2}{4\pi(1 + g^2 - 2g\cos\theta)^{3/2}}. \quad (5)$$

We solve the ERT in a two-dimensional plane of Ω with a finite-difference discrete-ordinates method, which yields the detector readings $\boldsymbol{p}$ on the tissue boundary given the distribution of optical parameters, μ_s and μ_a, inside the medium. A detailed description of this algorithm as well as experimental studies that validate the code can be found elsewhere [Klose02a].

2.2. Objective function

The difference between the actual measurements $\boldsymbol{m}$ and the predictions $\boldsymbol{p}$ for N_D source-detector pairs is mapped onto a scalar $\tilde{\varphi}$ by the objective function $\tilde{\Phi}(\boldsymbol{p})$. In this work we employ the widely applied *least-square error norm* as an objective function given by

$$\begin{aligned} &\tilde{\Phi} : \mathbb{R}^{N_D} \to \mathbb{R} \\ &\boldsymbol{p} \mapsto \tilde{\varphi} = \tilde{\Phi}(\boldsymbol{p}) = \frac{1}{2}\sum_{d=1}^{N_D}\left(\frac{p_d - m_d}{\kappa_d}\right)^2. \end{aligned} \quad (6)$$

The predictions $\boldsymbol{p}$ are determined for all N_D source-detector pairs, using the forward model F and given the N-dimensional vector $\boldsymbol{\mu}$ of optical parameters

$$\begin{aligned} F &: \mathbb{R}^N \to \mathbb{R}^{N_D} \\ \boldsymbol{\mu} &\mapsto \boldsymbol{p}(\boldsymbol{\mu}), \end{aligned} \tag{7}$$

where $N = 2 \times I \times J$. I and J denote the number of grid points of a finite-difference grid along the x- and y-axes, respectively. The vector $\boldsymbol{\mu}$ contains the distribution of the optical parameters, specifically the scattering (μ_s) and absorption (μ_a) coefficients. The parameter κ is used for normalizing the predictions and measurements. We usually set $\kappa_d = m_d$. Other objective functions can and have been defined and tested [Hielscher01]. However, in this work we focus on the effects of different optimization schemes rather than the advantages and disadvantages of different objective functions. Using definitions (6) and (7), we get the composite function $\Phi(\boldsymbol{\mu}) = \tilde{\Phi}(\boldsymbol{p}(\boldsymbol{\mu}))$. The objective function Φ is nonlinear because the predictions depend nonlinearly on the optical parameters. The goal of an optimization technique is now to determine a vector $\boldsymbol{\mu}$ that minimizes the objective function $\Phi(\boldsymbol{\mu})$. This vector will be a solution to the minimization problem and is displayed as a two-dimensional image.

2.3. *Quasi-Newton methods*

Most reconstruction algorithms for OT iteratively update the optical parameters that are used in the forward model starting with an initial estimation that is assumed to be close to the true distribution of optical parameters. CG methods have been popular [Saquib97, Arridge98b, Hielscher99, Klose99, Roy01], where the gradient of the objective function is calculated and updates are determined based on equation (1). In this work we suggest that QN methods, which make use of information on the second derivative (*Hessian* matrix), may lead to an improved image reconstruction algorithm.

QN methods are derived from *Newton's method*, which is often used for finding zeros of a nonlinear function. Applied to optimization schemes in OT one needs to find the zeros of the first derivative $\nabla_\mu \Phi(\boldsymbol{\mu})$. Given an estimate $\boldsymbol{\mu}_k$ of the solution, the nonlinear function is approximated by a linear function $r(\boldsymbol{u}_k)$ that consist of the first two terms of the Taylor series expansion of $\nabla_\mu \Phi(\boldsymbol{\mu}_k + \boldsymbol{u}_k)$ at $\boldsymbol{\mu}_k$

$$\nabla_\mu \Phi(\boldsymbol{\mu}_k + \boldsymbol{u}_k) = q(\boldsymbol{\mu}_k + \boldsymbol{u}_k) \approx q(\boldsymbol{\mu}_k) + \nabla_\mu q(\boldsymbol{\mu}_k)\boldsymbol{u}_k = r(\boldsymbol{u}_k). \tag{8}$$

Newton's method is derived by setting $r(\boldsymbol{u}_k) = 0$ and one can solve the resulting equation for the direction $\boldsymbol{u}_k$

$$\nabla_\mu q(\boldsymbol{\mu}_k)\boldsymbol{u}_k = -q(\boldsymbol{\mu}_k). \tag{9}$$

The Hessian matrix $\nabla_\mu q(\boldsymbol{\mu}_k) = \nabla_\mu^2 \Phi(\boldsymbol{\mu}_k)$ consists of second derivatives of the objective function with respect to the optical parameters. After determining $\boldsymbol{u}_k$ by means of equation (9), an updated set of optical properties $\boldsymbol{\mu}_{k+1}$ can be found with equation (1). However, Newton's method is rarely used in its 'classical' form for nonlinear programming problems, because it is often difficult to obtain the Hessian matrix for a given problem. To overcome this problem QN methods have been employed in many optimization problems.

QN methods are generalizations of the *secant method* for one-dimensional problems [Nash96]

$$f''(x_{k+1}) \approx \frac{f'(x_{k+1}) - f'(x_k)}{x_{k+1} - x_k}. \tag{10}$$

Applied to our multi-dimensional problem we arrive at the *secant condition*

$$\nabla_\mu^2 \Phi(\boldsymbol{\mu}_{k+1}) \cdot (\boldsymbol{\mu}_{k+1} - \boldsymbol{\mu}_k) \approx \nabla_\mu \Phi(\boldsymbol{\mu}_{k+1}) - \nabla_\mu \Phi(\boldsymbol{\mu}_k). \tag{11}$$

In the QN approach the Hessian matrix is replaced by a matrix $\boldsymbol{H}_{k+1}$ that approximates $\nabla_{\mu}^{2}\Phi(\boldsymbol{\mu}_{k+1})$ and can be obtained at a lower computational cost. We now have

$$\boldsymbol{H}_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k \tag{12}$$

where the vectors $\boldsymbol{s}_k$ and $\boldsymbol{y}_k$ are

$$\begin{aligned} \boldsymbol{s}_k &= \boldsymbol{\mu}_{k+1} - \boldsymbol{\mu}_k \\ \boldsymbol{y}_k &= \nabla_{\mu}\Phi(\boldsymbol{\mu}_{k+1}) - \nabla_{\mu}\Phi(\boldsymbol{\mu}_k). \end{aligned} \tag{13}$$

Furthermore, QN methods compute the matrix $\boldsymbol{H}_{k+1}$ from the previous matrix $\boldsymbol{H}_k$ in an iterative manner throughout the optimization process. For example the BFGS formula for the matrix $\boldsymbol{H}$ is given by [Nash96]

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k - \frac{(\boldsymbol{H}_k\boldsymbol{s}_k)(\boldsymbol{H}_k\boldsymbol{s}_k)^T}{\boldsymbol{s}_k^T\boldsymbol{H}_k\boldsymbol{s}_k} + \frac{\boldsymbol{y}_k\boldsymbol{y}_k^T}{\boldsymbol{y}_k^T\boldsymbol{s}_k}. \tag{14}$$

Typically, the problem is further simplified by using directly the inverse $\boldsymbol{A} = \boldsymbol{H}^{-1}$ as shown in [Nash96] and [Nocedal99]. In this case it is even easier to compute the search direction according to equation (2) without having to solve a system of linear equations (see equation (9)). An iterative way of calculating $\boldsymbol{A}_{k+1}$ is, for example, given by [Nash96]

$$\boldsymbol{A}_{k+1} = \boldsymbol{A}_k + \frac{(\boldsymbol{y}_k - \boldsymbol{A}_k\boldsymbol{s}_k)\boldsymbol{y}_k^T}{\boldsymbol{y}_k^T\boldsymbol{s}_k} - \frac{(\boldsymbol{y}_k - \boldsymbol{A}_k\boldsymbol{s}_k)^T\boldsymbol{s}_k}{(\boldsymbol{y}_k^T\boldsymbol{s}_k)^2}(\boldsymbol{y}_k\boldsymbol{y}_k^T). \tag{15}$$

This BFGS formula for the inverse matrix $\boldsymbol{A}$ together with equations (1) and (2) was implemented in this study. In addition, we also coded the lm-BFGS method. This method reduces the memory requirements for storing the matrix $\boldsymbol{A}_k$ which is the major disadvantage of the BFGS method, especially when large-scale problems with many unknowns are considered [Nash96]. In this case $\boldsymbol{A}_k$ in equation (15) is replaced with the identity matrix $\boldsymbol{I}$ and one obtains using equation (2) [Bishop97]

$$\boldsymbol{u}_k = -\nabla_{\mu}\Phi(\boldsymbol{\mu}_k) + \gamma\boldsymbol{s}_k + \lambda\boldsymbol{y}_k. \tag{16}$$

The scalars γ and λ are defined by

$$\begin{aligned} \gamma &= -\left(1 + \frac{\boldsymbol{y}_k^T\boldsymbol{y}_k}{\boldsymbol{s}_k^T\boldsymbol{y}_k}\right)\frac{\boldsymbol{s}_k^T\nabla_{\mu}\Phi(\boldsymbol{\mu}_{k+1})}{\boldsymbol{s}_k^T\boldsymbol{y}_k} + \frac{\boldsymbol{y}_k^T\nabla_{\mu}\Phi(\boldsymbol{\mu}_{k+1})}{\boldsymbol{s}_k^T\boldsymbol{y}_k}, \\ \lambda &= \frac{\boldsymbol{s}_k^T\nabla_{\mu}\Phi(\boldsymbol{\mu}_{k+1})}{\boldsymbol{s}_k^T\boldsymbol{y}_k}. \end{aligned} \tag{17}$$

QN methods require that the Hessian matrix $\boldsymbol{H}_k$ is positive definite. Only in this case, the descent property [Nocedal99]

$$\boldsymbol{u}_k^T(\boldsymbol{\mu}_k)\cdot\nabla_{\mu}\Phi(\boldsymbol{\mu}_k) < 0, \tag{18}$$

which is equivalent to the *curvature condition* [Nocedal99]

$$\boldsymbol{s}_k^T\cdot\boldsymbol{y}_k > 0, \tag{19}$$

is satisfied.

