

FREQUENCY DOMAIN OPTICAL TOMOGRAPHY BASED ON THE EQUATION OF RADIATIVE TRANSFER*

KUI REN[†], GUILLAUME BAL[†], AND ANDREAS H. HIELSCHER[‡]

Abstract. Optical tomography consists of reconstructing the spatial distribution of absorption and scattering properties of a medium from surface measurements of transmitted light intensities. Mathematically, this problem amounts to parameter identification for the equation of radiative transfer (ERT) with diffusion-type boundary measurements. Because they are posed in the phase-space, radiative transfer equations are quite challenging to solve computationally. Most past works have considered the steady-state ERT or the diffusion approximation of the ERT. In both cases, substantial cross-talk has been observed in the reconstruction of the absorption and scattering properties of inclusions. In this paper, we present an optical tomographic reconstruction algorithm based on the frequency-domain ERT. The inverse problem is formulated as a regularized least-squares minimization problem, in which the mismatch between forward model predictions and measurements is minimized. The ERT is discretized by using a discrete ordinates method for the directional variables and a finite volume method for the spatial variables. A limited-memory quasi-Newton algorithm is used to minimize the least-squares functional. Numerical simulations with synthetic data show that the cross-talk between the two optical parameters is significantly reduced in reconstructions based on frequency-domain data as compared to those based on steady-state data.

Key words. optical tomography, photon density waves, equation of radiative transfer, finite volume method, discrete ordinates method, generalized minimal residual algorithm, inverse problems, regularization, numerical optimization

AMS subject classifications. 65N21, 65Z05, 78M25, 78M50

DOI. 10.1137/040619193

1. Introduction. The inverse problem related to the equation of radiative transfer (ERT) has long been of interest in many fields of applied physics and mathematics. In this problem, one attempts to determine the spatial distribution of parameters (e.g., the spatial distribution of absorption and scattering coefficients) inside a medium from either transmission or reflectance data collected by illuminating the medium from many different directions. Applications include, for instance, astronomy [61], nuclear science [42, 43], and atmospheric science [12]; see the review [50]. Recent advances in this field have been fueled by an increased interest in medical applications. By performing optical measurements on the surface of the skin, optical tomographic techniques attempt to reconstruct the spatial distribution of absorption and scattering coefficients inside the body. These parameters can then be used for many diagnostic purposes [13].

Mathematically, optical tomography reduces to parameter identification problems (inverse problems) for the ERT, also referred to as the linear Boltzmann equation or the transport equation. Most early studies in the field deal with plane-parallel ge-

*Received by the editors November 18, 2004; accepted for publication (in revised form) March 15, 2006; published electronically September 15, 2006. The work of the first and third authors was supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant 2R01 AR046255) and by the National Institute of Biomedical Imaging and Bioengineering (grant R01 EB001900), which are both part of the National Institutes of Health. The work of the second author was supported by NSF grant DMS-0239097 and an Alfred P. Sloan Fellowship.

<http://www.siam.org/journals/sisc/28-4/61919.html>

[†]Department of Applied Physics & Applied Mathematics, Columbia University, New York, NY 10027 (kr2002@columbia.edu, gb2030@columbia.edu).

[‡]Department of Biomedical Engineering and Radiology, Columbia University, New York, NY 10027 (ahh2004@columbia.edu).

ometry [20, 24]. Inversion procedures were also developed in many more general yet still constrained geometries; see, for instance, [3, 42, 43, 61] and the review papers [49, 50]. Because of their geometrical specificities, most of these works do not directly apply to biomedical applications. In the mid-1990s several groups started developing schemes and algorithms that overcame these problems in the framework of the diffusion approximation to the ERT; see, for instance, [4, 34, 57, 63]. Many theoretical and experimental studies have shown that the diffusion approximation was valid only in specific contexts—for instance, when the scattering coefficient is sufficiently large and the absorption coefficient sufficiently small [16, 17, 23, 31]. While this is often the case in biomedical applications there remain many important applications, for which reconstructions based on the ERT are needed. Some examples include brain imaging, where the presence of low-absorbing and low-scattering cerebrospinal fluid leads to nondiffusive light propagation; imaging of joints, where clear synovial fluid has the same effects as cerebrospinal fluid in the brain; and small animal imaging, where small object dimensions (1–2 cm in diameter) invalidate the classical diffusion approximation [30, 32]; see, however, [6, 8] for ways to generalize the diffusion equation in certain situations.

Only quite recently have studies emerged that describe ERT-based reconstruction codes for use in biomedical optical tomography. First, a transport-backtransport method, a nonlinear inversion method, applied to the two-dimensional time-dependent equation of radiative transfer was reported in [18]. New algorithms were developed and experimentally tested for two- and three-dimensional cases using a time-independent ERT in [38, 39, 40, 41]. While these works, which address real-life three-dimensional problems, are an important step towards practical applications, they still suffer from considerable cross-talk between absorption and scattering reconstructions. What we mean by cross-talk is that purely scattering (or purely absorbing) inclusions are often reconstructed with unphysical absorption (or scattering) properties. This behavior is well-understood from the theoretical viewpoint: Different optical distributions inside the medium can lead to the same measurements collected at the surface of the medium [5, 33]. To avoid such cross-talk, which may lead to wrong diagnostics, we need different data. An experimental technique increasingly employed in recent years to obtain additional information is to use frequency domain measurements. In this case the source intensity is modulated (typically between 100–1000 MHz), leading to the propagation of so-called photon density waves. Since frequency domain measurements provide information about both the phase and the intensity of the waves (and not only the intensity as in steady-state measurements), it is expected that the frequency-domain techniques will allow for better separation of absorption and scattering effects [5, 48]. Numerical reconstructions based on the frequency-domain ERT, however, have not yet been developed in the literature. This is the major motivation for the present work.

We will propose in the following sections a numerical procedure that solves the frequency-domain optical tomography problem with angularly averaged measurement. The numerical procedure involves multiple techniques such as the finite volume method for transport discretization, limited memory BFGS for minimization, the adjoint state method for computing Fréchet derivatives, the matrix-free approach for evaluating matrix vector products, and the L-curve method for selecting regularization parameters. The combination of those techniques allows us to perform numerical reconstructions in geometrical settings that are of practical importance.

The remainder of the paper is organized as follows. In section 2 we formulate the inverse problem in frequency-domain optical tomography as a classical regularized

least-squares problem and describe the adjoint method for the computation of the gradient of the least-squares functional. In section 3 we present the details on the discretization of the forward and backward problems, which is based on using a discrete ordinates methods for the directional variables and a finite volume method for the spatial variables. The implementation of the minimization procedure, based on a quasi-Newton optimization algorithm, and the choice of the regularization parameter are described in section 4. Several numerical simulations illustrate the performance of the code in section 5. Reconstructions based on synthetic data indeed show a significant reduction of the cross-talk mentioned above when frequency-domain data are used. Concluding remarks are finally offered in section 6.

2. Problem formulation. We now formulate the optical tomography problem. Let $\mathcal{D} \subset \mathbb{R}^n$ be our domain of interest, with sufficiently regular boundary $\partial\mathcal{D}$. Then the frequency-domain equation of radiative transfer that describes the photon density in the phase space, i.e., as a function of position $\mathbf{x} \in \mathcal{D}$ and direction $\boldsymbol{\theta} \in S^{n-1}$ (unit sphere of \mathbb{R}^n) is given by [4]

$$(2.1) \quad \begin{aligned} \mathcal{T}u \equiv \left(\frac{i\omega}{v} + \boldsymbol{\theta} \cdot \nabla + \sigma_a(\mathbf{x}) \right) u(\mathbf{x}, \boldsymbol{\theta}) + Q(u)(\mathbf{x}, \boldsymbol{\theta}) &= 0 && \text{in } \mathcal{D} \times S^{n-1}, \\ u(\mathbf{x}, \boldsymbol{\theta}) &= f(\mathbf{x}, \boldsymbol{\theta}) && \text{on } \Gamma_-, \end{aligned}$$

where $i = \sqrt{-1}$, $n = 2, 3$ is the space dimension, $v \in \mathbb{R}^+$ is the speed of light in the medium, and ω is the modulation frequency of the boundary source $f(\mathbf{x}, \boldsymbol{\theta})$. The non-negative function $\sigma_a(\mathbf{x}) \in L^\infty(\mathcal{D})$ is the absorption coefficient. The unknown quantity, $u(\mathbf{x}, \boldsymbol{\theta})$, is the radiant power per unit solid angle per unit area perpendicular to the direction of propagation at \mathbf{x} in the direction $\boldsymbol{\theta}$. Note that $u(\mathbf{x}, \boldsymbol{\theta})$ depends also on ω although, for simplicity, we do not write this dependency explicitly. The boundary sets Γ_\pm are defined as

$$\Gamma_\pm = \{(\mathbf{x}, \boldsymbol{\theta}) \in \partial\mathcal{D} \times S^{n-1} \text{ s.t. } \pm \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) > 0\},$$

with $\boldsymbol{\nu}(\mathbf{x})$ the outward unit normal to the domain at $\mathbf{x} \in \partial\mathcal{D}$. The scattering operator Q is defined as

$$(2.2) \quad Q(u)(\mathbf{x}, \boldsymbol{\theta}) = \sigma_s(\mathbf{x}) \left(u(\mathbf{x}, \boldsymbol{\theta}) - \int_{S^{n-1}} k(\boldsymbol{\theta} \cdot \boldsymbol{\theta}') u(\mathbf{x}, \boldsymbol{\theta}') d\mu(\boldsymbol{\theta}') \right).$$

Here, $\sigma_s(\mathbf{x}) \in L^\infty(\mathcal{D})$ is the scattering coefficient and $d\mu$ is the surface measure on S^{n-1} normalized so that $\int_{S^{n-1}} d\mu(\boldsymbol{\theta}) = 1$. The ‘‘collision’’ kernel $k(\boldsymbol{\theta} \cdot \boldsymbol{\theta}')$, which describes the probability that photons traveling in direction $\boldsymbol{\theta}'$ scatter into direction $\boldsymbol{\theta}$, is a positive function independent of \mathbf{x} and satisfies the normalization condition:

$$(2.3) \quad \int_{S^{n-1}} k(\boldsymbol{\theta} \cdot \boldsymbol{\theta}') d\mu(\boldsymbol{\theta}') = 1.$$

The scattering kernel for light propagation in tissues is highly peaked forward and is chosen as the Henyey–Greenstein phase function [29, 62]

$$(2.4) \quad k(\boldsymbol{\theta} \cdot \boldsymbol{\theta}') = C \frac{1 - g^2}{(1 + g^2 - 2g \cos \phi)^{3/2}},$$

where ϕ is the angle between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, i.e., $\boldsymbol{\theta} \cdot \boldsymbol{\theta}' = \cos \phi$, and where $g \in [0, 1]$ is the anisotropy factor, which measures how peaked forward the phase function is.

The larger g is, the more forward the scattering. The anisotropy factor is often used to define the so-called effective scattering coefficient through $\sigma'_s = (1 - g)\sigma_s$. C is a normalization constant such that (2.3) hold. We mention that scattering kernels other than (2.4) have also been used in some situations [36] and that simplified (Fokker–Planck) models can also be used to analyze highly peaked scattering in biological tissues [37].

The optical tomography problem thus consists of reconstructing $\sigma_a(\mathbf{x})$ and $\sigma_s(\mathbf{x})$ in (2.1) from boundary current measurements; see (2.6) below. Our objective in this work is to present a numerical scheme that performs the reconstruction.

2.1. Forward problem. The absorption and scattering coefficients σ_a and σ_s cannot take negative values and have to be bounded. We thus introduce the following parameter space \mathcal{Q} :

$$\mathcal{Q} := \{(\sigma_a, \sigma_s) : \sigma_a \geq 0, \sigma_s \geq 0, \text{ and } (\sigma_a, \sigma_s) \in L^\infty(\mathcal{D}) \times L^\infty(\mathcal{D})\}.$$

We also introduce the functional spaces [2, 16]:

$$L^2_{\boldsymbol{\theta}, \nu}(\Gamma_\pm) := \left\{ u(\mathbf{x}, \boldsymbol{\theta}) : \int_{\Gamma_\pm} |u(\mathbf{x}, \boldsymbol{\theta})|^2 |\boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x})| d\sigma(\mathbf{x}) d\mu(\boldsymbol{\theta}) < +\infty \right\},$$

$$\mathcal{W}^2(\mathcal{D} \times S^{n-1}) := \{u(\mathbf{x}, \boldsymbol{\theta}) : u \in L^2(\mathcal{D} \times S^{n-1}) \text{ and } \boldsymbol{\theta} \cdot \nabla u \in L^2(\mathcal{D} \times S^{n-1})\}.$$

Adapting well-known results [2, 16] with complex-valued absorption coefficient $\sigma_a + \frac{i\omega}{\nu}$ in $L^\infty(\mathcal{D})$, we have the following statement about the forward problem.

