
PERGAMON

Journal of Quantitative Spectroscopy & Radiative Transfer 72 (2002) 715–732

Journal of Quantitative Spectroscopy & Radiative Transfer

www.elsevier.com/locate/jqsrt

Optical tomography using the time-independent equation of radiative transfer—Part 2: inverse model

Alexander D. Klose[a,b], Andreas H. Hielscher[a,c,*]

[a] *Department of Pathology, State University of New York Downstate Medical Center, 450 Clarkson Avenue, Brooklyn, NY 11203, USA*

[b] *Institut für Medizinische Physik, Klinikum Benjamin Franklin, Freie Universität Berlin, Fabeckstr. 60-62, 14195 Berlin, Germany*

[c] *Department of Electrical and Computer Engineering, Polytechnic University, 5 Metrotech Center, Brooklyn, NY 11201, USA*

Received 3 January 2001; accepted 20 April 2001

Abstract

Optical tomography is a novel imaging modality that is employed to reconstruct cross-sectional images of the optical properties of highly scattering media given measurements performed on the surface of the medium. Recent advances in this field have mainly been driven by biomedical applications in which near-infrared light is used for transillumination and reflectance measurements of highly scattering biological tissues. Many of the reconstruction algorithms currently utilized for optical tomography make use of model-based iterative image reconstruction (MOBIIR) schemes. The imaging problem is formulated as an optimization problem, in which an objective function is minimized. In the simplest case the objective function is a normalized-squared error between measured and predicted data. The predicted data are obtained by using a forward model that describes light propagation in the scattering medium given a certain distribution of optical properties.

In part I of this two-part study, we presented a forward model that is based on the time-independent equation of radiative transfer. Using experimental data we showed that this transport-theory-based forward model can accurately predict light propagation in highly scattering media that contain void-like inclusions. In part II we focus on the details of our image reconstruction scheme (inverse model). A crucial component of this scheme involves the efficient and accurate determination of the gradient of the objective function with respect to all optical properties. This calculation is performed using an adjoint differentiation algorithm that allows for fast calculation of this gradient. Having calculated this gradient, we minimize the objective function with a gradient-based optimization method, which results in the reconstruction of the spatial distribution of scattering and absorption coefficients inside the medium. In addition to presenting

*The authors are now with the Departments of Biomedical Engineering and Radiology, Columbia University, 416 Schapiro - CEPSR Bldg., Mail Code 8904, 530 West 120th Street, New York, NY 10027, USA. Tel. +1-212-854-4460; Fax: +1-212-854-8725.

E-mail address: ahielscher@downstate.edu (A.H. Hielscher).

PII: S0022-4073(01)00151-0

the mathematical and numerical background of our code, we present reconstruction results based on experimentally obtained data from highly scattering media that contain void-like regions. These types of media play an important role in optical tomographic imaging of the human brain and joints.

Keywords: Equation of radiative transfer; Transport theory; Inverse problems; Adjoint model; Adjoint method; Adjoint differentiation; Reverse differentiation; Optical tomography; Scattering media

1. Background and introduction

Optical tomography (OT) is concerned with reconstructing the spatial distribution of optical properties inside scattering and absorbing media by using near-infrared-light transmission measurements performed on the surface of the media [1]. Most of the currently employed imaging algorithms make use of so-called model-based iterative image reconstruction (MOBIIR) schemes. These schemes use a forward model to predict detector readings assuming a certain distribution of optical properties inside the medium. The predicted detector readings are compared to actual measurements by defining an appropriate objective function Φ. The objective function Φ is minimized by iteratively updating an initial guess of the distribution of optical properties. MOBIIR schemes in OT mainly differ in their choice of forward model and in what way the spatial distribution of optical properties is updated.

In part I of this study we focused on the importance of choosing a correct forward model in MOBIIR schemes. We introduced and experimentally tested an upwind-difference discrete-ordinates algorithm that computes numerical solutions of the time-independent equation of radiative transfer. Unlike the widely used diffusion approximation, this model is capable of accurately describing light propagation in media that contain void-like regions with very low scattering and absorption coefficients [2,3]. These types of media have drawn the attention of several research groups in recent years because void-like regions play an important role in several biomedical imaging applications. For example, in applications of OT to brain imaging one has to take into account the effects of the low-scattering and low-absorbing cerebrospinal fluid, which surrounds the brain. The effect of this layer on light propagation has been the subject of many studies and discussions [2–9]. Another example is the almost clear synovial fluid in joints, which plays an important role in the optical detection of rheumatoid arthritis [10–12].

In this part II we focus on how the radiative-transfer forward model can be used within gradient-based MOBIIR schemes. In recent years, several groups have embraced the concept of gradient-based MOBIIR schemes for optical tomography. Arridge et al. [13], Davis et al. [14], Hielscher et al. [15], and Roy et al. [16] have presented schemes for the diffusion equation, while Klose and Hielscher [17] have employed the gradient-based approach using the equation of radiative transfer. Gradient-based schemes use information about the gradient of the objective function with respect to the optical properties to find updates of the initial guess of the optical properties. For example, steepest-gradient-descent and conjugate-gradient schemes are well-established techniques in optimization theory [18]. Compared to other widely used

Newton-type algorithms, gradient-based schemes have the advantage that a full Jacobian matrix **J** neither needs to be explicitly created nor repeatedly inverted [1,15].

The major challenge concerning gradient-based MOBIIR algorithms, is to find a way to efficiently calculate the gradient $\nabla_\mu \Phi = \mathrm{d}\Phi/\mathrm{d}\mu$ of the objective function Φ with respect to the optical properties $\mu = [\mu_a(\mathbf{r}), \mu_s(\mathbf{r})]$. A straightforward approach would be to approximate the derivative with the method of divided difference, given by

$$\frac{\mathrm{d}\Phi}{\mathrm{d}\mu} \approx \frac{\Phi(\mu + \Delta\mu) - \Phi(\mu)}{\Delta\mu}. \tag{1}$$

If μ is a vector of n unknown optical properties, one has to run $n+1$ forward problems to obtain this gradient. Given that in optical tomography the number of unknowns typically is on the order of 10^3–10^5, this requirement results in an unacceptable computational burden. This is especially true in the case of transport-theory-based schemes, in which solving the forward model requires substantial computational time.

Instead of perturbing each component of the vector μ using Eq. (1), we employ an adjoint model to efficiently calculate the gradient $\nabla_\mu \Phi$. The power of this approach lies in its ability to calculate the derivatives with respect to all the variables in a CPU time that is comparable to one forward calculation. This method is widely used in meteorology and oceanography for sensitivity studies, data assimilation, and parameter estimation [19–21]. For a good overview of the adjoint model as applied to atmospherical sciences see, for example, Errico [20]. In atmospherical sciences a forward model B, for example an ocean circulation or climate model, is used to calculates the *output* parameters b given the *input* parameters a with

$$b = B(a). \tag{2}$$

The *output* parameters b are used in a forecast error $E(b)$. In general, it is of interest is how the forecast error depends on the input variables a. Researchers in atmospherical sciences typically refer to adjoint models, when they describe the equations that provide the sensitivity $\partial E/\partial a$ (gradient) as a function of the derivative $\partial E/\partial b$. Therefore, equations of the form

$$\frac{\partial E}{\partial a} = \left(\frac{\partial b}{\partial a}\right)^{\mathrm{T}} \frac{\partial E}{\partial b} \tag{3}$$

are called adjoint models, where now the *input* parameters are the derivatives $\partial E/\partial b$ and the *output* parameters are the derivatives $\partial E/\partial a$. In general $(\partial b/\partial a)^{\mathrm{T}}$ is the adjoint matrix of $\partial b/\partial a$, which in the case of real-valued parameters equals the transpose of $\partial b/\partial a$. In this work we have adapted this technique to OT and use it to calculate the gradient $\nabla_\mu \Phi = \mathrm{d}\Phi/\mathrm{d}\mu$, which is similar to a calculation of the gradient $\partial E/\partial a$. The *input* parameters of the forward model are the optical properties μ (a in Eq. (2)) and the *output* parameters are the detector readings **P** (b in Eq. (2)). On the other hand, the adjoint model uses the derivative of Φ with respect to the detector readings **P** as *input* parameters, and yields the derivative of Φ with respect to the optical parameters μ as *output* parameters (Eq. (3)).