In particular, if the initial estimate μ_0 is too far from the minimum the curvature condition might not be satisfied and $\boldsymbol{H}_k$ is not positive definitely.

2.4. Line search

In addition to finding a search direction $\boldsymbol{u}_k$, one needs to determine the step lengths α_k in order to use equation (1). Here a line search is employed along the search direction $\boldsymbol{u}_k$, which computes a sequence of step length α_k and accepts one when certain conditions are fulfilled. A simple condition is that the line search provides a new value of the objective function with $\Phi(\boldsymbol{\mu}_k + \alpha_k \boldsymbol{u}_k) < \Phi(\boldsymbol{\mu}_k)$. But this condition does not always lead to a sufficient decrease in Φ. A sufficient decrease in the objective function Φ is given by the inequality

$$\Phi(\boldsymbol{\mu}_k + \alpha_k \boldsymbol{u}_k) \leqslant \Phi(\boldsymbol{\mu}_k) + c_1 \alpha_k \nabla_\mu \Phi(\boldsymbol{\mu}_k)^T \boldsymbol{u}_k \tag{20}$$

for some constant $c_1 \in (0, 1)$, which is also called the *Armijo condition* or *sufficient decrease condition* [Nash96]. Furthermore, line searches are distinguished between exact and inexact searches depending on what method for calculating the step length α_k is employed. An exact line search performs a one-dimensional *line-minimization* $\Phi(\alpha_k) = \Phi(\boldsymbol{\mu}_k + \alpha_k \boldsymbol{u}_k)$ along the direction $\boldsymbol{u}_k$ to find the step length α_k. Therefore, α_k is iteratively changed until Φ is minimal along the direction $\boldsymbol{u}_k$. The CG method requires an exact line search for calculating the step size α_k to obtain the conjugate search directions [Shanno78, Nocedal92, Nash96, Bishop97, Nocedal99]. QN methods do not require such a search, and computationally less demanding inexact line searches, such as *backtracking methods* [Press92, Ruggirero96] can be employed. Backtracking algorithms start with $\alpha_k = 1$ and choose their candidate step length $\alpha_k = 1, 1/2, 1/4, 1/8, \ldots, 2^{-n}, \ldots$ until the sufficiently decreasing condition (20) is satisfied. In most cases $\alpha_k = 1$ is sufficient and one does not need to track back. We found the best results for c_1 in the range of 10^{-4} to 10^{-8} in equation (20). The inexact line search used with the QN methods requires only a fraction of the function evaluations of $\Phi(\boldsymbol{\mu})$ per given search direction as compared to CG methods.

3. Numerical studies

The main focus of our study was the comparison of computational speed, robustness and accuracy of the BFGS, lm-BFGS and CG methods in situations that are typically encountered in OT. The two most commonly encountered problems are noisy data $\boldsymbol{m}$ and an initial estimate μ_0 for the distribution of the optical parameters which is not very close to the actual parameter. Different levels of noise change the appearance of the objective function Φ and subsequently may lead to different reconstruction results. Furthermore, starting the optimization process from different initial estimates may also lead to different reconstruction results. The study was carried out on numerical examples that contained several heterogeneities.

3.1. Problem set-up and method

A two-dimensional numerical model of a scattering medium used in this study is shown in figure 1. The cross section to be reconstructed consists of three objects with scattering coefficients of $\mu_s = 2.9$ cm^{-1} (black), $\mu_s = 8.7$ cm^{-1} (gray) and $\mu_s = 11.6$ cm^{-1} (white). These are embedded in a 3×3 cm^2 background medium with $\mu_s = 5.8$ cm^{-1}. In this example, the absorption coefficient $\mu_a = 0.35$ cm^{-1} does not vary within the isotropically scattering medium ($g = 0$). The general character of the objective function is independent of the physical properties of the tissue-like medium, and primarily depends on the data type used, i.e. data in the time domain, the frequency domain or in the continuous wave domain [Schweiger99]. Therefore, other tissue samples with different optical parameters do not lead to qualitatively different results as studies with different types of media that vary in absorption and scattering properties have shown. The results are qualitatively the same as presented for this example.

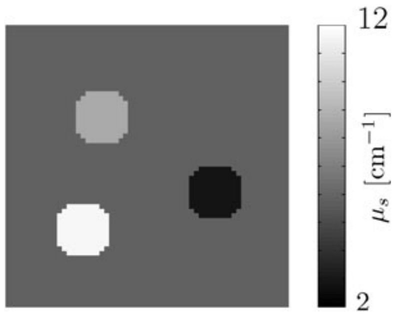

Figure 1. Scattering coefficients μ_s of original medium with dimensions of 3 cm $\times$ 3 cm containing three heterogeneities ($\mu_s = 2.9$ cm^{-1}, $\mu_s = 8.7$ cm^{-1} and $\mu_s = 11.6$ cm^{-1}). The bulk medium had a scattering coefficient of $\mu_s = 5.8$ cm^{-1}.

The medium was surrounded by 8 equally spaced sources and 116 equally spaced detectors (tow sources and 29 detectors were placed on each side of the medium). Detector readings on the same side as the source were not used for the reconstruction. Hence, we obtained a total of 8×87 source-detector pairs. The measured data at the detector positions were generated using the correct spatial distribution of the optical parameters. The forward calculations were done on a 61×61 grid with 16 ordinates using a finite-difference discrete-ordinates method based on the ERT [Klose99, Klose02a]. We used partly reflective boundary conditions based on Fresnel's law with a homogeneous refractive index $n = 1.54$ for the scattering medium. The synthetic data were used as input to the reconstruction algorithm. The reconstruction process was terminated after the normalized difference

$$\|(\Phi_{k+1} - \Phi_k)/\Phi_k\| < \epsilon \tag{21}$$

of the objective function between two subsequent iteration steps k and $(k+1)$ was smaller than $\epsilon = 10^{-3}$.

3.2. Definition of image accuracy

For evaluating the image accuracy of the reconstructed images, we determined the *correlation coefficient* ρ_a of the reconstructed image and the original medium (target image). This coefficient is defined as [Press92]

$$\rho_a = \frac{\sum_i^{IJ} (\mu_{s_i}^r - \bar{\mu}_s^r)(\mu_{s_i}^t - \bar{\mu}_s^t)}{(IJ - 1)\Delta\mu_s^r \Delta\mu_s^t}. \tag{22}$$

The *standard deviation* $\Delta\mu_s^t$ of the target medium and the standard deviation $\Delta\mu_s^r$ of the reconstructed image are given by

$$\Delta\mu_s^t = \sqrt{\frac{1}{IJ - 1} \sum_i^{IJ} (\mu_{s_i}^t - \bar{\mu}_s^t)^2} \tag{23}$$

and

$$\Delta\mu_s^r = \sqrt{\frac{1}{IJ - 1} \sum_i^{IJ} (\mu_{s_i}^r - \bar{\mu}_s^r)^2}. \tag{24}$$

The index $i \in [1, IJ]$ constitutes one pixel of the image or one element of the vector $\boldsymbol{\mu}_s$, respectively. The quantities $\bar{\mu}_s^t$ and $\bar{\mu}_s^r$ indicate the *mean values* of the scattering coefficients

Table 1. Image accuracy; a large ρ_a value and a small ρ_b value depicts a high image quality.

Example	Number of basic operations	Method	Correlation coefficient ρ_a	Deviation factor ρ_b
No noise	395	CG	0.80	0.62
	92	lm-BFGS	0.85	0.54
	172	BFGS	**0.87**	**0.50**
	92	CG	0.69	0.74
	92	lm-BFGS	0.85	0.54
	92	BFGS	**0.86**	**0.52**
20 dB SNR	365	CG	0.57	0.89
	26	lm-BFGS	**0.62**	**0.82**
	23	BFGS	0.61	0.83
	23	CG	0.56	0.87
	23	lm-BFGS	**0.62**	**0.82**
	23	BFGS	0.61	0.83
Initial estimate 30% higher	421	CG	0.72	0.72
	84	lm-BFGS	0.74	0.69
	213	BFGS	**0.82**	**0.59**
	84	CG	0.56	0.87
	84	lm-BFGS	0.74	0.69
	84	BFGS	**0.77**	**0.65**
Initial estimate 30% higher and 20 dB SNR	408	CG	0.45	1.24
	31	lm-BFGS	**0.55**	**1.01**
	41	BFGS	0.54	1.04
	31	CG	0.43	1.01
	31	lm-BFGS	**0.55**	**1.00**
	31	BFGS	0.54	1.02
μ_s (simultaneous)	237	CG	0.74	0.70
	95	lm-BFGS	0.81	0.61
	130	BFGS	**0.84**	**0.56**
μ_a (simultaneous)	237	CG	0.25	0.98
	95	lm-BFGS	0.27	0.97
	130	BFGS	**0.28**	**0.96**

of the target image and the reconstructed image, respectively. A large value of ρ_a shows a high correlation between the reconstructed and the target image and indicates a reconstructed image with a high level of accuracy. Almost no correlation between the target and the reconstructed image is present if ρ_a is very small. In addition to the correlation coefficient ρ_a, we also define a *deviation factor* ρ_b as

$$\rho_b = \frac{\sqrt{1/(IJ)\sum_i^{IJ}(\mu_{s_i}^r - \mu_{s_i}^t)^2}}{\Delta\mu_s^t}. \tag{25}$$

This parameter is a measure of the deviation of the reconstructed image from the target image. It is defined as the ratio of the χ-square error norm, which was obtained from the target and reconstructed image, and the standard deviation of the target image. A small value of ρ_b indicates a reconstructed image with high accuracy. The correlation coefficient and the deviation factor of all reconstructed images of the numerical study are shown in table 1.