PROPOSITION 2.1. *Assume that $(\sigma_a, \sigma_s) \in \mathcal{Q}$, the modulation frequency is finite $\omega < +\infty$, and $f \in L^2_{\boldsymbol{\theta}, \nu}(\Gamma_-)$. Then the forward problem (2.1) is well-posed and admits a unique solution $u(\mathbf{x}, \boldsymbol{\theta}) \in \mathcal{W}^2(\mathcal{D} \times S^{n-1})$.*

We can then define the following *albedo* operator (as well as its adjoint) [14, 51]:

$$(2.5) \quad \Lambda : \begin{matrix} f \longmapsto u|_{\Gamma_+} \\ L^2_{\boldsymbol{\theta}, \nu}(\Gamma_-) \longmapsto L^2_{\boldsymbol{\theta}, \nu}(\Gamma_+) \end{matrix}.$$

The albedo operator Λ maps the incoming flux on the boundary into the outgoing flux and is a functional of the optical parameters σ_a and σ_s .

A major difficulty in optical tomography comes from the fact that in practice only outgoing currents, which are angular averages of the outgoing flux and are similar to diffusion-type measurements, are available. This prevents us from using classical uniqueness and stability results in inverse transport theory [14]. In fact, the inverse problem we solve in this paper is very similar to the diffusion-based inverse problem [4], on which many more theoretical results exist. To date, we do not know of any theoretical result on the reconstruction of optical properties from outgoing currents for arbitrary geometries. This makes the development of numerical tools all the more important.

To be consistent with existing measurement technologies, we define the following “measurement operator”:

$$(2.6) \quad \mathcal{G}u|_{\Gamma_+} := \int_{S^{n-1}_+} \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) u|_{\Gamma_+} d\mu(\boldsymbol{\theta}) \equiv z(\mathbf{x}),$$

$$\mathcal{G} : L^2_{\boldsymbol{\theta}, \nu}(\Gamma_+) \longmapsto L^2(\partial\mathcal{D}) \equiv \mathcal{Z},$$

with $S_+^{n-1} := \{\boldsymbol{\theta} : \boldsymbol{\theta} \in S^{n-1} \text{ s.t. } \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) > 0\}$. We will call \mathcal{Z} the ‘‘measurement space.’’ Now the composite operator $\mathcal{G}\Lambda : f \mapsto z$ maps the incoming flux into the tomographic measurements. The adjoint operator \mathcal{G}^* of \mathcal{G} is defined via the identity

$$(2.7) \quad \langle \overline{\mathcal{G}^*g_1}, g_2 \rangle_{L^2_{\boldsymbol{\theta}, \boldsymbol{\nu}}(\Gamma_+)} = \langle \overline{g_1}, \mathcal{G}g_2 \rangle_{\mathcal{Z}}$$

for all $g_1 \in \mathcal{Z}$ and $g_2 \in L^2_{\boldsymbol{\theta}, \boldsymbol{\nu}}(\Gamma_+)$, where the symbol $\overline{\mathcal{Y}_1}$ denotes the complex conjugate of \mathcal{Y}_1 , and $\langle \cdot, \cdot \rangle_X$ is the usual inner product in a Hilbert space X . One observes that \mathcal{G}^* is nothing but the operation of multiplication by $\boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x})$.

2.2. Least-squares formulation. The inverse problem of optical tomography can be formulated as follows: determine $(\sigma_a, \sigma_s) \in \mathcal{Q}$ such that

$$(2.8) \quad \mathcal{G}\Lambda f = z$$

holds for all possible source-measurement pairs (f, z) . Here $z \in \mathcal{Z} \equiv L^2(\partial\mathcal{D})$ is the measured data corresponding to source f . This problem is in general severely ill-posed (assuming that uniqueness of reconstruction holds as in diffusion theory [4]) in the sense that when no regularization is applied, noise contained in the data z is more amplified during the inversion procedure than what would result from an arbitrary number of differentiations [21]. Another practical difficulty in solving (2.8) lies in the fact that the amount of available data may be quite limited [47]. For example, one may only be able to use a limited number (say, N_q) of light sources. After discretizing (2.8) on a reasonable mesh, we will end up with a very underdetermined nonlinear system. A classical way to resolve the lack of measurements is to turn to the following least-squares formulation: find (σ_a, σ_s) solving

$$(2.9) \quad \mathcal{F}(\sigma_a, \sigma_s) =: \frac{1}{2} \sum_{q=1}^{N_q} \|\mathcal{G}\Lambda f_q - z_q\|_{\mathcal{Z}}^2 \rightarrow \min.$$

Here, $1 \leq q \leq N_q$ denotes the light source number. For reasons we have mentioned earlier, the least-squares problem (2.9) is usually not stable [4]. To stabilize the problem, we impose additional smoothness restrictions on the coefficients we wish to reconstruct. In other words, we look for optical properties in a space that is much smaller than \mathcal{Q} . We call this space the space of *admissible parameters*:

$$\mathcal{Q}_{ad} := \{(\sigma_a, \sigma_s) : (\sigma_a, \sigma_s) \in [\sigma_a^l, \sigma_a^u] \times [\sigma_s^l, \sigma_s^u], \text{ and } (\sigma_a, \sigma_s) \in \mathcal{H}^1(\mathcal{D}) \times \mathcal{H}^1(\mathcal{D})\},$$

where σ_a^l (resp., σ_a^u) and σ_s^l (resp., σ_s^u) are lower (resp., upper) bounds of σ_a and σ_s , respectively, with $\sigma_a^l > 0$ and $\sigma_s^l > 0$. $\mathcal{H}^1(\mathcal{D})$ is the usual Hilbert space of $L^2(\mathcal{D})$ functions with first-order partial derivatives in $L^2(\mathcal{D})$:

$$(2.10) \quad \|\mathcal{Y}\|_{\mathcal{H}^1(\mathcal{D})}^2 := \|\mathcal{Y}\|_{L^2(\mathcal{D})}^2 + \|\nabla \mathcal{Y}\|_{L^2(\mathcal{D})}^2 \quad \text{for } \mathcal{Y} \in \mathcal{H}^1(\mathcal{D}).$$

It is known that \mathcal{Q}_{ad} is a closed and convex subset of $\mathcal{H}^1(\mathcal{D}) \times \mathcal{H}^1(\mathcal{D})$. We can thus introduce the following regularized least-squares functional:

$$(2.11) \quad \mathcal{F}_\beta(\sigma_a, \sigma_s) := \mathcal{F}(\sigma_a, \sigma_s) + \frac{\beta}{2} \mathcal{J}(\sigma_a, \sigma_s),$$

where the last term is a regularization term and β is the regularization parameter [21]. The method for choosing β will be described in section 4.3. We use the Tikhonov regularization functional in our problem:

$$(2.12) \quad \mathcal{J}(\sigma_a, \sigma_s) = \|\sigma_a - \sigma_a^0\|_{\mathcal{H}^1(\mathcal{D})}^2 + \epsilon \|\sigma_s - \sigma_s^0\|_{\mathcal{H}^1(\mathcal{D})}^2,$$

where σ_a^0 and σ_s^0 are initial guesses for the σ_a and σ_s profiles, and ϵ is a small constant. The choice of ϵ is addressed in section 4.3. We thus formulate the optical tomography problem as the following regularized least-squares problem:

$$(RLS) \quad \begin{aligned} & \min_{(\sigma_a, \sigma_s)} \mathcal{F}_\beta \\ & \sigma_a^l \leq \sigma_a \leq \sigma_a^u, \\ & \sigma_s^l \leq \sigma_s \leq \sigma_s^u. \end{aligned}$$

We first observe that problem (RLS) has at least one solution in the sense that the functional $\mathcal{F}_\beta(\sigma_a, \sigma_s)$ admits at least one minimizer. This existence result is classical and follows from the weak lower semicontinuity and coercivity of $\mathcal{F}_\beta(\sigma_a, \sigma_s)$ [46, 60]. However, we cannot show that $\mathcal{F}_\beta(\sigma_a, \sigma_s)$ is strictly convex and cannot conclude that the minimizer is unique [60].

Our implementation of the inverse problem of optical tomography is a gradient-based minimization approach. We thus need to compute the Fréchet derivative of the least-squares functional $\mathcal{F}_\beta(\sigma_a, \sigma_s)$. Direct estimates of the Fréchet derivatives being quite costly because the optical parameters are (at least at the continuous level) infinite-dimensional objects, we adopt the adjoint state (or costate) approach [60] to estimate the derivatives. We have the following result.

THEOREM 2.2 (Fréchet derivatives). *The functional $\mathcal{F}_\beta(\sigma_a, \sigma_s)$ is Fréchet differentiable with respect to σ_a and σ_s . The derivative at (σ_a, σ_s) in the direction (h_a, h_s) is given by*

$$(2.13) \quad \begin{pmatrix} \mathcal{F}'_\beta h_a \\ \mathcal{F}'_\beta h_s \end{pmatrix} = \begin{pmatrix} \operatorname{Re} \sum_{q=1}^{N_q} \left\langle \varphi_q, \left(\frac{\partial \mathcal{T}}{\partial \sigma_a} h_a \right) u_q \right\rangle_{L^2(\mathcal{D} \times S^{n-1})} + \beta \langle \sigma_a - \sigma_a^0, h_a \rangle_{\mathcal{H}^1(\mathcal{D})} \\ \operatorname{Re} \sum_{q=1}^{N_q} \left\langle \varphi_q, \left(\frac{\partial \mathcal{T}}{\partial \sigma_s} h_s \right) u_q \right\rangle_{L^2(\mathcal{D} \times S^{n-1})} + \beta \epsilon \langle \sigma_s - \sigma_s^0, h_s \rangle_{\mathcal{H}^1(\mathcal{D})} \end{pmatrix},$$

where \mathcal{T} is the transport operator defined in (2.1); u_q and φ_q are the solutions of the forward problem (2.1) with source f_q and its adjoint problem (2.16) (defined below), respectively. Re means taking the real part.

Proof. Let us denote by r_q the residual $\mathcal{G}\Lambda f_q - z_q = \mathcal{G}u_q|_{\Gamma_+} - z_q$. According to [18, 19], r_q is Fréchet differentiable with respect to both σ_a and σ_s . The L^2 -norm is Fréchet differentiable as shown in [46]. By the chain rule, $\|r_q\|_{\mathcal{Z}}^2$ is Fréchet differentiable. Since the summation is finite, we deduce that \mathcal{F} is differentiable. Together with the fact that \mathcal{J} is differentiable, we conclude that \mathcal{F}_β is Fréchet differentiable with respect to σ_a and σ_s .

We now compute these Fréchet derivatives. Let us compute the derivative with respect to σ_a :

$$(2.14) \quad \begin{aligned} \mathcal{F}'_\beta(\sigma_a, \sigma_s)h_a &= \operatorname{Re} \sum_{q=1}^{N_q} \left\langle \bar{r}_q, \mathcal{G} \left(\frac{\partial u_q|_{\Gamma_+}}{\partial \sigma_a} h_a \right) \right\rangle_{\mathcal{Z}} + \beta \langle \sigma_a - \sigma_a^0, h_a \rangle_{\mathcal{H}^1(\mathcal{D})} \\ &= \operatorname{Re} \sum_{q=1}^{N_q} \left\langle \overline{\mathcal{G}^* r_q}, \frac{\partial u_q|_{\Gamma_+}}{\partial \sigma_a} h_a \right\rangle_{L^2_{\theta, \nu}(\Gamma_+)} + \beta \langle \sigma_a - \sigma_a^0, h_a \rangle_{\mathcal{H}^1(\mathcal{D})}, \end{aligned}$$

where we have used the properties of the adjoint operator (2.7). On the other hand, differentiating the transport equation (2.1) for source f_q gives

$$(2.15) \quad \begin{aligned} \mathcal{T}\phi_q + \left(\frac{\partial \mathcal{T}}{\partial \sigma_a} h_a\right) u_q &= 0 && \text{in } \mathcal{D} \times S^{n-1}, \\ \phi_q &= 0 && \text{on } \Gamma_-, \end{aligned}$$

where $\phi_q \equiv \frac{\partial u_q}{\partial \sigma_a} h_a$, and \mathcal{T} is the transport operator defined in (2.1). We need also to introduce an adjoint variable φ_q of u_q which is the solution of the following adjoint transport equation:

$$(2.16) \quad \begin{aligned} \mathcal{T}^* \varphi_q \equiv \left(\frac{i\omega}{v} - \boldsymbol{\theta} \cdot \nabla + \sigma_a(\mathbf{x})\right) \varphi_q(\mathbf{x}, \boldsymbol{\theta}) + Q(\varphi_q)(\mathbf{x}, \boldsymbol{\theta}) &= 0 && \text{in } \mathcal{D} \times S^{n-1}, \\ \varphi_q(\mathbf{x}, \boldsymbol{\theta}) &= -\overline{\mathcal{G}^* r_q} && \text{on } \Gamma_+. \end{aligned}$$

Here we have used that $Q^* = Q$, which follows from the definition (2.2). Multiplying (2.15) by φ_q and (2.16) by ϕ_q , then integrating over $\mathcal{D} \times S^{n-1}$, we obtain

$$(2.17) \quad \langle \overline{\mathcal{G}^* r_q}, \phi_q \rangle_{L^2_{\boldsymbol{\theta}, \nu}(\Gamma_+)} = \left\langle \varphi_q, \left(\frac{\partial \mathcal{T}}{\partial \sigma_a} h_a\right) u_q \right\rangle_{L^2(\mathcal{D} \times S^{n-1})},$$

which leads to

$$(2.18) \quad \mathcal{F}'_{\beta}(\sigma_a, \sigma_s) h_a = \text{Re} \sum_{q=1}^{N_q} \left\langle \varphi_q, \left(\frac{\partial \mathcal{T}}{\partial \sigma_a} h_a\right) u_q \right\rangle_{L^2(\mathcal{D} \times S^{n-1})} + \beta \langle \sigma_a - \sigma_a^0, h_a \rangle_{\mathcal{H}^1(\mathcal{D})}.$$

The derivative with respect to σ_s can be computed similarly. □

This result shows that in order to compute the Fréchet derivative of the objective functional $\mathcal{F}_{\beta}(\sigma_a, \sigma_s)$, we need to solve one forward transport problem (2.1) and one adjoint transport problem (2.16).

