

REVIEW ARTICLE

Recent Developments in Numerical Techniques for Transport-Based Medical Imaging Methods

Kui Ren*

*Department of Mathematics, University of Texas at Austin, 1 University Station
C1200, Austin, TX 78712, USA.*

Received 22 May 2009; Accepted (in revised version) 20 January 2010

Available online 12 February 2010

Abstract. The objective of this paper is to review recent developments in numerical reconstruction methods for inverse transport problems in imaging applications, mainly optical tomography, fluorescence tomography and bioluminescence tomography. In those inverse problems, one aims at reconstructing physical parameters, such as the absorption coefficient, the scattering coefficient and the fluorescence light source, inside heterogeneous media, from partial knowledge of transport solutions on the boundaries of the media. The physical parameters recovered can be used for diagnostic purpose. Numerical reconstruction techniques for those inverse transport problems can be roughly classified into two categories: linear reconstruction methods and nonlinear reconstruction methods. In the first type of methods, the inverse problems are linearized around some known background to obtain linear inverse problems. Classical regularization techniques are then applied to solve those inverse problems. The second type of methods are either based on regularized nonlinear least-square techniques or based on gradient-driven iterative methods for nonlinear operator equations. In either case, the unknown parameters are iteratively updated until the solutions of the transport equations with the those parameters match the measurements to a certain extent.

We review linear and nonlinear reconstruction methods for inverse transport problems in medical imaging with stationary, frequency-domain and time-dependent data. The materials presented include both existing and new results. Meanwhile, we attempt to present similar algorithms for different problems in the same framework to make it more straightforward to generalize those algorithms to other inverse (transport) problems.

AMS subject classifications: 35Q20, 35R30, 65N21, 65Z05, 78A46

Key words: Inverse problem, radiative transport equation, optical tomography, bioluminescence tomography, fluorescence tomography, iterative reconstruction.

*Corresponding author. *Email address:* ren@math.utexas.edu (K. Ren)

Contents

1	Introduction	2
2	Inverse transport problems in medical imaging	4
3	Linearized reconstruction methods	12
4	Nonlinear iterative reconstructions	19
5	Techniques for features reconstructions	37
6	Summary and further remarks	40

1 Introduction

Inverse problems related to the radiative transport equation have been extensively studied in recent years; see for example the reviews [14, 88–90]. In those inverse problems, the objective is to reconstruct some of the physical parameters in the transport equation inside spatial domains from partial measurements of the transport solutions on the boundaries of the domains. Apart from their applications in medical imaging, which we will address below, inverse transport problems emerge from many other areas of scientific research, including intensity modulated radiation therapy (IMRT) [8, 24], imaging in random media [15, 18], semiconductor design [31], remote sensing [6, 17, 19, 50, 103, 113–115, 119, 125, 126, 132], reactor physics [3, 82, 83] as well as ocean and atmospheric optics [29, 46, 91, 118]. For readers interested in more applications, we refer to the references [20–22, 26–28, 102, 103, 127] and the references cited there.

In this paper, we are interested in the application of inverse transport problems in medical imaging. Three major applications that we will focus on are the fields of diffuse optical tomography (DOT), fluorescence tomography (FT) and bioluminescence tomography (BLT).

Diffuse optical tomography is a biomedical imaging modality that utilizes diffuse light as a probe of tissue structure and function [9]. In diffuse optical tomography, near infra-red light are sent into biological tissues. The outgoing photon current at the surfaces of the tissues are then measured. We then want to infer the optical properties of the tissues from the knowledge of those measurements. These optical properties can be used for diagnostic purposes. Applications of optical tomography include, but not limited to, brain [40], breast [85] and joint imaging [97, 104]. We refer interested reader to [9, 11, 55, 65] for recent developments on theoretical and experimental aspects of diffusion optical tomography.

In optical molecular imaging such as fluorescence and bioluminescence tomography [81], we seek to determine the spatial concentration distribution of biological light sources inside tissues from measurements of the light current on the surface of the tissue. The light sources can come from either the fluorescent biochemical markers that we injected into the biological object, or bioluminescent cells of the object. In the former case, the markers have to be excited by an external light source, while in the later case, the cells

emit light automatically. The two imaging strategies are called fluorescence tomography and bioluminescence tomography, respectively. The distribution information of either light source will serve as diagnostic tools to differentiate healthy and cancerous tissues. We refer interested reader to [32,38,58,60,68,81,99,117] and references therein for recent developments on optical molecular imaging.

In both diffuse optical tomography and optical molecular imaging, the law that governs the transport of near-infrared light in tissues, denoted by $u(\mathbf{x},\theta)$ (or $u(t,\mathbf{x},\theta)$ in time-dependent case), is the radiative transfer equation that we will introduce in next section; see, for example, Eq. (2.1) below. Optical properties of the tissues, for example the absorption coefficient (Σ_a in Eq. (2.1) below), the scattering coefficient (Σ_s in Eq. (2.1) below), and the bioluminescence source ($g(\mathbf{x})$ in Eq. (2.8) below) appear as parameters in the corresponding transport equation. The objective of diffuse optical tomography is thus to reconstruct the absorption and scattering coefficients from boundary measurement of photon current, and that of optical molecular imaging is to reconstruct the distribution of the fluorescence and bioluminescence sources from boundary measurement. Both problems are then formulated as inverse transport problems [12,13,34,35,42–44,73,120,122].

Let us mention that the advantage of optical imaging over traditional imaging modalities, such as X-ray imaging, is that optical imaging is non-invasive and it provides functional (rather than anatomical) information. Optical imaging devices are also less expensive in general. However, unlike in X-ray imaging, there is, in general, no analytical reconstruction formulas available for optical imaging, except in very specific geometrical settings such as those in [111]. We thus have to rely on numerical computations in most cases. At the same time, the inverse transport problems in optical tomography and optical molecular imaging are in general ill-posed, or even severely ill-posed sometimes, in the sense that assuming uniqueness holds, the stabilities of the reconstructions are very poor. The resolution of optical methods is thus not comparable to traditional X-ray imaging.

Numerical reconstruction techniques for inverse transport problems fall roughly into two categories: linear reconstruction methods and nonlinear reconstruction methods. In the first type of methods, the inverse problems are linearized around some known background states to obtain linear inverse problems that are of the form (3.6) below. Classical regularization techniques are then applied to solve those inverse problems. Let us mention that unlike in many inverse problems for partial differential equations, where the Green's function for the homogeneous background is known analytically, analytical Green's functions are almost never available for even very simple settings in transport theory. Those Green's functions that are needed have to be evaluated numerically to construct the forward map in the linear inverse problems. So linearization methods in inverse transport are also computationally expensive. This fact will be made clear in the following sections.

The second type of reconstruct methods works directly on the nonlinear inverse problems. Those methods are iterative in nature. In most cases, the inverse problems are formulated as the problems of minimizing the discrepancy between predictions by the

transport model and the measurements. Iterative methods, such as those of Newton's type, are then applied to solve those nonlinear optimization problems. Other iterative methods based on local linearization have also been introduced. In general, iterative reconstruction schemes are even more computationally expensive, not to mention that they converge only locally despite of globalization strategies, such as line search, often employed.

The objective of this paper is to review recent developments in numerical methods for inverse transport problems in medical imaging applications. Before we present details, let us remark that the reconstruction methods we discuss in the following sections can be applied to almost any kinds of discretization on the radiative transport equation. In other words, the inversion methods are independent of how accurate the forward problem are discretized numerically, although the quality of the reconstructions will certainly depend on how the accurate the forward problems are solved; see more discussion in Section 6. We will thus not review many results in transport discretizations but refer, in a very subjective way, interested readers to the references [4, 5, 75, 76, 84, 87, 106], knowing that the list is by no means complete. Let us also mention that since the radiative transport equation is posed in phase space, involving both spatial and angular variables, any reconstruction method, as long as it requires the numerical solution of the transport equation, will be extremely expensive.

The rest of the paper is organized as follows. In Section 2 we introduce briefly the transport models for optical tomography and optical molecular imaging. We also recall various types of measurements available in the community. We then review the basic procedure of linearized reconstruction in Section 3. Nonlinear iterative reconstruction schemes, including quasi-Newton methods such as Gauss-Newton, BFGS, Levenberg-Marquardt algorithms, that are based on least-square formulations and other methods such as the nonlinear Kaczmarz method that are not based least-square formulation, are presented in Section 4. In Section 5 we review some special methods used to reconstruct features in the unknowns. A summary and some concluding remarks are provided in Section 6.

2 Inverse transport problems in medical imaging

We review in this section the formulation of the inverse transport problems in optical tomography, fluorescence tomography and bioluminescence tomography, three major applications of the inverse radiative transport problems. We need the following notations. We denote by $\Omega \subset \mathbb{R}^n$ ($n = 2, 3$) the spatial domain of interest, with sufficiently regular boundary $\partial\Omega$. Although practical applications are all posed in three dimensional space ($n = 3$), inverse transport algorithms that we will describe are in general independent of spatial dimensions. We denote by S^{n-1} the unit sphere in \mathbb{R}^n , the space of light propagation directions, and $\theta \in S^{n-1}$ the unit vector on the sphere. We then denote by $\mathcal{X} \equiv \Omega \times S^{n-1}$ the phase space in which the radiative transport equation is posed. We

define the boundary sets of the phase space, Γ_{\pm} , as

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

with $\boldsymbol{\nu}(\mathbf{x})$ the unit outer normal vector at $\mathbf{x} \in \partial\Omega$. The measure one can then introduce on Γ_{\pm} is $d\zeta = |\boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x})| d\sigma(\mathbf{x}) d\boldsymbol{\theta}$ with $d\sigma(\mathbf{x})$ the surface measure on $\partial\Omega$.

2.1 Diffuse optical tomography

In diffuse optical tomography, near infrared lights are sent into tissues to be probed. The propagation of NIR light in tissues is governed by the following phase space radiative transport equation for photon density $u(\mathbf{x}, \boldsymbol{\theta})$ [9]:

$$\begin{aligned} \boldsymbol{\theta} \cdot \nabla u(\mathbf{x}, \boldsymbol{\theta}) + \Sigma(\mathbf{x})u(\mathbf{x}, \boldsymbol{\theta}) &= \Sigma_s(\mathbf{x})K(u)(\mathbf{x}, \boldsymbol{\theta}) & \text{in } \mathcal{X}, \\ u(\mathbf{x}, \boldsymbol{\theta}) &= g(\mathbf{x}, \boldsymbol{\theta}) & \text{on } \Gamma_-. \end{aligned} \quad (2.1)$$

Here the non-negative functions $0 < \Sigma(\mathbf{x}), \Sigma_s(\mathbf{x}) \in L^\infty(\Omega)$ are the total absorption coefficient and the scattering coefficients, respectively. The physical absorption coefficient is given by

$$\Sigma_a(\mathbf{x}) \equiv \Sigma(\mathbf{x}) - \Sigma_s(\mathbf{x}).$$

The function $g(\mathbf{x}, \boldsymbol{\theta}) \in L^1(\Gamma_-, d\zeta)$ is the incoming light source. The scattering operator K is defined as

$$K(u)(\mathbf{x}, \boldsymbol{\theta}) = \int_{S^{n-1}} \mathcal{K}(\boldsymbol{\theta} \cdot \boldsymbol{\theta}') u(\mathbf{x}, \boldsymbol{\theta}') d\boldsymbol{\theta}',$$

where the kernel $\mathcal{K}(\boldsymbol{\theta} \cdot \boldsymbol{\theta}')$ describes the way that photons traveling in direction $\boldsymbol{\theta}'$ getting scattered into direction $\boldsymbol{\theta}$. The normalization condition

$$\int_{S^{n-1}} \mathcal{K}(\boldsymbol{\theta} \cdot \boldsymbol{\theta}') d\boldsymbol{\theta}' = 1, \quad \forall \boldsymbol{\theta} \in S^{n-1}$$

should hold. In practical applications in biomedical optics, \mathcal{K} is often taken to be the Henyey-Greenstein phase function [9, 107]. Also, physical absorption always exists in biological tissues, so $\Sigma_a(\mathbf{x}) > 0$ (thus $\Sigma(\mathbf{x}) > \Sigma_s(\mathbf{x})$).

Measurements in optical tomography experiments are generally taken on the boundaries of the interested regions. We thus introduce the measurement operator, a bounded linear functional \mathcal{M} , acting on $u(\mathbf{x}, \boldsymbol{\theta})$. In typical optical fiber based measurement, the measurement operator is defined as [9]

$$(\mathcal{M}_1 u)(\mathbf{x}) \equiv \int_{S^{n-1}} \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) u|_{\Gamma_+} d\boldsymbol{\theta}. \quad (2.2)$$

This is nothing but the outgoing current of photons on the boundary of the domain. In CCD cameras based measurement, the measurement operator is defined as

$$(\mathcal{M}_2 u)(\mathbf{x}, \boldsymbol{\theta}) \equiv u(\mathbf{x}, \boldsymbol{\theta})|_{\Gamma_+}. \quad (2.3)$$

This is just the solution of the transport equation on the boundary of the domain in outgoing directions. It is clear that data in this type of measurement is more rich than those in (2.2) since angularly dependent information are collected.

The goal of optical tomography is thus to reconstruct the optical properties of tissues, the functions $\Sigma(\mathbf{x})$ and $\Sigma_s(\mathbf{x})$, from the full knowledge of the map $\Lambda: g \mapsto \mathcal{M}u$. In practice, however, only a finite source-detector pairs can be used in the measurement process. We thus have only partial information on the map Λ . $\{G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d)\}_{d=1}^{N_d}$ Let us denote by N_q the number of sources, N_d the number of detectors that can be used in an optical tomography experiment and N_k the number of directions in which measurements can be taken. Then the total data set we have is

$$\left\{ g_q, \left\{ z_{q,d} \right\}_{d=1}^{N_d} \right\}_{q=1}^{N_q} \quad \text{or} \quad \left\{ g_q, \left\{ \left\{ z_{q,d,k} \right\}_{k=1}^{N_k} \right\}_{d=1}^{N_d} \right\}_{q=1}^{N_q}, \quad (2.4)$$

depending on which measurement type we have. Here g_q is the q -th source, $z_{q,d} = (\mathcal{M}_1 u_q)(\mathbf{x}_d)$ is the measurement corresponding to the d -th detector and

$$z_{q,d,k} = (\mathcal{M}_2 u_q)(\mathbf{x}_d, \boldsymbol{\theta}_k).$$

The inverse transport problem in optical tomography can now be formulated formally as:

Problem 2.1. To reconstruct the functions $\Sigma_a(\mathbf{x})$ and $\Sigma_s(\mathbf{x})$ in the radiative transport equation (2.1) from the data set (2.4).

The numerical methods we will present in Section 3 depend slightly on what type of measurements are taken. The nonlinear reconstruction methods in Section 4, however, will be almost independent of the measurement operator. We will make those points more clear later.

There is extensive literature on optical tomography with the radiative transport equation. We refer interested readers to [14, 42, 56, 64, 80, 107, 111, 124] and the references cited there.

2.2 Fluorescence tomography

Fluorescence tomography (FT) is a molecular imaging technique in which fluorescent biochemical markers are injected into biological objects. The markers will then accumulate on target tissues and emit near-infrared light (at wavelength λ^m) upon excitation by an external light source (at a different wavelength which we denote by λ^x). Both the propagation of the external light source and the fluorescent light in the tissues are described by the radiative transport equation. So we have a system of two radiative transport equations in this case:

$$\begin{aligned} \boldsymbol{\theta} \cdot \nabla u^x(\mathbf{x}, \boldsymbol{\theta}) + \Sigma^x(\mathbf{x}) u^x(\mathbf{x}, \boldsymbol{\theta}) &= \Sigma_s^x K(u^x)(\mathbf{x}, \boldsymbol{\theta}) && \text{in } \mathcal{X}, \\ \boldsymbol{\theta} \cdot \nabla u^m(\mathbf{x}, \boldsymbol{\theta}) + \Sigma^m(\mathbf{x}) u^m(\mathbf{x}, \boldsymbol{\theta}) &= \Sigma_s^m K(u^m)(\mathbf{x}, \boldsymbol{\theta}) + \eta \Sigma_{af}^x(\mathbf{x}) E(u^x)(\mathbf{x}) && \text{in } \mathcal{X}, \\ u^x(\mathbf{x}, \boldsymbol{\theta}) &= g^x(\mathbf{x}, \boldsymbol{\theta}), \quad u^m(\mathbf{x}, \boldsymbol{\theta}) = 0 && \text{on } \Gamma_-. \end{aligned} \quad (2.5)$$

Here the superscripts x and m denote the quantities at the excitation wavelength and emission wavelength, respectively. The total absorption coefficients are

$$\Sigma^{x,m} = \Sigma_{ai}^{x,m} + \Sigma_{af}^{x,m} + \Sigma_s^{x,m}.$$

Here $\Sigma_{ai}^{x,m}$ and $\Sigma_{af}^{x,m}$ are absorption due to chromophores and fluorophores, respectively. $E(u^x)$ is the average of $u^x(\mathbf{x}, \boldsymbol{\theta})$ over direction variable $\boldsymbol{\theta}$. In other words, the operator E is defined as

$$E(u^x)(\mathbf{x}) = \int_{S^{n-1}} u^x(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (2.6)$$

In fluorescence imaging, the coefficients Σ_{ai}^m , Σ_{af}^m and $\Sigma_s^{x,m}$ are all assumed to be known already by other techniques (such as optical tomography). So the main interest is to reconstruct the coefficients Σ_{ai}^x and Σ_{af}^x (and thus the emission light source $\Sigma_{af}^x(\mathbf{x})E(u^x)(\mathbf{x})$). The data measured is the map $\Lambda: g^x \mapsto (\mathcal{M}u^x, \mathcal{M}u^m)$ with the measurement operator \mathcal{M} given by (2.2) or (2.3). In practice, the total data set we have is

$$\left\{ g_q^x, \left\{ z_{q,d}^x, z_{q,d}^m \right\}_{d=1}^{N_d} \right\}_{q=1}^{N_q} \quad \text{or} \quad \left\{ g_q^x, \left\{ \left\{ z_{q,d,k}^x, z_{q,d,k}^m \right\}_{k=1}^{N_k} \right\}_{d=1}^{N_d} \right\}_{q=1}^{N_q}, \quad (2.7)$$

where again g_q^x is the q -th excitation source,

$$z_{q,d}^{x,m} = (\mathcal{M}u_q^{x,m})(\mathbf{x}_d), \quad z_{q,d,k}^{x,m} = (\mathcal{M}u_q^{x,m})(\mathbf{x}_d, \boldsymbol{\theta}_k)$$

are measurement taken at wave length $\lambda^{x,m}$. Now the fluorescence inverse problem can be posed as:

Problem 2.2. To reconstruct the two absorption coefficients $\Sigma_{af}^x(\mathbf{x})$ and $\Sigma_{ai}^x(\mathbf{x})$ in the system (2.5) from the data set (2.7).