One can distinguish between three different approaches to the calculation of $\nabla_\mu \Phi$ utilizing an adjoint model (Fig. 1) [21]. In the first method one obtains the gradient $\nabla_\mu \Phi$ by using the solution of the adjoint equation of the forward model. The challenge in this approach is to derive the adjoint equation from a given forward model and to solve it. A good general

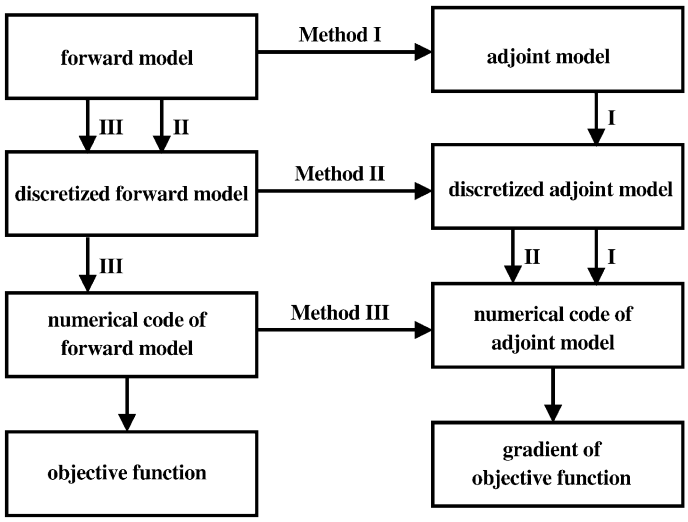

Fig. 1. There are basically three different ways to obtain the gradient of an objective function using adjoint schemes. All methods start from the formulation of the equation that presents the forward model (e.g. the diffusion equation or the radiative-transfer equation.) In Method I one derives first the adjoint equation, then discretizes the adjoint equation, implements a numerical code for the adjoint equation, and from that obtains the gradient of the objective function. In Method II one formulates the discretized equation first, then derives the discretized adjoint equation, and programs the numerical adjoint equation. In Method III, one formulates the discretized equation, programs the numerical code for the discretized equation, and implements a numerical code for the adjoint equation. A detailed discussion can be found in the main text.

overview on the theory of adjoint equations in dynamical systems can be found, for example, in works by Marchuk et al. or Cacuci [22–25]. Talagrand gives an example how this approach can be used for sensitivity calculation in meteorological applications [26]. Norton uses this method for computing the Frechet derivative for inverse scattering problems and compares it to other techniques [27]. In optical tomography no group has adapted the adjoint transport equation in the calculation of the gradient within a MOBIIR scheme. However, Dorn uses the time-dependent adjoint transport equation to determine the Frechet derivative of a residual that is proportional to the difference between predicted and measured data [28]. The resulting nonlinear system of equations is solved by a nonlinear generalization of the algebraic reconstruction technique (ART), where the optical parameters are iteratively updated. Dorn presents numerical results for scattering media with non-reentry boundary conditions. A similar approach is applied by Arridge et al. to the diffusion equation in OT [13]. Arridge et al. derived the gradient $\nabla_{\mu}\Phi$ from the solution to the diffusion equation for a source and from the solution to the adjoint diffusion equation for the boundary residual. The boundary residual is a function of the difference of the measured and predicted data. The solutions of the diffusion equation and its adjoint equation are obtained numerically by a finite element method (FEM). Since numerically solving the adjoint diffusion equation requires approximately the same time as solving the diffusion equation itself, Arridge et al. obtain the gradient in a time comparable to two forward calculations.

Several authors [21,29] have pointed out that the appropriate discretization scheme for the adjoint equation is in general different from the appropriate discretization of the forward equation. Therefore it is a priori not clear that the gradient obtained with a discretized version of the adjoint equation truly equals the gradient of the discretized version of the forward equation. Therefore Shah [30] and Talagrand et al. [31] have argued that it is favorable to derive the adjoint model from the discretized form of the forward equation. This is a second way to use adjoint schemes for the gradient calculation. This approach has been applied mainly to weather forecast models [32] and to ocean circulation models [33], but has not been pursued in optical tomography.

The third approach for calculating the gradient by means of an adjoint model does not require the formulation of an adjoint equation, neither of the continuous nor discretized version of the equation that describes the forward model. This method is usually referred to as computational differentiation in the adjoint or reverse mode, reverse differentiation, or adjoint differentiation [34]. Here, the numerical code of the forward model, which is a sequence of arithmetic operations, is directly differentiated to compute the gradient. The procedure used to find the derivatives of arbitrary algebraic functions, such as the gradient of an objective function, was first introduced by Wengert [35]. Over the last 15 years, Griewank has generalized and refined Wengert's initial ideas [36]. Again, the main applications so far lie outside the field of optical tomography, as for example, in geoscience [21,37,38]. The key to this method is the decomposition of a given forward model into a series of elementary differentiable function steps. By systematically applying the chain rule of differentiation to every single elementary step of the forward code in the reverse direction, a value for the gradient is obtained. The main advantage of this approach is that on the level of the elementary steps in the forward model code the gradient can be attained according to simple rules [21]. Thus the task of gradient determination can be handled independently from any knowledge of the nature of the main problem.

In optical tomography only the groups of Davis et al. [14], Roy et al. [16], and ours [15,17] have made use of the concept of adjoint differentiation. Davis and Roy have applied the adjoint differentiation technique to the time-independent diffusion equation and to diffusion fluorescent problems in the frequency domain. Hielscher et al. [15] described an adjoint differentiation scheme for the time-dependent diffusion equation. Most recently, Klose and Hielscher [17] employed adjoint differentiation in a reconstruction algorithm that uses the time-dependent transport equation. They considered isotropically scattering media with none-reentry boundary conditions and showed, using simulated data, reconstructions of the spatial distribution of the scattering coefficients μ_s.

Fig. 1 depicts the relationships among the three different approaches to the gradient calculation using an adjoint model. All methods start from the forward model, which in OT is based on the diffusion equation or radiative transfer equation. In method I, one first derives the adjoint equation of the forward model, which is then discretized and implemented as a computer code. Employing method II one first discretizes the equations used in the forward model, then derives the adjoint discretized model, and implements the discretized model as a computer program. Using method III, one first discretizes the forward model, implements a forward computer code, and from this derives the code for the adjoint model. In the last case, which is used in this paper, the gradient is directly determined from the forward computer code without explicit knowledge of the adjoint equation.