3. Discretization methods. There is a vast literature on the discretization of radiative transfer equations; see, for instance, [1, 25, 44]. In this paper, we have chosen to use the discrete ordinates method to discretize the directional variables and the finite volume method [22] to discretize the spatial variables.

3.1. The discrete ordinates formulation. In the discrete ordinates method [1, 44], we approximate the total scalar flux, defined as the integral of $u(\mathbf{x}, \boldsymbol{\theta})$ over S^{n-1} , by the following quadrature rule:

$$(3.1) \quad \int_{S^{n-1}} u(\mathbf{x}, \boldsymbol{\theta}) d\mu(\boldsymbol{\theta}) \approx \sum_{j=1}^J \eta_j u(\mathbf{x}, \boldsymbol{\theta}_j),$$

where $\boldsymbol{\theta}_j$ is the j th direction and η_j the associated weight for $1 \leq j \leq J$. Details on how to choose the set of directions $\{\boldsymbol{\theta}_j\}_{j=1}^J$ and the corresponding weights $\{\eta_j\}_{j=1}^J$ can be found in [44]. To ensure particle conservation, we impose that

$$(3.2) \quad \sum_{j=1}^J \eta_j = 1.$$

The equation of radiative transfer is now decomposed as a discrete set of J coupled differential equations that describe the photon flux field along J directions:

$$(3.3) \quad \nabla \cdot (\boldsymbol{\theta}_j u) + \left(\sigma_t + \frac{i\omega}{v} \right) u(\mathbf{x}, \boldsymbol{\theta}_j) = \sigma_s(\mathbf{x}) \sum_{j'=1}^J \eta_{j'} k_{jj'} u(\mathbf{x}, \boldsymbol{\theta}_{j'})$$

for $j = 1, 2, \dots, J$, where $k_{jj'} = k(\boldsymbol{\theta}_j \cdot \boldsymbol{\theta}_{j'})$, and where $\sigma_t = \sigma_a + \sigma_s$. We impose

$$(3.4) \quad \sum_{j=1}^J \eta_j k_{jj'} = 1, \quad 1 \leq j' \leq J,$$

so that the number of photons in the system is preserved by the scattering process.

3.2. Spatial discretization. We use a finite volume method to perform the spatial discretization. Finite volume methods [22] ensure the conservation of mass (or momentum, energy) in a discrete sense, which is important in transport calculations. They also have the advantage of easily handling complicated geometries by arbitrary triangulations, which we need in tomographic applications.

We implement a cell-centered version of the finite volume methods. Consider a mesh of \mathbb{R}^n , \mathcal{M} , consisting of polyhedral bounded convex subsets of \mathbb{R}^n , which covers our computational domain \mathcal{D} . Let $\mathcal{C} \in \mathcal{M}$ be a *control cell*, that is, an element of the mesh \mathcal{M} , $\partial\mathcal{C}$ its boundary, and $V_{\mathcal{C}}$ its Lebesgue measure. We assume that the unknown quantity, for example, $u(\mathbf{x}, \boldsymbol{\theta}_j)$, takes its averaged value in \mathcal{C} (and thus is constant). We denote this value by $u_j^{\mathcal{C}}$:

$$(3.5) \quad u_j^{\mathcal{C}} \equiv \frac{1}{V_{\mathcal{C}}} \int_{V_{\mathcal{C}}} u(\mathbf{x}, \boldsymbol{\theta}_j) d\mathbf{x}.$$

Integrating the above discrete ordinates equations (3.3) over cell \mathcal{C} and using the divergence theorem on the first term, we obtain the equations

$$(3.6) \quad \int_{\partial\mathcal{C}} \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}}(\mathbf{x}) u_j d\gamma(\mathbf{x}) + \left(\sigma_t^{\mathcal{C}} + \frac{i\omega}{v} \right) V_{\mathcal{C}} u_j^{\mathcal{C}} = V_{\mathcal{C}} \sigma_s^{\mathcal{C}} \sum_{j'=1}^J \eta_{j'} k_{jj'} u_{j'}^{\mathcal{C}}$$

for $1 \leq j \leq J$, where $\mathbf{n}_{\mathcal{C}}(\mathbf{x})$ denotes the outward normal to $\partial\mathcal{C}$ at point $\mathbf{x} \in \partial\mathcal{C}$, $d\gamma(\mathbf{x})$ denotes the surface Lebesgue measure on $\partial\mathcal{C}$, and $\sigma_s^{\mathcal{C}}$ ($\sigma_t^{\mathcal{C}}$) is the value of σ_s (σ_t) on cell \mathcal{C} .

Now we have to approximate the flux through the boundary of \mathcal{C} , i.e., the first integral term in (3.6). Let $\{\mathcal{C}_i\}_{i=1}^I$ be the set of neighboring cells of \mathcal{C} . We denote by $S_{\mathcal{C},i}$ the common edge of cell \mathcal{C} and \mathcal{C}_i , i.e., $S_{\mathcal{C},i} = \partial\mathcal{C} \cap \partial\mathcal{C}_i$. We then have

$$(3.7) \quad \int_{\partial\mathcal{C}} \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}}(\mathbf{x}) u_j d\gamma(\mathbf{x}) = \sum_i \int_{S_{\mathcal{C},i}} \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}}(\mathbf{x}) u_j d\gamma(\mathbf{x}).$$

The flux $\int_{S_{\mathcal{C},i}} \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}}(\mathbf{x}) u_j d\gamma(\mathbf{x})$ can be approximated by various numerical schemes. In this work, we take a first-order upwind scheme:

$$(3.8) \quad F_{j,i}^{\mathcal{C}} := \int_{S_{\mathcal{C},i}} \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}}(\mathbf{x}) u_j d\gamma(\mathbf{x}) = \begin{cases} \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}}|_{S_{\mathcal{C},i}} u_j^{\mathcal{C}} & \text{if } \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}} \geq 0, \\ \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}}|_{S_{\mathcal{C},i}} u_j^{\mathcal{C}_i} & \text{if } \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}} < 0, \end{cases}$$

where $|S_{\mathcal{C},i}|$ is the Lebesgue measure of $S_{\mathcal{C},i}$. We then obtain a full discretization of the discrete ordinates equations

$$(3.9) \quad \sum_i F_{j,i}^{\mathcal{C}} + \left(\sigma_t^{\mathcal{C}} + \frac{i\omega}{v} \right) V_{\mathcal{C}} u_j^{\mathcal{C}} = V_{\mathcal{C}} \sigma_s^{\mathcal{C}} \sum_{j'=1}^J \eta_{j'} k_{jj'} u_{j'}^{\mathcal{C}}$$

for $j = 1, 2, \dots, J$. Let N denote the total number of control cells. After collecting the discretized transport equation (3.9) on all control cells, we arrive at the following system of complex-valued algebraic equations:

$$(3.10) \quad \mathbf{A}\mathbf{U} = \mathbf{S}\mathbf{U} + \mathbf{G},$$

where $\mathbf{A} \in \mathbb{C}^{NJ \times NJ}$ and $\mathbf{S} \in \mathbb{C}^{NJ \times NJ}$ are the discretized streaming-collision and scattering operators, respectively. The boundary source $f(\mathbf{x}, \boldsymbol{\theta})$, which comes into the discretized system via the flux approximation (3.8), is denoted by \mathbf{G} . The vector $\mathbf{U} \in \mathbb{C}^{NJ \times 1}$, which contains the values of $u(\mathbf{x}, \boldsymbol{\theta})$ on the cell \mathcal{C} in the direction $\boldsymbol{\theta}_j$, is organized as

$$(3.11) \quad \mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_J \end{pmatrix}, \text{ with } \mathbf{U}_j = \begin{pmatrix} u_j^1 \\ \vdots \\ u_j^N \end{pmatrix} \in \mathbb{C}^N.$$

The matrices \mathbf{A} and \mathbf{S} have sparse structures. In fact, they are sparse block matrices. \mathbf{A} is a block diagonal matrix that can be written as

$$(3.12) \quad \mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_J \end{pmatrix} + \begin{pmatrix} \mathbf{C}_0 & & \\ & \ddots & \\ & & \mathbf{C}_0 \end{pmatrix},$$

where $\mathbf{A}_j \in \mathbb{C}^{N \times N}$ is the discretization of the advection operator \mathcal{A} defined by $\mathcal{A}u := \boldsymbol{\theta}_j \cdot \nabla u$. From (3.8) we can deduce that \mathbf{A}_j has no more than $N \times N_E$ nonzero elements, where N_E is the total number of edges (surfaces in three dimensions) each control cell has.

Matrix $\mathbf{C}_0 \in \mathbb{C}^{N \times N}$ is diagonal:

$$\mathbf{C}_0 = \begin{pmatrix} V_1(\sigma_t^1 + \frac{i\omega}{v}) & & \\ & \ddots & \\ & & V_N(\sigma_t^N + \frac{i\omega}{v}) \end{pmatrix},$$

where we recall $\sigma_t^i \equiv \sigma_a^i + \sigma_s^i$ ($i = 1, \dots, N$).

The matrix \mathbf{S} can be expressed as the direct product of two smaller matrices:

$$(3.13) \quad \mathbf{S} = \mathbf{K} \otimes \mathbf{D}_0,$$

with $\mathbf{D}_0 \in \mathbb{C}^{N \times N}$ a diagonal matrix given by

$$\mathbf{D}_0 = \begin{pmatrix} V_1 \sigma_s^1 & & \\ & \ddots & \\ & & V_N \sigma_s^N \end{pmatrix},$$

and $\mathbf{K} \in \mathbb{C}^{J \times J}$ a dense matrix with component $(\mathbf{K})_{jj'} = \eta_{j'} k_{jj'}$. In practical applications, the number of directions is much smaller than the number of spatial mesh elements ($J \ll N$). So although \mathbf{K} is dense, the scattering matrix \mathbf{S} is sparse. However, in general the matrix \mathbf{K} is not symmetric unless we choose η_j to be constant. The matrix $\mathbf{A} - \mathbf{S}$ is thus neither symmetric nor positive definite, which is the reason for us to choose a GMRES solver in section 4.2.

Let us remark here that our finite volume discretization reduces to an upwind finite difference scheme on usual finite difference grids. We refer to our earlier work [54] for some numerical tests on the finite volume discretization of the transport equation.

3.3. Discrete adjoint problem. We present in this section the numerical method we have employed to compute the gradient of discrete objective function with respect to the optical properties on each cell.

To simplify the notation, from now on we denote by $\sigma_a \in \mathbb{R}^{N \times 1}$ the absorption coefficient vector $(\sigma_a^1, \dots, \sigma_a^C, \dots, \sigma_a^N)^T$ and by $\sigma_s \in \mathbb{R}^{N \times 1}$ the scattering coefficient vector $(\sigma_s^1, \dots, \sigma_s^C, \dots, \sigma_s^N)^T$.

We want to minimize the discrepancy between model predictions and measurements over a set of source and detector pairs. Let N_q denote the number of sources used in an experiment, and let N_d denote the number of detectors used for each source. Then the following objective function we employed takes the form

$$(3.14) \quad \mathcal{F}_\beta(\sigma_a, \sigma_s) = \frac{1}{2} \sum_{q=1}^{N_q} \sum_{d=1}^{N_d} |\mathcal{P}_d \mathbf{U}^q - z_{q,d}^\delta|^2 + \frac{\beta}{2} \mathcal{J}(\sigma_a, \sigma_s),$$

where $z_{q,d}^\delta$ denote the d th measurement of the q th source. The superscript δ is used to denote the level of noise contained in the measurements. \mathbf{U}^q is a solution of the transport equation for the q th source. $\mathcal{P}_d \in \mathbb{R}^{1 \times N}$ is a discretized version of the measurement operator. It takes the outgoing flux at detector d and averages over S_+^{n-1} . The discretized regularization term is given by

$$(3.15) \quad \mathcal{J}(\sigma_a, \sigma_s) = \sum_{\mathcal{C}=1}^N \left(\sum_{\kappa=\{x,y,z\}} [\mathcal{D}_\kappa^{\mathcal{C}}(\sigma_a - \sigma_a^0)]^2 + (\sigma_a^{\mathcal{C}} - \sigma_a^{0,\mathcal{C}})^2 \right) + \epsilon \sum_{\mathcal{C}=1}^N \left(\sum_{\kappa=\{x,y,z\}} [\mathcal{D}_\kappa^{\mathcal{C}}(\sigma_s - \sigma_s^0)]^2 + (\sigma_s^{\mathcal{C}} - \sigma_s^{0,\mathcal{C}})^2 \right),$$

where $\mathcal{D}_\kappa^{\mathcal{C}} \in \mathbb{R}^{1 \times N}$ denotes the discretized partial differential operator at cell \mathcal{C} in the κ ($= x, y, z$) direction.