Note that in the fluorescence tomography problem, since the first transport equation only involves quantities at the excitation wavelength, the problem is reduced to optical tomography problem if we only have measurements at the excitation wavelength. If we only measure at the emission wavelength, we can at most reconstruct the product $\eta \Sigma_{af}^x(\mathbf{x})E(u^x)$, not $\Sigma_{af}^x(\mathbf{x})$ since $E(u^x)$ is not known. We thus have to measure at both wavelengths.

For recent development in fluorescence tomography, we refer interested readers to [69, 79, 105, 116, 117] and references therein.

2.3 Bioluminescence tomography

Bioluminescence tomography (BLT) is another optical molecular imaging technique in which we attempts to recover the distribution of light sources of bioluminescent cells. Since those bioluminescence cells emit light automatically, there is no need to have excitation light in this case. The same radiative transport equation can be used to describe the

propagation of bioluminescent photons in tissues, and since we do not need excitation, the non-incoming boundary condition has to be imposed. The equation reads

$$\begin{aligned} \boldsymbol{\theta} \cdot \nabla u(\mathbf{x}, \boldsymbol{\theta}) + \Sigma(\mathbf{x})u(\mathbf{x}, \boldsymbol{\theta}) &= \Sigma_s(\mathbf{x})K(u)(\mathbf{x}, \boldsymbol{\theta}) + g(\mathbf{x}), \\ u(\mathbf{x}, \boldsymbol{\theta}) &= 0, \quad (\mathbf{x}, \boldsymbol{\theta}) \in \Gamma_-, \end{aligned} \quad (2.8)$$

where both the scattering and absorption coefficients are now assumed to be known. The measurement now is the solution on the boundary of the domain, $h(\mathbf{x}, \boldsymbol{\theta}) = u(\mathbf{x}, \boldsymbol{\theta})|_{\Gamma_+}$, or in practice

$$\left\{ \left\{ z_{d,k} = h(\mathbf{x}_d, \boldsymbol{\theta}_k) \right\}_{k=1}^{N_k} \right\}_{d=1}^{N_d}, \quad (2.9)$$

where again N_k is the total number of directions at which we can measure the out-going photon densities.

The inverse transport problem in bioluminescence tomography can now be formulated formally as:

Problem 2.3. To reconstruct the bioluminescence source function $g(\mathbf{x})$ in equation (2.8) from the data set (2.9).

As we have mentioned above, what makes bioluminescence tomography unique is the fact that since no excitation is needed, there is very little background autofluorescence. So we can obtain a better signal to noise ratio in this case than in the case of fluorescence tomography. This feature makes bioluminescence tomography a promising imaging technique in probing molecular and cellular processes.

We refer interested reader to references [30, 36, 37, 61, 128] for recent advances in theoretical and practical studies on bioluminescence imaging. Note that most of the work on model-based bioluminescence imaging are based on the diffusion equation, not the radiative transport equation that we focus on in this review.

Let us finally remark that bioluminescence tomography is a linear inverse source problem for the transport equation. The fluorescence tomography is also an inverse source problem (for wavelength λ^m) but is coupled with an optical tomography problem for the coefficients (for wavelength λ^x). Fluorescence tomography is thus a non-linear inverse problem. Apart from fluorescence and bioluminescence tomography, inverse source problems for the radiative transport equation have also applications in other fields; see for example [22, 66, 86, 101, 110, 115].

2.4 Miscellaneous extensions

The optical tomography, fluorescence tomography and bioluminescence tomography problems we described in the previous sections can be generalized when different data types are available in practice.

The first generalization is to consider time-dependent case. In this case, the term $\frac{1}{c} \frac{\partial u}{\partial t}$ is added to the transport equations, with c denoting the speed of light in tissues. One

can thus measure the time evolution of the photon currents on the boundary of the domain. In terms of the notation before, the measurement we have for optical tomography is still (2.4) except that each data point $z_{q,d}$ (or $z_{q,d,k}(t)$) is now a function of $t \in (0, t_{\max})$:

$$\left\{ g_q(t), \{ z_{q,d}(t) \}_{d=1}^{N_d} \right\}_{q=1}^{N_q}, \quad \text{or} \quad \left\{ g_q(t), \{ \{ z_{q,d,k}(t) \}_{k=1}^{N_k} \}_{d=1}^{N_d} \right\}_{q=1}^{N_q}. \quad (2.10)$$

The time-dependent data for fluorescence tomography is now

$$\left\{ g_q^x(t), \{ z_{q,d}^x(t), z_{q,d}^m(t) \}_{d=1}^{N_d} \right\}_{q=1}^{N_q}, \quad \text{or} \quad \left\{ g_q^x(t), \{ \{ z_{q,d,k}^x(t), z_{q,d}^m(t) \}_{k=1}^{N_k} \}_{d=1}^{N_d} \right\}_{q=1}^{N_q}. \quad (2.11)$$

Reconstructions with time-dependent measurement for both optical tomography and fluorescence tomography have been investigated in the past. Most of those investigations use diffusion equation as their model of photon propagation [116, 117].

Frequency domain measurement is a cheaper alternative to the expensive time-dependent measurement. Essentially, this is Fourier domain technique. The intensity of the incoming source is modulated to some specific frequencies, so that we have the so called propagation of photon density waves. Mathematically, this is just a Fourier transform of the time-dependent problem. So we just need to replace the term $\frac{1}{c} \frac{\partial u}{\partial t}$ by $\frac{i\omega u}{c}$ with ω the modulation frequency used. Ideally if we can measure data for all ω , we can do an inverse Fourier transform to get time-dependent measurement. In practice, one can only measure data for a few (say, N_ω) frequencies. The data available in this case for optical tomography is thus, assuming measurement of type (2.2),

$$\left\{ g_q(\omega), \{ z_{q,d}(\omega) \}_{d=1}^{N_d} \right\}_{q=1}^{N_q}, \quad \omega \in \{\omega_1, \dots, \omega_{N_\omega}\}. \quad (2.12)$$

The same type of frequency-domain measurements can be realized in fluorescence tomography. Reconstructions with frequency-domain measurement for both optical tomography and fluorescence tomography have been investigated in the past; see for example [69, 107].

A recent attempt to improve bioluminescence tomography is to introduce multiple spectral measurements. The technique is thus called multispectral bioluminescence tomography [39, 61, 129]. Here it is assumed that one can measure light at different wavelengths. The transport equation in this case is

$$\begin{aligned} \boldsymbol{\theta} \cdot \nabla u(\mathbf{x}, \boldsymbol{\theta}, \lambda) + \Sigma(\mathbf{x}, \lambda) u(\mathbf{x}, \boldsymbol{\theta}, \lambda) &= \Sigma_s(\mathbf{x}, \lambda) K(u)(\mathbf{x}, \boldsymbol{\theta}, \lambda) + g(\mathbf{x}, \lambda), \\ u(\mathbf{x}, \boldsymbol{\theta}, \lambda) &= 0, \quad (\mathbf{x}, \boldsymbol{\theta}) \in \Gamma_-. \end{aligned} \quad (2.13)$$

The measured data is again $h(\mathbf{x}, \boldsymbol{\theta}, \lambda) = u(\mathbf{x}, \boldsymbol{\theta}, \lambda)|_{\Gamma_+}$. In practice, we can measure for several (say, N_λ) different λ , so the data set we have is

$$\left\{ \{ z_{d,k}(\lambda) = h(\mathbf{x}_d, \boldsymbol{\theta}_k, \lambda) \}_{k=1}^{N_k} \right\}_{d=1}^{N_d}, \quad \lambda \in \{\lambda_1, \dots, \lambda_{N_\lambda}\}. \quad (2.14)$$

It is obvious that although we enriched our data set by measuring at different wavelength, we also introduce an extra variable for the unknown function ($g(\mathbf{x})$ becomes $g(\mathbf{x}, \lambda)$). Now if there is no *a priori* information on the source $g(\mathbf{x}, \lambda)$, we are back to the usual bioluminescence problem for each wavelength. In fact, it is assumed in multispectral bioluminescence tomography that $g(\mathbf{x}, \lambda_i) = w(\lambda_i) \tilde{g}(\mathbf{x})$ with $w(\lambda_i)$ known and satisfies $\sum_i^{N_\lambda} w(\lambda_i) = 1$. Thus the multispectral data will help us to average out noises in the reconstruction of $\tilde{g}(\mathbf{x})$. We remark that the multispectral reconstruction can be roughly decomposed into N_λ single spectral reconstruction problem (i.e., to reconstruct $\{g(\mathbf{x}, \lambda_i)\}_{i=1}^{N_\lambda}$) plus a data processing problem (i.e., to recover $\tilde{g}(\mathbf{x})$ from $\{g(\mathbf{x}, \lambda_i)\}_{i=1}^{N_\lambda}$). So we will not cover the topic in the following sections.

2.5 Remarks on uniqueness and stability

Inverse transport problems in optical imaging have been extensively analyzed mathematically in recent years [16, 33–35, 121–123, 130]. For a very complete review on the subject, we refer to the reference [14] where uniqueness and stability results on various types of inverse transport problems are presented in more general settings than the cases we considered here. Here we only briefly mention the following results on a formal level, without being mathematically precise. To ensure the well-posedness of the forward transport problem (in appropriate function spaces, depending on the regularity of the source functions), we assume that

$$0 < \Sigma_a(\mathbf{x}), \Sigma_s(\mathbf{x}) \in L^\infty(\Omega), \quad \text{so that} \quad \Sigma(\mathbf{x}) \equiv \Sigma_a(\mathbf{x}) + \Sigma_s(\mathbf{x}) > \Sigma_s(\mathbf{x}).$$

We also assume that the scattering kernel $\mathcal{K}(\boldsymbol{\theta} \cdot \boldsymbol{\theta}')$ (> 0 a.e.) is regular enough (say, \mathcal{K} belongs to $L^1(S^{n-1})$) with respect to both $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, which is clearly true for the Henyey-Greenstein phase function that is commonly used.

For the inverse coefficient problem in optical tomography, it is shown [14] that, assuming \mathcal{K} is known:

(i) If we measure time-dependent, angularly-resolved data,

$$\Lambda : g(t, \mathbf{x}, \boldsymbol{\theta}) \mapsto u(t, \mathbf{x}, \boldsymbol{\theta})|_{\Gamma_+},$$

then both $\Sigma_a(\mathbf{x})$ and $\Sigma_s(\mathbf{x})$ can be uniquely and stably reconstructed in dimension $n \geq 2$;

(ii) If we measure time-independent, angularly-resolved data,

$$\Lambda : g(\mathbf{x}, \boldsymbol{\theta}) \mapsto u(\mathbf{x}, \boldsymbol{\theta})|_{\Gamma_+},$$

then $\Sigma_a(\mathbf{x})$ and $\Sigma_s(\mathbf{x})$ can be uniquely reconstructed when $n = 3$. When $n = 2$, $\Sigma_a(\mathbf{x})$ can be uniquely reconstructed if $\Sigma_s(\mathbf{x})$ is known. Moreover, the reconstruction of Σ_a is stable;

(iii) If we measure time-dependent, angularly-averaged data,

$$\Lambda : |\nu(\mathbf{x}) \cdot \boldsymbol{\theta}| g(t, \mathbf{x}) \mapsto (\mathcal{M}u)(t, \mathbf{x})|_{\partial\Omega},$$

then $\Sigma_a(\mathbf{x})$ and $\Sigma_s(\mathbf{x})$ can be uniquely reconstructed when $n \geq 2$; (iv) If we measure time-independent, angularly-averaged data,

$$\Lambda: |\nu(\mathbf{x}) \cdot \boldsymbol{\theta}| g(\mathbf{x}) \mapsto (\mathcal{M}u)(\mathbf{x})|_{\partial\Omega},$$

only Σ_a can be uniquely reconstructed and the reconstruction is severely ill-posed. For the bioluminescence inverse source problem in Section 2.3, it is shown under some conditions on \mathcal{K} that $g(\mathbf{x})$ can be uniquely reconstructed from measurement of $u(\mathbf{x}, \boldsymbol{\theta})|_{\Gamma_+}$.

These statements, although not mathematically precise, provide a flavor of uniqueness results in optical tomography and bioluminescence tomography. In fact, many of the uniqueness statements above can be generalized to more complicated absorption, scattering coefficients and the kernel \mathcal{K} ; see the review paper [14] for more precise mathematical presentation of the above uniqueness results. Note that these uniqueness results are obtained with infinite number of source detector pairs, i.e., with all possible sources and measure at each point on the boundary. In practice, we never have such measurements since our number of source-detector pairs is always finite (and often small).

For the fluorescence problem in Section 2.2, it is clear that to uniquely reconstruct the product $\eta \Sigma_{af}^x(\mathbf{x})$, we need at least data of the following type:

$$\Lambda: |\nu(\mathbf{x}) \cdot \boldsymbol{\theta}| g(\mathbf{x}) \mapsto (\mathcal{M}u^x(\mathbf{x})|_{\partial\Omega}, u^m(\mathbf{x}, \boldsymbol{\theta})|_{\Gamma_+}).$$

With these data, we can reconstruct uniquely $\Sigma_{ai}^x + \Sigma_{af}^x$ and $u^x(\mathbf{x}, \boldsymbol{\theta})$ from the first equation with measurement at wavelength λ^x . The second equation and the measurement at wavelength λ^m would allow us to reconstruct uniquely $\eta \Sigma_{af}^x E(u^x)$. It is then clear that Σ_{ai}^x and Σ_{af}^x can be reconstructed uniquely provided that η is known.

For the multispectral bioluminescence tomography problem (2.13), we can deduce from the above-mentioned results that full measurement at wavelength λ_i will allow us to reconstruct uniquely the quantity

$$g(\mathbf{x}, \lambda_i) = w(\lambda_i) \tilde{g}(\mathbf{x}).$$

If the weight $w(\lambda_i)$ is known, then we can of course obtain uniquely $\tilde{g}(\mathbf{x})$. When noisy measurement are used, we can first perform reconstructions at all N_λ wavelengths to get N_λ estimates of $\tilde{g}(\mathbf{x})$. We then take an average over the N_λ reconstructions to get a better estimate of $\tilde{g}(\mathbf{x})$, $(1/N_\lambda) \sum_{i=1}^{N_\lambda} g(\mathbf{x}, \lambda_i) / w(\lambda_i)$. This is why multispectral measurements are useful. In fact, it is shown in [36, 61] that one can obtain better reconstruction of $\tilde{g}(\mathbf{x})$ even in the case where $\{w(\lambda_i)\}_{i=1}^{N_\lambda}$ are not known exactly.

As we have mentioned above, the stability of reconstructions in optical imaging with angularly averaged measurements is usually very low. In fact, theory shows that the reconstruction in this regime is exponentially unstable [14], very similar to the case of optical imaging with diffusion equations. More detailed analysis [111] shows that the instability mainly happens in the depth direction. In other words, the resolution of diffuse optical tomography in the depth direction is extremely low, while the resolution in the

transverse direction is controlled by sampling. We will not discuss the issue further in the following sections but refer interested readers to the above-mentioned references.

Let us conclude Section 2 by the following two remarks. The traditional imaging techniques such as X-ray tomography uses high-energy so that scattering effect can be neglected [95,96]. Thus, the transport equations involved there for the propagation of X-rays are the free radiative transport equation without the scattering term (i.e., the operator K). The inversion for those transport equations are thus, although still highly nontrivial, significantly simpler (computationally) than the cases we considered in this work; see the reviews in [93,94].

In practical applications, all the data we measured are polluted by noise of various sources. Dealing with noisy data is itself a challenging research topic. In inverse transport community, this issue is often addressed by introducing regularization techniques. In the presentation below, we will assume that the data we used are noisy as well and incorporate simple regularization strategies in the algorithms we present. We will not address the question of how to choose optimal regularization parameters, such as the β in (3.24) below, but refer interested readers to the classical monograph [47] and references therein.

3 Linearized reconstruction methods

We have seen from the previous section that both optical tomography and fluorescence tomography are nonlinear inverse transport problems. For those nonlinear inverse problems, it is often desirable to linearize the problem around some known (not necessarily constant) background to obtain linear inverse problems. In fact, this is often what is done in the field [25,74]. We study the procedure of linearization of nonlinear inverse transport problem in this section. Let us start with optical tomography.