3.3. Impact of noise

The detector readings, obtained from an experimental setting, are typically corrupted by noise to various degrees. In our numerical study we determined the effects of different noise levels based on the assumption that shot noise[1] and thermal noise[2] from the source (laser diode) and the detector (avalanche photodiode) are the dominant noise contribution during the measurement process. Therefore, we defined the noise level σ_{m_d} as the standard deviation of the *Gaussian distribution* around the signal m_d of the dth source-detector pair of the measurement vector $\boldsymbol{m}$. The signal-to-noise ratio (SNR) of the measurement data in units of dB is obtained from

$$\mathrm{SNR} = 10 \log_{10} \frac{m_d}{\sigma_{m_d}}. \tag{26}$$

Typical SNRs in an experimental setting are within 15–25 dB. Given that range of SNRs we determined the noise level or standard deviation σ_{m_d} from equation (26). Subsequently, a Gaussian distributed random number was calculated [Press92] using this standard deviation σ_{m_d}. A new noise-corrupted measurement value was obtained by adding this random number to the measurement value m_d.

We carried out numerical studies on four examples with different SNRs present in the synthetic measurement data (SNR $= \infty$ (no noise), SNR $= 45$, 20 and 15 dB) using the test medium as explained in section 3.1. First, the noise-free case was considered. We reconstructed the scattering coefficients μ_s starting from an initial estimate μ_{s_0} that had the same optical parameters as the background medium. The reconstruction was performed using the CG method, lm-BFGS method and the BFGS method. The objective functions of all three methods are depicted in figure 2(a).

The value of the objective function is displayed as a function of the number of basic operations. A basic operation is either a forward or a gradient calculation. The gradient was obtained by an adjoint differentiation (AD) technique applied to the numerical forward model based on the ERT. The AD technique derives the gradient of the objective function by applying the chain rule of differentiation to the forward model. With employing the AD technique the time required to calculate the gradient is approximately the same as that needed for one forward calculation (for more details see [Klose02b]). Therefore, the abscissa can be interpreted as the execution time of the optimization process[3]. Furthermore, we define the number of iterations as the number of gradient calculations.

As can be seen in figure 2(a) the lm-BFGS method reached the stopping criterion (equation (21)) after 92 basic operations, consisting of 54 forward and 38 gradient calculations[4](54 + 38). The BFGS method converged after 172 basic operations (98 + 74). The CG method took the longest time to reach the stopping criterion, requiring 395 basic operations (363+32). The absolute number of forward runs and gradient calculations differed, depending on the method used. Usually the CG method made more forward runs and fewer gradient calculations than the QN methods, because the CG method relies on an exact line search (line-minimization, see section 2.4). The BFGS method and the lm-BFGS method only require an inexact line search (backtracking method, see section 2.4), which use less forward runs for each search direction. The final value $\tilde{\varphi}$ of the objective function was different for all three methods. The BFGS method had the smallest value $\log_{10}(\tilde{\varphi}) = -5.32$, whereas the CG method had the largest value $\log_{10}(\tilde{\varphi}) = -4.48$.

[1] Shot noise is the time-dependent fluctuation in electrical current associated with the discreteness of the charge carrier in semiconductor devices. The charge carriers create photons by recombination processes in laser diodes. Since noise represents randomly fluctuating events, we must use statistical distributions to characterize noise. When the photon numbers are large a continuous probability distribution such as the Gaussian distribution is needed.

[2] Thermal noise is a quantum statistical phenomenon where the thermal motion of charge carriers cause macroscopic fluctuations in electrical state of system. The thermal noise is described by a Gaussian distribution.

[3] The processing time for one forward calculation was approximately 40–60 s using an *Intel Pentium III Xeon* 1.26 GHz processor.

[4] From now on we will write the number of forward and gradient calculations within parentheses (number of forward calculations + number of gradient calculations).

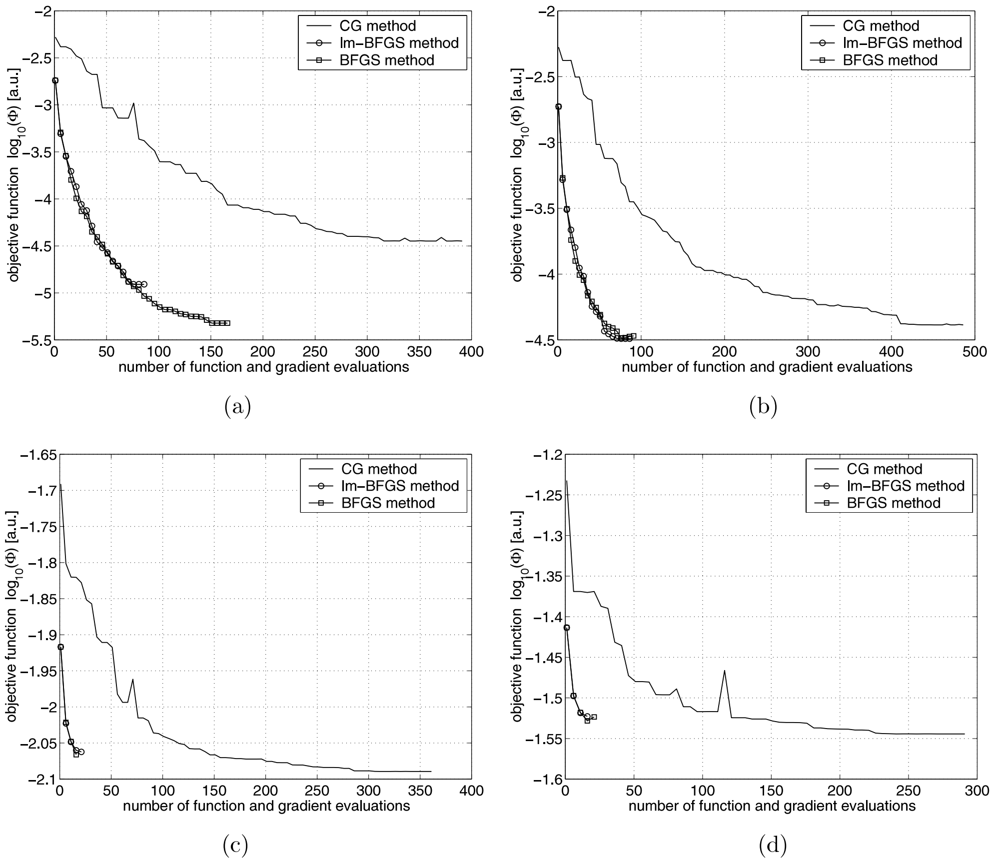

Figure 2. Objective functions for different SNRs of the synthetic measurement data. Note that in some iteration steps the objective function actually increases. This can be best understood when considering that the minimization of the objective function consists of two types of iteration cycles. One iteration cycle consists of updating the search direction according to equation (2). After a new search direction is determined, a second iteration cycle (a line search as described in section 2.4) seeks to find iteratively a new step length α along that search direction. However, a new update α might be too large sometimes at the start of the line search and leads subsequently to an increase in the objective function. (a) SNR $= \infty$ (no noise), (b) SNR = 45 dB, (c) SNR = 20 dB and (d) SNR = 15 dB.

The resulting image reconstructions of μ_s are displayed in figure 3. In figure 4 we show reconstructions of all three methods for the same number of basic iterations. The interval between adjacent isolines is $\mu_s = 1\ \mathrm{cm}^{-1}$. The image accuracy is highest (ρ_a is highest and ρ_b is smallest) for the image obtained by the BFGS method ($\rho_a^{\mathrm{BFGS}} > \rho_a^{\mathrm{lm\text{-}BFGS}} > \rho_a^{\mathrm{CG}}$ and $\rho_b^{\mathrm{BFGS}} < \rho_b^{\mathrm{lm\text{-}BFGS}} < \rho_b^{\mathrm{CG}}$), see also table 1.

In a second example, we added noise to the synthetic measurement data resulting in a SNR of 45 dB. The objective functions are shown in figure 2(b). The lm-BFGS method needed 91 basic operations (55 + 36), and the BFGS method took 98 basic operations (54 + 44). The CG method took 495 basic operations (454 + 41) for completion. The final values after termination are $\log_{10}(\tilde{\varphi}) = -4.48$ for the BFGS and lm-BFGS methods, and $\log_{10}(\tilde{\varphi}) = -4.38$ for the CG method.