We now start to compute the gradient of objective function (3.14) with respect to optical properties on each mesh element. It is straightforward to check that

$$(3.16) \quad \frac{\partial \mathcal{F}_\beta}{\partial \sigma_a^{\mathcal{C}}} = \left[\sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \bar{\mathbf{r}}_d^q \mathcal{P}_d \frac{\partial \mathbf{U}^q}{\partial \sigma_a^{\mathcal{C}}} \right]_{\text{Re}} + \frac{\beta}{2} \frac{\partial \mathcal{J}}{\partial \sigma_a^{\mathcal{C}}},$$

with $\mathbf{r}_d^q = \mathcal{P}_d \mathbf{U}^q - z_{q,d}^\delta$, and $[\cdot]_{\text{Re}}$ denotes the real part of $[\cdot]$.

At the same time, we notice from (3.10) that

$$(3.17) \quad \frac{\partial \mathbf{A}}{\partial \sigma_a^{\mathcal{C}}} \mathbf{U}^q + \mathbf{A} \frac{\partial \mathbf{U}^q}{\partial \sigma_a^{\mathcal{C}}} = \frac{\partial \mathbf{S}}{\partial \sigma_a^{\mathcal{C}}} \mathbf{U}^q + \mathbf{S} \frac{\partial \mathbf{U}^q}{\partial \sigma_a^{\mathcal{C}}}$$

for source $q = 1, \dots, N_q$, which is equivalent to saying that

$$(3.18) \quad \frac{\partial \mathbf{U}^q}{\partial \sigma_a^c} = -(\mathbf{A} - \mathbf{S})^{-1} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_a^c} \mathbf{U}^q,$$

since $\mathbf{A} - \mathbf{S}$ is invertible. It is very important to note that the matrices \mathbf{A} and \mathbf{S} are independent of the source used. Thus, there are no superscripts q associated with them. We thus have

$$(3.19) \quad \frac{\partial \mathcal{F}_\beta}{\partial \sigma_a^c} = - \left[\sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \bar{\mathbf{r}}_d^q \mathcal{P}_d (\mathbf{A} - \mathbf{S})^{-1} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_a^c} \mathbf{U}^q \right]_{\text{Re}} + \frac{\beta}{2} \frac{\partial \mathcal{J}}{\partial \sigma_a^c}.$$

We now introduce a new state variable $\Psi^q \in \mathbb{C}^{N \times 1}$ (called the adjoint variable of \mathbf{U}^q) given by

$$(3.20) \quad - \sum_{d=1}^{N_d} \bar{\mathbf{r}}_d^q \mathcal{P}_d (\mathbf{A} - \mathbf{S})^{-1} = \Psi^{qT},$$

where Ψ^{qT} denotes the transpose of Ψ^q . We then say that Ψ^q is the solution of the following *adjoint* equation of (3.10):

$$(3.21) \quad (\mathbf{A} - \mathbf{S})^T \Psi^q = - \sum_{d=1}^{N_d} \bar{\mathbf{r}}_d^q \mathcal{P}_d^T.$$

One then arrives at

$$(3.22) \quad \frac{\partial \mathcal{F}_\beta}{\partial \sigma_a^c} = \left[\sum_{q=1}^{N_q} \Psi^{qT} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_a^c} \mathbf{U}^q \right]_{\text{Re}} + \frac{\beta}{2} \frac{\partial \mathcal{J}}{\partial \sigma_a^c},$$

with

$$\frac{\partial \mathcal{J}}{\partial \sigma_a^c} = 2 \left(\sum_{\kappa=\{x,y,z\}} \mathcal{D}_\kappa^c (\sigma_a - \sigma_a^0) (\mathcal{D}_\kappa^c \mathcal{I}_c) + (\sigma_a^c - \sigma_a^{0,c}) \right),$$

where the unit direction vector $\mathcal{I}_c \in \mathbb{R}^{N \times 1}$ is a vector whose c th element is 1 and all other components are zero.

Very similar computation leads to the fact that the derivatives of the objective functional with respect to σ_s^c are given by

$$(3.23) \quad \frac{\partial \mathcal{F}_\beta}{\partial \sigma_s^c} = \left[\sum_{q=1}^{N_q} \Psi^{qT} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_s^c} \mathbf{U}^q \right]_{\text{Re}} + \frac{\beta}{2} \frac{\partial \mathcal{J}}{\partial \sigma_s^c},$$

with

$$\frac{\partial \mathcal{J}}{\partial \sigma_s^c} = 2\epsilon \left(\sum_{\kappa=\{x,y,z\}} \mathcal{D}_\kappa^c (\sigma_s - \sigma_s^0) (\mathcal{D}_\kappa^c \mathcal{I}_c) + (\sigma_s^c - \sigma_s^{0,c}) \right).$$

Formulas (3.22) and (3.23) are what we used to compute the derivatives of objective function with respect to optical properties on each element. Note that we did not

form explicitly the matrix $\frac{\partial(\mathbf{A}-\mathbf{S})}{\partial\sigma_a^c}$ (resp., $\frac{\partial(\mathbf{A}-\mathbf{S})}{\partial\sigma_s^c}$) in the evaluations of $\Psi^{qT} \frac{\partial(\mathbf{A}-\mathbf{S})}{\partial\sigma_a^c} \mathbf{U}^q$ (resp., $\Psi^{qT} \frac{\partial(\mathbf{A}-\mathbf{S})}{\partial\sigma_s^c} \mathbf{U}^q$) because this matrix has a very simple sparse structure according to (3.12) and (3.13). Instead, a matrix-free method was adopted. In fact, since

$$(3.24) \quad \left[\frac{\partial(\mathbf{A}-\mathbf{S})}{\partial\sigma_a^c} \right]_{ij} = \begin{cases} V_c, & i = j \text{ and } \text{mod}(i, N) = C, \\ 0 & \text{otherwise,} \end{cases}$$

where we recall that N is the total number of volume cells that cover our computational domain, we have

$$(3.25) \quad \Psi^{qT} \frac{\partial(\mathbf{A}-\mathbf{S})}{\partial\sigma_a^c} \mathbf{U}^q = \sum_{j=1}^J \Psi_{(j-1) \times N + C}^q V_c \mathbf{U}_{(j-1) \times N + C}^q.$$

Note that here $\Psi_{(j-1) \times N + C}^q$ (resp., $\mathbf{U}_{(j-1) \times N + C}^q$) denotes the $[(j-1) \times N + C]$ th element of Ψ^q (resp., \mathbf{U}^q).

The same observation on $\frac{\partial(\mathbf{A}-\mathbf{S})}{\partial\sigma_s^c}$ leads to

$$(3.26) \quad \Psi^{qT} \frac{\partial(\mathbf{A}-\mathbf{S})}{\partial\sigma_s^c} \mathbf{U}^q = \sum_{j=1}^J \Psi_{(j-1) \times N + C}^q V_c \mathbf{U}_{(j-1) \times N + C}^q - \sum_{j'=1}^J \sum_{j=1}^J (\mathbf{K})_{j'j} \Psi_{(j'-1) \times N + C}^q V_c \mathbf{U}_{(j-1) \times N + C}^q,$$

where the (j', j) th component of matrix \mathbf{K} , $(\mathbf{K})_{jj'} = \eta_{j'} k_{jj'}$, is given in (3.13). We can thus evaluate (3.22) and (3.23) without forming any intermediate matrices.

4. Numerical implementation. We have implemented the quasi-Newton optimization algorithm to solve the regularized least-squares problem (RLS) introduced in section 2.2. We have found in practice that this method converged much faster (in terms of function evaluations) than the nonlinear conjugate gradient method with either the Fletcher–Reeves or the Polak–Ribière updating formula [52]. This is expected from theory [52] and is consistent with practical applications tested in [41]. We have also implemented a Gauss–Newton method [52] to solve the least-squares problem (without the bounds constraints) and found that the method converges extremely slowly in our case. This is probably due to the fact that our problem is highly nonlinear and the Gauss–Newton method usually does not work well in this kind of situation [27, 52]. Detail comparison between various methods of solving the least-squares reconstruction problem is an ongoing project.

In this work, we employ the BFGS update rule [52] of the inverse Hessian matrix for our quasi-Newton method. The usual BFGS method, however, requires the explicit construction of the Hessian matrix, which is unrealistic for large problems. The memory size required to store the Hessian matrix is roughly proportional to the square of the memory used for the unknown parameters. We have thus resorted to a limited-memory version of the BFGS method which avoids the explicit construction of the inverse Hessian matrix.

4.1. Numerical optimization. The BFGS algorithm can be viewed as a special case of the quasi-Newton method [52]. With σ denoting the vector of discretized

optical properties, the quasi-Newton methods can be characterized by the following iterative process:

$$(4.1) \quad \boldsymbol{\sigma}_{k+1} = \boldsymbol{\sigma}_k + \alpha_k \mathbf{p}_k, \quad k \in \mathbb{N}^+,$$

where \mathbf{p}_k is a descent direction vector and α_k is the step length. The BFGS algorithm chooses \mathbf{p}_k to be the solution of an approximated solution of Newton-type optimality equation, i.e.,

$$(4.2) \quad \mathbf{p}_k = H_k \mathbf{g}_k,$$

where \mathbf{g}_k is the gradient of the least-squares functional, $\mathbf{g}_k = -\nabla_{\boldsymbol{\sigma}} \mathcal{F}_\beta(\boldsymbol{\sigma}_k)$. H_k is the inverse Hessian matrix of \mathcal{F}_β at step k . Instead of computing real inverse Hessian matrices, which is very time consuming, the BFGS algorithm chooses to approximate H_k by the following updating rule:

$$(4.3) \quad H_{k+1} = W_k^T H_k W_k + \rho_k \mathbf{s}_k \mathbf{s}_k^T,$$

with $W_k = \mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^T$, $\mathbf{s}_k = \boldsymbol{\sigma}_{k+1} - \boldsymbol{\sigma}_k$, $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$, and $\rho_k = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}$. \mathbf{I} is the identity matrix. As we mentioned above, forming (4.3) takes a tremendous amount of computer memory for large problems. To overcome this shortcoming, the limited-memory version of BFGS stores only the vectors \mathbf{y}_k and \mathbf{s}_k obtained in the last m ($3 \leq m \leq 7$ usually) iterations [35] and discards the rest. Thus after the first m iterations, (4.3) can be expressed as

$$(4.4) \quad \begin{aligned} H_{k+1} &= (W_k^T \cdots W_{k-m}^T) H_{k+1}^0 (W_{k-m} \cdots W_k) \\ &+ \rho_{k-m} (W_k^T \cdots W_{k-m+1}^T) \mathbf{s}_{k-m} \mathbf{s}_{k-m}^T (W_{k-m+1} \cdots W_k) \\ &+ \rho_{k-m+1} (W_k^T \cdots W_{k-m+2}^T) \mathbf{s}_{k-m+1} \mathbf{s}_{k-m+1}^T (W_{k-m+2} \cdots W_k) \\ &\vdots \\ &+ \rho_k \mathbf{s}_k \mathbf{s}_k^T, \end{aligned}$$

with the sparse initial guess H_{k+1}^0 given by $H_{k+1}^0 = \frac{\mathbf{y}_{k+1}^T \mathbf{s}_{k+1}}{\mathbf{y}_{k+1}^T \mathbf{y}_{k+1}} \mathbf{I}$.

We refer interested readers to [35, 52] for more details on the limited-memory BFGS algorithms, and to [41] for applications of those algorithms to optical tomographic problems. Convergence of BFGS algorithms has been proved under certain conditions and has been tested on many applications [11, 52].

To impose bounds on optical parameters, we have to modify the relation (4.2) slightly. We adopt a gradient projection method [11, 35, 52] to do this. At the beginning of each iteration, we use the gradient projection method to find a set of active bounds. We then solve a subminimization problem,

$$(4.5) \quad \min_{\boldsymbol{\sigma}} Q_k(\boldsymbol{\sigma}) \equiv \mathcal{F}_\beta(\boldsymbol{\sigma}_k) + \mathbf{g}_k^T (\boldsymbol{\sigma} - \boldsymbol{\sigma}_k) + \frac{1}{2} (\boldsymbol{\sigma} - \boldsymbol{\sigma}_k)^T H_k^{-1} (\boldsymbol{\sigma} - \boldsymbol{\sigma}_k),$$

on the set of free variables to find an approximation solution $\bar{\boldsymbol{\sigma}}_{k+1}$, treating the active bounds as equality constraints by Lagrange multipliers. After we find an approximation solution $\bar{\boldsymbol{\sigma}}_{k+1}$, a line search along $\mathbf{p}_k = \bar{\boldsymbol{\sigma}}_{k+1} - \boldsymbol{\sigma}_k$ is done to find the step length α_k in (4.1). We use a line search method that enforces the Wolfe conditions [52]. That is, we look for an α_k that solves

$$(4.6) \quad \min_{\alpha_k > 0} \mathcal{F}_\beta(\boldsymbol{\sigma}_k + \alpha_k \mathbf{p}_k)$$

and satisfies

$$(4.7) \quad \mathcal{F}_\beta(\boldsymbol{\sigma}_k + \alpha_k \mathbf{p}_k) \leq \mathcal{F}_\beta(\boldsymbol{\sigma}_k) + c_1 \alpha_k \nabla \mathcal{F}_\beta^T(\boldsymbol{\sigma}_k) \mathbf{p}_k,$$

$$(4.8) \quad \nabla \mathcal{F}_\beta^T(\boldsymbol{\sigma}_k + \alpha_k \mathbf{p}_k) \mathbf{p}_k \geq c_2 \nabla \mathcal{F}_\beta^T(\boldsymbol{\sigma}_k) \mathbf{p}_k,$$

where $c_1 = 10^{-4}$, $c_2 = 0.1$ in our case. More details on how to impose bound constraints in BFGS algorithms can be found in [35, sect. 5.5] and [11, 64].