3.1 DOT with stationary measurement

Let us first consider the situation where we have measurement of type (2.2), i.e., assuming that we have data (2.4). Let us assume that the physical absorption coefficient is what we are looking for. We consider the case where $\Sigma_a(\mathbf{x})$ can be written as superposition of a *known* background Σ_a^0 (not necessarily constant) and a perturbation $\tilde{\Sigma}_a(\mathbf{x})$ from the background. In other words,

$$\Sigma_a(\mathbf{x}) = \Sigma_a^0 + \tilde{\Sigma}_a(\mathbf{x}). \quad (3.1)$$

We then linearize the problem around the background Σ_a^0 . We denote by $U_q^0(\mathbf{x}, \boldsymbol{\theta})$ ($1 \leq q \leq N_q$) the solution of the transport equation (2.1) with the known background Σ_a^0 for source g_q . Then $U_q^0(\mathbf{x}, \boldsymbol{\theta})$ solves the transport equation

$$\begin{aligned} \boldsymbol{\theta} \cdot \nabla U_q^0(\mathbf{x}, \boldsymbol{\theta}) + \Sigma_a^0(\mathbf{x}) U_q^0(\mathbf{x}, \boldsymbol{\theta}) &= \Sigma_s(\mathbf{x}) K(U_q^0)(\mathbf{x}, \boldsymbol{\theta}) && \text{in } \mathcal{X}, \\ U_q^0(\mathbf{x}, \boldsymbol{\theta}) &= g_q(\mathbf{x}, \boldsymbol{\theta}) && \text{on } \Gamma_-, \end{aligned} \quad (3.2)$$

with $\Sigma^0(\mathbf{x}) = \Sigma_a^0(\mathbf{x}) + \Sigma_s(\mathbf{x})$, where both $\Sigma_a^0(\mathbf{x})$ and $\Sigma_s(\mathbf{x})$ are assumed known.

The solution of the full problem can then be written as

$$u_q(\mathbf{x}, \boldsymbol{\theta}) = U_q^0(\mathbf{x}, \boldsymbol{\theta}) + \tilde{u}_q(\mathbf{x}, \boldsymbol{\theta}), \quad (3.3)$$

where $\tilde{u}_q(\mathbf{x}, \boldsymbol{\theta})$ is the perturbation in the solution caused by the perturbation $\tilde{\Sigma}_a(\mathbf{x})$ in the coefficient. The equation satisfied by the perturbation $\tilde{u}(\mathbf{x}, \boldsymbol{\theta})$, to the first order, is

$$\begin{aligned} \boldsymbol{\theta} \cdot \nabla \tilde{u}_q(\mathbf{x}, \boldsymbol{\theta}) + \Sigma^0(\mathbf{x}) \tilde{u}_q(\mathbf{x}, \boldsymbol{\theta}) &= \Sigma_s(\mathbf{x}) K(\tilde{u}_q)(\mathbf{x}, \boldsymbol{\theta}) - \tilde{\Sigma}_a(\mathbf{x}) U_q^0(\mathbf{x}, \boldsymbol{\theta}) & \text{in } \mathcal{X}, \\ \tilde{u}_q(\mathbf{x}, \boldsymbol{\theta}) &= 0 & \text{on } \Gamma_-. \end{aligned} \quad (3.4)$$

We now introduce the adjoint (in L^2 sense) Green's function $G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d)$ for the homogeneous problem

$$\begin{aligned} -\boldsymbol{\theta} \cdot \nabla G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) + \Sigma^0(\mathbf{x}) G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) &= \Sigma_s(\mathbf{x}) K(G)(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) & \text{in } \mathcal{X}, \\ G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) &= \delta(\mathbf{x} - \mathbf{x}_d) & \text{on } \Gamma_+, \end{aligned} \quad (3.5)$$

for $\mathbf{x}_d \in \partial\Omega$. Note that since the transport operator is not self-adjoint, the boundary condition is now put on Γ_+ .

If we multiply (3.4) by $G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d)$ and integrate over phase space \mathcal{X} , and multiply (3.5) by $\tilde{u}_q(\mathbf{x}, \boldsymbol{\theta})$ and integrate over the phase space, we obtain

$$\int_{\Omega} \tilde{\Sigma}_a(\mathbf{x}) \left(\int_{S^{n-1}} U_q^0(\mathbf{x}, \boldsymbol{\theta}) G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) d\boldsymbol{\theta} \right) d\mathbf{x} = - \int_{S^{n-1}} \mathbf{n}(\mathbf{x}_d) \cdot \boldsymbol{\theta} \tilde{u}_q(\mathbf{x}_d, \boldsymbol{\theta}) |_{\Gamma_+} d\boldsymbol{\theta}. \quad (3.6)$$

The right hand side of Eq. (3.6) is nothing but the difference between measured data and prediction from the background problem, i.e.,

$$\tilde{z}_{q,d} = - (z_{q,d} - (\mathcal{M}_1 U_q^0 |_{\Gamma_+})(\mathbf{x}_d)),$$

which is now treated as the new data. The kernel for this linear integral equation, $\int_{S^{n-1}} U_q^0 G d\boldsymbol{\theta}$ is known since it only involves transport solution (and Green's function) for the background medium. We have thus constructed a linear map that maps the perturbation of absorption coefficient, $\tilde{\Sigma}_a(\mathbf{x})$, to measured data. The inverse problem can now be solved by inverting the linear map (3.6).

In practice, let us assume that we have a numerical procedure, say a quadrature rule, to discretize the integral equation (3.6), and assume that we discretize $\tilde{\Sigma}_a(\mathbf{x})$ on a mesh of N_{Ω} nodes, $\{\mathbf{y}_k\}_{k=1}^{N_{\Omega}}$. Then, after collecting the discretization for all N_q sources, we obtain a linear system of algebraic equation of the form

$$\mathcal{A} \tilde{\Sigma}_a = \mathcal{Z}, \quad (3.7)$$

with the matrix $\mathcal{A} \in \mathbb{R}^{(N_q N_d) \times N_{\Omega}}$ and the column vector $\mathcal{Z} \in \mathbb{R}^{(N_q N_d) \times 1}$ are of the form

$$\mathcal{A} = [\mathcal{A}_1^{\mathfrak{z}}, \dots, \mathcal{A}_{N_q}^{\mathfrak{z}}]^{\mathfrak{z}}, \quad \mathcal{Z} = [\mathcal{Z}_1^{\mathfrak{z}}, \dots, \mathcal{Z}_{N_q}^{\mathfrak{z}}]^{\mathfrak{z}}, \quad (3.8)$$

with

$$\mathcal{A}_q \in \mathbb{R}^{N_d \times N_\Omega}, \quad \mathcal{Z}_q = [z_{q,1}, \dots, z_{q,N_d}]^{\mathfrak{T}} \in \mathbb{R}^{N_d \times 1}, \quad \leq q \leq N_q.$$

The superscript \mathfrak{T} is used to denote the transpose of a quantity. The $\tilde{\Sigma}_a \in \mathbb{R}^{N_d \times 1}$ now denotes the column vector that contains the value of the function $\tilde{\Sigma}_a$ on the mesh nodes. The elements of the matrix \mathcal{A} is given by

$$(\mathcal{A}_q)_{dk} = \zeta_k \int_{S^{n-1}} U_q^0(\mathbf{y}_k, \boldsymbol{\theta}) G(\mathbf{y}_k, \boldsymbol{\theta}; \mathbf{x}_d) d\boldsymbol{\theta}, \quad (3.9)$$

with ζ_k ($1 \leq k \leq N_\Omega$) the weight of the quadrature on the k th element.

3.2 Generalization to other measurement types

We can generalize the linearization method to problems with other measurements. For example, if instead of considering angularly averaged measurement (2.2) we consider measurement of the form (2.3), we can construct a similar linear problem by following the same procedure as above. The only difference would be to replace the source $\delta(\mathbf{x} - \mathbf{x}_d)$ in the equation for the Green's function, i.e., Eq. (3.5), by $\delta(\mathbf{x} - \mathbf{x}_d)\delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$; see Eq. (3.22) below. Let us denote the Green's function in this case by $G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d, \hat{\boldsymbol{\theta}})$. Then the same procedure would provide us the following linearized problem

$$\int_{\Omega} \tilde{\Sigma}_a(\mathbf{x}) \left(\int_{S^{n-1}} U_q^0(\mathbf{x}, \boldsymbol{\theta}) G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d, \hat{\boldsymbol{\theta}}) d\boldsymbol{\theta} \right) d\mathbf{x} = -\hat{\boldsymbol{\theta}} \cdot \mathbf{n}(\mathbf{x}_d) \tilde{u}_q(\mathbf{x}_d, \hat{\boldsymbol{\theta}}). \quad (3.10)$$

Here again the right hand side is nothing but the measured data (multiplied by the known factor $-\hat{\boldsymbol{\theta}} \cdot \mathbf{n}(\mathbf{x}_d)$ which, when the measurement direction is taken to be $\mathbf{n}(\mathbf{x}_d)$, is just -1). The integral equation (3.10) can be discretized to obtain a linear algebraic system that is very similar to (3.7).

Frequency-domain data (2.12) are often used when both the absorption and the scattering coefficients have to be recovered. In this case, linearization techniques can also be developed. Assuming we want to recover both absorption and scattering perturbations, we can follow the same procedure as before to obtain

$$\begin{aligned} & \int_{\Omega} \tilde{\Sigma}_a(\mathbf{x}) \left(\int_{S^{n-1}} U_q^0(\omega, \mathbf{x}, \boldsymbol{\theta}) G(\omega, \mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) d\boldsymbol{\theta} \right) d\mathbf{x} \\ & + \int_{\Omega} \tilde{\Sigma}_s(\mathbf{x}) \left(\int_{S^{n-1}} (U_q^0 - K(U_q^0)) G(\omega, \mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) d\boldsymbol{\theta} \right) d\mathbf{x} \\ & = - \int_{S^{n-1}} \mathbf{n}(\mathbf{x}_d) \cdot \boldsymbol{\theta} \tilde{u}_q(\omega, \mathbf{x}_d, \boldsymbol{\theta})|_{\Gamma_+} d\boldsymbol{\theta}. \end{aligned} \quad (3.11)$$

where $U_q^0(\omega, \mathbf{x}, \boldsymbol{\theta})$ solves the frequency domain transport equation, i.e., Eq. (3.2) with a term $\frac{i\omega}{c} U_0$ on the left hand side, with background coefficients Σ_a^0 and Σ_s^0 , and $G(\omega, \mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d)$ is the corresponding adjoint Green's function. Now since the integral equation (3.11) is

complex, we can split it into two separate equations for the real and imaginary parts. After discretization, we can obtain system of equations of the form (3.7)

$$\begin{aligned} \mathcal{A}_r \tilde{\Sigma}_a + \mathcal{B}_r \tilde{\Sigma}_s &= \mathcal{Z}_r, \\ \mathcal{A}_i \tilde{\Sigma}_a + \mathcal{B}_i \tilde{\Sigma}_s &= \mathcal{Z}_i. \end{aligned} \quad (3.12)$$

Here the subscripts r and i are used to denote the real and imaginary part, respectively. The matrices \mathcal{A}_r and \mathcal{A}_i come from the discretization of the first integral in (3.11). Matrices \mathcal{B}_r and \mathcal{B}_i come from the discretization of the second integral. It is hoped that when the modulation frequency ω is non-zero, one can solve (3.12) to recover both $\tilde{\Sigma}_a$ and $\tilde{\Sigma}_s$. This separation between the two coefficients is not possible when only stationary angularly-averaged measurements are available.

To generalize the linearization method to time-dependent problems, let us assume that we can have measurements up to time t_{\max} . We thus consider the following initial and boundary value problem for the transport equation

$$\begin{aligned} \frac{\partial u}{\partial t} + \boldsymbol{\theta} \cdot \nabla u(t, \mathbf{x}, \boldsymbol{\theta}) + \Sigma(\mathbf{x})u(t, \mathbf{x}, \boldsymbol{\theta}) &= \Sigma_s(\mathbf{x})K(u)(t, \mathbf{x}, \boldsymbol{\theta}) \quad \text{in } (0, t_{\max}) \times \mathcal{X}, \\ u(0, \mathbf{x}, \boldsymbol{\theta}) &= 0 \quad \text{in } \Omega, \quad u(t, \mathbf{x}, \boldsymbol{\theta}) = g(t, \mathbf{x}, \boldsymbol{\theta}) \quad \text{on } (0, t_{\max}) \times \Gamma_-. \end{aligned} \quad (3.13)$$

Note that it is not crucial to assume the zero initial condition. Follow the procedure in Section 3.1, we obtain, with the same notation as before,

$$\begin{aligned} &\int_{\Omega} \tilde{\Sigma}_a(\mathbf{x}) \left(\int_0^{t_{\max}} \int_{S^{n-1}} U_q^0(t, \mathbf{x}, \boldsymbol{\theta}) G(t, \mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) d\boldsymbol{\theta} dt \right) d\mathbf{x} \\ &= - \int_0^{t_{\max}} \int_{S^{n-1}} \mathbf{n}(\mathbf{x}_d) \cdot \boldsymbol{\theta} \tilde{u}_q(t, \mathbf{x}_d, \boldsymbol{\theta})|_{\Gamma_+} d\boldsymbol{\theta} dt. \end{aligned} \quad (3.14)$$

Here the adjoint Green's function is the solution of

$$\begin{aligned} -\frac{\partial G}{\partial t} - \boldsymbol{\theta} \cdot \nabla G + \Sigma^0(\mathbf{x})G(t, \mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) &= \Sigma_s(\mathbf{x})K(G)(t, \mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) \quad \text{in } (0, t_{\max}) \times \mathcal{X}, \\ G(t_{\max}, \mathbf{x}, \boldsymbol{\theta}) &= 0 \quad \text{in } \mathcal{X}, \quad G(t, \mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) = \delta(\mathbf{x} - \mathbf{x}_d) \quad \text{on } (0, t_{\max}) \times \Gamma_+, \end{aligned} \quad (3.15)$$

which is an evolution equation that starts from $t = t_{\max}$, not $t = 0$.

The linear integral equation (3.14) can be discretized again to obtain a system of linear equation of the same form as (3.7), with the elements of the coefficient matrix given by

$$(\mathcal{A}_q)_{dk} = \tilde{\zeta}_k \int_0^{t_{\max}} \int_{S^{n-1}} U_q^0(t, \mathbf{y}_k, \boldsymbol{\theta}) G(t, \mathbf{y}_k, \boldsymbol{\theta}; \mathbf{x}_d) d\boldsymbol{\theta} dt. \quad (3.16)$$

This indicates that the solution of the linear inverse problem will depend on t_{\max} (since $(\mathcal{A}_q)_{dk}$ is a function of t_{\max}).

3.3 Linearization in fluorescence tomography

The nonlinear fluorescence tomography problem can also be linearized to obtain linear problems around some known background. The procedure is very similar to the linearization we have just presented. However, due to the fact that fluorescence tomography involves a coupled system of two transport equations, extra efforts are needed. With similar notations as before, the equations for the perturbations are

$$\begin{aligned} \boldsymbol{\theta} \cdot \nabla \tilde{u}_q^x + \Sigma^{x,0}(\mathbf{x}) \tilde{u}_q^x &= \Sigma_s^x(\mathbf{x}) K(\tilde{u}_q^x) - (\tilde{\Sigma}_{ai}^x + \tilde{\Sigma}_{af}^x) U_q^{x,0} && \text{in } \mathcal{X}, \\ \boldsymbol{\theta} \cdot \nabla \tilde{u}_q^m + \Sigma^m(\mathbf{x}) \tilde{u}_q^m &= \Sigma_s^m(\mathbf{x}) K(\tilde{u}_q^m) + \eta \tilde{\Sigma}_{af}^x E(U_q^{x,0}) + \eta \Sigma_{af}^{x,0} E(\tilde{u}_q^x) && \text{in } \mathcal{X}, \\ \tilde{u}_q^x(\mathbf{x}, \boldsymbol{\theta}) &= 0, \quad \tilde{u}_q^m(\mathbf{x}, \boldsymbol{\theta}) = 0 && \text{on } \Gamma_-, \end{aligned} \quad (3.17)$$

where $U_q^{x,0}$ and $U_q^{m,0}$ are the solutions with background optical parameters. Let us introduce the adjoint boundary Green's functions, $G^x(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d)$ and $G^m(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d)$, that solve

$$\begin{aligned} -\boldsymbol{\theta} \cdot \nabla G^x + \Sigma^{x,0}(\mathbf{x}) G^x &= \Sigma_s^x(\mathbf{x}) K(G^x)(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) && \text{in } \mathcal{X}, \\ -\boldsymbol{\theta} \cdot \nabla G^m + \Sigma^m(\mathbf{x}) G^m &= \Sigma_s^m(\mathbf{x}) K(G^m)(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) && \text{in } \mathcal{X}, \\ G^x(\mathbf{x}, \boldsymbol{\theta}) &= \delta(\mathbf{x} - \mathbf{x}_d), \quad G^m(\mathbf{x}, \boldsymbol{\theta}) = \delta(\mathbf{x} - \mathbf{x}_d) && \text{on } \Gamma_+. \end{aligned} \quad (3.18)$$

Note that if measurements are only taken at wavelength λ^m , then the $\delta(\mathbf{x} - \mathbf{x}_d)$ source for G^x will not be necessary. We are now ready to obtain

$$\begin{aligned} &\int_{\Omega} \left(\tilde{\Sigma}_{ai}^x(\mathbf{x}) + \tilde{\Sigma}_{af}^x(\mathbf{x}) \right) \left(\int_{S^{n-1}} U_q^{x,0} G^x d\boldsymbol{\theta} \right) d\mathbf{x} \\ &= - \int_{S^{n-1}} \mathbf{n}(\mathbf{x}_d) \cdot \boldsymbol{\theta} \tilde{u}_q^x(\mathbf{x}_d, \boldsymbol{\theta})|_{\Gamma_+} d\boldsymbol{\theta} \end{aligned} \quad (3.19)$$

and

$$\begin{aligned} &\int_{\Omega} \eta \tilde{\Sigma}_{af}^x(\mathbf{x}) \left(E(U_q^{x,0}) E(G^m) \right) d\mathbf{x} \\ &= - \int_{S^{n-1}} \mathbf{n}(\mathbf{x}_d) \cdot \boldsymbol{\theta} \tilde{u}_q^x(\mathbf{x}_d, \boldsymbol{\theta})|_{\Gamma_+} d\boldsymbol{\theta} + \int_{\Omega} \eta \Sigma_{af}^{x,0} E(G^m) E(\tilde{u}_q^x) d\mathbf{x}. \end{aligned} \quad (3.20)$$

The first equation, Eq. (3.19), is the same as (3.6) but for wavelength λ^x . The second equation, Eq. (3.20), is very different. The right hand side involves not only measured data at wavelength λ^m but also the solutions of the perturbed equations inside the domain. In order to solve for the unknowns $\tilde{\Sigma}_{ai}^x$ and $\tilde{\Sigma}_{af}^x$, we have to combine the linear integral equations, Eq. (3.19) and Eq. (3.20), with the equations for perturbations at wavelength λ^x , i.e., the first equation in (3.17) to form an enlarged least square problem. We can also solve the problem sequentially as follows. We first solve (3.19) to obtain the summation $\tilde{\Sigma}_{ai}^x + \tilde{\Sigma}_{af}^x$. We then solve the first equation in (3.17) to obtain $u_q^x(\mathbf{x}, \boldsymbol{\theta})$. After that, we can solve (3.20) to obtain $\tilde{\Sigma}_{af}^x$ (and thus $\tilde{\Sigma}_{ai}^x$ assuming that η is known). Note that since the parameter η (possibly as a function of space also) and Σ_{af}^x appear as a product in the equation, it is not possible to reconstruct both of them, but only the product.