In part I of this work we described how our previous transport code [17] was extended to include anisotropic scattering and Fresnel boundary conditions at the air–tissue interface. Therefore, the new code considers a refractive index mismatch between tissue and air. With this additions it is for the first time possible to test the algorithm with experimental data from well-characterized tissue phantoms. Furthermore, as described in detail in the companion manuscript, we solve the discretized equation of radiative transfer with a successive over-relaxation (SOR) method instead of using the Jacobi method, which was employed in our previous work [17]. The use of the SOR method considerably improves the convergence rate of the forward algorithm.

In part II we now focus on how adjoint differentiation can be used to obtain the gradient of the objective function. The use of the SOR method in the forward code results in a different adjoint code from the one presented in Ref. [17]. First, we show how the calculation of the objective function is decomposed into sub-functions of elementary computational units. Then we describe how these sub-functions are differentiated by applying the adjoint differentiation technique. This yields the gradient of the objective function, which in turn is used to optimize the objective function. To illustrate the performance of our code, we analyze experimental data obtained from media that contain void-like regions. As already mentioned, these types of media have drawn considerable attention in recent years, since they are significant in several areas of optical tomography [4–12].

2. Numerical methods

First, a forward model is employed to provide predictions of detector readings as a function of certain system parameters. (In OT these system parameters are the optical properties.) Second, an objective function is evaluated that compares the predicted with measured data. And Third, an updating scheme is used to provide a new set of system parameters for the next iteration. The forward model was described and tested in detail in part I of this work. Here we focus on the second and third component, namely the objective function and the updating scheme, in which adjoint differentiation is used for the gradient calculation.

2.1. Objective function

OT can be viewed as an optimization problem with a nonlinear objective function. The discrepancy between the actual measurements, represented by the vector $\mathbf{M}$, and the predictions, given by the vector $\mathbf{P}$, for m source–detector pairs is defined by a scalar-valued objective function $\Phi(\mathbf{P})$. The predictions $\mathbf{P}$ are mapped onto a scalar φ using the χ^2-error norm

$$\Phi : \Re^m \to \Re, \tag{4a}$$

$$\mathbf{P} \mapsto \varphi = \frac{1}{2}\sum_{i}^{m} (P_i - M_i)^2, \tag{4b}$$

where $\Re^m$ is the m-dimensional space of real numbers.

The predictions $\mathbf{P}$ are determined for all m source–detector pairs by evaluating the forward model F, given the spatial distribution of optical parameters μ (see part I):

$$F : \Re^n \to \Re^m, \tag{5a}$$

$$\mu \mapsto P(\mu). \tag{5b}$$

The vector μ is of length $n = 2IJ$, and contains the optical parameters μ_s and μ_a. The parameters I and J denote the number of grid points of a finite-difference grid along the x-axis and y-axis, respectively.

The goal in image reconstruction is to determine a vector μ that minimizes the objective function Φ. As already mentioned in the introduction, gradient-based optimization techniques, such as the conjugate-gradient method, employ the gradient $\nabla_\mu \Phi(\mu)$ to calculate a sequence of points $\mu_0, \mu_1, \ldots, \mu_i$ that lead to ever-improving values of Φ. This iterative process is terminated when $|(\Phi(\mu_{i+1}) - \Phi(\mu_{i+1}))|$ becomes smaller than a predefined value ε. A crucial task within this process is the computationally efficient calculation of the gradient $\nabla_\mu \Phi(\mu_i)$. In the following section we will describe in detail how this can be done using an adjoint differentiation scheme.

2.2. Gradient calculation with adjoint differentiation

In adjoint differentiation schemes the numerical code that calculates the objective function Φ is directly differentiated to obtain the gradient $\nabla_\mu \Phi(\mu_i)$. To apply this method one has to decompose the objective function into a series of elementary differentiable function steps. By systematically applying the chain rule of differentiation to each of these elementary steps in the reverse direction of the forward model computation, a value for the gradient is obtained.

In our particular problem we notice that $\Phi = \Phi(P(\mu))$ which can be decomposed into Z sub-functions F^z in the following way:

$$\begin{aligned}\Phi(F(\mu)) &= \Phi(F^Z(F^{Z-1}(F^{Z-2}(\ldots(F^2(F^1(\mu),\mu))\ldots)\mu)\mu)) \\ &:= (\Phi \circ F^Z(\mu) \circ F^{Z-1}(\mu) \circ F^{Z-2}(\mu) \circ \cdots \circ F^2(\mu) \circ F^1)(\mu).\end{aligned} \tag{6}$$

The sub-functions F^z are defined by the SOR-method that is used to solve the forward model (see part I). The SOR-method is an iterative approach and the zth iteration yields the intermediate result $\Psi^z_{k,i,j}$. The radiance vector Ψ is of length $p = K \times I \times J$ with $k \in [1,K]$, $i \in [1,I]$, and $j \in [1,J]$. The detector readings $\mathbf{P}(\mu)$ are the angular-dependent radiances $\Psi^Z_{k,i,j}$ at the last iteration step Z at detector positions (i,j) on the boundary. The computational graph of the forward model is depicted in Fig. 2a. Starting with the optical parameters μ as the input variables, the sub-function F^1 produces the intermediate result and output variable Ψ^1. The output variable Ψ^z of F^z and the optical parameters μ always serve as input variables for the next sub-function F^{z+1} for all subsequent steps of the decomposition. The step F^z calculates the predictions $\mathbf{P}$, which become the input to the final step of the objective function Φ determination, which is the calculation of the scalar φ.

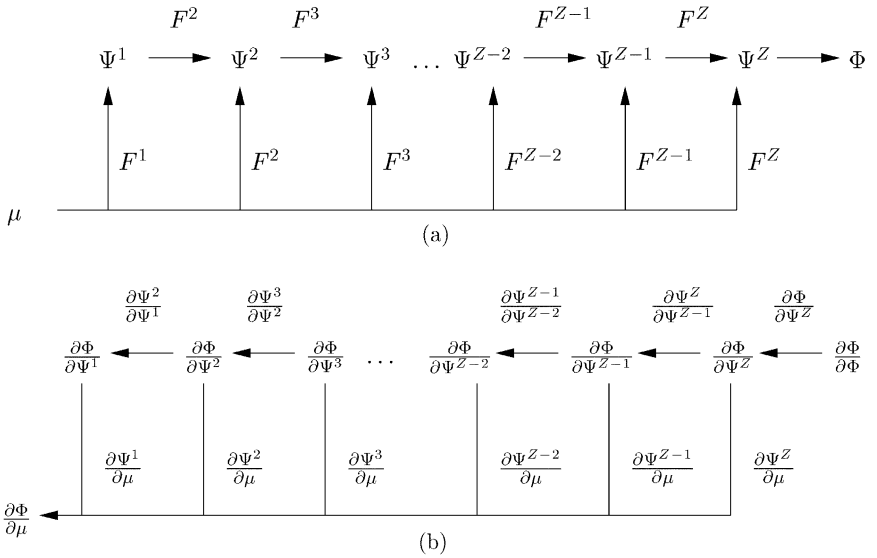

Fig. 2. (a) Computational graph of the forward model calculating the objective function. (b) Computational graph of the adjoint differentiation technique.