4.2. Solving algebraic systems. As we have mentioned before, at each step of the minimization process, we have to solve both a discretized transport equation (3.10) and its adjoint problem (3.21) to compute the Fréchet derivatives (3.22) and (3.23) of the objective functional, forming the gradient vector \mathbf{g}_k in (4.2). In fact, almost all of the computational time in the reconstruction process is devoted to the solution of these transport equations. In this work, instead of using the popular source iteration method, which converges very slowly in diffusive regimes unless it is properly accelerated [1], we choose to solve the forward problems by a preconditioned GMRES(n) algorithm [55, 56], where n denotes the number of iterative steps after which GMRES is restarted. Our general principle is to choose n large when the problem size is small and n small when the problem size is large. The implementation of the algorithm is based on the template provided in [9]. The preconditioner we employ is the zero fill-in incomplete LU factorization (ILU(0)) [53, 55] that has been proved to be efficient in transport calculations [53]. Details about this factorization can be found in [55]. In all of the numerical examples in section 5, we pick $n = 7$, and the GMRES algorithm is stopped if the relative residual is small enough. For example, the stopping criteria $\|\mathbf{G} - (\mathbf{A} - \mathbf{S})\mathbf{U}_k\|_{l^2} / \|\mathbf{G} - (\mathbf{A} - \mathbf{S})\mathbf{U}_0\|_{l^2} \leq 10^{-10}$ is used to solve (3.10). Here \mathbf{U}_0 is the initial guess and \mathbf{U}_k is the \mathbf{U} value at the k th GMRES iteration.

4.3. Selecting regularization parameter. To choose the optimal regularization parameter β in (2.11), we adopt the L -curve method in this study. Although there exist proofs that the L -curve method fails to convergence for certain classes of inverse problems [59], we have observed satisfactory results in our applications. We plot the log of the regularization functional against the squared norm of the regularized residual, say, r_β , for a range of values of the regularization parameter. The right parameter β is the one at which the L -curve reaches the maximum of its curvature [28, 60]. One can show that the right β maximizes the curvature function [60]

$$(4.9) \quad \kappa(\beta) = -\frac{R(\beta)S(\beta)[\beta R(\beta) + \beta^2 S(\beta)] + [R(\beta)S(\beta)]/S'(\beta)}{[R^2(\beta) + \beta^2 S^2(\beta)]^{3/2}},$$

where $R(\beta)$ and $S(\beta)$ are defined by

$$R(\beta) := \|r_\beta\|_{L^2}^2, \quad S(\beta) := \mathcal{J}(\sigma_a, \sigma_s).$$

We recall that β is not included in $\mathcal{J}(\sigma_a, \sigma_s)$. One notices immediately that the L -curve method requires several reconstructions for any single problem, and thus is very time consuming. A simple *continuation* method is suggested in [27] to reduce the computational cost of regularization parameter selection. In this method, one starts the first reconstruction with a relatively large β . The result of this reconstruction is then taken to be the initial guess of next reconstruction with a smaller β . If the two

β are not dramatically different from each other, then the two reconstructions should converge to similar results. Thus, the reconstruction with smaller β is supposed to converge quickly since its initial guess is chosen to be close enough to its real solution. The process can be repeated to perform reconstructions with several values of β . In practice, this continuation method saves tremendous computational time for finding optimal regularization parameters. It now takes only about 80% extra work to find the optimal parameters in most cases. We adopt this continuation method in our three-dimensional numerical example (Example 4) in the next section. We present in Figures 5.2 (B) and 5.6 (B) the L -curve we have used in Examples 1 and 3, respectively, to choose the optimal regularization parameter β . Note that in the Figure 5.2 (B), $\mathcal{J}(\sigma_a, \sigma_s)$ simplifies to $\|\sigma_a\|_{\mathcal{H}^1}^2$ since we reconstruct only σ_a and have chosen $\sigma_a^0 = 0$ in that case.

Another important issue is to choose an appropriate weight ϵ in the regularization functional defined in (2.12). This weight is necessary because, in practice, σ_s takes values that are about two orders of magnitude larger than σ_a . The weight is used to bring the two terms in the regularization functional to the same level so that the regularization term has an effect on both σ_a and σ_s . In all our numerical simulations in section 5, we choose ϵ to be the ratio $(\sigma_a^b/\sigma_s^b)^2$, where σ_a^b and σ_s^b are optical properties of background media.

Note that the main purpose for introducing the scaling factor ϵ is to reduce computational complexity in choosing optimal regularization parameters. It is by no means the only way to perform regularizations on both absorption and scattering coefficients. For example, one can regularize σ_a and σ_s separately, which will involve selecting multiple regularization parameters such as those in [10, 26, 45].

We remark finally that the \mathcal{H}^1 -norm we use in the regularization functional can be replaced by other norms or seminorms. For example, we have performed reconstructions on numerical Example 1 in section 5.3 with stricter bounds on σ_a and $\|\nabla\sigma_a\|_{L^2}$ (instead of $\|\sigma_a\|_{\mathcal{H}^1}$) as the regularization functional. We have obtained very similar results (although the optimal regularization parameter changes). The main reason for us to use the \mathcal{H}^1 -norm is that in many practical applications, we want to find solutions near some reference (σ_a^0, σ_s^0) , for example, some known background.

4.4. Cost of the numerical method. The computational cost of our method consists of two main parts. The first part is the evaluation of the objective function and its gradient in the optimization process. The second part is the updating of the BFGS matrices and vectors.

The costs of the function evaluation and of its gradient scale linearly with the number of optical sources N_q . Since each forward problem and its corresponding adjoint problem cost about the same, each gradient calculation (about $2N_q$ forward solves) is approximately twice as expensive as a function evaluation (about N_q forward solves). The cost in updating BFGS matrices and vectors can be neglected compared to function and gradient evaluations. The reason is that BFGS vectors (in \mathbb{R}^{2N}) are dramatically smaller than the vectors appearing in the forward and adjoint problems (in \mathbb{C}^{JN}).

In our computations, we store the nonzero elements of the matrix $\mathbf{A} - \mathbf{S}$ by using the compressed row storage scheme [9] whenever it is possible to do so. When it is not possible to store $\mathbf{A} - \mathbf{S}$, we store \mathbf{A} , \mathbf{C}_0 , \mathbf{K} , and \mathbf{D}_0 defined in (3.12) and (3.13). This requires much less memory, but with the price that extra effort has to be paid to evaluate matrix vector products in GMRES. We use the following procedure to compute $\mathbf{Y} \equiv (\mathbf{A} - \mathbf{S})\mathbf{X}$ for any vector \mathbf{X} with the same structure as \mathbf{U} in (3.11):

1. For $j = 1, \dots, J$,
 - Compute $\mathbf{X}'_j = \mathbf{D}_0 \mathbf{X}_j$;
 - Compute $\mathbf{X}''_j = \mathbf{C}_0 \mathbf{X}_j$;
 - Compute $\mathbf{X}'''_j = \mathbf{X}''_j + \mathbf{A}_j \mathbf{X}_j$;
2. For $j = 1, \dots, J$, $\mathbf{Y}_j = \mathbf{X}'''_j - \sum_{j'} \mathbf{K}_{jj'} \mathbf{X}'_{j'}$.

We prefer to store the matrix $\mathbf{A} - \mathbf{S}$ because it saves computational time when matrix-vector products are calculated. In all the numerical examples shown in the following section, we were able to store $\mathbf{A} - \mathbf{S}$. Note that the storage requirement does not increase with the number of sources (N_q) because we solve the transport equation (and its adjoint) with different sources sequentially. The storage cost of BFGS vectors can be neglected compared with the storage of the forward and adjoint matrices and vectors.

5. Numerical examples. We provide in this section four numerical examples that illustrate the performance of our numerical method. In the first example, we reconstruct the spatial distribution of the absorption coefficient while keeping the scattering coefficient fixed. In the second example, the spatially varying scattering coefficient is reconstructed, while the absorption coefficient is fixed. We then show an example in which both optical properties are reconstructed simultaneously. All the first three examples are done in two-dimensional settings. In the fourth example, we show three-dimensional simultaneous reconstructions of both optical properties.

5.1. Domain partition and measurement setup. For our two-dimensional simulations, we consider a computational domain of size 2×2 cm², denoted by $\overline{\mathcal{D}} \equiv \mathcal{D} \cup \partial\mathcal{D} = [0, 2] \times [0, 2]$. We cover the domain by 80×80 cells of uniform size whose nodes are given by

$$\mathcal{D}_h = \{\mathbf{x}_{i,j} = (x_i, y_j), x_i = i\Delta x, y_j = j\Delta y, i, j = 0, 1, \dots, 80\},$$

with $\Delta x = \Delta y = 0.025$. The direction space is discretized into 128 uniformly distributed (over $[0, 2\pi)$) directions with identical quadrature weight:

$$S_{\Delta\theta}^1 = \{\boldsymbol{\theta}_i : \boldsymbol{\theta}_i = (i-1) * \Delta\theta, i = 1, \dots, 128\},$$

where $\Delta\theta = 2\pi/128$. The above discretizations yield a total number of 819,200 unknowns for one forward problem (solving for \mathbf{U} in (3.10)), which is also true for the corresponding adjoint problem ($\boldsymbol{\Psi}$ in (3.21)). In all two-dimensional simulations, four sources ($N_q = 4$) are used. They are located at $(0, 1)$, $(1, 0)$, $(2, 1)$, and $(1, 2)$, respectively. For each sources, 20 detectors ($N_d = 20$) are used. The detectors are uniformly distributed along the sides of the square.

For the three-dimensional simulation, we consider a cylindrical domain given by

$$\overline{\mathcal{D}} := \{(x, y, z) : x^2 + y^2 \leq 1; 0 \leq z \leq 2\}.$$

We cover the domain by 22,022 tetrahedral elements. For the integration over S^2 , we employ the full level symmetric S_8 discrete ordinate set of [44]. This set consists of 80 directions. A total number of 16 sources ($N_q = 16$) are used in the simulation. The sources are uniformly distributed on the two circles defined by $\Gamma_i^q = \{(x, y, z) : x^2 + y^2 = 1; z = z_i\}$ ($i = 1, 2$), where $z_1 = 0.5$, $z_2 = 1.5$. We arrange 8 layers of detectors on the boundary of the domain. Each layer has 16 detectors. Those detectors are located at $\Gamma_j^d = \{(x, y, z) : x^2 + y^2 = 1; z = z_j\}$ ($j = 1, \dots, 8$), with $z_j = 0.3 + (j-1) * 0.2$. Note that some detectors are placed on the same xy -plane as sources (but they do not overlap each other).

5.2. Generating synthetic data. In all the numerical reconstructions presented in section 5, the “measurements” are synthetic rather than obtained from real experimental data. A straightforward way of generating synthetic data is to use the same discretization for the forward model and for the inversion procedure. This may lead to somewhat simplified inversions of the finite-dimensional problem, which is often referred to as *inverse crimes* in the inverse problems community [15, p. 133]. To avoid committing these “crimes,” the data are generated with a discretization about twice as fine (in all variables) as the discretization used in the inversions.

In the following sections, our “exact data” will thus be those obtained from the fine mesh calculations. By noisy data, we mean the “exact data” polluted by additional multiplicative noise. Let z_d be the exact data; the “noisy” data are simulated according to the rule $z_d^\delta = (1 + \delta * \text{random})z_d$, where “random” is a uniformly distributed random variable in $[-1, 1]$ and $\delta \geq 0$ will vary in our numerical simulations.

It should be noted that the “exact data” seen on the coarse grid used in the inversion actually already contain some “noise” because they were generated by the fine mesh calculations.

5.3. Single parameter reconstructions. We start with a simpler case where only one optical parameter needs to be reconstructed. Such reconstructions are often useful in practical applications. For example, it is generally believed that changes in the oxygenation of tissues correspond mainly to changes in the absorption property of tissues [58].

Example 1: Reconstructing σ_a . We first reconstruct a small absorbing disc centered at (1.15 cm, 1.15 cm) of radius 0.2 cm embedded in the two-dimensional computational domain. The optical parameters for the background and the disc are $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 80 \text{ cm}^{-1}$ and $\sigma_a = 0.2 \text{ cm}^{-1}$, $\sigma_s = 80 \text{ cm}^{-1}$, respectively. The anisotropy factor $g = 0.90$. Those are typical parameter values as they are encountered in biomedical applications. The modulation frequency of the source is $\omega = 600$ MHz. Each reconstruction here takes approximately 5 hours on a 2.4 GHz Pentium XEON processor.