When time-dependent or frequency-domain measurements [116, 117] are available in fluorescence tomography, we can generalize, in a straightforward way, the above linearization technique to use those data. We will not repeat the procedure here.

3.4 Bioluminescence problem

Unlike optical tomography and fluorescence tomography, bioluminescence tomography itself is a linear inverse source problem. There is thus no linearization procedure needed. The map between the unknown source $g(\mathbf{x})$ and the measurement on the boundary is just the solution restriction of the solution operator for (2.8) on Γ_+ , which we can express as

$$\int_{\Omega} g(\mathbf{x}) \left(\int_{S^{n-1}} G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d, \hat{\boldsymbol{\theta}}) d\boldsymbol{\theta} \right) d\mathbf{x} = u(\mathbf{x}_d, \hat{\boldsymbol{\theta}}), \quad (\mathbf{x}_d, \hat{\boldsymbol{\theta}}) \in \Gamma_+, \quad (3.21)$$

where the Green's function $G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d, \hat{\boldsymbol{\theta}})$ solves

$$\begin{aligned} -\boldsymbol{\theta} \cdot \nabla G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d, \hat{\boldsymbol{\theta}}) + \Sigma(\mathbf{x}) G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d, \hat{\boldsymbol{\theta}}) &= \Sigma_s(\mathbf{x}) K(G)(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d, \hat{\boldsymbol{\theta}}) && \text{in } \mathcal{X}, \\ G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d, \hat{\boldsymbol{\theta}}) &= \delta(\mathbf{x} - \mathbf{x}_d) \delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) && \text{on } \Gamma_+. \end{aligned} \quad (3.22)$$

The kernel of the integral operator in (3.21) thus involves only the Green's function, which is quite different from the linearization problem in optical tomography, (3.6), and fluorescence tomography, (3.19) and (3.20).

When multispectral measurements (2.14) are available, the bioluminescence tomography problem (2.13) can be re-formulated as

$$\int_{\Omega} g(\mathbf{x}, \lambda) \left(\int_{S^{n-1}} G(\mathbf{x}, \boldsymbol{\theta}, \lambda; \mathbf{x}_d, \hat{\boldsymbol{\theta}}) d\boldsymbol{\theta} \right) d\mathbf{x} = u(\mathbf{x}_d, \hat{\boldsymbol{\theta}}, \lambda), \quad (\mathbf{x}_d, \hat{\boldsymbol{\theta}}) \in \Gamma_+. \quad (3.23)$$

Note that the Green's function used here depends on wavelength λ . It solves Eq. (3.22) with λ -dependent absorption and scattering coefficients.

3.5 Computational complexities

In all the problems we have discussed about in this section, we end up with a linear system of algebraic equations of the form (3.7). In practice, those are either underdetermined or overdetermined system, depending on the amount of data available. Besides, as we have mentioned at the end of Section 2, the measured data usually contain noise. Thus, the linearized problems, for example Eq. (3.7), are usually solved in regularized least-square sense. For example, when Tikhonov regularization is used, $\tilde{\Sigma}_a$ is found as the solution to

$$\min_{\tilde{\Sigma}_a} \frac{1}{2} \|\mathcal{A}\tilde{\Sigma}_a - \mathcal{Z}\|_2^2 + \frac{\beta}{2} \|\tilde{\Sigma}_a\|_2^2. \quad (3.24)$$

The minimizer of (3.24) is the solution of the normal equation

$$(\mathcal{A}^{\mathfrak{T}}\mathcal{A} + \beta\mathcal{I})\tilde{\Sigma}_a = \mathcal{A}^{\mathfrak{T}}\mathcal{Z}, \quad (3.25)$$

that is,

$$\tilde{\Sigma}_a = (\mathcal{A}^\mathfrak{T} \mathcal{A} + \beta \mathcal{I})^{-1} \mathcal{A}^\mathfrak{T} \mathcal{Z}. \quad (3.26)$$

In practice, when the number of unknowns (discretized optical properties or sources) is large, the inverse matrix $(\mathcal{A}^\mathfrak{T} \mathcal{A} + \beta \mathcal{I})^{-1}$ is usually not formed directly. Instead, iterative methods are used to solve (3.25).

The major computation cost for linearized reconstruction methods that we presented above is devoted to the formation of the linear system such as (3.7). This is because of the fact that analytical formulas for the adjoint Green's functions are not available. We thus have to solve the transport equations to get those Green's functions and background solutions. For the optical tomography problem with measurement type (2.2) (resp. (2.3)), we need to solve N_q transport problems to get $\{U_q^0\}_{q=1}^{N_q}$. We then need to solve N_d (resp. $N_d \times N_k$) adjoint transport problems to get the Green's functions $\{G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d)\}_{d=1}^{N_d}$ (resp. $\{\{G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d, \boldsymbol{\theta}_k)\}_{d=1}^{N_d}\}_{k=1}^{N_k}$). The cost in the solution of the least square problem (3.24) is small compared to a transport solver. So the total computational cost is roughly $N_q + N_d$ (resp. $N_q + N_d \times N_k$) transport solvers. This is also true for time-dependent problems in which cases $N_q + N_d$ (resp. $N_q + N_d \times N_k$) time-dependent transport problems have to be solved. If frequency-domain data are used, the total computational cost will be roughly $(N_q + N_d) \times N_\omega$ (resp. $(N_q + N_d \times N_k) \times N_\omega$) complex transport solvers.

For the fluorescence tomography with measurement of type (2.2) (resp. (2.3)), to solve each background forward problem, we need to solve two transport equations. We also need to solve two adjoint transport equations to get the Green's function for each detector. So we need totally $2(N_q + N_d)$ (resp. $2(N_q + N_d \times N_k)$) transport solvers to build the linear system (3.19) and (3.20). To solve the linear system, however, we need to solve N_q transport equations for the perturbation u_q^x as we have discussed in Section 3.3. Thus the total computational cost in this case is roughly $3N_q + 2N_d$ (resp. $3N_q + 2N_d \times N_k$) transport solvers.

For the bioluminescence tomography problem in Section 3.4 with data (2.9), the computational cost is $N_d \times N_k$ transport solvers. The computational cost for multispectral BLT is roughly $N_\lambda \times N_d \times N_k$ if the total number of wavelength used is N_λ . This is because the absorption and scattering coefficients depend on wavelength, so that Green's functions for different wavelength is different.

3.6 Further remarks

The computational advantage of the linearization method over nonlinear reconstruction methods is not very obvious in inverse transport problems. The main reason is the lack of analytical results on transport Green's functions, even in very simplified cases. We have to compute those Green's functions when needed, which makes the linearization method still computational expensive as we have discussed above. For example, in all those linearized reconstruction methods we have presented, the computational costs of

the methods depend on not only the number of sources N_q , but also the number of detectors N_d and the number of directions N_k . In some of the nonlinear reconstruction methods we will introduce in Section 4, the computational cost will be independent of the number of detectors (N_d) and the number of directions N_k .

In general, however, linearization methods are indeed computationally cheaper than nonlinear iterative methods (such as those in Section 4) which require a considerable number of iterations to converge, starting from an initial guess. Also, since the majority of the computational costs is spent on forming the linearized problem, not on solving the linearized problem by the least-square method, the cost of the whole reconstruction process is (almost) independent of the value of the regularization parameter β . It is much less expensive to choose the optimal regularization parameter in linearized reconstruction methods than in nonlinear reconstruction methods. For more details on how to choose the optimal regularization parameter, we refer to [47].

When we are in the linearization regime, the quality of the reconstructions are comparable to those obtained by more advanced reconstruction techniques. Indeed, it has been observed in many applications that the linearization method works even when the perturbation in the absorption coefficient is not very small. However, when the perturbation is very large, the linearization method fails. Let us emphasize finally that the background around which we linearize the problem does not have to be constant but have to be *known a priori*.

In recent years, it is of great interests to solve inverse transport problems with very large data sets. In terms of the linearization method, this basically means that we have to solve $N_q \sim 10^3$ forward transport problems and $N_d \sim 10^3$ adjoint transport problems to form the matrix (3.7). Note that for all the N_q forward transport problems, only the source terms in the equations are different. All other parameters (including absorption and scattering coefficients) are the same. This is all true for the N_d adjoint problems. In other words, we have to solve the same transport equation for N_q different source terms. In terms of linear algebra, we want to solve linear systems with multiple right-hand-sides. It would be of great interest if one can adapt fast algorithms such as those proposed in [71, 72] (for different forward models) to solve the system efficiently so that we can solve the inversion problem in a reasonable computational time.

4 Nonlinear iterative reconstructions

The majority of the reconstruction methods that have been developed in transport-based medical imaging are nonlinear in nature since those inverse transport problems are mostly nonlinear inverse problems. All the methods that we will review here are iterative where starting from some initial guesses, the unknowns are iteratively updated so that the discrepancy between the physical measurements and model predictions can be reduced.

To simplify the presentation, we will consider the reconstruction problems for the

transport equations on the discretized level. The way how the transport equations are discretized is not very relevant. We will mainly focus on methods for optical tomography with stationary measurements in Sections 4.1, 4.2 and 4.3. We will then attempt to extend those methods to time-dependent measurements and fluorescence tomography in Section 4.4.

Let us assume that the discretized transport equation (2.1) for source q ($1 \leq q \leq N_q$) takes the form

$$\mathbf{T}(\boldsymbol{\Sigma})\mathbf{U}_q = \mathbf{G}_q, \quad (4.1)$$

where $\mathbf{T}(\boldsymbol{\Sigma}) \in \mathbb{R}^{N_\Omega N_\theta \times N_\Omega N_\theta}$ denote the discretized transport operator, depending on the discretized optical parameter $\boldsymbol{\Sigma} \in \mathbb{R}^{N_\Omega \times 1}$, $\mathbf{U}_q \in \mathbb{R}^{N_\Omega N_\theta \times 1}$ is the discretized photon flux and $\mathbf{G}_q \in \mathbb{R}^{N_\Omega N_\theta \times 1}$ denote the discretized source function, coming from $g_q(\mathbf{x}, \boldsymbol{\theta})$. As before, N_Ω denote the number of nodes in spatial mesh and N_θ denote the number of directions used in the discretization of the direction variable. For the detector located at a mesh node $\mathbf{x}_d \in \partial\Omega$, we denote by $\mathbf{M}_d^\Sigma: \mathbb{R}^{N_\Omega N_\theta \times 1} \mapsto \mathbb{R}$ the discretized version of the measurement operator at \mathbf{x}_d . We can now introduce the nonlinear map between optical parameter $\boldsymbol{\Sigma}$ and the measured data

$$\mathbf{F}_{q,d}(\boldsymbol{\Sigma}) \equiv \mathbf{M}_d^\Sigma \mathbf{U}_q(\boldsymbol{\Sigma}) = z_{q,d}, \quad 1 \leq q \leq N_q, \quad 1 \leq d \leq N_d. \quad (4.2)$$

The objective is now inverting this map to find the optical parameter $\boldsymbol{\Sigma}$.

In practice, the system (4.2) is either overdetermined or underdetermined. Besides, the physical measurements always contain noise. So, as in the linearized inversion case we discussed in the previous section, (4.2) is often also inverted in (scaled) least-square sense. Roughly speaking, we attempt to recover the unknowns by minimizing the discrepancy between measured data $z_{q,d}$ and model predictions $\mathbf{M}_d^\Sigma \mathbf{U}_q$. More precisely, with data (2.4), the objective function to be minimized in optical tomography is defined as

$$\Phi(\boldsymbol{\Sigma}) = \frac{1}{2} \sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \frac{|\mathbf{M}_d^\Sigma \mathbf{U}_q(\boldsymbol{\Sigma}) - z_{q,d}|^2}{|z_{q,d}|^2} + \beta \mathcal{R}(\boldsymbol{\Sigma}), \quad (4.3)$$

where $\mathcal{R}(\boldsymbol{\Sigma})$ is the regularization functional and β is the regularization parameter. It is important to note that the objective function Φ takes into account solutions \mathbf{U}_q of the forward problem for all N_q sources simultaneously. We emphasize that, although theoretically the normalization factor $1/|z_{q,d}|^2$ does not make any impact since we assume we can minimize the objective function to arbitrary accuracy, it is very important to keep the normalization in practice. This is because optical signals measured at different locations can have values that are different by one or more order of magnitude. The normalization process thus re-weights the mismatch terms coming from different source-detector pairs so that they are on the same scale. Let us also mention that, when the data used contain noise, it is useful to rescale the objective function using the covariance matrix of the noisy data. The algorithms we present below, however, are independent of both rescaling strategies.

Now on the discretized level, optical tomography can be formulated as a minimization problem subject to the constraints (4.1) for N_q different sources:

$$\begin{aligned} \min_{(\boldsymbol{\Sigma}, \{\mathbf{U}_q\}_{q=1}^{N_q})} \quad & \Phi\left(\boldsymbol{\Sigma}, \{\mathbf{U}_q\}_{q=1}^{N_q}\right), \\ \text{subject to} \quad & \mathbf{T}(\boldsymbol{\Sigma})\mathbf{U}_q - \mathbf{G}_q = 0, \quad 1 \leq q \leq N_q. \end{aligned} \quad (4.4)$$

We remark that in this constrained minimization formulation, we view the function Φ as a function of both $\boldsymbol{\Sigma}$ and $\{\mathbf{U}_q\}_{q=1}^{N_q}$. The two variables $\boldsymbol{\Sigma}$ and $\{\mathbf{U}_q\}_{q=1}^{N_q}$ are linked together by the transport equations in the constraints. If we solve the transport equations to obtain $\{\mathbf{U}_q\}_{q=1}^{N_q}$ (as functions of $\boldsymbol{\Sigma}$) and substitute back in Φ , we can then view Φ as a function of $\boldsymbol{\Sigma}$ only. This is the perspective of unconstrained minimization; see discussion in next section for more details. In the presentations in the following sections, we will use both the notation $\Phi(\boldsymbol{\Sigma})$ and the notation $\Phi(\boldsymbol{\Sigma}, \{\mathbf{U}_q\}_{q=1}^{N_q})$, depending on the context of the presentation. We ask the readers to be alert about this difference.

To make the notations consistent, for any scalar function $f(\mathbf{X}, \mathbf{Y})$, we will regard its gradients, $\nabla_{\mathbf{X}}f$ and $\nabla_{\mathbf{Y}}f$, as column vectors, and define its Hessian matrix (matrix of second order derivatives) as $\nabla_{\mathbf{X}\mathbf{Y}}^2 f = \nabla_{\mathbf{Y}}((\nabla_{\mathbf{X}}f)^{\mathbb{T}})$.