The sub-functions F^z map the intermediate variables Ψ^{z-1} and the input values of μ onto the intermediate result $\Psi^z = F^z(\Psi^{z-1}, \mu)$ for all iteration steps of the transport forward model

$$F^z : \Re^p \times \Re^n \to \Re^p, \tag{7a}$$

$$\begin{pmatrix} \Psi^{z-1} \\ \mu \end{pmatrix} \mapsto \Psi^z. \tag{7b}$$

For the ordinates with $\xi_k > 0$, $\eta_k > 0$ this mapping is given explicitly by (see part I)

$$\Psi^z_{k,i,j} = (1-\rho)\Psi^{z-1}_{k,i,j} + \rho \frac{\{S_{k,i,j} + \mu_s \sum_{k'=1}^{K} a_{k'} P_{k,k'} \Psi^{z-1}_{k',i,j} + (\xi_k/\Delta x)\Psi^z_{k,i-1,j} + (\eta_k/\Delta y)\Psi^z_{k,i,j-1}\}}{\{\xi_k/\Delta x + \eta_k/\Delta y + (\mu_a + \mu_s)\}}. \tag{8}$$

For the first iteration step $z = 1, F^z$ only maps the input variables μ:

$$F^1 : \Re^n \to \Re^p, \tag{9a}$$

$$\mu \mapsto \Psi^1 \tag{9b}$$

and we have

$$\Psi^1_{k,i,j} = \rho \frac{\{S_{k,i,j} + (\xi_k/\Delta x)\Psi^1_{k,i-1,j} + (\eta_k/\Delta y)\Psi^1_{k,i,j-1}\}}{\{\xi_k/\Delta x + \eta_k/\Delta y + (\mu_a + \mu_s)\}}. \tag{10}$$

Eqs. (8) and (10) are the smallest computational units in the forward code and play an important role in the adjoint differentiation step, which we discuss next.

To obtain the gradient of the objective function, we start by differentiating Eq. (6) with respect to the optical properties μ. We derived the following equation for the gradient (see Appendix A):

$$\nabla_\mu \Phi := \left(\frac{\mathrm{d}\Phi}{\mathrm{d}\mu}\right)^{\mathrm{T}} = \left(\sum_z \left(\frac{\partial \Phi}{\partial \Psi^z}\right)^{\mathrm{T}} \frac{\partial \Psi^z}{\partial \mu}\right) + \left(\frac{\partial \Phi}{\partial \mu}\right)^{\mathrm{T}} \tag{11}$$

which contains three distinct terms. The last term $(\partial\Phi/\partial\mu)^{\mathrm{T}}$ is zero, because Φ is not an explicit function of the optical properties. The middle term $\partial\Psi^z/\partial\mu$ can be calculated from Eq. (8) of the forward model. For the derivatives with respect to μ_a and μ_s we obtain

$$\left(\frac{\partial \Psi^z}{\partial \mu_\mathrm{s}}\right)_{k,i,j} = \rho \frac{\sum_{k'=1}^{K} a_{k'}\, p_{k,k'}\, \Psi^{z-1}_{k',i,j}}{\{\xi_k/\Delta x + \eta_k/\Delta y + (\mu_\mathrm{a} + \mu_\mathrm{s})\}} - \rho \frac{\{S_{k,i,j} + \mu_\mathrm{s} \sum_{k'=1}^{K} a_{k'}\, p_{k,k'}\, \Psi^{z-1}_{k',i,j} + (\xi_k/\Delta x)\Psi^z_{k,i-1,j} + (\eta_k/\Delta y)\Psi^z_{k,i,j-1}\}}{\{\xi_k/\Delta x + \eta_k/\Delta y + (\mu_\mathrm{a} + \mu_\mathrm{s})\}^2}, \tag{12a}$$

$$\left(\frac{\partial \Psi^z}{\partial \mu_\mathrm{a}}\right)_{k,i,j} = -\rho \frac{\{S_{k,i,j} + \mu_\mathrm{s} \sum_{k'=1}^{K} a_{k'}\, p_{k,k'}\, \Psi^{z-1}_{k',i,j} + (\xi_k/\Delta x)\Psi^z_{k,i-1,j} + (\eta_k/\Delta y)\Psi^z_{k,i,j-1}\}}{\{\xi_k/\Delta x + \eta_k/\Delta y + (\mu_\mathrm{a} + \mu_\mathrm{s})\}^2}. \tag{12b}$$

At this point we have not yet used the adjoint differentiation technique, since we have not stepped backward through the forward code. This procedure comes into play in the calculation of the first term $(\partial\Phi/\partial\Psi^z)^{\mathrm{T}}$ in Eq. (11). This can be best understood while looking at the computational graph of the adjoint differentiation technique in Fig. 2b. Starting with the last step (output) of the forward code, which is the calculation of the objective function given the predictions $P = \Psi^Z$, we differentiate Φ with respect to Ψ^Z and multiply it with $(\partial\Phi/\partial\Phi)^{\mathrm{T}} = 1$. The result is the difference between the measured and predicted data

$$\left(\frac{\partial \Phi}{\partial \Psi^z}\right)^{\mathrm{T}} = \left(\Psi^Z - \mathbf{M}\right)^{\mathrm{T}}. \tag{13}$$

This is the *input* parameter to our adjoint model, which will eventually provide the *ouput* parameter $\nabla_\mu\Phi$ (see also Eq. (3)). More specifically, continuing to step backward through the forward code we calculate $(\partial\Phi/\partial\Psi^{Z-1})^{\mathrm{T}}$, which is given by

$$\left(\frac{\partial \Phi}{\partial \Psi^{z-1}}\right)^{\mathrm{T}} = \left(\frac{\partial \Psi^z}{\partial \Psi^{z-1}}\right)^{\mathrm{T}} \left(\frac{\partial \Phi}{\partial \Psi^z}\right)^{\mathrm{T}}. \tag{14}$$

The remaining derivatives $(\partial\Phi/\partial\Psi^z)^{\mathrm{T}}$ of all intermediate steps in Eq. (11) are computed using the previously calculated derivatives $(\partial\Phi/\partial\Psi^{z+1})^{\mathrm{T}}$, such that

$$\begin{aligned}\left(\frac{\partial\Phi}{\partial\Psi^z}\right)^{\mathrm{T}} &= \frac{\partial(\Phi\circ\cdots\circ\Psi^{z+1})(\Psi^z)^{\mathrm{T}}}{\partial\Psi^z}\\ &= \left(\frac{\partial\Psi^{z+1}}{\partial\Psi^z}\right)^{\mathrm{T}}\frac{\partial(\Phi\circ\cdots\circ\Psi^{z+2})(\Psi^{z+1})^{\mathrm{T}}}{\partial\Psi^{z+1}}\\ &= \left(\frac{\partial\Psi^{z+1}}{\partial\Psi^z}\right)^{\mathrm{T}}\left(\frac{\partial\Phi}{\partial\Psi^{z+1}}\right)^{\mathrm{T}}.\end{aligned} \tag{15}$$

This step, in which $(\partial\Phi/\partial\Psi^z)^{\mathrm{T}}$ is calculated from $(\partial\Phi/\partial\Psi^{z+1})^{\mathrm{T}}$, constitutes the adjoint differentiation step in our code. The matrix $(\partial\Psi^{z+1}/\partial\Psi^z)^{\mathrm{T}}$ is calculated by analytically differentiating the $(z+1)$th SOR-iteration step, given in Eq. (8), with respect to Ψ^z. We get, for example in the case of ordinates with $\xi_k>0, \eta_k>0$:

$$\frac{\partial\Psi^{z+1}_{k,i,j}}{\partial\Psi^z_{k',i',j'}} = (1-\rho)\delta_{k,i,j} + \rho\frac{\left\{\mu_{\mathrm{s}}a_{k'}\,p_{k,k'}\delta_{i,j} + \dfrac{\xi_k}{\Delta x}\delta_{k,i-1,j} + \dfrac{\eta_k}{\Delta y}\delta_{k,i,j-1}\right\}}{\left\{\dfrac{\xi_k}{\Delta x} + \dfrac{\eta_k}{\Delta y}(\mu_{\mathrm{a}}+\mu_{\mathrm{s}})\right\}} \tag{16}$$

with $\delta_{a,b,c} = \delta_a\delta_b\delta_c$ and

$$\delta_x = \begin{cases} 1 & \text{if } x = x',\\ 0 & \text{if } x \neq x'.\end{cases}$$