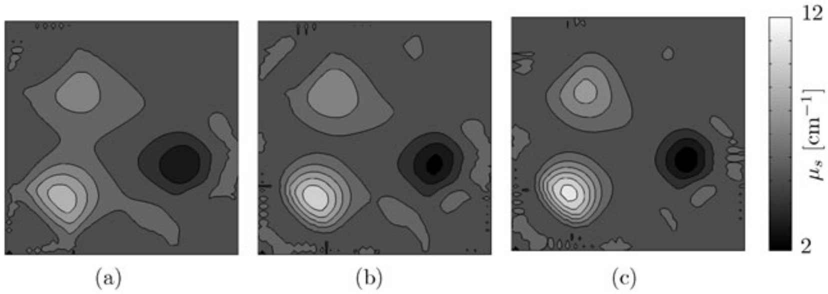

Figure 3. The final image reconstructions of μ_s. No noise was present in the synthetic measurement data. The distance between the adjacent isolines is 1 cm^{-1}. (a) The CG method, after 395 basic operations, (b) the lm-BFGS method, after 92 basic operations and (c) the BFGS method, after 172 basic operations.

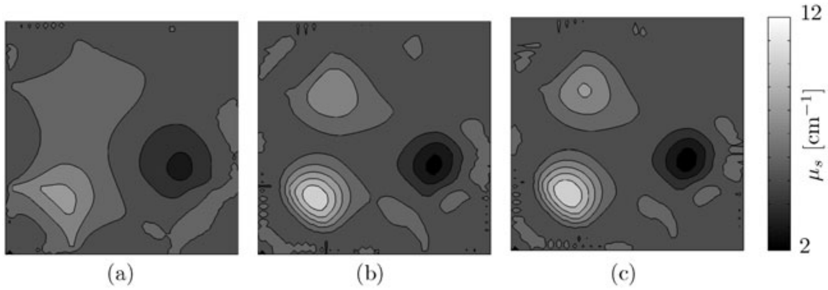

Figure 4. The image reconstructions of μ_s after 92 basic operations. No noise was present in the synthetic measurement data. The distance between the adjacent isolines is 1 cm^{-1}. (a) The CG method, after 92 basic operations, (b) the lm-BFGS method, after 92 basic operations and (c) the BFGS method, after 92 basic operations.

The SNR in the third example was 20 dB. This value is typical for experimental data that involve human tissues. The objective functions of all three optimization techniques are presented in figure 2(c), and the image reconstructions are shown in figure 5. The final reconstructions were obtained after 26 basic operations (15 + 11) (lm-BFGS), 23 basic operations (12 + 11) (BFGS) and 365 basic operations (343 + 22) (CG), respectively. The final value of the objective function was $\log_{10}(\tilde{\varphi}) = -2.07$ for the BFGS and lm-BFGS methods, and $\log_{10}(\tilde{\varphi}) = -2.08$ for the CG method. The image accuracy is highest for both the lm-BFGS and BFGS methods (see table 1).

In figure 6 we show reconstructed images of all three methods after 23 basic operations. During these 23 basic operations the CG method only completed one line search along one gradient, which required 22 forward calculations. The BFGS and limited-BFGS methods performed 11 and 10 inexact line searches, respectively, which on average only required one forward calculation each. As can be clearly seen, the image accuracy is highest for the images obtained by the QN methods (see table 1).

In a fourth example we decreased the SNR of the synthetic measurement data to 15 dB. The BFGS method needed 27 basic operations (15 + 12), the lm-BFGS method took 22 basic operations (12 + 10) and the CG method terminated after 297 basic operations (281 + 16). The objective functions are shown in figure 2(d) with the final values $\log_{10}(\tilde{\varphi}) = -1.53$ for the BFGS and lm-BFGS methods and $\log_{10}(\tilde{\varphi}) = -1.54$ for the CG method.

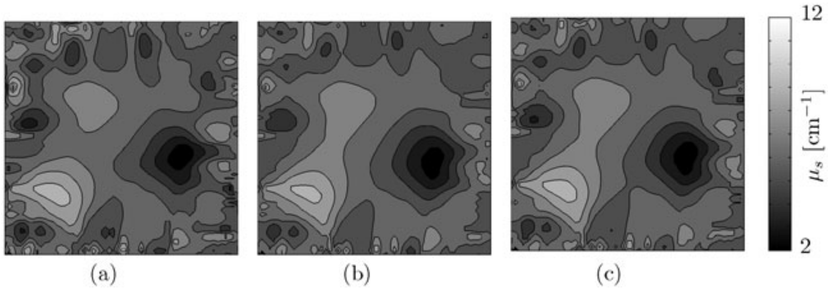

Figure 5. The final image reconstructions of μ_s. The SNR of the synthetic measurement data was 20 dB. The distance between the adjacent isolines is 1 cm^{-1}. (a) The CG method, after 365 basic operations, (b) the lm-BFGS method, after 26 basic operations and (c) the BFGS method, after 23 basic operations.

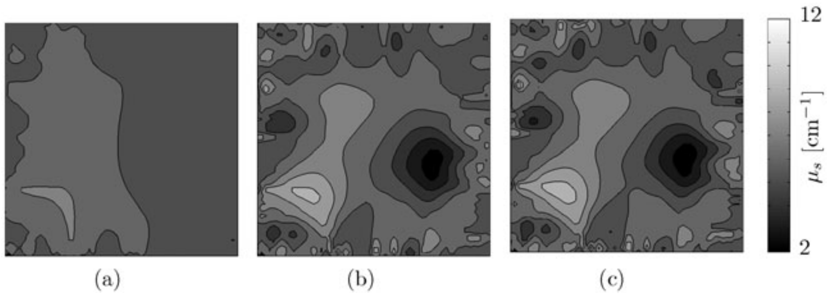

Figure 6. The image reconstructions of μ_s after 23 basic operations. The SNR of the synthetic measurement data was 20 dB. The distance between the adjacent isolines is 1 cm^{-1}. (a) The CG method, after 23 basic operations, (b) the lm-BFGS method, after 23 basic operations and (c) the BFGS method, after 23 basic operations.

In summary, we found that the BFGS and the lm-BFGS methods lead to smaller values $\tilde{\varphi}$ of the objective function when no noise is present in the measurement data. Furthermore, the QN methods required fewer basic operations than the CG method to satisfy the stopping criterion. The CG method needed at least twice as many basic operations (see figure 2(a)). The image accuracy, represented by the correlation coefficient ρ_a and deviation factor ρ_b, is highest for the BFGS and lm-BFGS methods (see table 1). The advantages of QN methods over CG methods were diminished once noise was added to the measurement data. All three (CG, lm-BFGS and BFGS) methods reached approximately the same final value $\tilde{\varphi}$ of the objective function for a SNR < 45 dB. The image accuracy of the final reconstructed images is not significantly different. However, the lm-BFGS and BFGS methods were considerably faster than the CG method, as they needed 10–15 times fewer basic operations than the CG method (see figures 2(c) and (d)).

3.4. Impact of different initial estimates

Optimization schemes require an initial estimate of the optical properties as a starting point for the iterative minimization. Usually this estimate is a homogeneous medium, described by a spatially independent optical parameter. To study the influence of different homogeneous initial estimates μ_{s_0} on the reconstruction results, we chose three different examples of μ_{s_0} with a 20, 30 and 50% higher scattering coefficient as compared to the original background

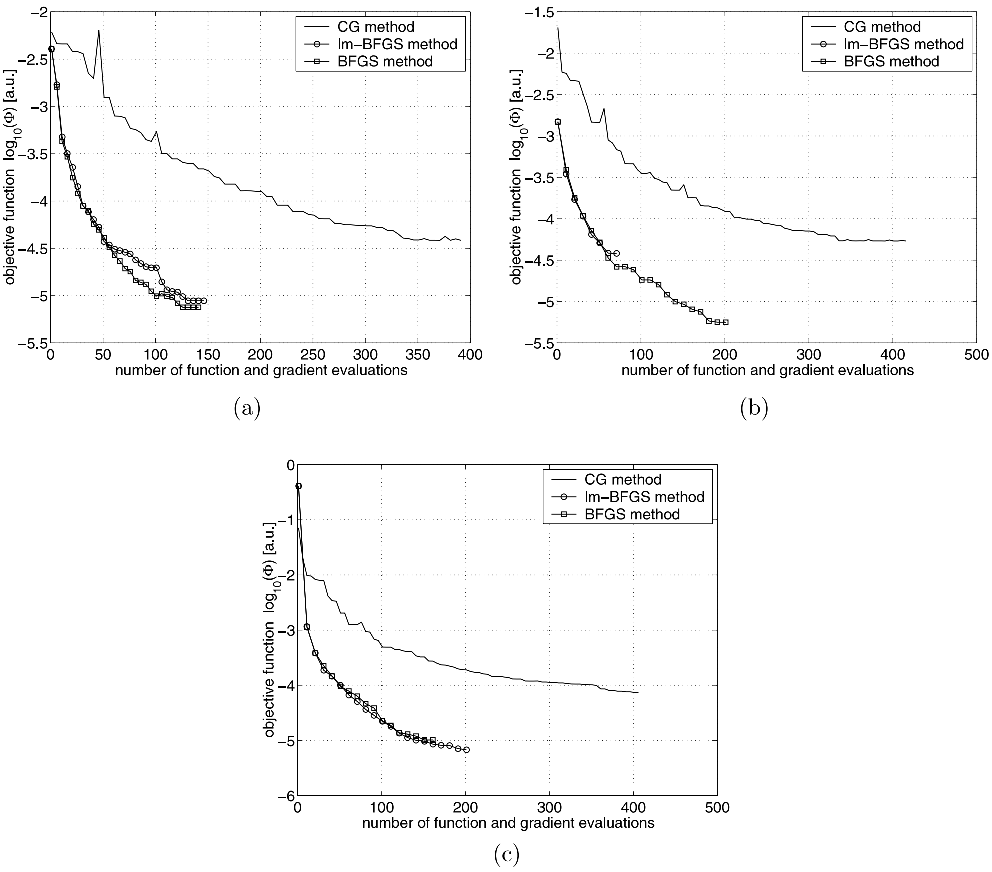

Figure 7. The objective functions for the different initial estimate μ_{s0}. (a) The initial estimate μ_{s0} was 20% higher than the background scattering of the original medium. (b) The initial estimate μ_{s0} was 30% higher than the background scattering of the original medium. (c) The initial estimate μ_{s0} was 50% higher than the background scattering of the original medium.

medium. The values of the objective function as a function of the basic operations are shown in figure 7 for all three examples.