4.1 Methods of Newton type with line search

To solve the minimization problem (4.4), let us first consider methods of Newton type. There are a few versions of the Newton's method that have been developed in transport-based optical tomography [78, 107]. We present them in the same framework here. To do that, let us first introduce the Lagrangian function $\mathcal{L}: \mathbb{R}^N \times \mathbb{R}^{N_\Omega N_\theta \times N_q} \times \mathbb{R}^{N_\theta N_\Omega \times N_q} \mapsto \mathbb{R}$ for the constrained optimization problem, defined by

$$\mathcal{L}\left(\boldsymbol{\Sigma}, \{\mathbf{U}_q\}_{q=1}^{N_q}; \{\mathbf{V}_q\}_{q=1}^{N_q}\right) = \Phi\left(\boldsymbol{\Sigma}, \{\mathbf{U}_q\}_{q=1}^{N_q}\right) + \sum_{q=1}^{N_q} \mathbf{V}_q^{\mathbb{T}} (\mathbf{T}(\boldsymbol{\Sigma})\mathbf{U}_q - \mathbf{G}_q), \quad (4.5)$$

where $\mathbf{V}_q \in \mathbb{R}^{N_\theta N_\Omega \times 1}$ is the adjoint variable to \mathbf{U}_q , $1 \leq q \leq N_q$. The solution to the optimization problem (4.4), $(\boldsymbol{\Sigma}^*, \{\mathbf{U}_q^*\}_{q=1}^{N_q})$, satisfies the first order optimality conditions, also known as the Karush-Kuhn-Tucker (KKT) conditions, of (4.5),

$$\nabla_{\boldsymbol{\Sigma}} \mathcal{L} = \mathbf{0}, \quad \nabla_{[\mathbf{U}]} \mathcal{L} = \mathbf{0}, \quad \nabla_{[\mathbf{V}]} \mathcal{L} = \mathbf{0},$$

where, to save space, we have introduced

$$[\mathbf{U}] \equiv (\mathbf{U}_1^{\mathbb{T}}, \dots, \mathbf{U}_q^{\mathbb{T}}, \dots, \mathbf{U}_{N_q}^{\mathbb{T}})^{\mathbb{T}}.$$

Those conditions can be explicitly written as:

$$\nabla_{\Sigma} \mathcal{L} = \beta \nabla_{\Sigma} \mathcal{R} + \sum_{q=1}^{N_q} (\mathbf{T}_{\Sigma} \mathbf{U}_q)^{\mathfrak{T}} \mathbf{V}_q = \mathbf{0}, \quad (4.6)$$

$$\nabla_{\mathbf{U}_q} \mathcal{L} = \mathbf{T}^{\mathfrak{T}}(\Sigma) \mathbf{V}_q + \sum_{d=1}^{N_d} \frac{\mathbf{M}_d^{\mathfrak{T}} \mathbf{U}_q - z_{q,d}}{|z_{q,d}|^2} \mathbf{M}_d = \mathbf{0}, \quad 1 \leq q \leq N_q, \quad (4.7)$$

$$\nabla_{\mathbf{V}_q} \mathcal{L} = \mathbf{T}(\Sigma) \mathbf{U}_q - \mathbf{G}_q = \mathbf{0}, \quad 1 \leq q \leq N_q, \quad (4.8)$$

with the quantity

$$\mathbf{T}_{\Sigma} \equiv \nabla_{\Sigma} \mathbf{T}$$

understood as a tensor in $\mathbb{R}^{N_{\theta} N_{\Omega} \times N_{\Omega} \times N_{\theta} N_{\Omega}}$ so that $(\mathbf{T}_{\Sigma} \mathbf{U}_q)$ is in $\mathbb{R}^{N_{\theta} N_{\Omega} \times N_{\Omega}}$ and $(\mathbf{T}_{\Sigma} \mathbf{U}_q)^{\mathfrak{T}} \mathbf{V}_q$ is a column vector in $\mathbb{R}^{N_{\Omega} \times 1}$. The k th component of $(\mathbf{T}_{\Sigma} \mathbf{U}_q)^{\mathfrak{T}} \mathbf{V}_q$ is given by

$$\mathbf{U}_q^{\mathfrak{T}} \frac{\partial \mathbf{T}^{\mathfrak{T}}}{\partial \Sigma_k} \mathbf{V}_q$$

with Σ_k the k th component of Σ . Theory of constrained optimization says that if $(\Sigma^*, \{\mathbf{U}_q^*\}_{q=1}^{N_q})$ provides an optimal solution of (4.4), then there exist $\mathbf{V}_q^* \in \mathbb{R}^{N_{\theta} N_{\Omega} \times 1}$, $1 \leq q \leq N_q$, such that $(\Sigma^*, \{\mathbf{U}_q^*\}_{q=1}^{N_q}, \{\mathbf{V}_q^*\}_{q=1}^{N_q})$ is a stationary point of the Lagrangian function (4.5).

The first order KKT conditions, (4.6), (4.7) and (4.8) forms a system of nonlinear equations for Σ , $\{\mathbf{U}_q\}_{q=1}^{N_q}$ and $\{\mathbf{V}_q\}_{q=1}^{N_q}$. It can be solved iteratively starting from an initial guess $(\Sigma^0, [\mathbf{U}^0], [\mathbf{V}^0])$. Assuming that we have the unknowns at iteration k , $(\Sigma^k, [\mathbf{U}^k], [\mathbf{V}^k])$, the Newton method updates them according to

$$(\Sigma^{k+1}, [\mathbf{U}^{k+1}], [\mathbf{V}^{k+1}]) = (\Sigma^k, [\mathbf{U}^k], [\mathbf{V}^k]) + l_k (\tilde{\Sigma}^k, [\tilde{\mathbf{U}}^k], [\tilde{\mathbf{V}}^k]), \quad (4.9)$$

where $(\tilde{\Sigma}^k, [\tilde{\mathbf{U}}^k], [\tilde{\mathbf{V}}^k])$ is the update direction and l_k is the step length in the that direction. l_k is usually obtained by a line search method or other globalization technique. The update direction is obtained by solving the following second order KKT system

$$\begin{pmatrix} \nabla_{[\mathbf{U}][\mathbf{U}]}^2 \mathcal{L} & \nabla_{[\mathbf{U}][\mathbf{V}]}^2 \mathcal{L} & \nabla_{[\mathbf{U}]\Sigma}^2 \mathcal{L} \\ \nabla_{[\mathbf{V}][\mathbf{U}]}^2 \mathcal{L} & \nabla_{[\mathbf{V}][\mathbf{V}]}^2 \mathcal{L} & \nabla_{[\mathbf{V}]\Sigma}^2 \mathcal{L} \\ \nabla_{\Sigma[\mathbf{U}]}^2 \mathcal{L} & \nabla_{\Sigma[\mathbf{V}]}^2 \mathcal{L} & \nabla_{\Sigma\Sigma}^2 \mathcal{L} \end{pmatrix} \begin{pmatrix} [\tilde{\mathbf{U}}^k] \\ [\tilde{\mathbf{V}}^k] \\ \tilde{\Sigma}^k \end{pmatrix} = - \begin{pmatrix} \nabla_{[\mathbf{U}]} \mathcal{L} \\ \nabla_{[\mathbf{V}]} \mathcal{L} \\ \nabla_{\Sigma} \mathcal{L} \end{pmatrix}, \quad (4.10)$$

where all the derivatives are evaluated at $(\Sigma^k, [\mathbf{U}^k], [\mathbf{V}^k])$. In terms of transport operators,

we have explicitly

$$\begin{aligned}
\nabla_{[\mathbf{U}][\mathbf{U}]}^2 \mathcal{L} &= \mathbf{I}_{N_q} \otimes \left(\sum_{d=1}^{N_d} \frac{\mathbf{M}_d \mathbf{M}_d^{\mathfrak{T}}}{|z_{q,d}|^2} \right), \\
\nabla_{[\mathbf{U}][\mathbf{V}]}^2 \mathcal{L} &= (\nabla_{[\mathbf{V}][\mathbf{U}]}^2 \mathcal{L})^{\mathfrak{T}} = \mathbf{I}_{N_q} \otimes \mathbf{T}^{\mathfrak{T}}, \\
\nabla_{[\mathbf{U}]\Sigma}^2 \mathcal{L} &= (\nabla_{\Sigma[\mathbf{U}]}^2 \mathcal{L})^{\mathfrak{T}} = \begin{pmatrix} \mathbf{T}_{\Sigma}^{\mathfrak{T}} \mathbf{V}_1 \\ \vdots \\ \mathbf{T}_{\Sigma}^{\mathfrak{T}} \mathbf{V}_{N_q} \end{pmatrix}, \quad \nabla_{[\mathbf{V}][\mathbf{V}]}^2 \mathcal{L} = \mathbf{0}, \\
\nabla_{\Sigma\Sigma}^2 \mathcal{L} &= \beta \nabla_{\Sigma\Sigma}^2 \mathcal{R}, \quad \nabla_{[\mathbf{V}]\Sigma}^2 \mathcal{L} = (\nabla_{\Sigma[\mathbf{V}]}^2 \mathcal{L})^{\mathfrak{T}} = \begin{pmatrix} \mathbf{T}_{\Sigma} \mathbf{U}_1 \\ \vdots \\ \mathbf{T}_{\Sigma} \mathbf{U}_{N_q} \end{pmatrix},
\end{aligned} \tag{4.11}$$

where $\mathbf{I}_{N_q} \in \mathbb{R}^{N_q \times N_q}$ is the identity matrix and \otimes denotes tensor product. As before, the quantities $\mathbf{T}_{\Sigma} \mathbf{U}_q$ ($1 \leq q \leq N_q$) (or respectively $\mathbf{T}_{\Sigma}^{\mathfrak{T}} \mathbf{V}_q$) ($1 \leq q \leq N_q$) are understood as matrices in $\mathbb{R}^{N_{\theta} N_{\Omega} \times N_{\Omega}}$ whose k th rows are given by $\frac{\partial \mathbf{T}}{\partial \Sigma_k} \mathbf{U}_q$ (or respectively $\frac{\partial \mathbf{T}^{\mathfrak{T}}}{\partial \Sigma_k} \mathbf{V}_q$).

In principle, if we can solve (4.10) to get the update direction, we can use the iterative scheme (4.9) to find the final solution assuming that such an iterative process converges. In practice, however, system (4.10) is a very large, $(2N_{\theta} N_{\Omega} N_q + N_{\Omega})^2$, and ill-conditioned. This full space approach has thus not being explored so far in optical imaging community. All existing algorithms in optical tomography convert the minimization problem (4.4) into unconstrained problem where only the unknown optical property are iteratively updated. This is done as follows. Suppose we have Σ^k in the k -th Newton iteration. We then solves the N_q forward transport problems in (4.8) with Σ^k exactly to get $[\mathbf{U}^k]$. We then solve the N_q adjoint problem in (4.7) to get $[\mathbf{V}^k]$. After that, we have

$$\nabla_{[\mathbf{U}]} \mathcal{L}(\Sigma^k, [\mathbf{U}^k]; [\mathbf{V}^k]) = \mathbf{0}, \quad \nabla_{[\mathbf{V}]} \mathcal{L}(\Sigma^k, [\mathbf{U}^k]; [\mathbf{V}^k]) = \mathbf{0},$$

so the update equation can now be simplifies to

$$\begin{pmatrix} \nabla_{[\mathbf{U}][\mathbf{U}]}^2 \mathcal{L} & \nabla_{[\mathbf{U}][\mathbf{V}]}^2 \mathcal{L} & \nabla_{[\mathbf{U}]\Sigma}^2 \mathcal{L} \\ \nabla_{[\mathbf{V}][\mathbf{U}]}^2 \mathcal{L} & \mathbf{0} & \nabla_{[\mathbf{V}]\Sigma}^2 \mathcal{L} \\ \nabla_{\Sigma[\mathbf{U}]}^2 \mathcal{L} & \nabla_{\Sigma[\mathbf{V}]}^2 \mathcal{L} & \nabla_{\Sigma\Sigma}^2 \mathcal{L} \end{pmatrix} \begin{pmatrix} [\tilde{\mathbf{U}}^k] \\ [\tilde{\mathbf{V}}^k] \\ \tilde{\Sigma}^k \end{pmatrix} = - \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \nabla_{\Sigma} \mathcal{L} \end{pmatrix}, \tag{4.12}$$

where we have used the fact that

$$\nabla_{[\mathbf{V}][\mathbf{V}]}^2 \mathcal{L} = \mathbf{0}.$$

It is now straightforward to perform a block Gauss elimination to eliminate $[\tilde{\mathbf{U}}^k]$ and $[\tilde{\mathbf{V}}^k]$. We thus obtain:

$$\mathbf{H}^k \tilde{\Sigma}^k = -\nabla_{\Sigma} \mathcal{L}, \tag{4.13}$$

where the reduced Hessian operator is given by

$$\begin{aligned} \mathbf{H}^k = & \nabla_{\Sigma\Sigma}^2 \mathcal{L} - \nabla_{\Sigma[\mathbf{U}]}^2 \mathcal{L} (\nabla_{[\mathbf{V}][\mathbf{U}]}^2 \mathcal{L})^{-1} \nabla_{[\mathbf{V}]\Sigma}^2 \mathcal{L} - \nabla_{\Sigma[\mathbf{V}]}^2 \mathcal{L} (\nabla_{[\mathbf{U}][\mathbf{V}]}^2 \mathcal{L})^{-1} \nabla_{[\mathbf{U}]\Sigma}^2 \mathcal{L} \\ & + \nabla_{\Sigma[\mathbf{V}]}^2 \mathcal{L} (\nabla_{[\mathbf{U}][\mathbf{V}]}^2 \mathcal{L})^{-1} \nabla_{[\mathbf{U}][\mathbf{U}]}^2 \mathcal{L} (\nabla_{[\mathbf{V}][\mathbf{U}]}^2 \mathcal{L})^{-1} \nabla_{[\mathbf{V}]\Sigma}^2 \mathcal{L}, \end{aligned} \quad (4.14)$$

and the reduced gradient $\nabla_{\Sigma} \mathcal{L}$ is evaluated at $(\Sigma^k, [\mathbf{U}^k], [\mathbf{V}^k])$,

$$\nabla_{\Sigma} \mathcal{L} = \beta \nabla_{\Sigma} \mathcal{R}(\Sigma^k) + \sum_{q=1}^{N_q} (\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{T}} \mathbf{V}_q^k. \quad (4.15)$$

The updating formula (4.9) is also simplified to

$$\Sigma^{k+1} = \Sigma^k + l_k \tilde{\Sigma}^k. \quad (4.16)$$

The Newton type of methods that have been developed are all characterized by (4.16) and (4.13) with line search method to determine l_k . The main difference between different implementations is how the reduced Hessian matrix and the reduced gradient are approximated in practice.

4.1.1 The Gauss-Newton method

In Gauss-Newton version of the implementation, we drop the terms that involve $\nabla_{\Sigma[\mathbf{U}]}^2 \mathcal{L}$ and $\nabla_{[\mathbf{U}]\Sigma}^2 \mathcal{L}$ to ensure that \mathbf{H}^k is non-negative in the sense that

$$\mathbf{X}^{\mathfrak{T}} \mathbf{H}^k \mathbf{X} \geq 0, \quad \forall \mathbf{X} \in \mathbb{R}^{N_{\Omega} \times 1}.$$

This is important since we want to find the minimizer of the Lagrangian, not the maximizer. We thus obtain, after considering (4.11),

$$\mathbf{H}_{GN}^k = \beta \nabla_{\Sigma\Sigma}^2 \mathcal{R} + \sum_{q=1}^{N_q} \left((\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{T}} \mathbf{T}^{-\mathfrak{T}} \right) \left(\sum_{d=1}^{N_d} \frac{1}{|z_{q,d}^2|} \mathbf{M}_d \mathbf{M}_d^{\mathfrak{T}} \right) \left(\mathbf{T}^{-1} (\mathbf{T}_{\Sigma} \mathbf{U}_q^k) \right). \quad (4.17)$$

If we introduce a new matrix $\mathbf{B} \in \mathbb{R}^{N_{\Omega} N_{\theta} \times N_d}$ defined as

$$\mathbf{B} = \left[\frac{\mathbf{M}_1}{|z_{q,1}|}, \dots, \frac{\mathbf{M}_d}{|z_{q,d}|}, \dots, \frac{\mathbf{M}_{N_d}}{|z_{q,N_d}|} \right], \quad (4.18)$$

then the reduced Hessian can be written as

$$\mathbf{H}_{GN}^k = \beta \nabla_{\Sigma\Sigma}^2 \mathcal{R} + \sum_{q=1}^{N_q} \mathbf{J}_q \mathbf{J}_q^{\mathfrak{T}}, \quad \mathbf{J}_q \equiv (\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{T}} \mathbf{T}^{-\mathfrak{T}} \mathbf{B}. \quad (4.19)$$

There are also some different versions of the Gauss-Newton method. The most frequently used one replaces the Hessian matrix (4.17) with

$$\mathbf{H}_{GN2}^k = \nabla_{\Sigma} \Phi(\Sigma^k) \left(\nabla_{\Sigma} \Phi(\Sigma^k) \right)^{\mathfrak{T}}, \quad (4.20)$$

where Φ is the objective function defined in (4.3). It turns out that the gradient of the objective function, $\nabla_{\Sigma}\Phi$, is the same as $\nabla_{\Sigma}\mathcal{L}$. To see this, we first differentiate Φ with respect to Σ to obtain

$$\nabla_{\Sigma}\Phi(\Sigma) = \frac{1}{2} \sum_{q=1}^{N_q} \sum_{d=1}^{N_d} (\nabla_{\Sigma}\mathbf{U}_q^{\bar{\zeta}}) \mathbf{M}_d \frac{\mathbf{M}_d^{\bar{\zeta}} \mathbf{U}_q - z_{q,d}}{|z_{q,d}|^2} + \beta \nabla_{\Sigma} \mathcal{R}. \quad (4.21)$$

We then differentiate the transport equation (4.1), after a transpose, with respect to Σ to obtain

$$\nabla_{\Sigma} \mathbf{U}_q^{\bar{\zeta}} \mathbf{T}^{\bar{\zeta}} + \mathbf{U}_q^{\bar{\zeta}} \mathbf{T}_{\Sigma}^{\bar{\zeta}} = \mathbf{0}. \quad (4.22)$$

We can solve for $\nabla_{\Sigma}\mathbf{U}_q$ and substitute into (4.21) to get

$$\begin{aligned} \nabla_{\Sigma}\Phi &= - \sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \mathbf{U}_q^{\bar{\zeta}} \mathbf{T}_{\Sigma}^{\bar{\zeta}} \mathbf{T}^{-\bar{\zeta}} \mathbf{M}_d \frac{\mathbf{M}_d^{\bar{\zeta}} \mathbf{U}_q - z_{q,d}}{|z_{q,d}|^2} + \beta \nabla_{\Sigma} \mathcal{R} \\ &= \sum_{q=1}^{N_q} (\mathbf{T}_{\Sigma} \mathbf{U}_q)^{\bar{\zeta}} \mathbf{V}_q + \beta \nabla_{\Sigma} \mathcal{R}, \end{aligned} \quad (4.23)$$

which is exactly $\nabla_{\Sigma}\mathcal{L}$. Here \mathbf{V}_q is again the solution to the adjoint problem, i.e., the second equation in the KKT conditions, Eq. (4.7).