Here, we made the approximations

$$\frac{\xi_k}{\Delta x}\frac{\partial\Psi^{z+1}_{k,i-1,j}}{\partial\Psi^z_{k',i',j'}} \cong \frac{\xi_k}{\Delta x}\delta_{k,i-1,j}$$

and

$$\frac{\eta_k}{\Delta y}\frac{\partial\Psi^{z+1}_{k,i,j-1}}{\partial\Psi^z_{k',i',j'}} \cong \frac{\eta_k}{\Delta y}\delta_{k,i,j-1}$$

for the relevant terms on the right-hand side of Eq. (16).

As can be seen, the gradient of the objective function is calculated by stepping backward through all previously calculated iteration steps of the forward model without solving an entirely new numerical adjoint equation of radiative transfer. Furthermore, the particular underlying physical system does not have to be known, because the derivative is computed directly from the elementary steps of the forward model code (Eqs. (8) and (10)). A disadvantage of the reverse mode of differentiation is that all intermediate results of the forward calculation have to be stored for subsequent use in the backward pass.

2.3. Gradient-based optimization

Once the gradient $\nabla_\mu \Phi(\mu_0)$ for a point μ_0 is obtained, a one-dimensional line minimization along the given gradient direction is performed until a minimum, $\Phi(\mu_1)$ is found. Various techniques can be applied to perform such one-dimensional minimizations [39,40]. Our approach is to start with three points, the initial guess $\mu_0, \mu_1 = \mu_0 + \Delta\mu$, and $\mu_2 = \mu_0 + 2\Delta\mu$ chosen along the direction of the gradient. After calculating the objective function at these three points, the largest calculated value is neglected, and a new point μ_3 is determined using the golden section rule until a minimum is bracketed [40]. At that time a parabola is fit through these three points. The smallest value of the objective function on this parabola is determined to be the minimum. Once the minimum is found, a new gradient calculation $\nabla_\mu \Phi(\mu_1)$ is performed at the new minimum and a new direction is chosen in a conjugate-gradient framework. In our work we employed a Polak–Riberie conjugate-gradient scheme [39,40].

3. Tissue phantoms and experimental setup

The inverse model was tested using experimental data from scattering phantoms that contained void-like perturbations. As mentioned in the Introduction, these types of media play an important role in optical tomography for two reasons. First, void-containing media are encountered in several important applications, such as brain or joint imaging. Second, the widely applied diffusion theory fails to describe light propagation in void-containing media and the equation of radiative transfer needs to be employed. We constructed two phantoms of different size and different void-like perturbations.

Phantom A measured $4 \times 4 \times 14\ \mathrm{cm}^3$ and contained a void-like ring filled with clear water. The water-ring had a thickness of 0.3 cm and an inner diameter of 2.4 cm. The background material of the phantom was a mixture of silicon-dioxide monospheres and ink embedded in clear epoxy resin. The optical properties of the background medium were $\mu_s' = 11.6 \pm 0.3\ \mathrm{cm}^{-1}$, $\mu_a = 0.35 \pm 0.1\ \mathrm{cm}^{-1}$, $g = 0.8$, and $n = 1.56$ (see part I). The cross-sectional shape of the phantom is shown in Fig. 3a.

Phantom B had dimensions of $3 \times 3 \times 14\ \mathrm{cm}^3$ and contained three cylindrical holes with a diameter of 0.5 cm. The holes were filled with clear water. In this case the optical properties of the background medium were $\mu_s' = 7.5 \pm 0.3\ \mathrm{cm}^{-1}$, $\mu_a = 0.62 \pm 0.12$, and $g = 0.85$, and $n = 1.56$. The cross-sectional shape of the phantom is shown in Fig. 3b.

Both phantoms were illuminated with near-infrared light at 678 nm using a laser diode with a collimated beam. The applied power on the phantom surface was approximately 10 mW. To detect the light we used an avalanche photodiode (APD) with a lock-in detection system. A frequency generator provided a 1014 Hz sinusoidal modulation of the laser diode. The detection area at the boundary of the phantom was limited by a pinhole, which had a diameter of 0.1 cm. The detector aperture was 45°. At each detection spot the acquisition time was approximately 10 s yielding a signal-to-noise ration of approximately 30 dB. The detector was translated along the two sides adjacent to, and one side opposite to, the source position. The outgoing fluences on each of the surfaces of the phantom were normalized to the mean values.

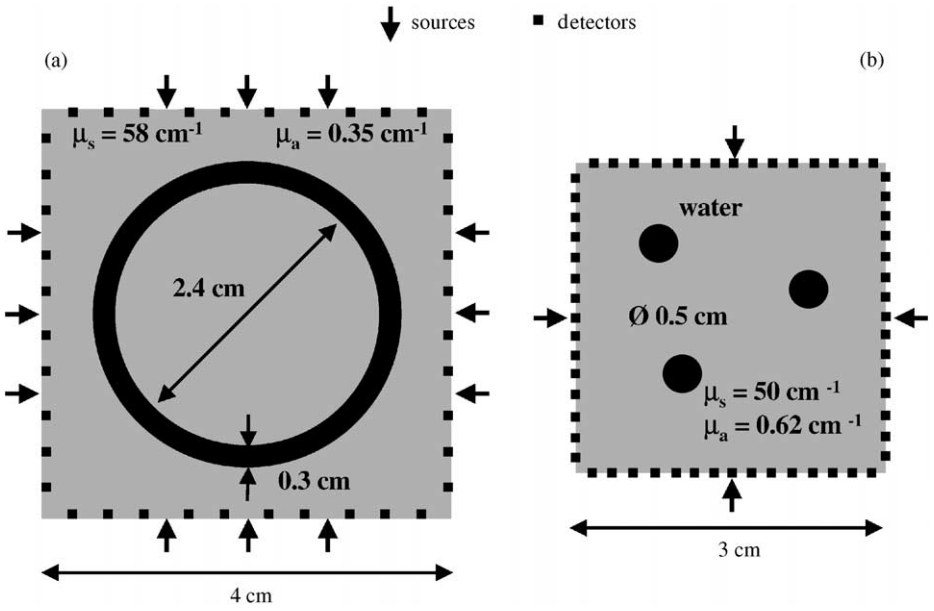

Fig. 3. Geometry and source–detector arrangements for the media with water-filled, void-like inclusions that are used in this study. Phantom A contains a water-filled ring (Fig. 3a), while phantom B contains three water-filled cylindrical inclusions (Fig. 3b).