Figure 7(a) shows the results of the first example where a 20% higher scattering coefficient for the initial estimate was chosen. The QN methods were faster than the CG method, and also reached a smaller value $\tilde{\varphi}$ of the objective function when the stopping criterion was satisfied. The BFGS method needed 148 basic operations $(87 + 61)$ and the lm-BFGS method completed after 150 basic operations $(89 + 61)$ with $\log_{10}(\tilde{\varphi}) = -5.12$ and $\log_{10}(\tilde{\varphi}) = -5.05$, respectively. The CG required 398 basic operations $(366 + 32)$ with $\log_{10}(\tilde{\varphi}) = -4.4$.

The second example with a 30% higher scattering coefficient (see figure 7(b)) leads to similar results. The lm-BFGS method needed 84 basic operations $(51 + 33)$ with $\log_{10}(\tilde{\varphi}) = -4.42$ to converge. The BFGS method converged after 213 basic operations $(126 + 87)$ with the smallest value of $\log_{10}(\tilde{\varphi}) = -5.25$ compared to the other techniques. The CG method took the longest time to satisfy the stop criterion. It finished after 421 basic operations $(389 + 32)$ with $\log_{10}(\tilde{\varphi}) = -4.25$. The final reconstructed images are shown in figure 8. Additionally, we also compared all three methods after 84 basic operations in figure 9, at which the lm-BFGS method was the first to satisfy the stopping criterion. During this time the lm-BFGS method

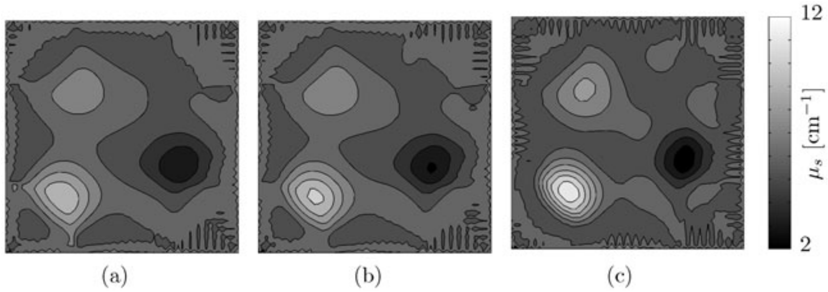

Figure 8. The final image reconstructions of μ_s. The initial estimate μ_{s_0} was 30% higher than the background scattering of the original medium. The distance between the adjacent isolines is 1 cm^{-1}. (a) The CG method, after 421 basic operations, (b) the lm-BFGS method, after 84 basic operations and (c) the BFGS method, after 213 basic operations.

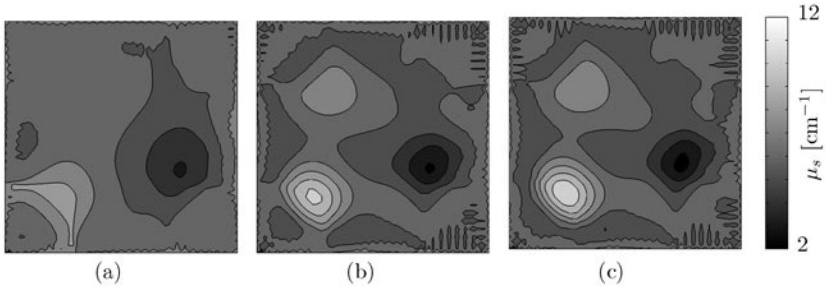

Figure 9. The image reconstructions of μ_s after 84 basic operations. The initial estimate μ_{s_0} was 30% higher than the background scattering of the original medium. The distance between the adjacent isolines is 1 cm^{-1}. (a) The CG method, after 84 basic operations, (b) the lm-BFGS method, after 84 basic operations and (c) the BFGS method, after 84 basic operations.

performed 33 inexact line searches requiring 33 gradient calculations and the BFGS method determined 40 gradients. The CG method only performed six exact line searches requiring six gradient calculations. The BFGS method yielded images with the highest image accuracy (ρ_a was the highest and ρ_b was the smallest), see also table 1.

This last example used a starting point μ_{s_0} with a 50% higher scattering coefficient (see figure 7(c)). The lm-BFGS method needed 175 basic operations (98 + 77). The BFGS method finished after 217 basic operations (118 + 99). The CG technique took the longest time to reach the stop criterion. It finished after 412 basic operations (381 + 31). The final values of the objective function were $\log_{10}(\tilde{\varphi}) = -5.19$ for the BFGS method, $\log_{10}(\tilde{\varphi}) = -4.99$ for the lm-BFGS method and $\log_{10}(\tilde{\varphi}) = -4.13$ for the CG method.

Furthermore, the last reconstruction example illustrates a particular point to be considered when using QN methods in OT. The Hessian $\boldsymbol{H}_k$ at the starting point μ_{s_0} was not positive definite and the curvature condition (see equation (19)) was not satisfied. If this happens the approximated inverse Hessian $\boldsymbol{A}_k$ in equation (2) has to be replaced with a positive definite matrix, in order to assure a descent direction. In this and similar cases we replaced the inverse Hessian matrix with the positive definite identity matrix until a point was found for which the curvature condition holds.

In conclusion, we found that when the initial estimate of the optical properties was chosen to be different from the background medium, the CG method always needed at least twice as

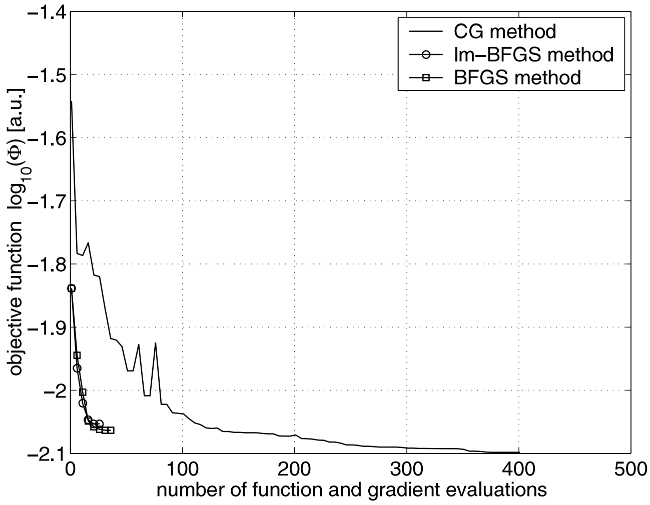

Figure 10. Objective functions starting from an initial estimate μ_{s0} that was 30% higher than the background scattering of the original medium. Additionally, the synthetic measurement data were corrupted by noise with a SNR of 20 dB.

many basic operations than the QN methods. Moreover, the BFGS and the lm-BFGS methods found smaller values of $\tilde{\varphi}$ than the CG method resulting in reconstructed images with higher image accuracy (see table 1).

3.5. Impact of noise and initial estimate

In practice, we typically encounter the situation where the measurement data are corrupted by noise and the initial estimate does not closely match the background medium. To illustrate the performance of all three optimization techniques in this case, we generated a data set with SNR = 20 dB (see figures 2(c) and 5) and started the reconstruction process with an initial estimate μ_{s_0} that was 30% higher than the scattering coefficient of the background medium (see figures 7(b) and 8).

The lm-BFGS method took 31 basic operations (18 + 13) and the BFGS method needed 41 basic operations (25 + 16), with $\log_{10}(\tilde{\varphi}) = -2.05$ and -2.06, respectively. The longest reconstruction time was again required by the CG method with 408 basic operations (382 + 26) and with $\log_{10}(\tilde{\varphi}) = -2.09$. The objective functions are shown in figure 10. The reconstructed images are shown in figures 11 and 12. Again, we find that the image accuracy was highest for the QN methods (see table 1), while the final value of the objective functions after the termination of the reconstruction process are close to each other.

3.6. Simultaneous reconstruction of μ_s and μ_a

As a final example we reconstructed simultaneously the scattering coefficient, μ_s, and the absorption coefficient, μ_a. We replaced the scattering perturbation with $\mu_s = 8.7$ cm^{-1} in figure 1 with a low-absorbing heterogeneity with $\mu_a = 0.1$ cm^{-1}. We started the reconstruction process with an initial estimate of $\mu_{s_0} = 5.8$ cm^{-1} and $\mu_{a_0} = 0.35$ cm^{-1} and measured the performance of all three optimization methods by determining the image accuracy, the number of basic operations and the final values of the objective function.