We refer [124] for detailed implementation of the Gauss-Newton method for optical tomography with the radiative transport equation.

4.1.2 The BFGS method

In the BFGS implementation of the Newton's method (4.16) and (4.13), we approximate the Hessian matrix \mathbf{H}^k in a different way. Denote by

$$\mathbf{s}_k = \Sigma^{k+1} - \Sigma^k, \quad \mathbf{y}_k = \nabla_{\Sigma}\Phi(\Sigma^{k+1}) - \nabla_{\Sigma}\Phi(\Sigma^k),$$

the BFGS updating rule for the Hessian matrix is

$$\mathbf{H}_B^{k+1} = \mathbf{H}_B^k - \frac{\mathbf{H}_B^k \mathbf{s}_k \mathbf{s}_k^{\bar{\zeta}} \mathbf{H}_B^k}{\mathbf{s}_k^{\bar{\zeta}} \mathbf{H}_B^k \mathbf{s}_k} + \frac{\mathbf{y}_k \mathbf{y}_k^{\bar{\zeta}}}{\mathbf{y}_k^{\bar{\zeta}} \mathbf{s}_k}, \quad (4.24)$$

starting from guess for Hessian, \mathbf{H}_B^0 , which is often a scalar multiple of the identity matrix. To reduce storage requirement for the Hessian matrix, which could be very large in some cases, the limited memory version of the BFGS method chooses to form inverse Hessian directly (in which case, (4.13) can be solved just by applying the inverse Hessian to the right hand side). In this case, the updating rule for inverse Hessian is

$$(\mathbf{H}_B^{k+1})^{-1} = \left(\mathbf{I}_{N_{\Omega}} - \frac{\mathbf{y}_k \mathbf{s}_k^{\bar{\zeta}}}{\mathbf{y}_k^{\bar{\zeta}} \mathbf{s}_k} \right) (\mathbf{H}_B^k)^{-1} \left(\mathbf{I}_{N_{\Omega}} - \frac{\mathbf{y}_k \mathbf{s}_k^{\bar{\zeta}}}{\mathbf{y}_k^{\bar{\zeta}} \mathbf{s}_k} \right) + \frac{\mathbf{y}_k \mathbf{y}_k^{\bar{\zeta}}}{\mathbf{y}_k^{\bar{\zeta}} \mathbf{s}_k}. \quad (4.25)$$

As we mentioned above, forming (4.25) takes tremendous computer memory for large problems. To overcome this shortcoming, the limited-memory version of BFGS only stores the vector \mathbf{y}_k and \mathbf{s}_k obtained in the last m ($3 \leq m \leq 7$ usually) iterations [70] and discards the rest. Thus after first m iterations, (4.25) can be expressed as:

$$\begin{aligned}
(\mathbf{H}_B^{k+1})^{-1} &= (\mathbf{Q}_k^{\bar{\Sigma}} \cdots \mathbf{Q}_{k-m}^{\bar{\Sigma}}) (\mathbf{H}_{B,0}^{k+1})^{-1} (\mathbf{Q}_{k-m} \cdots \mathbf{Q}_k) \\
&+ \rho_{k-m} (\mathbf{Q}_k^{\bar{\Sigma}} \cdots \mathbf{Q}_{k-m+1}^{\bar{\Sigma}}) \mathbf{s}_{k-m} \mathbf{s}_{k-m}^{\bar{\Sigma}} \times (\mathbf{Q}_{k-m+1} \cdots \mathbf{Q}_k) \\
&+ \rho_{k-m+1} (\mathbf{Q}_k^{\bar{\Sigma}} \cdots \mathbf{Q}_{k-m+2}^{\bar{\Sigma}}) \mathbf{s}_{k-m+1} \mathbf{s}_{k-m+1}^{\bar{\Sigma}} \times (\mathbf{Q}_{k-m+2} \cdots \mathbf{Q}_k) \\
&\vdots \\
&+ \rho_k \mathbf{s}_k \mathbf{s}_k^{\bar{\Sigma}},
\end{aligned} \tag{4.26}$$

with the sparse initial guess given by

$$(\mathbf{H}_{B,0}^{k+1})^{-1} = \frac{\mathbf{y}_{k+1}^{\bar{\Sigma}} \mathbf{s}_{k+1}}{\mathbf{y}_{k+1}^{\bar{\Sigma}} \mathbf{y}_{k+1}} \mathbf{I}_{N_\Omega}, \quad \rho_k = \frac{1}{\mathbf{y}_k^{\bar{\Sigma}} \mathbf{s}_k}, \quad \mathbf{Q}_k = \mathbf{I}_{N_\Omega} - \mathbf{y}_k \mathbf{s}_k^{\bar{\Sigma}}.$$

We refer interested readers to [98] for more details on the limited-memory BFGS algorithms, and to reference [1, 57, 80, 100, 107, 112] for applications of those algorithms to optical tomography with transport equation.

4.1.3 The Levenberg-Marquardt method

The Levenberg-Marquardt method can be also be viewed as a special case of the Newton's method. In this case, the Hessian matrix is taken as

$$\mathbf{H}_{LM}^k = \nu_k \mathbf{I}_{N_\Omega} + \beta \nabla_{\Sigma}^2 \mathcal{R} + \sum_{q=1}^{N_q} \mathbf{J}_q \mathbf{J}_q^{\bar{\Sigma}}, \tag{4.27}$$

with \mathbf{J}_q ($1 \leq q \leq N_q$) given by (4.19). Note that (4.27) is the same as the Gauss-Newton approximation (4.17) except that there is an extra term $\nu_k \mathbf{I}_{N_\Omega}$. In other words, the Levenberg-Marquardt method can be viewed as a special case of the Gauss-Newton method with the regularization functional

$$\beta \mathcal{R}(\Sigma) + \frac{\nu_k}{2} \Sigma^{\bar{\Sigma}} \Sigma.$$

The parameter ν is chosen to control the nonlinearity of the inverse problem and is chosen such that $\nu_k \rightarrow 0$ as $k \rightarrow \infty$. This means that the Levenberg-Marquardt method solves the original regularized least-square problem when converges after infinite number of iterations. In practice, the algorithm has to be terminated in a finite number of steps (before ν_k to be zero), so the method regularizes the inversion more than designed.

The reduced gradient in the Levenberg-Marquardt method is taken as in (4.15) but without the regularization term. In other words,

$$(\nabla_{\Sigma} \mathcal{L})_{LM} = \sum_{q=1}^{N_q} (\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\bar{\Sigma}} \mathbf{V}_q^k. \tag{4.28}$$

We remark that there are also different implementations of the Levenberg-Marquardt method in the literature. To make a connection between them, let us consider the original inverse problem in terms of (4.2). A local linearization at the k th iteration will give us the following linear equations

$$\left(\nabla_{\Sigma} \mathbf{F}_{q,d}(\Sigma^k)\right)^{\mathfrak{F}} (\Sigma^{k+1} - \Sigma^k) = -(\mathbf{F}_{q,d}(\Sigma^k) - z_{q,d}), \quad 1 \leq q \leq N_q, \quad 1 \leq d \leq N_d. \quad (4.29)$$

Treating $\Sigma^{k+1} - \Sigma^k$ as the unknown, we can solve this linear equation, after rescaling by the factor $1/|z_{q,d}|$, by regularized least square method to obtain

$$\begin{aligned} & \left[\frac{1}{|z_{q,d}|} \nabla_{\Sigma} \mathbf{F}_{q,d}(\Sigma^k) \left(\frac{1}{|z_{q,d}|} \nabla_{\Sigma} \mathbf{F}_{q,d}(\Sigma^k) \right)^{\mathfrak{F}} + \nu_k \mathbf{I}_{N_{\Omega}} + \beta \nabla_{\Sigma\Sigma}^2 \mathcal{R} \right] (\Sigma^{k+1} - \Sigma^k) \\ & = - \frac{1}{|z_{q,d}|} \nabla_{\Sigma} \mathbf{F}_{q,d}(\Sigma^k) \frac{\mathbf{F}_{q,d}(\Sigma^k) - z_{q,d}}{|z_{q,d}|}, \end{aligned} \quad (4.30)$$

for $1 \leq q \leq N_q$, $1 \leq d \leq N_d$, with ν_k as the regularization parameter. We can now add the equations for all source-detector pairs to obtain, still denoting by β and ν_k (instead of $(N_q + N_d)\beta$ and $(N_q + N_d)\nu_k$) the parameters,

$$\begin{aligned} & \left[\sum_{q=1}^{N_q} (\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{F}} \mathbf{T}^{-\mathfrak{F}} \left(\sum_{d=1}^{N_d} \frac{1}{|z_{q,d}|^2} \mathbf{M}_d \mathbf{M}_d^{\mathfrak{F}} \right) \mathbf{T}^{-1} (\mathbf{T}_{\Sigma} \mathbf{U}_q^k) + \nu_k \mathbf{I}_{N_{\Omega}} + \beta \nabla_{\Sigma\Sigma}^2 \mathcal{R} \right] (\Sigma^{k+1} - \Sigma^k) \\ & = - \sum_{q=1}^{N_q} (\mathbf{T}_{\Sigma} \mathbf{U}_q)^{\mathfrak{F}} \mathbf{V}_q, \end{aligned} \quad (4.31)$$

where we have used the fact that

$$\nabla_{\Sigma} \mathbf{F}_{q,d} = (\mathbf{T}_{\Sigma} \mathbf{U}_q)^{\mathfrak{F}} \mathbf{T}^{-\mathfrak{F}} \mathbf{M}_d.$$

We now arrive at the following iteration for Σ^k

$$\Sigma^{k+1} = \Sigma^k + (\mathbf{H}_{LM}^k)^{-1} \left(- \sum_{q=1}^{N_q} (\mathbf{T}_{\Sigma} \mathbf{U}_q)^{\mathfrak{F}} \mathbf{V}_q \right). \quad (4.32)$$

This iteration is just the Newton method characterized in (4.13) and (4.16) with a fixed step length, $l_k = 1$, in line search. In practical implementations, we can of course use a line search to find variable step length l_k . That is why we treat the Levenberg-Marquardt method as a special case of the Newton method with line search.

We can modify the above procedure slightly to get another implementation. Let us first add together Eq. (4.29) for different detectors, again after rescaling by $1/|z_{q,d}|$, to

obtain

$$\begin{aligned} & \sum_{d=1}^{N_d} \frac{1}{|z_{q,d}|} \mathbf{M}_d^{\mathfrak{T}} \mathbf{T}^{-1} (\mathbf{T}_{\Sigma} \mathbf{U}_q) (\boldsymbol{\Sigma}^{k+1} - \boldsymbol{\Sigma}^k) \\ &= - \sum_{d=1}^{N_d} \frac{1}{|z_{q,d}|} (\mathbf{M}_d^{\mathfrak{T}} \mathbf{U}_q - z_{q,d}), \quad 1 \leq q \leq N_q. \end{aligned} \quad (4.33)$$

We now again reformulate this equation in regularized least-square sense, and add the results for all sources together, to get

$$\mathbf{H}_{LM2}^k (\boldsymbol{\Sigma}^{k+1} - \boldsymbol{\Sigma}^k) = -(\nabla_{\Sigma} \mathcal{L})_{LM2}, \quad (4.34)$$

with

$$\begin{aligned} \mathbf{H}_{LM2}^k &= v_k \mathbf{I}_{N_{\Omega}} + \beta \nabla_{\Sigma\Sigma}^2 \mathcal{R} \\ &+ \sum_{q=1}^{N_q} \left(\sum_{d=1}^{N_d} (\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{T}} \mathbf{T}^{-\mathfrak{T}} \frac{\mathbf{M}_d}{|z_{q,d}|} \right) \left(\sum_{d=1}^{N_d} (\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{T}} \mathbf{T}^{-\mathfrak{T}} \frac{\mathbf{M}_d}{|z_{q,d}|} \right)^{\mathfrak{T}}, \end{aligned} \quad (4.35)$$

and

$$(\nabla_{\Sigma} \mathcal{L})_{LM2} = \sum_{q=1}^{N_q} \left[\left(\sum_{d=1}^{N_d} (\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{T}} \mathbf{T}^{-\mathfrak{T}} \frac{\mathbf{M}_d}{|z_{q,d}|} \right) \sum_{d=1}^{N_d} \frac{1}{|z_{q,d}|} (\mathbf{M}_d^{\mathfrak{T}} \mathbf{U}_q - z_{q,d}) \right]. \quad (4.36)$$

If we introduce \mathbf{W}_q ($1 \leq q \leq N_q$) that solves

$$\mathbf{T}^{\mathfrak{T}} \mathbf{W}_q = \sum_{d=1}^{N_d} \frac{1}{|z_{q,d}|} \mathbf{M}_d, \quad (4.37)$$

then \mathbf{H}_{LM2}^k and $(\nabla_{\Sigma} \mathcal{L})_{LM2}$ can be simplified as

$$\begin{aligned} \mathbf{H}_{LM2}^k &= v_k \mathbf{I}_{N_{\Omega}} + \beta \nabla_{\Sigma\Sigma}^2 \mathcal{R} + \sum_{q=1}^{N_q} \left((\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{T}} \mathbf{W}_q \right) \left((\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{T}} \mathbf{W}_q \right)^{\mathfrak{T}}, \\ (\nabla_{\Sigma} \mathcal{L})_{LM2} &= \sum_{q=1}^{N_q} \left[\left((\mathbf{T}_{\Sigma} \mathbf{U}_q^k)^{\mathfrak{T}} \mathbf{W}_q \right) \sum_{d=1}^{N_d} \frac{1}{|z_{q,d}|} (\mathbf{M}_d^{\mathfrak{T}} \mathbf{U}_q - z_{q,d}) \right]. \end{aligned} \quad (4.38)$$

The iteration defined in (4.34) is again a Newton method. It is identical to the iteration defined in (4.32) except that the Hessian matrix and the reduced gradient are slightly different.

The Levenberg-Marquardt method has been implemented for inverse transport problems recently in [48,52–54]. We refer to those references for more details on the algorithm.

4.1.4 Computational considerations

The methods of Newton type with line search that we have just introduced in this section can be implemented as follows.

Algorithm 4.1: Newton's method with line search

-
- **FOR** $k=0,1,2,\dots,\text{MAXIT}$
 1. **For** $q=1,2,\dots,N_q$
 - Solve the forward problem (4.8);
 - Solve the adjoint problem (4.7);
 - End**
 2. Compute the reduced gradient and the Hessian matrix using
 - (4.15) and (4.17) (or resp. (4.23) and (4.20)) for the Gauss-Newton method,
 - (4.23) and (4.24) for the BFGS method, or
 - (4.28) and (4.27) (or resp. (4.36) and (4.35)) for Levenberg-Marquardt method,
depending on the method selected;
 3. Solve (4.13) for the update direction $\tilde{\Sigma}^k$;
 4. Perform a line search in direction $\tilde{\Sigma}^k$ to determine l_k :
$$\min_{l_k > 0} \Phi(\Sigma^k + l_k \tilde{\Sigma}^k);$$
 5. Update Σ^k according to (4.16);
 6. **If** (stopping criteria satisfied)
 - Stop and take Σ^{k+1} as the final solution;
 - Else**
 - Set $\Sigma^{k+1} = \Sigma^k$;
 - End**
 - **END**
-

Different kinds of line search methods can be used in the algorithm to find $l_k > 0$. Usually, one impose the Wolfe conditions [98] on l_k to ensure the convergence of the algorithm:

$$\Phi(\Sigma^k + l_k \tilde{\Sigma}^k) \leq \Phi(\Sigma^k) + c_1 l_k [\nabla_{\Sigma} \Phi(\Sigma^k)]^{\top} \tilde{\Sigma}^k, \quad (4.39)$$

$$[\nabla_{\Sigma} \Phi(\Sigma^k + l_k \tilde{\Sigma}^k)]^{\top} \tilde{\Sigma}^k \geq c_2 [\nabla_{\Sigma} \Phi(\Sigma^k)]^{\top} \tilde{\Sigma}^k, \quad (4.40)$$

where c_1 and c_2 are two small constants that can be tuned. The algorithm is usually stopped when either the objective function (relative to its initial value, $\Phi(\Sigma^k)/\Phi(\Sigma^0)$)

is small enough or its gradient (in appropriate norm) is small enough. For details on the implementation in the setting of inverse transport problems, including methods for choosing the regularization parameter β , we refer to [48, 53, 54, 80, 107, 124].

The Newton type of methods, although straightforward to implement, are computational expensive. At each Newton iteration, the major costs are spent on the evaluation of the objective function (or residual), the reduced gradient and the Hessian matrix. To evaluate the residual, we need to solve N_q forward transport problems. We then have to N_q adjoint transport problems to evaluate the reduced gradient. We thus need $2N_q$ transport solvers. In the second Gauss-Newton method (characterized by (4.20)), the BFGS method, and the second Levenberg-Marquardt method (characterized by (4.35)), the reduced gradient can be used to form the Hessian matrices directly. So we do not need extra transport solvers. The computational cost of those versions of the Newton's method is independent of the number of detectors used in the measurements. To form the Hessian matrix (4.17) in the first Gauss-Newton and the Hessian matrix (4.27) in the first Levenberg-Marquardt method, however, we need to solve N_d extra adjoint transport equations to evaluate quantities of the form $\mathbf{T}^{-\top} \mathbf{M}_d$ ($1 \leq d \leq N_d$). So we need totally $2N_q + N_d$ transport solvers to evaluate the residual, the reduced gradient and the Hessian matrix. The computational costs for these two methods thus depend on the number of detectors used.