The outgoing light was measured on phantom A for 12 different source positions, three symmetrically placed on each side of the phantom, with a separation of 0.8 cm. Ten detector points with a separation of 0.4 cm were recorded on each side, except the source side. This was repeated for each of the 12 different source positions yielding a total of 12×30 measurement points. For phantom B we used fewer sources but more detectors. Only one source was located at the center of each of the 4 sides of the phantom. We placed 15 detectors on each side with a separation of 0.2 cm yielding a total of 4×45 measurement points. A more detailed description of the experimental setup can be found in the companion manuscript (part I).

4. Results

The experimental data was used as input for the reconstruction algorithm, which calculated cross-sectional images of the μ_s and μ_a distribution in a horizontal section through the phantoms. As initial guesses for the reconstructions we chose homogeneous media with optical properties equal to those of the background medium. Therefore reconstructions for phantom A started from $\mu_a = 0.35\ cm^{-1}$, $\mu_s = 58\ cm^{-1}$ and $g = 0.8$, while reconstructions for phantom B started from $\mu_a = 0.62\ cm^{-1}$, $\mu_s = 50\ cm^{-1}$, and $g = 0.85$. The reconstruction process was stopped when the change of the objective function in subsequent iterations, $|(\Phi(\mu_{i+I}) - \Phi(\mu_i))/\Phi(\mu_{i+1})|$, was smaller than $\varepsilon = 10^{-3}$.

The reconstructions for phantom A were done on a 81×81 grid with a grid point separation of 0.05 cm. For the angular discretization we chose 16 ordinates. Figs. 4a and b show

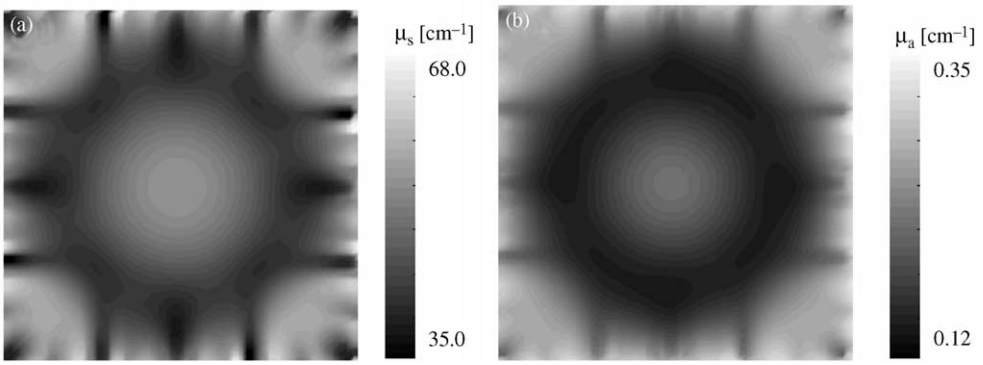

Fig. 4. Reconstruction results that are based on experimental data obtained with phantom A, which contains a water-filled void-like ring. Fig. 4a shows the results for the scattering coefficient and Fig. 4b shows the results for the absorption coefficient.

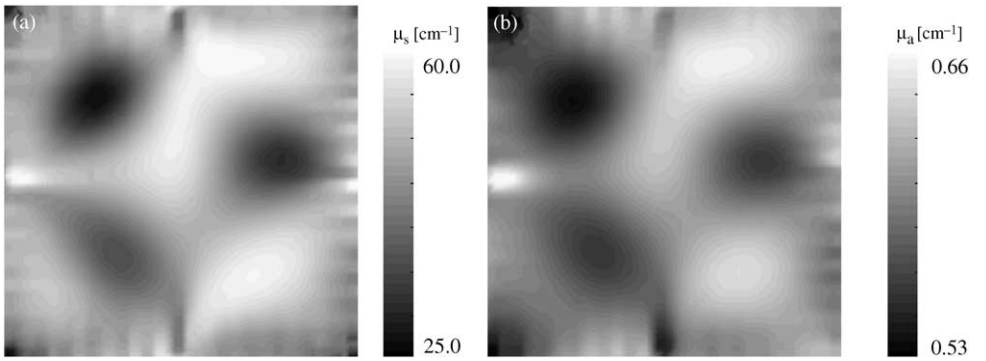

Fig. 5. Reconstruction results that are based on experimental data obtained with phantom B, which contains three water-filled perturbations. Fig. 5a shows the results for the scattering coefficient and Fig. 5b shows the results for the absorption coefficient.

the reconstruction results for μ_s and μ_a after 4 iterations, which took 98 h on a LINUX workstation with a 700 MHz Pentium III Xenon processor (ASL Workstations, Milpitas, CA). It can be seen that the ring-structure was found in both reconstructions. The lowest μ_s values in Fig. 4a form a ring with a diameter of 2.65 cm, which agrees well with that of the ring of the original (diameter 2.7 cm). However, the width of the ring appears broadened. The absolute values within the ring zone only decrease from the initial guesses of $\mu_s = 58.0$ to $35.0\,\mathrm{cm}^{-1}$, and $\mu_a = 0.35$ to $0.12\,\mathrm{cm}^{-1}$. That equals a relative change from the initial guesses of 39.7% for μ_s and 65.7% for μ_a.

The reconstructions for phantom B were done on a 61×61 grid with a grid point separation of 0.05 cm. For the angular discretization we chose 16 ordinates. The results of the reconstruction for μ_s and μ_a are shown in Figs. 5a and b. The algorithm found the locations of all three water perturbations in both the μ_a reconstruction and the μ_s reconstruction. The location of the

smallest μ_s value in each reconstructed perturbation deviates by less than 0.15 cm from the center of the original location of the perturbation. The circular shape of the perturbations is slightly distorted in both reconstructed cross-sectional images. The perturbation in the upper-left corner, which is closest to the boundary, is reconstructed more accurately than the other two perturbations. We find that areas closer to the boundary, where the sources and detectors are located, are very sensitive to changes in the optical parameters whereas changes in the center of the phantom contribute less to the objective function. Therefore, perturbations close to the boundary are reconstructed with a higher accuracy than perturbations closer to the center of the phantom. Finally we observe that the smallest reconstructed scattering coefficient in the image is $\mu_s = 25\ \mathrm{cm}^{-1}$ and the smallest absorption coefficient is $\mu_a = 0.53\ \mathrm{cm}^{-1}$. Compared to the initial guesses of $\mu_s = 50\ \mathrm{cm}^{-1}$ and $\mu_a = 0.62\ \mathrm{cm}^{-1}$ this equals relative changes of 50.0% and 14.5% for μ_s and μ_a, respectively.

In both cases, void-like ring and three void-like cylinders, we find that while the position of the voids are accurately determined, the absolute values of the optical properties inside the voids are overestimated. This is most likely caused by the fact that the initial guess is very far away from the true value in these areas. It is well known that OT is an ill-posed problem and that different distribution of optical properties can yield similarly low values of the objective function [41–43]. For example, in the case of the void-like ring, smaller changes in optical properties in a broader area yield almost the same objective function as a stronger change in optical properties in a smaller area. This problem possibly could be overcome by using time-resolved or frequency domain measurements, which contain more information. Schweiger and Arridge recently showed that different data types may lead to a better defined minimum of the objective function [44]. In addition, better reconstruction results may be obtained when regularization schemes [42,43] are used, which we have not considered in this study.

Furthermore, we find that the scattering images show larger artifacts than the absorption images. For example in the case of the void-like ring (Fig. 4a), we can see streaks of low scattering areas that seem to originate from the sources. The reasons for this behavior are currently unclear, but may be explained by the particular use of the upwind-difference discrete-ordinates method and highly anisotropic sources. Other groups [45,46] that employ dicrete-ordinates methods have observed similar "ray effects", especially in void-like areas that are close to a source.