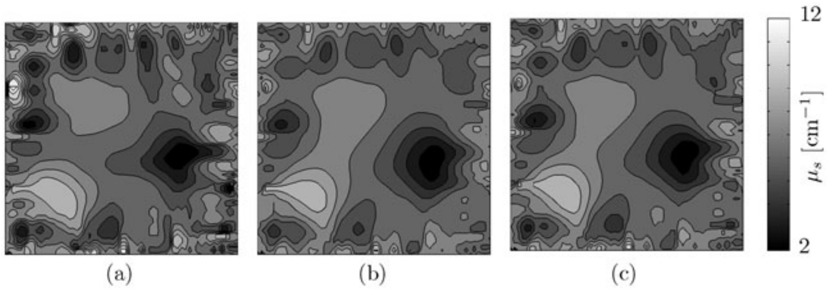

Figure 11. The final image reconstructions of μ_s. The initial estimate μ_{s_0} was 30% higher than the background scattering of the original medium. The SNR of the synthetic measurement data was 20 dB. The distance between the adjacent isolines is 1 cm^{-1}. (a) The CG method, after 408 basic operations, (b) the lm-BFGS method, after 31 basic operations and (c) the BFGS method, after 41 basic operations.

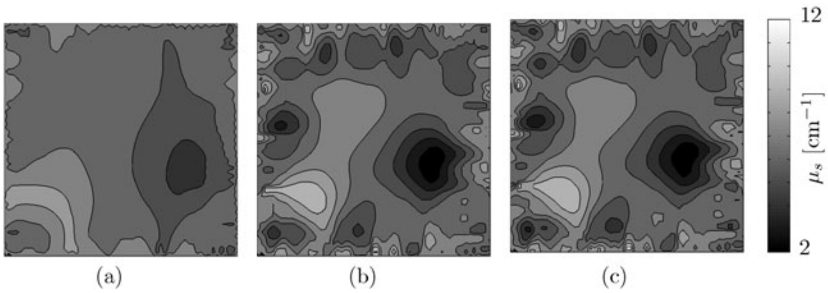

Figure 12. The image reconstructions of μ_s after 31 basic operations. The initial estimate μ_{s_0} was 30% higher than the background scattering of the original medium. The SNR of the synthetic measurement data was 20 dB. The distance between adjacent isolines is 1 cm^{-1}. (a) The CG method, after 31 basic operations, (b) the lm-BFGS method, after 31 basic operations and (c) the BFGS method, after 31 basic operations.

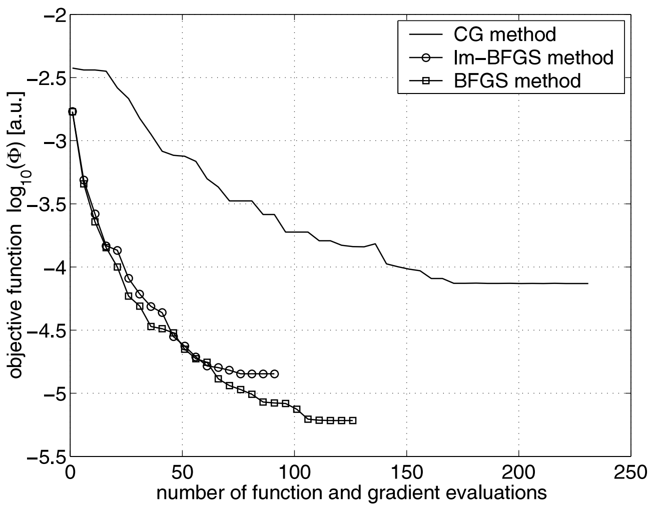

Figure 13. Objective function for simultaneous reconstruction of μ_s and μ_a.

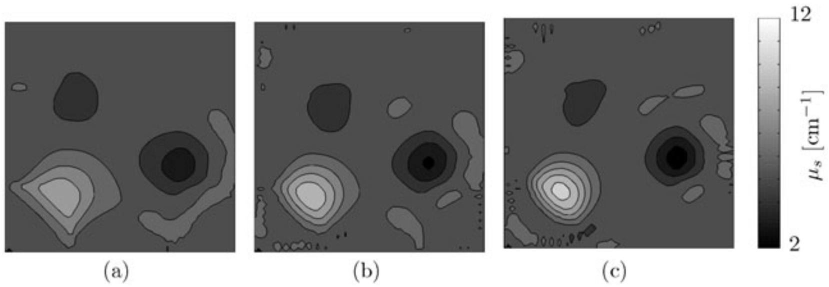

Figure 14. The final image reconstructions of μ_s. The initial estimate was $\mu_{s_0} = 5.8$ cm^{-1} and $\mu_{a_0} = 0.35$ cm^{-1}. The distance between the adjacent isolines is 1 cm^{-1}. Both the scattering heterogeneities are clearly present in the images. A slight cross talk of the absorption perturbation can be observed on the upper left side. (a) The CG method, after 237 basic operations, (b) the lm-BFGS method, after 95 basic operations and (c) the BFGS method, after 130 basic operations.

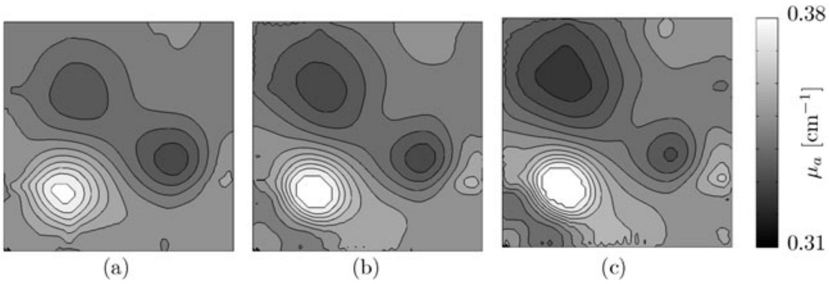

Figure 15. The final image reconstructions of μ_a. The initial estimate was $\mu_{s_0} = 5.8$ cm^{-1} and $\mu_{a_0} = 0.35$ cm^{-1}. The distance between the adjacent isolines is 0.005 cm^{-1}. A weak absorbing heterogeneity is reconstructed in the upper left side of the images. However, strong cross talk of μ_s is observed. Both scattering heterogeneities show up as absorbing perturbations in the lower part of the images. (a) The CG method, after 237 basic operations, (b) the lm-BFGS method, after 95 basic operations and (c) the BFGS method, after 130 basic operations.

The lm-BFGS method took only 95 basic operations (59 + 36) and the BFGS method needed 130 basic operations (75 + 55), with the final values of $\log_{10}(\tilde{\varphi}) = -4.84$ and $\log_{10}(\tilde{\varphi}) = -5.21$, respectively. The CG method again required the longest reconstruction time with 237 basic operations (218 + 19) with $\log_{10}(\tilde{\varphi}) = -4.12$. Figure 13 shows the objective functions of all three optimization techniques. The reconstructed images of μ_s and μ_a are shown in figures 14 and 15.

The image accuracy of μ_s, as shown in table 1, is similar to that of the first example of our numerical study where no noise was present in the synthetic measurement data (see also figure 3). The objective functions of both examples behave in the same way as seen in figures 13 and 2(a).

However, we observed some cross-talk between μ_s and μ_a due to the illposedness of the optical image reconstruction problem. The cross-talk is more pronounced in the absorption images. Several authors have already reported on the cross-talk of both optical parameters [Arridge98a, Schweiger99, McBride01, Xu02]. We find that the image accuracy of the μ_a images is lowest for all reconstructions done so far. The correlation coefficient ρ_a is much smaller than in other reconstruction examples. That can be explained with the observed cross-talk between μ_s and μ_a, where the scattering perturbations appear as heterogeneities in the absorption image.

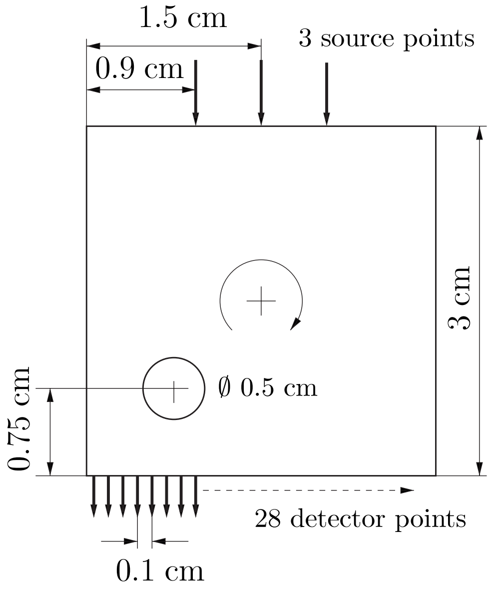

Figure 16. Schematic and source-detector configuration of the phantom that contained a single scattering heterogeneity. The phantom was illuminated from all four sides with three sources on each side. The measurements were taken on the sides opposite the sources at 28 points.

4. Experimental studies

4.1. Problem set-up

In addition to the numerical studies, we also compared the BFGS, lm-BFGS and CG methods for experimental data. Experiments were carried out on a scattering phantom illuminated with near infrared light. The phantom was composed of clear epoxy resin into which silicon-dioxide (SiO_2) monospheres and ink were mixed. The scattering properties were adjusted by varying the concentration of the monospheres, while the absorption properties were controlled by the concentration of the ink. The g-factor could be varied by using spheres with different diameters. The phantom had dimensions of $3 \times 3 \times 14$ cm^3 and contained a cylindrical hole with a diameter of 0.5 cm (figure 16). The hole was filled with *Intralipid* [Flock89a, Flock89b], a scattering fluid with $\mu_s' = (1-g)\mu_s = 23.2 \pm 5$ cm^{-1} and $\mu_a = 0.006\,75 \pm 0.003$ cm^{-1} for the measurement wavelength. The optical parameters of the bulk medium were determined to be $\mu_s = 58 \pm 5$ cm^{-1}, $\mu_a = 0.35 \pm 0.3$ cm^{-1} and $g = 0.8 \pm 0.08$.