The storage requirement for the Newton's method is mainly the transport matrix \mathbf{T} (identical for all N_q sources) and the Hessian matrix \mathbf{H} . The matrix \mathbf{T} is in general sparse, while \mathbf{H} is not necessarily so. To circumvent storage limitations, one can use methods such as the limited-memory version of the BFGS method [107]. A more general way is to employ Krylov subspace methods (such as the GMRES method [109]) to solve the reduced KKT system (4.13). Those methods are "matrix-free" in the sense that they do not ask for the matrix explicitly but require only matrix-vector product.

Let us finally remark that the N_q transport equations (or resp. the N_q adjoint problems) for different sources are independent of each other. Thus they can be solved on different processors simultaneously when multiple-processor computational resources are available. Also, Newton type of methods converge faster (in terms of function and gradient evaluations needed before the algorithms converge) than conjugate gradient methods such as the one developed in [77].

4.2 Method of augmented Lagrangian

In the Newton type of methods we just introduced, the constrained minimization problem (4.4) is converted to an unconstrained minimization problem by solving the forward and adjoint problems exactly at each Newton step. This method thus requires solving the forward problem for some approximations of Σ over and over again in each reconstruction. The speed of the algorithm depends strongly on how fast and accurately the forward problems are solved.

In fact, (4.4) can be solved directly as a constrained problem. The augmented La-

grangian method is an iterative method for this purpose that is easy to implement [2]. Still denote by \mathcal{L} the Lagrangian function defined in (4.5), the method defines an augmented Lagrangian function by

$$\mathcal{L}_A(\boldsymbol{\Sigma}, \{\mathbf{U}_q\}_{q=1}^{N_q}; \{\mathbf{V}_q\}_{q=1}^{N_q}) = \mathcal{L}(\boldsymbol{\Sigma}, \{\mathbf{U}_q\}_{q=1}^{N_q}; \{\mathbf{V}_q\}_{q=1}^{N_q}) + \frac{1}{2A} \sum_{q=1}^{N_q} \|\mathbf{T}(\boldsymbol{\Sigma})\mathbf{U}_q - \mathbf{G}_q\|_2^2, \quad (4.41)$$

where the last term is used as a penalty for violating the constraints $\mathbf{T}(\boldsymbol{\Sigma})\mathbf{U}_q - \mathbf{G}_q = 0$, $1 \leq q \leq N_q$. The strength parameter A will be updated during the iterative process. The augmented Lagrangian method will look for a stationary point of the function \mathcal{L}_A instead of \mathcal{L} . Note that in the limit the constraints are satisfied exactly, $\mathcal{L}_A = \mathcal{L}$. More details on the method can be found in reference [98].

Let us assume that at the k th iteration, we have an approximation $(\boldsymbol{\Sigma}^k, \{\mathbf{U}_q^k\}_{q=1}^{N_q}; \{\mathbf{V}_q^k\}_{q=1}^{N_q})$ to the stationary point $(\boldsymbol{\Sigma}^*, \{\mathbf{U}_q^*\}_{q=1}^{N_q}; \{\mathbf{V}_q^*\}_{q=1}^{N_q})$ of the Lagrangian function. We fix the current estimates of the Lagrangian multipliers $\{\mathbf{V}_q^k\}_{q=1}^{N_q}$ and a penalty parameter A_k . Minimizing $\mathcal{L}_{A_k}(\boldsymbol{\Sigma}, \{\mathbf{U}_q\}_{q=1}^{N_q}; \{\mathbf{V}_q^k\}_{q=1}^{N_q})$ with respect to $\boldsymbol{\Sigma}$ and $\{\mathbf{U}_q\}_{q=1}^{N_q}$ yields the following system for the minimizers:

$$\beta \nabla_{\boldsymbol{\Sigma}} \mathcal{R}(\boldsymbol{\Sigma}) + \sum_{q=1}^{N_q} (\mathbf{T}_{\boldsymbol{\Sigma}} \mathbf{U}_q)^{\top} \left[\mathbf{V}_q^k - \frac{1}{A_k} (\mathbf{T} \mathbf{U}_q - \mathbf{G}_q) \right] = \mathbf{0}, \quad (4.42)$$

$$\mathbf{T}^{\top}(\boldsymbol{\Sigma}) \left[\mathbf{V}_q^k - \frac{1}{A_k} (\mathbf{T} \mathbf{U}_q - \mathbf{G}_q) \right] + \sum_{d=1}^{N_d} \frac{\mathbf{M}_d^{\top} \mathbf{U}_q - z_{q,d}}{|z_{q,d}|^2} \mathbf{M}_d = \mathbf{0}, \quad 1 \leq q \leq N_q. \quad (4.43)$$

Let $(\boldsymbol{\Sigma}^k, \{\mathbf{U}_q^k\}_{q=1}^{N_q})$ be approximate solution of this system, i.e., an approximate minimizer of the augmented Lagrangian $\mathcal{L}_{A_k}(\boldsymbol{\Sigma}, \{\mathbf{U}_q\}_{q=1}^{N_q}; \{\mathbf{V}_q^k\}_{q=1}^{N_q})$. We thus conclude, by comparing this system with the optimality conditions of the Lagrangian, Eq. (4.6) and Eq. (4.7), that $\mathbf{V}_q^k - (1/A_k)(\mathbf{T}(\boldsymbol{\Sigma}^k)\mathbf{U}_q^k - \mathbf{G}_q)$ approximates \mathbf{V}_q^* :

$$\mathbf{V}_q^* \approx \mathbf{V}_q^k - (1/A_k)(\mathbf{T}(\boldsymbol{\Sigma}^k)\mathbf{U}_q^k - \mathbf{G}_q), \quad 1 \leq q \leq N_q. \quad (4.44)$$

This formula can be rearranged to produce an estimate of $\mathbf{V}_q^k - (1/A_k)(\mathbf{T}(\boldsymbol{\Sigma}^k)\mathbf{U}_q^k - \mathbf{G}_q)$:

$$\mathbf{T}(\boldsymbol{\Sigma}^k)\mathbf{U}_q^k - \mathbf{G}_q \approx A_k(\mathbf{V}_q^k - \mathbf{V}_q^*). \quad (4.45)$$

Hence, we deduce that if \mathbf{V}_q^k is close to the optimal Lagrangian multiplier \mathbf{V}_q^* , and A_k is small enough, then the pair $(\boldsymbol{\Sigma}^k, \{\mathbf{U}_q^k\}_{q=1}^{N_q})$ satisfies the corresponding constraint with a high accuracy. Formula (4.44) prompts a rule for iterative updating of the Lagrangian multipliers (the adjoint variables):

$$\mathbf{V}_q^{k+1} = \mathbf{V}_q^k - \frac{1}{A_k} (\mathbf{T}(\boldsymbol{\Sigma}^k)\mathbf{U}_q^k - \mathbf{G}_q), \quad 1 \leq q \leq N_q. \quad (4.46)$$

The augmented Lagrangian method is thus an iterative method that update simultaneously the optical property Σ and the solutions of the forward (and adjoint) transport problems. The algorithm can be implemented as follows.

Algorithm 4.2: Augmented Lagrangian algorithm

Initially choose A_0 , $\tau_0 > 0$ and maximum iteration step MAXIT. Also choose initial guess $(\tilde{\Sigma}^0, \{\tilde{\mathbf{U}}_q^0\}_{q=1}^{N_q}, \{\mathbf{V}_q^0\}_{q=1}^{N_q})$.

• **FOR** $k=0,1,2,\dots,\text{MAXIT}$

1. Solve the sub-minimization problem

$$\min_{\Sigma, \{\mathbf{U}_q^0\}_{q=1}^{N_q}} \mathcal{L}_{A_0}(\Sigma, \{\mathbf{U}_q^0\}_{q=1}^{N_q}; \{\mathbf{V}_q^k\}_{q=1}^{N_q}) \quad (4.47)$$

to find the minimizer $(\Sigma^k, \{\mathbf{U}_q^k\}_{q=1}^{N_q})$ by an iterative method that

- starts from initial value $(\tilde{\Sigma}^k, \{\tilde{\mathbf{U}}_q^k\}_{q=1}^{N_q})$;
- terminates when $\|\nabla_{\Sigma} \mathcal{L}_{A_k}\|_{l^2} + \sum_{q=1}^{N_q} \|\nabla_{\mathbf{U}_q} \mathcal{L}_{A_k}\|_{l^2} \leq \tau_k$ is satisfied;

2. **If** (stopping criteria reached)

Stop and take $(\Sigma^k, \{\mathbf{U}_q^k\}_{q=1}^{N_q})$ as the final solution;

End

3. Update the Lagrangian multipliers according to (4.46).

4. Choose a new penalty parameter $A_{k+1} \in (0, A_k)$ and new a parameter τ_k ;

5. Set starting point for the next iteration:

$$(\tilde{\Sigma}^{k+1}, \{\tilde{\mathbf{U}}_q^{k+1}\}_{q=1}^{N_q}) = (\Sigma^k, \{\mathbf{U}_q^k\}_{q=1}^{N_q})$$

• **END**

To solve the sub-optimization problem (4.47) in the above algorithm, we can use any iterative method such as the BFGS algorithm. The gradients of the objective function, $\mathcal{L}_{A_k}(\Sigma, \{\mathbf{U}_q\}_{q=1}^{N_q}; \{\mathbf{V}_q^k\}_{q=1}^{N_q})$, with respect to Σ and $\{\mathbf{U}_q\}_{q=1}^{N_q}$ are available analytically; see for example, the left hand sides of (4.42) and (4.43). This makes the minimization procedure more-or-less straightforward.

For initial guess set $(\tilde{\Sigma}^0, \{\tilde{\mathbf{U}}_q^0\}_{q=1}^{N_q}, \{\mathbf{V}_q^0\}_{q=1}^{N_q})$, we only choose an initial guess Σ^0 . The $\{\tilde{\mathbf{U}}_q^0\}_{q=1}^{N_q}$ and $\{\mathbf{V}_q^0\}_{q=1}^{N_q}$ are chosen as the solution to the forward and adjoint transport problems with Σ^0 , respectively.

One advantage of the augmented Lagrangian method is that it can be easily parallelized. For example, in the sub-minimization problem, the gradient of the augmented

Lagrangian function has an analytical form that involves only the summation of local matrix-vector and vector-vector products. The computation of this gradient can thus be done on separate processors and then collected. Also, the update of the Lagrangian multipliers can be done on parallel processors.

We refer to references [23, 67, 98] for more detailed discussion of the augmented Lagrangian method, including the choice of parameters A_k and τ_k . It is shown in [2] that with appropriate choice of various algorithmic parameters, the augmented Lagrangian method can speed up the reconstruction process in optical tomography significantly. The problem is exactly that there is no general theory on how to tune those parameters for specific problems. One has to perform test reconstructions to choose efficient parameters.

4.3 The nonlinear Kaczmarz method

Besides those methods that are closely related to optimization theory, there are also a few other types of iterative methods that have been implemented for inverse transport problems. The nonlinear Kaczmarz method [94, 95] has been implemented in a few different settings [42].

In the Kaczmarz method, at each iteration, the data from different sources are used in sequential to update the unknowns. This is very different from the methods we presented above where, at each iteration, the measured data for all source-detector pairs are used simultaneously to update the unknown. To present the method, let us first rewrite the discretized nonlinear operator equation (4.2) into N_q small groups of nonlinear equations

$$\mathbf{F}_q(\boldsymbol{\Sigma}) = \mathbf{z}_q, \quad 1 \leq q \leq N_q, \quad (4.48)$$

with the notation

$$\mathbf{F}_q = \begin{pmatrix} \mathbf{F}_{q,1} \\ \vdots \\ \mathbf{F}_{q,d} \\ \vdots \\ \mathbf{F}_{q,N_d} \end{pmatrix}, \quad \mathbf{z}_q = \begin{pmatrix} z_{q,1} \\ \vdots \\ z_{q,d} \\ \vdots \\ z_{q,N_d} \end{pmatrix}. \quad (4.49)$$

Then the nonlinear Kaczmarz method is characterized by the following double iterative process, starting with $\boldsymbol{\Sigma}^0$

$$\hat{\boldsymbol{\Sigma}}^{k,0} = \boldsymbol{\Sigma}^k, \quad (4.50a)$$

$$\hat{\boldsymbol{\Sigma}}^{k,q} = \hat{\boldsymbol{\Sigma}}^{k,q-1} + \omega (\nabla_{\boldsymbol{\Sigma}} \mathbf{F}_q^{\mathfrak{T}}(\hat{\boldsymbol{\Sigma}}^{k,q-1})) \mathbf{C}_q^{-1} (\mathbf{z}_q - \mathbf{F}_q(\hat{\boldsymbol{\Sigma}}^{k,q-1})), \quad q = 1, \dots, N_q, \quad (4.50b)$$

$$\boldsymbol{\Sigma}^{k+1} = \hat{\boldsymbol{\Sigma}}^{k,N_q}, \quad (4.50c)$$

with

$$\mathbf{C}_q = \left(\nabla_{\boldsymbol{\Sigma}} \mathbf{F}_q^{\mathfrak{T}}(\hat{\boldsymbol{\Sigma}}^{k,q-1}) \right)^{\mathfrak{T}} \left(\nabla_{\boldsymbol{\Sigma}} \mathbf{F}_q^{\mathfrak{T}}(\hat{\boldsymbol{\Sigma}}^{k,q-1}) \right) \in \mathbb{R}^{N_d \times N_d}, \quad (4.51)$$

and $0 < \omega < 2$ an algorithmic parameter.

The nonlinear Kaczmarz method (4.50) can be implemented straightforwardly. What we need to pay attention to is the fact that the matrix \mathbf{C}_q is very different from the Hessian matrices in Newton type of method we have discussed above. $\mathbf{C}_q \in \mathbb{R}^{N_d \times N_d}$ is of different size to the Hessian matrices. The component of the matrix is given by

$$(\mathbf{C}_q)_{d_1 d_2} = \left((\mathbf{T}_\Sigma \mathbf{U}_q)^{\mathfrak{T}} \mathbf{T}^{-\mathfrak{T}} \mathbf{M}_{d_1} \right)^{\mathfrak{T}} \left((\mathbf{T}_\Sigma \mathbf{U}_q)^{\mathfrak{T}} \mathbf{T}^{-\mathfrak{T}} \mathbf{M}_{d_2} \right). \quad (4.52)$$

To evaluate the matrix \mathbf{C}_q we thus have to solve the N_d adjoint transport equations to compute $\mathbf{T}^{-\mathfrak{T}} \mathbf{M}_d$ ($1 \leq d \leq N_d$). The computational cost of the method is thus dependent not only on the number of sources used but also on the number of the detectors employed. The stop

It has not been discussed very much about how to impose regularization within the framework of the Kaczmarz method besides stopping the iteration prematurely. One possible way is to apply a weak low-pass filter, say \mathcal{F} , on the iteration (4.50). In other words, we replace the last step in the iteration by $\Sigma^{k+1} = \mathcal{F}(\hat{\Sigma}^{k, N_q})$. It would be interesting to see some theoretical analysis on how to choose optimal filters in this case.

4.4 Extensions and remarks

The nonlinear methods we have presented so far for optical tomography with stationary data can be generalized to other data types and inverse transport problems.

4.4.1 Frequency-domain data

Frequency-domain algorithms can be constructed in the same way as before. The transport matrix \mathbf{T} is now understood as the discretization of the transport operator T defined as

$$Tu = \frac{i\omega}{c}u + \boldsymbol{\theta} \cdot \nabla u + \Sigma(\mathbf{x})u - \Sigma_s(\mathbf{x})K(u). \quad (4.53)$$

The transport solutions (\mathbf{U}_q , $1 \leq q \leq N_q$) and the adjoint variables (\mathbf{V}_q , $1 \leq q \leq N_q$) are complex instead of being real. The transpose (i.e., adjoint) operation has to be understood in complex (Hermitian) sense also. In other words, the transpose $\mathbf{X}^{\mathfrak{T}}$ is replaced by $\overline{\mathbf{X}^{\mathfrak{T}}}$; see for example [107] for more details. An alternative choice is to split the complex equation into real and imaginary parts to get a set of two (coupled) transport equations; see similar discussion in Section 4.4.3.

Let us remark that when frequency domain data are available, it is desirable sometimes to rescale the phase and amplitude information separately. In minimization based

algorithms, the objective function can be chosen as

$$\Phi(\Sigma) = \frac{1}{2} \sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \left\{ \frac{|\operatorname{Re} [\ln(\mathbf{M}_d^\Sigma \mathbf{U}_q) - \ln(z_{q,d})]|^2}{|\operatorname{Re} [\ln(z_{q,d})]|^2} + \frac{|\operatorname{Im} [\ln(\mathbf{M}_d^\Sigma \mathbf{U}_q) - \ln(z_{q,d})]|^2}{|\operatorname{Im} [\ln(z_{q,d})]|^2} \right\} + \beta \mathcal{R}(\Sigma), \quad (4.54)$$

where $\operatorname{Re}X$ and $\operatorname{Im}X$ denote the real part (amplitude) and imaginary part (phase) of X , respectively. The rescaling factors can be chosen differently, depending on the needs of concrete applications; see for example discussions in [124].