5. Summary

We have developed an image reconstruction algorithm for optical tomography that is based on the equation of radiative transfer. This algorithm consists of three major parts. First, an upwind-difference discrete-ordinates formulation of the time-independent equation of radiative transfer is used to predict detector readings on the surface of a medium, given the location of the sources and an estimate of the spatial distribution of the optical parameters μ_s, μ_a, and g inside the medium. Second, the predicted detector readings are compared to actual measurements by defining a χ^2-error norm. Third, an update of the estimated spatial distribution of the optical properties is obtained by calculating the gradient of the objective function with respect to all optical properties. This gradient calculation is performed by employing an adjoint differentiation

scheme. Unlike other adjoint schemes currently employed in optical tomography our scheme does not require the formulation of an adjoint equation of the forward model, either in its continuous or its discretized form. Using this algorithm, we presented reconstructions based on experimental data obtained from media that contain void-like regions. It has been shown previously that diffusion-theory based algorithms cannot be used in these cases. Our algorithm converges in the presence of voids and provides information about their locations. While there are still some artifact issues to overcome, the results obtained are very promising. To our knowledge the presented results are the first experimentally obtained reconstructions of media that contain void-like areas.

Acknowledgements

The authors would like to thank Uwe Netz and Professor Jürgen Buethan, Freie Universitat Berlin, for providing experimental data from tissue phantoms. Furthermore, we would like to thank Avraham Bluestone and Dr. Harry Graber, State University of New York—Downstate Medical Center (SUNY DMC), for their helpful comments concerning this manuscript. This work was supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, a part of the National Institutes of Health (grant # R01 AR46255-01), the Whitaker Foundation (grant # 98-0244), the City of New York Council Speaker's Fund for Biomedical Research: Towards the Science of Patient Care, and the Dean's Office of the College of Medicine at the SUNY DMC.

Appendix A. Derivation of gradient $\nabla_\mu \Phi$

The total derivative $\nabla_\mu \Phi$ can be obtained by systematically applying the chain rule of differentiation to Eq. (4). The total derivative $\nabla_\mu \Phi$ is a composition of derivatives of all intermediate steps of the forward model due to the explicit dependence of the intermediate results on the optical parameters. Starting from the last step Z of the forward code, in which the objective function Φ is calculated given the predicted detector readings $P = \Psi^Z$, the total derivative is given by the chain rule as

$$\left(\frac{\mathrm{d}\Phi}{\mathrm{d}\mu}\right)^{\mathrm{T}} = \left(\frac{\partial \Phi}{\partial \Psi^Z}\right)^{\mathrm{T}} \frac{\mathrm{d}\Psi^Z}{\mathrm{d}\mu} + \left(\frac{\partial \Phi}{\partial \mu}\right)^{\mathrm{T}}. \tag{A.1}$$

The derivative $\mathrm{d}\Psi^z/\mathrm{d}\mu$ is obtained by again applying the chain rule of differentiation:

$$\frac{\mathrm{d}\Psi^Z}{\mathrm{d}\mu} = \frac{\partial \Psi^Z}{\partial \Psi^{Z-1}} \frac{\mathrm{d}\Psi^{Z-1}}{\mathrm{d}\mu} + \frac{\partial \Psi^Z}{\partial \mu}. \tag{A.2}$$

Eqs. (A.1) and (A.2) yield together

$$\left(\frac{\mathrm{d}\Phi}{\mathrm{d}\mu}\right)^{\mathrm{T}} = \left(\frac{\partial \Phi}{\partial \Psi^Z}\right)^{\mathrm{T}} \frac{\partial \Psi^Z}{\partial \Psi^{Z-1}} \frac{\mathrm{d}\Psi^{Z-1}}{\mathrm{d}\mu} + \left(\frac{\partial \Phi}{\partial \Psi^Z}\right)^{\mathrm{T}} \frac{\partial \Psi^Z}{\partial \mu} + \left(\frac{\partial \Phi}{\partial \mu}\right)^{\mathrm{T}}. \tag{A.3}$$

In the next step the total derivative $\mathrm{d}\Psi^{Z-1}/\mathrm{d}\mu$ is replaced again. This process is repeated for all total derivatives down to the first step $\mathrm{d}\Psi^1/\mathrm{d}\mu = \partial\Psi^1/\partial\mu$, and we obtain

$$\left(\frac{\mathrm{d}\Phi}{\mathrm{d}\mu}\right)^{\mathrm{T}} = \left\{\left(\frac{\partial\Phi}{\partial\Psi^Z}\right)^{\mathrm{T}} \frac{\partial\Psi^Z}{\partial\Psi^{Z-1}} \cdots \frac{\partial\Psi^2}{\partial\Psi^1}\frac{\partial\Psi^1}{\partial\mu}\right\} + \left\{\left(\frac{\partial\Phi}{\partial\Psi^Z}\right)^{\mathrm{T}} \frac{\partial\Psi^Z}{\partial\Psi^{Z-1}} \cdots \frac{\partial\Psi^3}{\partial\Psi^2}\frac{\partial\Psi^2}{\partial\mu}\right\} + \cdots + \left\{\left(\frac{\partial\Phi}{\partial\Psi^Z}\right)^{\mathrm{T}} \frac{\partial\Psi^Z}{\partial\mu}\right\} + \left(\frac{\partial\Phi}{\partial\mu}\right). \tag{A.4}$$

Using the short hand notation

$$\left\{\left(\frac{\partial\Phi}{\partial\Psi^Z}\right)^{\mathrm{T}} \frac{\partial\Psi^Z}{\partial\Psi^{Z-1}} \cdots \frac{\partial\Psi^{z+1}}{\partial\Psi^z}\right\} = \frac{\partial(\Phi \circ \Psi^Z \circ \cdots \circ \Psi^{z+1})^{\mathrm{T}}}{\partial\Psi^z} = \left(\frac{\partial\Phi}{\partial\Psi^z}\right)^{\mathrm{T}}, \tag{A.5}$$

we rewrite Eq. (A.4):

$$\left(\frac{\mathrm{d}\Phi}{\mathrm{d}\mu}\right)^{\mathrm{T}} = \left(\frac{\partial\Phi}{\partial\Psi^1}\right)^{\mathrm{T}} \frac{\partial\Psi^1}{\partial\mu} + \left(\frac{\partial\Phi}{\partial\Psi^2}\right)^{\mathrm{T}} \frac{\partial\Psi^2}{\partial\mu} + \cdots + \left(\frac{\partial\Phi}{\partial\Psi^z}\right)^{\mathrm{T}} \frac{\partial\Psi^z}{\partial\mu} + \cdots + \left(\frac{\partial\Phi}{\partial\Psi^Z}\right)^{\mathrm{T}} \frac{\partial\Psi^Z}{\partial\mu} + \left(\frac{\partial\Phi}{\partial\mu}\right)^{\mathrm{T}} \tag{A.6}$$

or

$$\nabla_\mu\Phi := \left(\frac{\partial\Phi}{\partial\mu}\right)^{\mathrm{T}} = \left(\sum_z \left(\frac{\partial\Phi}{\partial\Psi^z}\right)^{\mathrm{T}} \frac{\partial\Psi^z}{\partial\mu}\right) + \left(\frac{\partial\Phi}{\partial\mu}\right)^{\mathrm{T}} \tag{A.7}$$

which equals Eq. (11) in the main text.