The phantom was continuously illuminated with a laser diode (*Laser 2000 GmbH, Germany, LAS-670-20*) at $\lambda = 678$ nm. Measurements were taken with the source positioned at 12 different locations around the phantom. We used an avalanche photodiode (APD; *Hamamatsu, C5460-01*) to measure the fluence $\phi(x, y)$ at 28 points on the side opposite the source. Therefore, only transmitted measurement data were used. The distance between two adjacent measurement points was 0.1 cm. The detector could be translated around the phantom. The detection area at the boundary of the phantom was limited by a pinhole, which had a diameter of 0.1 cm. We used a lock-in technique (*Stanford Research Systems, model SR 830*) to improve the signal-noise ratio. For this purpose, a frequency generator (*Hewlett Packard, Waveform Generator 33120A*) provided a sinusoidal modulation of the laser diode input with a frequency at 1014 Hz. For more details on the experimental set-up see Klose *et al* [Klose02a].

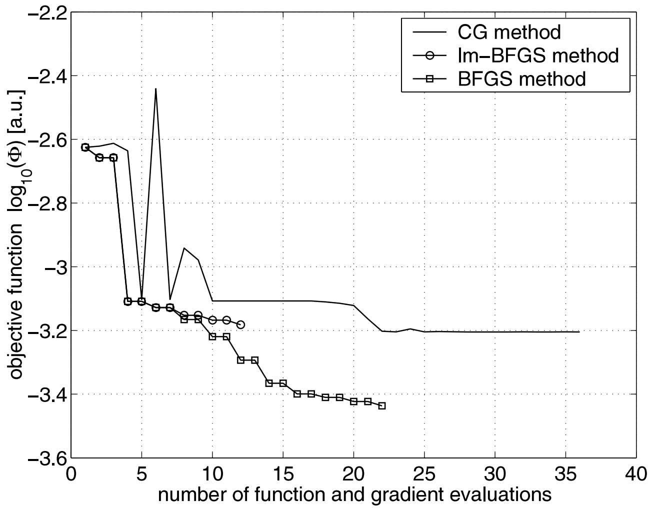

Figure 17. The objective function of the experimental data.

The forward calculations were performed on a 61×61 grid with 16 ordinates. The refractive index was $n = 1.54$. We assumed a constant anisotropy factor $g = 0.86$ throughout the optimization process. The reconstruction was terminated after the normalized difference $\|(\Phi_{k+1} - \Phi_k)/\Phi_k\|$ of the objective function between two subsequent iteration steps k and $(k + 1)$ was smaller than $\epsilon = 10^{-3}$.

4.2. Experimental results

The cross sectional images of the scattering coefficients were reconstructed using the CG, lm-BFGS and the BFGS methods given the near-infrared measurements on the boundary of the scattering phantom (see figure 16). We started with an initial estimate of a homogeneous medium of $\mu_{s0} = 50\,\text{cm}^{-1}$ and $\mu_{a0} = 0.45\,\text{cm}^{-1}$. In figure 17 we show the objective function of all three methods throughout the optimization process. The CG optimization was terminated after two iterations, yielding a total of 36 combined forward and gradient calculations (34+2). The lm-BFGS method finished after six iterations with a total of 12 combined forward and gradient calculations (6+6), and the BFGS method finished after 11 iterations (11+11). With respect to the value of the objective function ($\log_{10}(\tilde{\varphi}) = -3.2$) the BFGS method achieves the same result as the CG method after 10 combined forward and gradient calculations compared to 36 calculations for the CG method. The same can be said about the lm-BFGS method, however, no further decrease in the objective function can be achieved. The optimization process stops after 12 forward and gradient calculations.

In figure 18 the reconstructed images of the scattering coefficients are shown. The distance between the two subsequent isolines in the images is $\mu_s = 2\ \text{cm}^{-1}$. The absolute scattering coefficients do not differ much in all images. All three methods localized the scattering perturbation in the phantom, but it varies in its size. The largest scattering coefficient of the perturbation is 20% off the background medium.

These experimental studies confirm our numerical investigations that reconstruction results can be achieved in less computational time by using QN methods. If the SNR is large enough then QN methods can also find smaller values of the objective function as compared to CG methods.

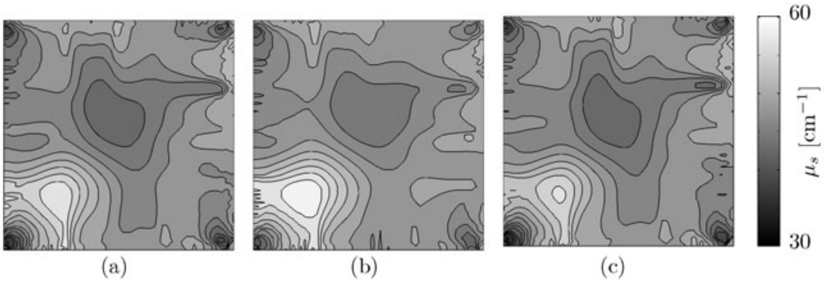

Figure 18. The reconstructed μ_s values of the phantom. The lm-BFGS method achieves the highest image accuracy after 12 basic operations. (a) The CG method, after 36 basic operations, (b) the lm-BFGS method, after 12 basic operations and (c) the BFGS method, after 22 basic operations.

5. Summary and conclusion

OT is used to determine the cross sectional distribution of optical parameters of highly scattered biological tissue. The image reconstruction process can be viewed as an optimization problem, in which an objective function that compares predicted values with actual measurements, is minimized. Typically, optimization techniques start from an initial estimate of optical parameters and determine iteratively new updates of these parameters along search directions until the minimum value of the objective function is found. The final distribution of the optical parameters is displayed in an image.

The computational speed and performance of the reconstruction process crucially depends on the effectiveness of the updating scheme. In general, optimization techniques employ either the first derivative (e.g. CG methods) or the first derivative in combination with some approximation of the second derivative (QN methods) of the objective function for calculating the update. QN methods, which have proven to be computationally superior to CG methods in many fields, have so far not been applied to OT. In this work, we compared the performance of QN techniques (BFGS and lm-BFGS methods) with the already widely used CG method.

We found that in general the QN methods outperform the CG method. Numerical studies with synthetic data showed that for data with (SNR > 45 dB) and an initial estimate of optical properties that is equal to that of the background medium the objective function always reaches a smaller value when QN methods are used as compared to when CG methods are employed.

Furthermore, using QN rather than CG methods the minimum is reached 2–10 times faster. When the measurement data are increasingly corrupted by noise (SNR < 45 dB), we observe that the objective function and consequently the image quality are about the same for QN and CG methods. However, the advantage of faster convergence towards the minimum remains. This result was also confirmed with the experimental data obtained from a tissue phantom model.

The reason for the better performance of the QN method compared to the CG method appears to be twofold. First, the QN methods follow a better search direction by using the inverse of the approximated Hessian. Second, the QN methods use an inexact line search, whereas the CG method requires an exact line search (line minimization). The exact line search typically requires far more forward calculations for each gradient calculation. For certain problems, as reported in the literature [Shanno78], CG methods might also work without an exact line search. However, we did not observe a better performance with inexact line searches but instead found a premature convergence of the reconstruction process.

As the reason for the advantages of the QN methods compared to the CG methods does not depend on the form of the objective function, QN methods may also be invaluable when additional regularization terms are added to the objective function. Regularization terms have shown to improve the overall performance of image reconstruction codes when carefully applied [Hielscher01].

Furthermore, we observed that the approximated Hessian matrix is not always positive definite and the optimization process leads consequently to a premature convergence. This problem, for example, might occur if the initial estimate is too far from the solution. Here, the linear function $r(\boldsymbol{u}_k)$, that consists of the first two terms of the Taylor series expansion of $\nabla_\mu \Phi(\boldsymbol{\mu})$ at $\boldsymbol{\mu}_k$, is a poor approximation of the nonlinear function $\nabla_\mu \Phi(\boldsymbol{\mu})$. We could solve this problem by forcing the Hessian to be positive definite and replacing it with the identity matrix.

A disadvantage of using the BFGS method is the memory requirement for storing the matrix $\boldsymbol{A}_k$ in order to calculate an update of that matrix at the next iteration step. The storage space can be quite large and leads to a computational burden for large-scale problems. However, our studies showed that the lm-BFGS method can alleviate this problem, while maintaining the image reconstruction speed and image quality.

A difficulty for all three methods remains the cross-talk between μ_s and μ_a when both parameters are reconstructed simultaneously. This problem is well known to the optical imaging community and might be alleviated with modified objective functions that contain *prior* knowledge about the scattering medium. We observed a strong cross-talk of scattering heterogeneities in the absorption images, whereas in the opposite case this effect is negligible. Again, both QN methods outperformed the CG method in terms of computational speed and reached a smaller value of the objective function at the final iteration step.

Acknowledgments

The authors would like to thank Dr Harry Graber (SUNY Downstate Medical Centre Brooklyn, NY) and Dr Kenneth M Hanson (Los Alamos National Laboratory, NM) for many discussions on optimization techniques and QN methods. This research was supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), a part of the National Institute of Health (grant #R01 AR46255), the New York City Council Speaker's Fund for Biomedical Research: towards the Science of Patient Care, the Whitaker Foundation and the Ernst-Schering Research Foundation (Germany). Alexander D Klose was supported by the Ernst-Schering Research Foundation, Germany.