Figure 5.1 shows the real absorption map and the reconstructed maps for different noise levels. We also provide in Figure 5.2 (A) (solid line) the evolution of the normalized objective function versus the iteration step in the case of reconstructions using noise-free synthetic data. Note that although the total number of BFGS iterations can be reduced by using a stricter line search scheme in the optimization algorithm, the total computational time remains almost constant. We have tested many classical line search algorithms. The results shown here use the one that works best for our application.

The quality of the reconstructions is measured as follows. Denote by $M^o \in \mathbb{R}^N$ ($M^r \in \mathbb{R}^N$) an exact (reconstructed) quantity, which can be either the absorption or the scattering map. We then define the relative l^2 error between M^o and M^r by

$$(5.1) \quad \mathcal{E}_{l^2} = \frac{\|M^r - M^o\|_{l^2}}{\|M^o\|_{l^2}} := \frac{\sqrt{\sum_{i=1}^N (M_i^r - M_i^o)^2}}{\sqrt{\sum_{i=1}^N (M_i^o)^2}}.$$

The quality of the reconstruction in the case of an absorbing disc is given by the parameters shown in the second column of Table 5.1. In Figure 5.2 (B), we display the L -curve we use to choose the optimal parameter β in the above reconstructions. We mention again that due to the acquisition of our data presented in section 5.2, the noise-free data actually contain “noise” on the coarse grid.

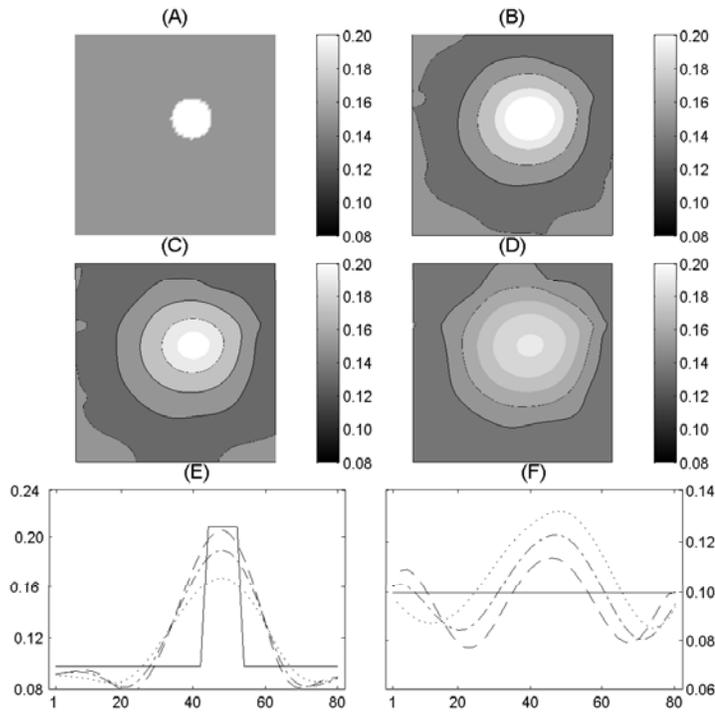


FIG. 5.1. Maps of the reconstructed absorption coefficient σ_a [cm^{-1}] in Example 1. (A) Real absorption map. (B) Reconstructed absorption map with exact synthetic data. (C) Reconstruction with 10% random noise. (D) Reconstruction with 20% random noise. (E) Cross sections of map (A) (solid line), (B) (dashed line), (C) (dash-dotted line), and (D) (dotted line) along the bottom left to top right diagonal. (F) Same as (E) except that the cross section is along $y = 0.4$.

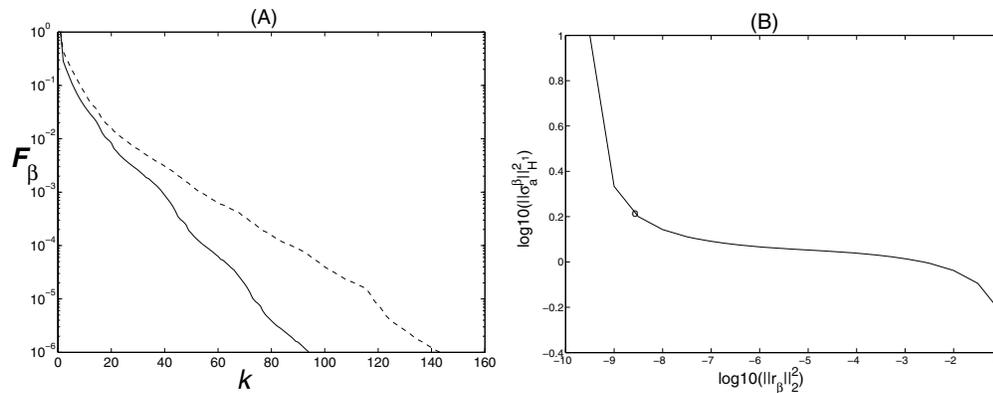


FIG. 5.2. (A) Evolution of normalized objective functional \mathcal{F}_β with respect to the number of iteration steps k for Example 1. Solid line: reconstruction of an absorbing disc; dashed line: reconstruction of a scattering disc. (B) L-curve used to choose optimal regularization parameter β for reconstruction with noise-free data in the reconstruction of an absorbing disc. The circle (\circ) denotes the place where β is chosen. Note that here $\mathcal{J} = \|\sigma_a\|_{\mathcal{H}^1}^2$ because we reconstruct only σ_a and we have set $\sigma_a^0 = 0$.

TABLE 5.1

Optimal regularization parameters β and errors in reconstructions for different cases in Examples 1 and 2.

Cases	Absorbing disc			Scattering disc		
Noise level	0%	10%	20%	0%	10%	20%
$\beta \times 10^8$	1.0	2.8	4.6	1.3	2.9	5.7
\mathcal{E}_{l_2}	0.064	0.069	0.072	0.065	0.074	0.080

The reconstructions are classical examples of what we may expect in the field. Because of the severe ill-posedness of the inverse problem, it is difficult to reconstruct localized objects unless additional information is included in the reconstruction (which we do not want to do here). Note that the center and the integral of the absorption inclusion are more or less well reconstructed, even if the localized absorption map is somewhat smeared over a relatively large domain (whose size increases as the noise level, whence the regularization parameter β , increases). This is consistent with results obtained from asymptotic theories; see, for instance, [7].

Example 2: Reconstructing σ_s . We now replace the highly absorbing disc in the previous case with a highly scattering disc in the same position and with the same size. The optical parameters for the background and the disc are $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 70 \text{ cm}^{-1}$ and $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 80 \text{ cm}^{-1}$, respectively. Again, the anisotropy factor $g = 0.9$, and the modulation frequency $\omega = 600 \text{ MHz}$. Each reconstruction here takes approximately 6 hours on a 2.4 GHz Pentium XEON processor. Figure 5.3 shows the exact scattering map and the reconstructed maps for different noise levels. Error estimates are presented in the third column of Table 5.1.

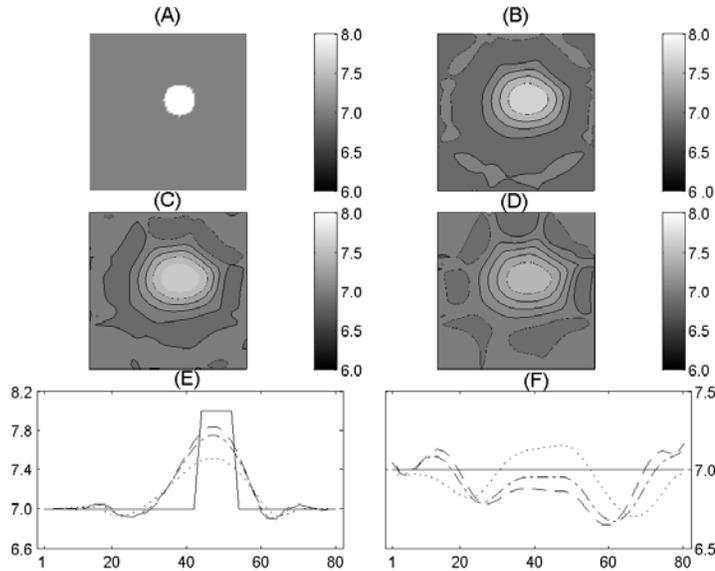


FIG. 5.3. Maps of reconstructed reduced scattering coefficients $\sigma'_s = (1 - g)\sigma_s \text{ [cm}^{-1}\text{]}$ for Example 1. (A) Real scattering map. (B) Reconstructed scattering map with noise-free synthetic data. (C) Reconstruction with 10% random noise. (D) Reconstruction with 20% random noise. (E) Cross sections of map (A) (solid line), (B) (dashed line), (C) (dash-dotted line), and (D) (dotted line) along the bottom left to top right diagonal. (F) Same as (E) except that the cross section is along $y = 0.4$.

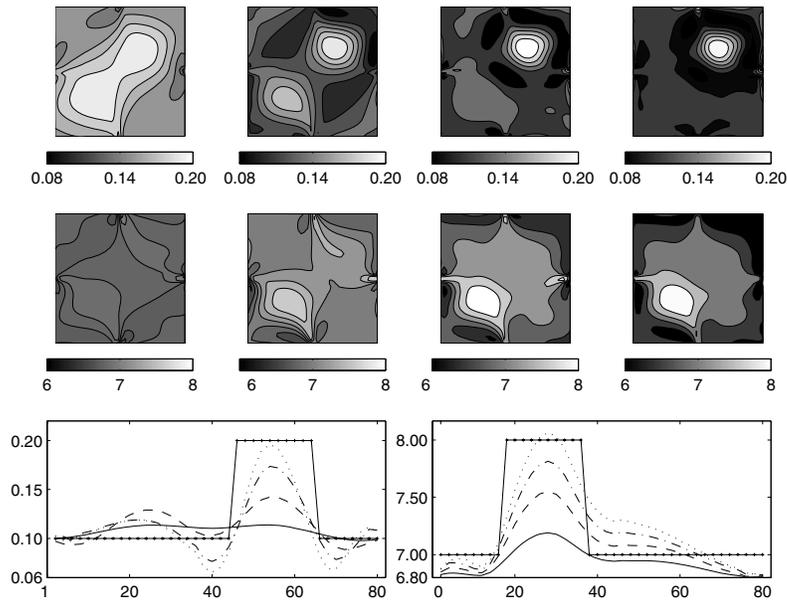


FIG. 5.4. *Top row: Maps of reconstructed absorption coefficients σ_a [cm^{-1}] at BFGS iteration $k = 40, 80, 120,$ and 156 (final), respectively, for the frequency domain reconstruction in Example 3. Middle row: same as the top row but for the reduced reconstructed scattering coefficients $\sigma'_s = (1-g)\sigma_s$ [cm^{-1}]. Bottom row: cross section of real (solid line with +) and reconstructed absorption maps (left), reduced scattering maps (right) along the diagonal at iterations $k = 40$ (solid line), 80 (dashed line), 120 (dash-dotted line), and 156 (dotted line). The reconstructions are done with noise-free synthetic data.*

The quality of the reconstructions is very similar in the above two examples and is quite satisfactory. In spite of the fact that stronger regularizations have to be imposed as the noise level increases, the localization and the estimate for the optical parameters in the presence of moderate noise indeed allow us to obtain reasonably accurate information toward diagnostic purposes.

5.4. Frequency-domain versus steady-state. One of the main reasons for introducing frequency-domain reconstructions is that they allow for a better separation between the scattering and absorption properties of the inclusions [48]. Diffusion-based theories show that both coefficients cannot be reconstructed simultaneously without additional geometrical hypotheses [5, 33]. We now show in two numerical examples that frequency-domain data indeed substantially improve the reconstruction of both coefficients.

Example 3: Two-dimensional simultaneous reconstruction. We reconstruct here in the square domain two small discs of radius 0.2 cm and centered at $(1.35$ cm, 1.35 cm) and $(0.65$ cm, 0.65 cm), respectively. The first disc is highly absorbing and the second one is highly scattering. Optical properties for the two discs are $\sigma_a = 0.2$ cm^{-1} , $\sigma_s = 70$ cm^{-1} and $\sigma_a = 0.1$ cm^{-1} , $\sigma_s = 80$ cm^{-1} , respectively. The background parameters are $\sigma_a = 0.1$ cm^{-1} and $\sigma_s = 70$ cm^{-1} . As before, $g = 0.9$ and $\omega = 600$ MHz. Each reconstruction takes approximately 8 hours on a 2.4 GHz Pentium XEON processor.