References

[1] Arridge SR. Optical tomography in medical imaging. Inverse Problems 1999;15:R41–93.
[2] Hielscher AH, Alcouffe RE, Barbour RL. Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues. Phys Med Biol 1998;43:1285–302.
[3] Firbank M, Arridge SR, Schweiger M, Delpy DT. An investigation of light transport through scattering bodies with non-scattering regions. Phys Med Biol 1996;41:767–83.
[4] Ripoll J, Nieto-Vesperinas M, Arridge SR, Dehghani H. Boundary conditions for light propagation in diffusive media with nonscattering regions. J Opt Soc Am A 2000;17(9):1671–82.
[5] Dehghani H, Arridge SR, Schweiger M, Delpy DT. Optical tomography in the presence of void regions. J Opt Soc Am A 2000;17(9):1659–71.
[6] Arridge SR, Dehghani H, Schweiger M, Okada E. The finite element model for the propagation of light in scattering media: a direct method for domains with nonscattering regions. Med Phys 2000;27(1):252–65.
[7] Okada E, Firbank M, Schweiger M, Arridge SR, Cope M, Delpy DT. Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head. Appl Opt 1997;36(1):21–32.
[8] Dehghani H, Delpy DT, Arridge SR. Photon migration in non-scattering tissue and the effects on image reconstruction. Phys Med Biol 1999;44(12):2897–906.
[9] Schweiger M, Arridge SR. Optical tomographic reconstruction in a complex head model using a priori region boundary information. Phys Med Biol 1999;44(11):2703.
[10] Prapavat V, Runge W, Mans J, Krause A, Beuthan J, Müller G. The development of a finger joint phantom for the optical simulation of early inflammatory rheumatic changes. Biomed Tech 1997;42:319–26.

[11] Klose A, Prapavat V, Minet O, Beuthan J, Müller G. RA diagnostics applying optical tomography in frequency-domain. Proc SPIE 1997;3196:194–204.
[12] Klose A, Hielscher AH, Hanson KM, Beuthan J. Three-dimensional optical tomography of a finger joint model for diagnostic of rheumatoid arthritis. Proc SPIE 1998;3566:151–60.
[13] Arridge SR, Schweiger M. A gradient-based optimisation scheme for optical tomography. Opt Express 1998;2(6):213–26.
[14] Davies AJ, Christianson DB, Dixon LCW, Roy R, van der Zee P. Reverse differentiation and the inverse diffusion problem. Adv Eng Software 1997;28:217–21.
[15] Hielscher AH, Klose AD, Hanson KM. Gradient-based iterative image reconstruction scheme for time-resolved optical tomography. IEEE Trans Med Imag 1999;18(3):262–71.
[16] Roy R, Sevick-Muraca EM. Truncated Newton's optimization scheme for absorption and fluorescence optical tomography. Part I: theory and formulation. Opt Express 1999;4(10):353–71.
[17] Klose AD, Hielscher AH. Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer. Med Phys 1999;26(8):1698–707.
[18] Luenberger DG. Linear and nonlinear programming. Boston: Addison-Wesley, 1984.
[19] Giering R. Tangent linear and adjoint biogeochemical models. Inverse methods in global biogeochemical cycles. Geophys Monogr 2000;114:33–48.
[20] Errico ME. What is an adjoint model? Bull Am Meteor Soc 1997;78(11):2577–91.
[21] Kaminski T, Heimann M. A coarse grid three-dimensional global inverse model of atmospheric transport. 1. Adjoint model and Jacobian matrix. J Geophys Res 1999;104(D15):18535–53.
[22] Marchuk GI. Adjoint equations and analysis of complex systems. Dordrecht: Kluwer Academic Publishers, 1995.
[23] Marchuk GI, Agoshkov VI, Shutyaev VP. Adjoint equations and perturbation algorithms in nonlinear problems. New York: CRC Press, 1996.
[24] Cacuci DG. Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach. J Math Phys 1981;22(12):2794–802.
[25] Cacuci DG. Sensitivity theory for nonlinear systems. II. Extension to additional classes of responses. J Math Phys 1981;22(12):2803–12.
[26] Talagrand O. Adjoint models. Numerical methods in atmospheric models. Reading, European Centre for Medium Range Weather Forecasts, 1991. p. 73–92.
[27] Norton SJ. Iterative inverse scattering algorithms: methods of computing Frechet derivatives. J Acoust Soc Am 1999;106(5):2653–60.
[28] Dorn O. A transport-backtransport method for optical tomography. Inverse Problems 1998;14:1107–30.
[29] Griewank A. On automatic differentiation. In: Iri M, Tanabe K, editors. Mathematical programming: recent developments and applications. Dordrecht: Kluwer Academic Publishers, 1989. p. 83–108.
[30] Shah P. Application of adjoint equations to estimation of parameters in distributed dynamic systems. In: Griewank A, Corliss GF, editors. Automatic differentiation of algorithms: theory, implementation, and application. Philadelphia: SIAM, 1991. p. 169–80.
[31] Talagrand O, Courtier P. Variational assimilation of meterological observations with the adjoint vorticity equation. I: theory. Q J R Meteorol Soc 1987;113:1311–28.
[32] Courtier P, Talagrand O. Variational assimilation of meteorological observations with the adjoint equation, II, numerical results. Q J R Meteorol Soc 1987;113:1329–47.
[33] Thacker WC, Long RB. Fitting dynamics to data. J Geophys Res 1988;93(C2):1227–40.
[34] Rall LB, Corliss GF. An introduction to automatic differentiation. In: Griewank A, Corliss GF, editors. Automatic differentiation of algorithms: theory, implementation, and application. Philadelphia: SIAM, 1991. p. 1–18.
[35] Wengert RE. A simple automatic derivative evaluation program. Commun ACM 1964;7(8):463–4.
[36] Griewank A. Evaluating derivatives. Principles and techniques of algorithmic differentiation. Philadelphia: SIAM, 2000.
[37] Talagrand O. The use of adjoint equations in numerical modeling of the atmospheric circulation. In: Griewank A, Corliss GF, editors. Automatic differentiation of algorithms: theory, implementation, and application. Philadelphia: SIAM, 1991. p. 169–80.

[38] Thacker WC. Automatic differentiation from an oceanographer's perspective. In: Griewank A, Corliss GF, editors. Automatic differentiation of algorithms: theory, implementations, and applications. Philadelphia: SIAM, 1991.
[39] Nash SG, Sofer A. Linear and nonlinear programming. New York: McGraw-Hill, 1996.
[40] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C. New York: Cambridge University Press, 1992.
[41] Hielscher AH, Klose AD, Beuthan J. Evolution strategies for optical tomographic characterization of homogeneous media. Opt Express 2000;7(13):507–18.
[42] Hielscher AH, Bartel S. Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography. J Biomed Opt 2001;6(2):183–92.
[43] Pogue BW, McBride TO, Prewitt J, Osterberg UL, Paulsen KD. Spatially variant regularization improves diffuse optical tomography. Appl Opt 1999;38:2950–62.
[44] Schweiger M, Arridge SR. Application of temporal filters to time-resolved data in optical tomography. Phys Med Biol 1999;44:1699–717.
[45] Lathrop KD. Ray effects in discrete ordinates equations. Nucl Sci Eng 1968;32:357–69.
[46] Ackroyd RT. Treatment of voids in finite element transport methods. Prog Nucl Energy 1996;18:85–9.