References

[Arridge95] Arridge S R and Schweiger M 1995 Photon-measurement density function: part II. Finite-element-method calculations *Appl. Opt.* **34** 8026–37

[Arridge97] Arridge S R and Hebden J C 1997 Optical imaging in medicine: II. Modelling and reconstruction *Phys. Med. Biol.* **42** 841–53

[Arridge98a] Arridge S R and Lionheart W 1998 Nonuniqueeness in diffusion-based optical tomography *Opt. Lett.* **23** 882–4

[Arridge98b] Arridge S R and Schweiger M 1998 A gradient-based optimisation scheme for optical tomography *Opt. Express* **2** 213–26

[Arridge99] Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** R41–93

[Benaron94] Benaron D A, Ho D C, Spilman S D, Van Houten J P and Stevenson D K 1994 Non-recursive linear algorithms for optical imaging in diffusive media *Oxygen Transport to Tissue (Adv. Exp. Med. Biol. vol 16)* (New York: Plenum) pp 215–22

[Benaron00] Benaron D A *et al* 2000 Noninvasive functional imaging of human brain using light *J. Cereb. Blood Flow Metab.* **20** 469–77

[Bishop97] Bishop Ch M 1997 *Neural Networks for Pattern Recognition* (Oxford: Clarendon)

[Broyden65] Broyden C G 1965 A class of methods for solving nonlinear simultaneous equations *Math. Comput.* **19** 577–93

[Broyden67] Broyden C G 1967 Quasi-Newton methods and their application to function minimization *Math. Comput.* **21** 368–81

[Case67] Case K M and Zweifel P F 1967 *Linear Transport Theory* (Reading, MA: Addison-Wesley)

[Chandrasekhar60] Chandrasekhar S 1960 *Radiative Transfer* (London: Oxford Univerity Press)

[Chang96] Chang J, Graber H L, Barbour R L and Aronson R 1996 Recovery of optical cross section perturbations in dense-scattering media by transport-theory-based imaging operators and steady-state simulated data *Appl. Opt.* **35** 3963–78

[Colak97] Colak S B, Papaioannou D G, 't Hooft G W, van der Mark M B, Schomberg H, Paasschens J C J, Melissen J B M and van Asten N A A J 1997 Tomographic image reconstruction from optical projections in light-diffusing media *Appl. Opt.* **36** 180–213

[Colak99] Colak S B, van der Mark M B, 't Hooft G W, Hoogenraad J H, van der Linden E S and Kuijpers F A 1999 Clinical optical tomography and NIR spectroscopy for breast cancer detection *IEEE J. Sel. Top. Quantum Electron.* **5** 1143–58

[Davidon91] Davidon W C 1991 Variable metric method for minimization *SIAM J. Optim.* **1** 1–17

[Dennis77] Dennis J E and More J J 1977 Quasi-Newton methods, motivation and theory *SIAM Rev.* **19** 46–89

[Dorn98] Dorn O 1998 A transport–backtransport method for optical tomography *Inverse Problems* **14** 1107–30

[Fletcher70] Fletcher R 1970 A new approach to variable metric algorithms *Comput. J.* **13** 317–22

[Fletcher87] Fletcher R 1987 *Practical Methods of Optimization* (Chichester: Wiley)

[Flock89a] Flock S T, Patterson M S, Wilson B C and Wyman D R 1989 Monte Carlo Modelling of light propagation in highly scattering tissue: 1. Model predictions and comparison with difusion theory *IEEE Trans. Biomed. Eng.* **36** 1162–8

[Flock89b] Flock S T, Wilson B C and Patterson M S 1989 Monte Carlo Modelling of light propagation in highly scattering tissue: 2. Comparison with measurements in phantoms *IEEE Trans. Biomed. Eng.* **36** 1169–73

[Goldfarb70] Goldfarb D 1970 A family of variable metric methods derived by variational means *Math. Comput.* **24** 23–6

[Hanson98] Hanson K M, Cunningham G S and Saquib S S 1998 Inversion based on computational simulations *Maximum Entropy and Bayesian Methods* ed G J Erickson, J T Rychert and C R Smith (Dordrecht: Kluwer–Academic) pp 121–35

[Herman80] Herman G T 1980 *Image Reconstruction from Projections. The Fundamentals of Computerized Tomography* (New York: Academic)

[Hielscher99] 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

[Hielscher01] Hielscher A H and Bartel S 2001 Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography *J. Biomed. Opt.* **6** 183–92

[Jiang96] Jiang H, Paulsen K D, Österberg U L, Pogue B W and Patterson M S 1996 Optical image reconstruction using frequency-domain data: simulations and experiments *J. Opt. Soc. Am.* A **13** 253–67

[Klose99] Klose A D and Hielscher A H 1999 Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer *Med. Phys.* **26** 1698–707

[Klose02a] Klose A D, Netz U, Beuthan J and Hielscher A H 2002 Optical tomography using the time-independent equation of radiative transfer: part I. Forward model *J. Quant. Spectrosc. Radiat. Transfer* **72** 691–713

[Klose02b] Klose A D and Hielscher A H 2002 Optical tomography using the time-independent equation of radiative transfer: part II. Inverse model *J. Quant. Spectrosc. Radiat. Transfer* **72** 715–32

[Liu89] Liu D C and Nocedal J 1989 On the limited memory BFGS method for large scale optimization *Math. Programme.* **45** 503–28

[Luenberger84] Luenberger D G 1984 *Linear and Nonlinear Programming* (Reading, MA: Addison-Wesley)

[Martinez00] Martinez J M 2000 Practical quasi-Newton methods for solving nonlinear systems *J. Comput. Appl. Math.* **124** 97–121

[Matson97] Matson C L, Clark N, McMackin L and Fender J S 1997 Three-dimensional tumor localization in thick tissue with the use of diffuse photon-density waves *Appl. Opt.* **36** 214–20

[McBride01] McBride T O, Pogue B W, Jiang S, Osterberg U L, Paulsen K D and Poplack S P 2001 Initial studies of *in vivo* absorbing and scattering heterogeneity in near-infrared tomographic breast imaging *Opt. Lett.* **26** 822–4

[Nash96] Nash S G 1996 *Linear and Nonlinear Programming* (New York: McGraw-Hill)

[Natterer99] Natterer F 1999 Mathematical models for medical imaging *Computational Radiology and Imaging: Therapy and Diagnostics (The IMA Volumes in Mathematics and its Applications)* ed Ch Boergers and F Natterer (New York: Springer) pp 17–32

[Netz01] Netz U, Beuthan J, Capius H J, Koch H C, Klose A D and Hielscher A H 2001 Imaging of rheumatoid arthritis in finger joints by sagittal optical tomography *Med. Laser Appl.* **16** 306–10

[Nocedal80] Nocedal J 1980 Updating quasi-Newton matrices with limited storage *Math. Comput.* **35** 773–82

[Nocedal92] Nocedal J 1992 Theory of algorithms for unconstrained optimization *Acta Numer.* **1** 199–242

[Nocedal99] Nocedal J and Wright S J 1999 *Numerical Optimization* (New York: Springer)

[OLeary95] O'Leary M A, Boas D A, Chance B and Yodh A G 1995 Experimental images of heterogeneous turbid media by frequency domain diffusing photon tomography *Opt. Lett.* **20** 426–8

[Patterson91] Patterson M S, Wilson B C and Wyman R 1991 The propagation of optical radiation in tissue: I. Models of radiationtransport and their application *Lasers Med. Sci.* **6** 155–67

[Paulsen95] Paulsen K D and Jiang H 1995 Spatially-varying optical property reconstruction using a finite element diffusion equation approximation *Med. Phys.* **22** 691–701

[Press92] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in C* (New York: Cambridge University Press)

[Radon17] Radon J H 1917 Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten *Ber. Verh. Sächs. Akad. Wiss.* **69** 262–77

[Roy01] Roy R and Sevick-Muraca E M 2001 A numerical study of gradient-based nonlinear optimization methods for contrast enhanced optical tomography *Opt. Express.* **9** 49–65

[Ruggirero96] Gomes-Ruggiero M A and Kozakevich D N 1996 A numerical study on large-scale nonlinear solvers *Comput. Math. Appl.* **32** 1–13

[Saquib97] Saquib S S, Hanson K M and Cunningham G S 1997 model-based image reconstruction from time-resolved diffusion data *Medical Imaging. Image Processing (Proc. SPIE)* vol 3034, ed K M Hanson (Bellingham, WA: SPIE Optical Engineering Press) pp 369–80

[Schottland93] Schottland J C, Haselgrove J C and Leigh J S 1993 Photon hitting density *Appl. Opt.* **32** 448–53

[Schweiger99] Schweiger M and Arridge S R 1999 Application of temporal filters to time resolved data in optical tomography *Phys. Med. Biol.* **44** 1699–717

[Shanno70] Shanno D 1970 Conditioning of quasi-Newton methods for function minimization *Math. Comput.* **24** 647–57

[Shanno78] Shanno D 1978 Conjugate gradient methods with inexact searches *Math. Oper. Res.* **3** 244–56

[Xu02] Xu Y, Gu X, Khan T and Jiang H 2002 Absorption and scattering images of heterogeneous scattering media can be simultaneously reconstructed by use of dc data *Appl. Opt.* **41** 5427–37

[Yao97] 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

[Ye99] Ye J C, Webb K J, Millane R P and Downar T J 1999 Modified distorted Born iterative method with an approximate Frechet derivative for optical tomography *J. Opt. Soc. Am.* A **16** 1814–26