We compare the reconstructions based on the frequency-domain ERT with those based on the steady-state ERT. The latter is obtained by setting the frequency $\omega = 0$

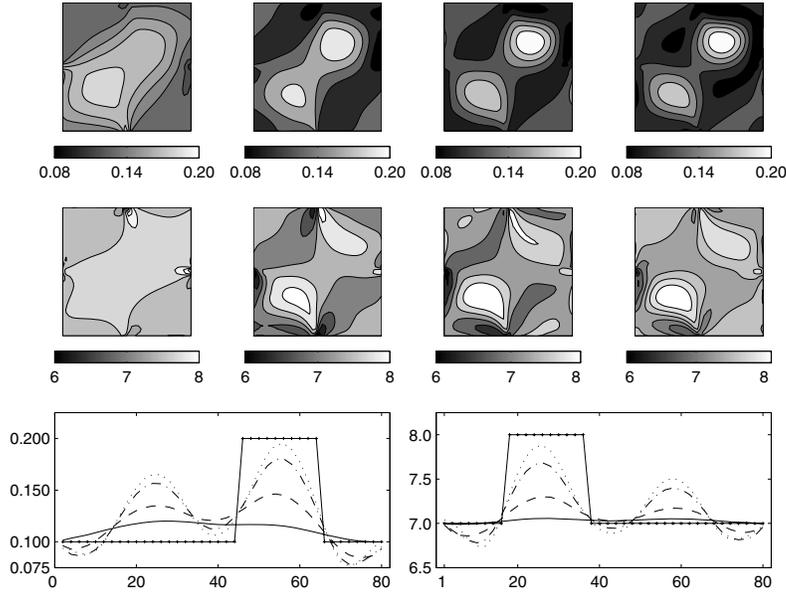


FIG. 5.5. Top row: Maps of reconstructed absorption coefficients σ_a [cm^{-1}] at BFGS iteration $k = 40, 80, 120,$ and 354 (final), respectively, for the steady-state reconstruction in Example 3. Middle row: same as the top row but for the reduced reconstructed scattering coefficients $\sigma'_s = (1-g)\sigma_s$ [cm^{-1}]. Bottom row: cross section of real (solid line with +) and reconstructed absorption maps (left), reduced scattering maps (right) along the diagonal at iterations $k = 40$ (solid line), 80 (dashed line), 120 (dash-dotted line), and 354 (dotted line). The reconstructions are done with noise-free synthetic data.

TABLE 5.2

Error estimates for the reconstructions of Example 3 (E3) and Example 4 (E4) for several iteration steps (k) in the optimization process. Here, “f” refers to frequency-domain calculations and “s” to steady-state calculations.

		$k = 40$		$k = 80$		$k = 120$		Final	
		σ_a	σ_s	σ_a	σ_s	σ_a	σ_s	σ_a	σ_s
E3	$\mathcal{E}_{l_2}(f)$	0.121	0.144	0.092	0.112	0.080	0.092	0.063	0.076
	$\mathcal{E}_{l_2}(s)$	0.181	0.224	0.127	0.132	0.113	0.112	0.094	0.106
E4	$\mathcal{E}_{l_2}(f)$	0.194	0.252	0.173	0.208	0.147	0.189	0.131	0.171
	$\mathcal{E}_{l_2}(s)$	0.342	0.422	0.287	0.366	0.245	0.322	0.210	0.292

in our formulation and keeping everything else the same. Note that we observe a slight difference (below 5%) in the computational times required for the case when $\omega = 0$ and the case when $\omega \neq 0$. We present in Figures 5.4 and 5.5 the reconstructions obtained by the frequency-domain method and the steady-state method, respectively. We also list the parameters which measure the quality of the reconstructions at different iteration steps in Table 5.2.

Example 4: Three-dimensional simultaneous reconstruction. In the last numerical test, we show simultaneous reconstructions of two optical properties in a three-dimensional setting. We try to reconstruct a small cylinder $\overline{\mathcal{D}}_s = \{(x, y, z) : (x - 0.5)^2 + y^2 \leq 0.2^2, 0 \leq z \leq 2\}$ embedded in the cylindrical domain defined in section 5.1. Both the absorption and the scattering coefficients of the small cylinder are

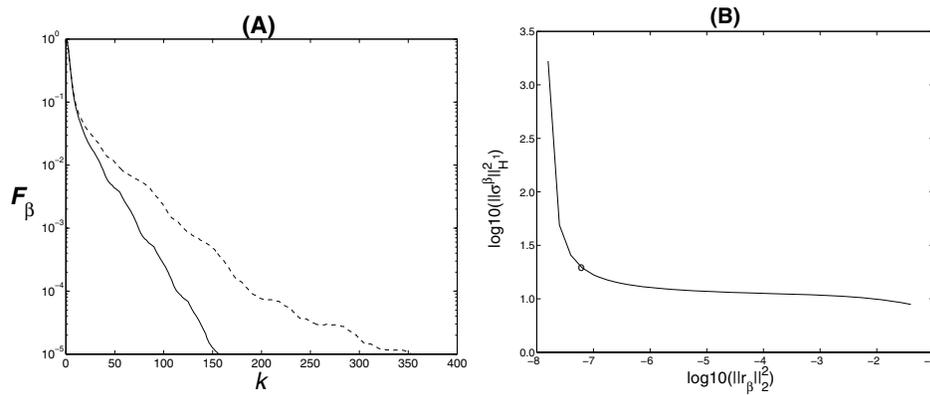


FIG. 5.6. (A) Evolution of the normalized objective functional \mathcal{F}_β with respect to the number of iteration steps k for Example 3. Solid line: frequency domain reconstruction of both coefficient simultaneously; dashed line: steady-state reconstruction of both coefficients simultaneously. (B) L-curve used to choose optimal regularization parameter β for reconstruction with noise-free data in the frequency domain simultaneous reconstruction of absorbing and scattering coefficients. The circle (\circ) denotes the place where β is chosen. Note that $\mathcal{J} \equiv \|\sigma^\beta\|_{\mathcal{H}^1}^2 := \|\sigma_a^\beta\|_{\mathcal{H}^1}^2 + \epsilon \|\sigma_s^\beta\|_{\mathcal{H}^1}^2$ because we have set $\sigma_a^0 = \sigma_s^0 = 0$.

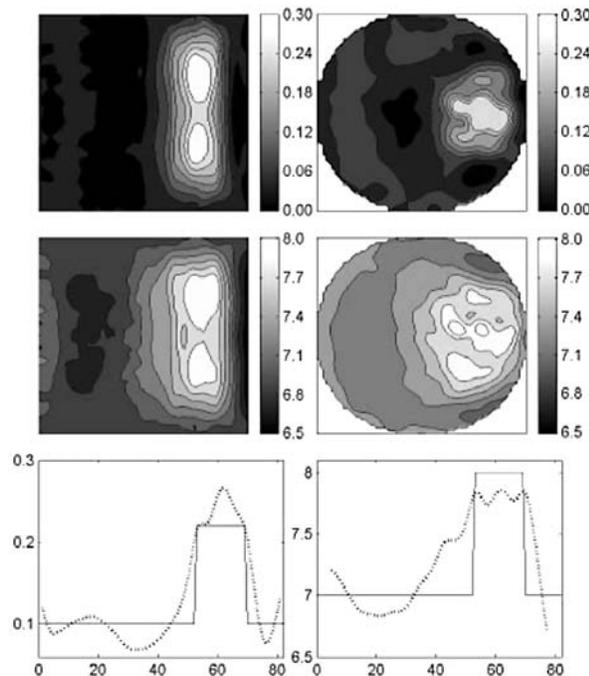


FIG. 5.7. Top row: xy cross section (at $z = 1$) and xz cross section (at $y = 0$) of the reconstructed absorption coefficient σ_a [cm $^{-1}$] for the frequency-domain reconstruction in Example 4. Middle row: same as the top row but for the reduced reconstructed scattering coefficients $\sigma'_s = (1 - g)\sigma_s$ [cm $^{-1}$]. Bottom left: real (solid) and reconstructed (dotted) absorption coefficient along line $y = 0, z = 1$. Bottom right: same as bottom left but for reconstructed reduced scattering coefficient.

different from those of the background. This is different from the case in Example 3 where absorption and scattering anomalies are located at different places. Optical properties for the small cylinder are $\sigma_a = 0.2 \text{ cm}^{-1}$ and $\sigma_s = 80 \text{ cm}^{-1}$, while those for the background are $\sigma_a = 0.1 \text{ cm}^{-1}$ and $\sigma_s = 70 \text{ cm}^{-1}$. The anisotropy factor $g = 0.9$ and the modulation frequency $\omega = 600 \text{ MHz}$. Each reconstruction takes approximately 22 hours on a 3 GHz Pentium XEON processor. As in Example 3, we compare the frequency-domain reconstructions with the steady-state reconstructions. Cross sections of reconstructions obtained by the frequency-domain method and the steady-state method are presented in Figures 5.7 and 5.8, respectively. Qualities of the reconstructions at different iteration steps are again listed in Table 5.2.

We first observe that in both the two-dimensional (Example 3) and the three-dimensional (Example 4) reconstructions, the frequency-domain reconstruction converges faster (in terms of BFGS iterations) than the steady state reconstruction; see, for example, the results in Figure 5.6 (A) and Table 5.2. This has been confirmed in many other geometrical settings we have tested: the speed of convergence of the steady-state reconstruction presented here is one of the most favorable we have obtained, whereas the speed of convergence of the frequency-domain reconstructions was very often similar to what we have presented here.

As far as quality of the reconstruction is concerned, we observe a significant improvement in the frequency domain reconstructions compared to the steady-state

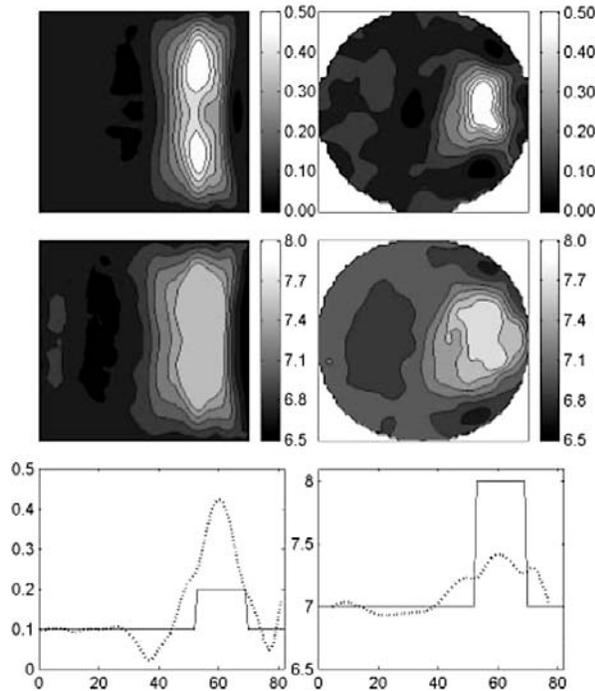


FIG. 5.8. Top row: xy cross section (at $z = 1$) and xz cross section (at $y = 0$) of the reconstructed absorption coefficient σ_a [cm^{-1}] for the steady-state reconstruction in Example 4. Middle row: same as the top row but for the reduced reconstructed scattering coefficients $\sigma'_s = (1 - g)\sigma_s$ [cm^{-1}]. Bottom left: real (solid) and reconstructed (dotted) absorption coefficient along line $y = 0$, $z = 1$. Bottom right: same as bottom left but for reconstructed reduced scattering coefficient.

reconstructions. In all simulations, the stopping criterion is the same:

$$\frac{\mathcal{F}_\beta^k(\sigma_a, \sigma_s)}{\mathcal{F}_\beta^0(\sigma_a, \sigma_s)} \leq 10^{-5}.$$

Although the L^2 errors may not enjoy a dramatic improvement (see Table 5.2), they are still significantly reduced. More importantly, the last rows in Figures 5.4 and 5.5 show dramatic reductions (at least by a factor of 2) of the cross-talk between the absorption and scattering reconstructions: the spurious bumps (left of the left picture on the bottom row and right of the right picture on the bottom row in Figures 5.4 and 5.5) are clearly much stronger in the steady-state calculations than in the frequency-domain calculations. This is the major advantage of the frequency-domain calculations. Very similar phenomena are observed in the three-dimensional simulations; see Figures 5.7 and 5.8. Although the absorption coefficient is still overestimated in the frequency-domain case (last row of Figure 5.7), it is better than the situation in the steady-state case (last row of Figure 5.8) where the absorption coefficient σ_a is severely overestimated, while the reduced scattering coefficient σ'_s is severely underestimated. As predicted by theory, we have observed that an increase in ω led to reduced cross-talk. How much this effect depends on the choice of the frequency (as well as on possible combinations of different frequencies) and on the geometrical setting will be explored in future works.

6. Concluding remarks. We have formulated an inverse problem in optical tomography as a regularized least-squares problem based on the frequency-domain equation of radiative transfer to model light propagation in biological tissues. In the inversion procedure, the forward model is discretized by using a finite volume method and a discrete ordinates method. We solve the regularized least-squares problem by using a limited-memory quasi-Newton method with BFGS-type updating rule for the Hessian matrix and have incorporated positivity constraints and L^∞ bounds on the optical parameters. Numerical reconstructions based on synthetic data provide results that are in agreement with the expected reconstructions. Notably, the crosstalk between the two optical parameters is significantly reduced in frequency-domain reconstructions.

The method presented here also overcomes several of the shortcomings of diffusion-equation-based optical tomography [4, 16], which provides a very useful tool in many problems but fails to adequately model strongly absorbing regions (e.g., large blood-filled spaces such as brain hematoma), low-scattering void-like inclusions (e.g., spaces filled with cerebrospinal fluid, amniotic fluid, or synovial fluid), and optically relatively thin media such as fingers and small animals. Main domains of application of the proposed method will be functional imaging of rheumatoid arthritis of human finger joints and small animal imaging [30, 32], where the diffusion equation typically fails to generate accurate forward predictions. Extensive numerical reconstructions in practically relevant geometries for these applications and based on possibly multifrequency experimental data will be performed in the near future.

Acknowledgment. We would like to thank the two anonymous referees for their many constructive comments and suggestions.

REFERENCES

- [1] M. L. ADAMS AND E. W. LARSEN, *Fast iterative methods for discrete-ordinates particle transport calculations*, Prog. Nucl. Energy, 40 (2002), pp. 3–150.

- [2] V. AGOSHKOV, *Boundary Value Problems for the Transport Equations*, Birkhäuser, Boston, 1998.
- [3] R. ARONSON, R. L. BARBOUR, J. LUBOWSKY, AND H. GRABER, *Application of transport theory to infra-red medical imaging*, in *Modern Mathematical Methods in Transport Theory*, W. Greenberg and J. Polewczak, eds., Birkhäuser, Basel, 1991, pp. 64–75.
- [4] S. R. ARRIDGE, *Optical tomography in medical imaging*, *Inverse Problems*, 15 (1999), pp. R41–R93.
- [5] S. R. ARRIDGE AND W. R. B. LIONHEART, *Nonuniqueness in diffusion-based optical tomography*, *Opt. Lett.*, 23 (1998), pp. 882–884.
- [6] G. BAL, *Transport through diffusive and nondiffusive regions, embedded objects, and clear layers*, *SIAM J. Appl. Math.*, 62 (2002), pp. 1677–1697.
- [7] G. BAL, *Optical tomography for small volume absorbing inclusions*, *Inverse Problems*, 19 (2003), pp. 371–386.
- [8] G. BAL AND K. REN, *Generalized diffusion model in optical tomography with clear layers*, *J. Opt. Soc. Amer. A*, 20 (2003), pp. 2355–2364.
- [9] R. BARRETT, M. BERRY, T. F. CHAN, J. DEMMEL, J. DONATO, J. DONGARRA, V. ELJKHOUT, R. POZO, C. ROMINE, AND H. VAN DER VORST, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, Philadelphia, 1994.
- [10] M. BELGE, M. KILMER, AND E. MILLER, *Efficient determination of multiple regularization parameters in a generalized L-curve framework*, *Inverse Problems*, 18 (2002), pp. 1161–1183.
- [11] R. H. BYRD, P. LU, J. NOCEDAL, AND C. ZHU, *A limited memory algorithm for bound constrained optimization*, *SIAM J. Sci. Comput.*, 16 (1995), pp. 1190–1208.
- [12] M. T. CHAHINE, *Inverse problems in radiative transfer: Determination of atmospheric parameters*, *J. Atmos. Sci.*, 27 (1970), pp. 960–967.
- [13] B. CHANCE, R. R. ALFANO, B. J. TROMBERG, AND A. KATZIR, EDS., *Optical Tomography and Spectroscopy of Tissue*, Vol. V, SPIE, Bellingham, WA, 2003.
- [14] M. CHOULLI AND P. STEFANOV, *Reconstruction of the coefficients of the stationary transport equation from boundary measurements*, *Inverse Problems*, 12 (1996), pp. L19–L23.
- [15] D. COLTON AND R. KRESS, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer-Verlag, New York, 1998.
- [16] R. DAUTRAY AND J.-L. LIONS, *Mathematical Analysis and Numerical Methods for Science and Technology*, Vol. 6, Springer-Verlag, Berlin, 1993.
- [17] H. DEGHANI, D. T. DELPY, AND S. R. ARRIDGE, *Photon migration in non-scattering tissue and the effects on image reconstruction*, *Phys. Med. Biol.*, 44 (1999), pp. 2879–2906.
- [18] O. DORN, *A transport-backtransport method for optical tomography*, *Inverse Problems*, 14 (1998), pp. 1107–1130.
- [19] O. DORN, *Scattering and absorption transport sensitivity functions for optical tomography*, *Optics Express*, 7 (2000), pp. 492–506.
- [20] R. A. ELLIOTT, T. DURACZ, N. J. MCCORMICK, AND D. R. EMMONS, *Experimental test of a time-dependent inverse radiative transfer algorithm for estimating scattering parameters*, *J. Opt. Soc. Amer. A*, 5 (1988), pp. 366–373.
- [21] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic, Dordrecht, The Netherlands, 1996.
- [22] R. EYMARD, T. GALLOWËT, AND R. HERBIN, *Finite volume methods*, in *Handbook of Numerical Analysis VII*, P. Ciarlet and J. L. Lions, eds., North-Holland, Amsterdam, 2000, pp. 713–1020.
- [23] M. FIRBANK, S. R. ARRIDGE, M. SCHWEIGER, AND D. T. DELPY, *An investigation of light transport through scattering bodies with non-scattering regions*, *Phys. Med. Biol.*, 41 (1996), pp. 767–783.
- [24] L. FUKSHANSKY, N. FUKSHANSKY-KAZARINOVA, AND A. M. V. REMISOWSKY, *Estimation of optical properties in a living tissue by solving the inverse problem of the multiframe radiative transfer*, *Appl. Opt.*, 30 (1991), pp. 3145–3153.
- [25] F. GOLSE, S. JIN, AND C. D. LEVERMORE, *The convergence of numerical transfer schemes in diffusive regimes I: Discrete-ordinate method*, *SIAM J. Numer. Anal.*, 36 (1999), pp. 1333–1369.
- [26] C. GU AND G. WAHBA, *Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method*, *SIAM J. Sci. Statist. Comput.*, 12 (1991), pp. 383–398.
- [27] E. HABER, U. ASCHER, AND D. OLDENBURG, *On optimization techniques for solving nonlinear inverse problems*, *Inverse Problems*, 16 (2000), pp. 1263–1280.
- [28] P. C. HANSEN AND D. P. O’LEARY, *The use of the L-curve in the regularization of discrete ill-posed problems*, *SIAM J. Sci. Comput.*, 14 (1993), pp. 1487–1503.

- [29] L. G. HENYEV AND J. L. GREENSTEIN, *Diffuse radiation in the galaxy*, *Astrophys. J.*, 90 (1941), pp. 70–83.
- [30] A. HIELSCHER, A. BLUESTONE, G. ABDOULAEV, A. KLOSE, J. LASKER, M. STEWART, U. NETZ, AND J. BEUTHAN, *Near-infrared diffuse optical tomography*, *Disease Markers*, 18 (2002), pp. 313–337.
- [31] A. H. HIELSCHER, R. E. ALCOUFFE, AND R. L. BARBOUR, *Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissue*, *Phys. Med. Biol.*, 43 (1998), pp. 1285–1302.
- [32] A. H. HIELSCHER, A. D. KLOSE, A. SCHEEL, B. MOA-ANDERSON, M. BACKHAUS, U. NETZ, AND J. BEUTHAN, *Sagittal laser optical tomography for imaging of rheumatoid finger joints*, *Phys. Med. Biology*, 49 (2004), pp. 1147–1163.
- [33] V. ISAKOV, *Inverse Problems for Partial Differential Equations*, Springer-Verlag, New York, 1998.
- [34] H. JIANG, K. D. PAULSEN, U. L. ÖSTERBERG, B. W. POGUE, AND M. S. PATTERSON, *Simultaneous reconstruction of optical absorption and scattering maps turbid media from near-infrared frequency-domain data*, *Opt. Lett.*, 20 (1995), pp. 2128–2130.
- [35] C. T. KELLEY, *Iterative Methods for Optimization*, *Frontiers Appl. Math.* 18, SIAM, Philadelphia, 1999.
- [36] A. KIENLE, F. K. FORSTER, AND R. HIBST, *Influence of the phase function on determination of the optical properties of biological tissue by spatially resolved reflectance*, *Opt. Lett.*, 26 (2001), pp. 1571–1573.
- [37] A. D. KIM AND J. B. KELLER, *Light propagation in biological tissue*, *J. Opt. Soc. Amer. A*, 20 (2003), pp. 92–98.
- [38] A. D. KLOSE, U. NETZ, J. BEUTHAN, AND A. H. HIELSCHER, *Optical tomography using the time-independent equation of radiative transfer. Part 1: Forward model*, *J. Quant. Spectrosc. Radiat. Transfer*, 72 (2002), pp. 691–713.
- [39] A. D. KLOSE AND A. H. HIELSCHER, *Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer*, *Med. Phys.*, 26 (1999), pp. 1698–1707.
- [40] A. D. KLOSE AND A. H. HIELSCHER, *Optical tomography using the time-independent equation of radiative transfer. Part 2: Inverse model*, *J. Quant. Spectrosc. Radiat. Transfer*, 72 (2002), pp. 715–720.
- [41] A. D. KLOSE AND A. H. HIELSCHER, *Quasi-Newton methods in optical tomographic image reconstruction*, *Inverse Problems*, 19 (2003), pp. 387–409.
- [42] E. W. LARSEN, *Solution of the inverse problem in multigroup transport theory*, *J. Math. Phys.*, 22 (1981), pp. 158–160.
- [43] E. W. LARSEN, *Solution of three dimensional inverse transport problems*, *Transport Theory Statist. Phys.*, 17 (1988), pp. 147–167.
- [44] E. E. LEWIS AND W. F. MILLER, *Computational Methods of Neutron Transport*, American Nuclear Society, La Grange Park, IL, 1993.
- [45] A. LI, E. MILLER, M. KILMER, T. BRUKILACCHIO, T. CHAVES, J. STOTT, Q. ZHANG, T. WU, M. CHORLTON, R. MOORE, D. KOPANS, AND D. BOAS, *Tomographic optical breast imaging guided by 3-D mammography*, *Appl. Opt.*, 42 (2003), pp. 5181–5190.
- [46] E. H. LIEB AND M. LOSS, *Analysis*, 2nd ed., AMS, Providence, RI, 2001.
- [47] V. MARKEL AND J. SCHOTLAND, *Effects of sampling and limited data in optical tomography*, *App. Phys. Lett.*, 81 (2002), pp. 1180–1182.
- [48] T. O. MCBRIDE, B. W. POGUE, U. L. ÖSTERBERG, AND K. D. PAULSEN, *Separation of absorption and scattering heterogeneities in NIR tomographic imaging of tissue*, in *OSA Technical Digest, Biomedical Topical Meetings, Optical Society of America, Washington, DC*, 2000.
- [49] N. J. McCORMICK, *Recent developments in inverse scattering transport methods*, *Transport Theory Statist. Phys.*, 13 (1984), pp. 15–28.
- [50] N. J. McCORMICK, *Inverse radiative transfer problems: A review*, *Nuclear Sci. Engrg.*, 112 (1992), pp. 185–198.
- [51] M. MOKHTAR-KHARROUBI, ED., *Mathematical Topics in Neutron Transport Theory: New Aspects*, Ser. Adv. Math. Appl. Sci. 46, World Scientific, Singapore, 1997.
- [52] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer-Verlag, New York, 1999.
- [53] B. W. PATTON AND J. P. HOLLOWAY, *Application of preconditioned GMRES to the numerical solution of the neutron transport equation*, *Ann. Nuclear Energy*, 29 (2002), pp. 109–136.
- [54] K. REN, G. S. ABDOULAEV, G. BAL, AND A. H. HIELSCHER, *Algorithm for solving the equation of radiative transfer in the frequency domain*, *Opt. Lett.*, 29 (2004), pp. 578–580.
- [55] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
- [56] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, *SIAM J. Sci. Statist. Comput.*, 7 (1986), pp. 856–869.

- [57] J. C. SCHOTLAND, *Continuous-wave diffusion imaging*, J. Opt. Soc. Amer. A, 14 (1997), pp. 275–279.
- [58] B. J. TROMBERG, N. SHAH, R. LANNING, A. CERUSSI, J. ESPINOZA, T. PHAM, L. SVAASAND, AND J. BUTLER, *Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy*, Neoplasia, 2 (2000), pp. 26–40.
- [59] C. R. VOGEL, *Non-convergence of the L-curve regularization parameter selection method*, Inverse Problems, 12 (1996), pp. 535–547.
- [60] C. R. VOGEL, *Computational Methods for Inverse Problems*, Frontiers Appl. Math. 23, SIAM, Philadelphia, 2002.
- [61] A. P. WANG AND S. UENO, *An inverse problem in a three-dimensional radiative transfer*, Astrophys. Space Sci., 155 (1989), pp. 105–111.
- [62] A. J. WELCH AND M. J. C. VAN-GEMERT, *Optical-Thermal Response of Laser Irradiated Tissue*, Plenum Press, New York, 1995.
- [63] A. G. YODH AND B. CHANCE, *Spectroscopy and imaging with diffusing light*, Phys. Today, 48 (1995), pp. 34–40.
- [64] C. ZHU, R. H. BYRD, P. LU, AND J. NOCEDAL, *L-BFGS-B-FORTRAN subroutines for large-scale bound constrained optimization*, ACM Trans. Math. Software, 23 (1997), pp. 550–560.