4.4.2 Time-dependent data

Time-dependent data can be Fourier transformed into frequency domain to use the algorithms described above. We can also generalize those nonlinear reconstruction algorithms to use data in time-domain directly. To see that, let us consider the Newton method. The semi-discretized transport initial-value problem is now

$$\frac{1}{c} \frac{\partial \mathbf{U}_q}{\partial t} + \mathbf{T}(\Sigma) \mathbf{U}_q = \mathbf{G}_q, \quad \mathbf{U}_q(t=0) = \mathbf{0}, \quad 1 \leq q \leq N_q, \quad (4.55)$$

where the zero initial condition can be replaced with other ones, depending on physical applications. Let us suppose that we can measure data in the interval $(0, t_{\max})$. The objective function to be minimized in this case is

$$\Phi(\Sigma, \{\mathbf{U}_q\}_{q=1}^{N_q}) = \frac{1}{2} \sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \int_0^{t_{\max}} \frac{|\mathbf{M}_d^\Sigma \mathbf{U}_q(t) - z_{q,d}(t)|^2}{|z_{q,d}(t)|^2} dt + \beta \mathcal{R}(\Sigma), \quad (4.56)$$

where $\mathbf{U}_q(t)$ is an implicit function of Σ as before. Optical tomography with time-dependent data can now be formulated into the same constrained minimization problem (4.4), except that constraints are now the equations in (4.55). The Lagrangian function can be redefined as

$$\begin{aligned} & \mathcal{L}(\Sigma, \{\mathbf{U}_q\}_{q=1}^{N_q}; \{\mathbf{V}_q\}_{q=1}^{N_q}) \\ &= \Phi(\Sigma, \{\mathbf{U}_q\}_{q=1}^{N_q}) + \sum_{q=1}^{N_q} \int_0^{t_{\max}} \mathbf{V}_q^\Sigma(t) \left(\frac{1}{c} \frac{\partial \mathbf{U}_q}{\partial t} + \mathbf{T} \mathbf{U}_q - \mathbf{G}_q \right) dt + \sum_{q=1}^{N_q} \mathbf{V}_q^\Sigma(0) \mathbf{U}_q(0). \end{aligned} \quad (4.57)$$

The first order KKT conditions are now

$$\beta \nabla_{\Sigma} \mathcal{R}(\Sigma) + \sum_{q=1}^{N_q} \int_0^{t_{\max}} (\mathbf{T}_{\Sigma} \mathbf{U}_q)^{\top} \mathbf{V}_q dt = \mathbf{0}, \quad (4.58)$$

$$-\frac{1}{c} \frac{\partial \mathbf{V}_q}{\partial t} + \mathbf{T}^{\top}(\Sigma) \mathbf{V}_q + \sum_{d=1}^{N_d} \frac{\mathbf{M}_d^{\top} \mathbf{U}_q - z_{q,d}}{|z_{q,d}|^2} \mathbf{M}_d = \mathbf{0}, \quad \mathbf{V}_q(t_{\max}) = \mathbf{0}, \quad 1 \leq q \leq N_q, \quad (4.59)$$

$$\frac{1}{c} \frac{\partial \mathbf{U}_q}{\partial t} + \mathbf{T}(\Sigma) \mathbf{U}_q - \mathbf{G}_q = \mathbf{0}, \quad \mathbf{U}_q(0) = \mathbf{0}, \quad 1 \leq q \leq N_q, \quad (4.60)$$

Attention has to be paid here to the fact that the adjoint problems are now transport equations that evolves to $t=0$ starting from $t=t_{\max}$. In other words, those are final-value problems.

We can follow the steps in Section 4.1 to derive the whole algorithm. We will omit the procedure but just mention that the second order operators involved in the reduced Hessian operator are now given by

$$\begin{aligned} (\nabla_{[\mathbf{U}][\mathbf{U}]}^2 \mathcal{L})[\tilde{\mathbf{U}}] &= \left[\mathbf{I}_{N_q} \otimes \left(\sum_{d=1}^{N_d} \frac{\mathbf{M}_d \mathbf{M}_d^{\top}}{|z_{q,d}|^2} \right) \right] [\tilde{\mathbf{U}}], \\ (\nabla_{[\mathbf{V}][\mathbf{V}]}^2 \mathcal{L})[\tilde{\mathbf{V}}] &= \mathbf{0}, \quad (\nabla_{\Sigma}^2 \mathcal{L}) \tilde{\Sigma} = \beta \nabla_{\Sigma}^2 \mathcal{R} \tilde{\Sigma}, \\ (\nabla_{[\mathbf{U}][\mathbf{V}]}^2 \mathcal{L})[\tilde{\mathbf{V}}] &= \left(\mathbf{I}_{N_q} \otimes \left(-\frac{1}{c} \frac{\partial}{\partial t} + \mathbf{T}^{\top} \right) \right) [\tilde{\mathbf{V}}], \quad [\tilde{\mathbf{V}}](t_{\max}) = \mathbf{0}, \\ (\nabla_{[\mathbf{V}][\mathbf{U}]}^2 \mathcal{L})[\tilde{\mathbf{U}}] &= \left(\mathbf{I}_{N_q} \otimes \left(\frac{1}{c} \frac{\partial}{\partial t} + \mathbf{T} \right) \right) [\tilde{\mathbf{U}}], \quad [\tilde{\mathbf{U}}](0) = \mathbf{0}, \\ (\nabla_{[\mathbf{U}][\Sigma]}^2 \mathcal{L}) \tilde{\Sigma} &= \begin{pmatrix} \mathbf{T}_{\Sigma}^{\top} \mathbf{V}_1 \\ \vdots \\ \mathbf{T}_{\Sigma}^{\top} \mathbf{V}_{N_q} \end{pmatrix} \tilde{\Sigma}, \quad (\nabla_{\Sigma[\mathbf{U}]}^2 \mathcal{L})[\tilde{\mathbf{U}}] = [\mathbf{V}_1^{\top} \mathbf{T}_{\Sigma} \cdots \mathbf{V}_{N_q}^{\top} \mathbf{T}_{\Sigma}] [\tilde{\mathbf{U}}], \\ (\nabla_{[\mathbf{V}][\Sigma]}^2 \mathcal{L}) \tilde{\Sigma} &= \begin{pmatrix} \mathbf{T}_{\Sigma} \mathbf{U}_1 \\ \vdots \\ \mathbf{T}_{\Sigma} \mathbf{U}_{N_q} \end{pmatrix} \tilde{\Sigma}, \quad (\nabla_{\Sigma[\mathbf{V}]}^2 \mathcal{L})[\tilde{\mathbf{V}}] = [\mathbf{U}_1^{\top} \mathbf{T}_{\Sigma}^{\top} \cdots \mathbf{U}_{N_q}^{\top} \mathbf{T}_{\Sigma}^{\top}] [\tilde{\mathbf{V}}]. \end{aligned}$$

We observe from (4.59) that to solve the adjoint problems, we need the forward solutions $\mathbf{U}_q(t)$ for all time $t \in (0, t_{\max})$. This means that we have to store full time-dependent forward solutions. This is a nontrivial requirement since the forward solutions are high-dimensional objects, the discretization of $u_q(t, \mathbf{x}, \boldsymbol{\theta})$ ($1 \leq q \leq N_q$).

4.4.3 Fluorescence problem

Nonlinear reconstruction methods for fluorescence tomography is not very different from those we have developed in the previous sections. To see that, let us consider the fluores-

cence tomography in discrete level

$$\begin{pmatrix} \mathbf{T}^x & \mathbf{0} \\ \mathbf{E}^x & \mathbf{T}^m \end{pmatrix} \begin{pmatrix} \mathbf{U}^x \\ \mathbf{U}^m \end{pmatrix} = \begin{pmatrix} \mathbf{G}^x \\ \mathbf{0} \end{pmatrix}, \quad (4.61)$$

where, as before, the superscript x and m denote quantities depending on wavelength λ^x and λ^m , respectively. The matrix \mathbf{E}^x is the discretization of the E operator defined in (2.6). Equation (4.61) is of the same form as (4.1) if we introduce the notations

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}^x & \mathbf{0} \\ \mathbf{E}^x & \mathbf{T}^m \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} \mathbf{U}^x \\ \mathbf{U}^m \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{G}^x \\ \mathbf{0} \end{pmatrix}.$$

Everything else follows immediately.

5 Techniques for features reconstructions

Inverse transport problems we have discussed in this paper are all ill-posed. In the algorithms we presented in Section 3 and Section 4, we treat the optical parameters as function of space and we attempts to reconstruct the full information about the optical parameters. Those are inverse problems that are ill-posed, and the reconstructions are not stable. In practice, it is always helpful when we have extra information about the unknowns to be recovered so that we can use those information to improve the reconstructions. We thus want to incorporate *a priori* information into the algorithms we have developed. The techniques we introduce below are exactly for this purpose.

5.1 Parameterized reconstructions

In practical applications, we often have very limited amount of data that can be used. We thus want to reduce the number of unknowns to be reconstructed so that the inverse problem is not very underdetermined. One technique to reduce the number of unknowns is to represent the unknown function with a basis under which the coefficients in the representation decay fast enough so that the first a few coefficients will be enough to represent the function accurately. In the cases where we know that the unknown function is smooth enough, we can just use the Fourier representation. In regular domains, we just look for a small number of Fourier coefficients of the unknown function. In a general domain, we can parameterize the unknown as

$$\Sigma(\mathbf{x}) = \sum_{k=0}^M \hat{\Sigma}_k \phi_k(\mathbf{x}), \quad (5.1)$$

with $\{\phi_k\}_{k=0}^{\infty}$ a global basis. One possible choice is of course the eigenfunctions of the Laplace operator in domain Ω . In other words, $\{\phi_k\}_{k=0}^{\infty}$ are the solution of the following eigenvalue problem

$$-\Delta\phi(\mathbf{x}) = \lambda\phi(\mathbf{x}), \quad \text{in } \Omega, \quad (5.2)$$

with appropriate boundary conditions. If we use this parameterization to the linearized inverse problem (3.6), we obtain, assuming that we are interested in only the first M Fourier modes,

$$\sum_{k=1}^M a_k(q,d) \hat{\Sigma}_k = z_{q,d}, \quad (5.3)$$

where the $a_k(q,d)$ are defined as

$$a_k(q,d) = \int_{\Omega} \phi_k(\mathbf{x}) \left(\int_{S^{n-1}} U_q^0(\mathbf{x}, \boldsymbol{\theta}) G(\mathbf{x}, \boldsymbol{\theta}; \mathbf{x}_d) d\boldsymbol{\theta} \right) d\mathbf{x}. \quad (5.4)$$

Collecting for all source and detector pairs, we get a linear system that we can solve to recover the coefficients $\{\hat{\Sigma}_k\}_{k=1}^M$. The benefit of doing this is of course to reduce the under-determinacy of the inverse problem. The reconstruction is thus more stable.

The parameterization can be also incorporated into nonlinear reconstruction schemes we presented in Section 4. The unknowns are now the Fourier coefficients. For example, in minimization-based methods, we now optimize with respect to $\{\hat{\Sigma}_k\}_{k=1}^M$. So the space in which the minimizer is sought is much smaller than the original problem. The iterative reconstruction algorithms can be directly used except that we have to compute the gradient $\mathbf{T}_{\hat{\Sigma}}$ instead of \mathbf{T}_{Σ} . In fact, by chain rule, we obtain

$$\mathbf{T}_{\hat{\Sigma}} = \mathbf{T}_{\Sigma} \nabla_{\hat{\Sigma}} \Sigma.$$

Here $\nabla_{\hat{\Sigma}} \Sigma$ can be analytically computed from the representation in (5.1).

The parameterization method is a regularization (with prior knowledge) method in the following sense. The number of modes kept, say M , play the role of the regularization parameter. When M is small ($1/M$ is large), the regularization is strong since we are recovering very few parameters. When M is large ($1/M$ is small), the regularization effect is small. So $1/M$ has the same function as the parameter β in the regularized objective function we have seen in the previous sections. The parameterized reconstruction method has been used in [59] for a two-dimensional rectangular domain where the basis functions are chosen as cosines. It has been shown there that, with the parameterization method, one can obtain reconstructions of very similar quality to full reconstructions, but with less computational expenses and data.

5.2 Shape reconstructions

In many applications of optical tomography, the objective optical properties that we are interested in consists of a few regions of constant values. In other words, we look for localized changes in optical properties. Assuming that there are M localized objects in regions $\{\Omega_{Inc}^k\}_{k=1}^M$, $\Omega_{Inc}^i \cap \Omega_{Inc}^j = \emptyset$ if $i \neq j$, we can parameterize the unknown function as

$$\Sigma(\mathbf{x}) = \Sigma_0 + \sum_{k=1}^M \hat{\Sigma}_k \chi_{\Omega_{Inc}^k}(\mathbf{x}), \quad (5.5)$$

where $\chi_A(\mathbf{x})$ denotes the characteristic function of the set A . The objective is now of course to look for the M regions, $\{\Omega_{Inc}^k\}_{k=1}^M$, and the coefficients in those regions, $\{\hat{\Sigma}_k\}_{k=1}^M$.

In bioluminescence tomography, to reduce the ill-posedness of the problem, it is popular to reconstruct features of the source instead of the general sources. For example, we can assume that the source is localized, which result in the following parameterization

$$g(\mathbf{x}) = \bar{g}\chi_{\Omega_{src}}(\mathbf{x}). \quad (5.6)$$

When multiple localized sources exist, we end up with a parameterization that is similar to (5.5). Note that this parameterization, looking attractive, is nonlinear in nature. It transforms the originally linear inverse problem into a nonlinear inverse problem to recover the intensity \bar{g} and the support of the source Ω_{src} .

Currently, most available results solve the inverse problem by solving the following minimization problem (with the same notation as before)

$$\min_{\{\hat{\Sigma}_k, \Omega_{Inc}^k\}_{k=1}^M} \frac{1}{2} \sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \frac{|\mathbf{M}_d^{\hat{\Sigma}_k} \mathbf{U}_q - z_{q,d}|^2}{|z_{q,d}|^2}. \quad (5.7)$$

The difficulty lies in the fact that any gradient-based iterative method will require the computation of derivatives with respect to the geometric objects $\{\Omega_{Inc}^k\}_{k=1}^M$. In reference [10], an efficient method is proposed to compute such derivatives. The method parameterize further the boundary of the regions by Fourier coefficients in two-dimensional case or spherical harmonics in three-dimensional case. For example, in two-dimensional case, the boundary of a region, say $(\theta, r(\theta))$ in polar coordinate, is decomposed into the superposition of a few Fourier modes:

$$r(\theta) = \sum_{k=-M}^M c_k e^{-ik\theta}, \quad c_{-k} = \bar{c}_k. \quad (5.8)$$

It is then not very hard to compute the derivatives of the objective function with respect to the Fourier coefficients $\{c_k\}_{k=1}^M$.

The level set approach provides another way to solve the minimization problem (5.7). We will not cover details here but refer interested readers to [44] for the implementation of level set method for shape reconstruction problem in inverse transport applications.

5.3 Sparsity constraints

Very recently, there is a new type of *a priori* information that have been proposed to improve reconstructions in bioluminescence imaging: the sparsity prior [53, 54]. The ideas is based on the observation that the bioluminescence source we intend to recover is often very localized. In other words, the source function only take nonzero values in a small subregion of the domain. On discrete level, this means that the source vector, to be recovered has sparse structure: there is only very small number (compared to the length

of the vector) of nonzero entries. Recent study shows this type of sparse signals can be recovered stably with only a small number of measurements [53]. Let us assume that we have discretized the integral equation (3.21) into the following form

$$\mathbf{G}\mathbf{g} = \mathbf{u} \quad (5.9)$$

with \mathbf{g} the unknown source vector, \mathbf{G} the system matrix and \mathbf{u} the measurement vector. We then recover the source vector by minimize the difference between prediction and measurement with l_1 regularization on the unknown:

$$\min_{\mathbf{g} \geq 0} \|\mathbf{G}\mathbf{g} - \mathbf{u}\|_2^2 + \beta \|\mathbf{g}\|_1. \quad (5.10)$$

It has been shown numerically in [53] that this minimization problem select sparse solutions to the underdetermined problem (5.9), which thus helps to recover localized sources. Based on the results in [53], it is proposed in [54] that one can go one step further by changing the data fidelity term to l_1 norm also to obtain the following minimization problem

$$\min_{\mathbf{g} \geq 0} \|\mathbf{G}\mathbf{g} - \mathbf{u}\|_1 + \beta \|\mathbf{g}\|_1. \quad (5.11)$$

The benefit of using l_1 data fidelity (instead of the l_2 one) is that it allows one to recover the right source function with data set that contain *outliers* (data points that are very distant from the rest of the data), a claim that is verified by the numerical simulation in [54] in a slightly more complicated situation.

Note that the minimization problems (5.10) and (5.11) are now non-smooth problems, and it is non-trivial to solve those minimization problems. Most importantly, the problems are now nonlinear so iterative methods have to be used. The computational cost of solving the inverse problem thus increase dramatically compared to linear least-square techniques such as (3.24).

6 Summary and further remarks

We have reviewed reconstruction algorithms developed for optical tomography, fluorescence tomography and bioluminescence tomography based on the radiative transport equations. We constructed both linearized algorithms and nonlinear iterative algorithms and discussed briefly the properties of those algorithms. Since the inverse boundary value problems we considered in this paper are in general (severely) ill-posed, we do not expect very high accuracy in the numerical solution of those inverse problems. Most of algorithms provide reconstructions of very similar quality; see for example the results in Fig. 1. We are thus more interested in the speed of the reconstruction (which is not a huge problem for many traditional imaging scheme such as X-ray tomography). Iterative schemes of Newton type presented here are all local convergent with order of convergence close to 2, better than conjugate gradient type of schemes. The advantage of

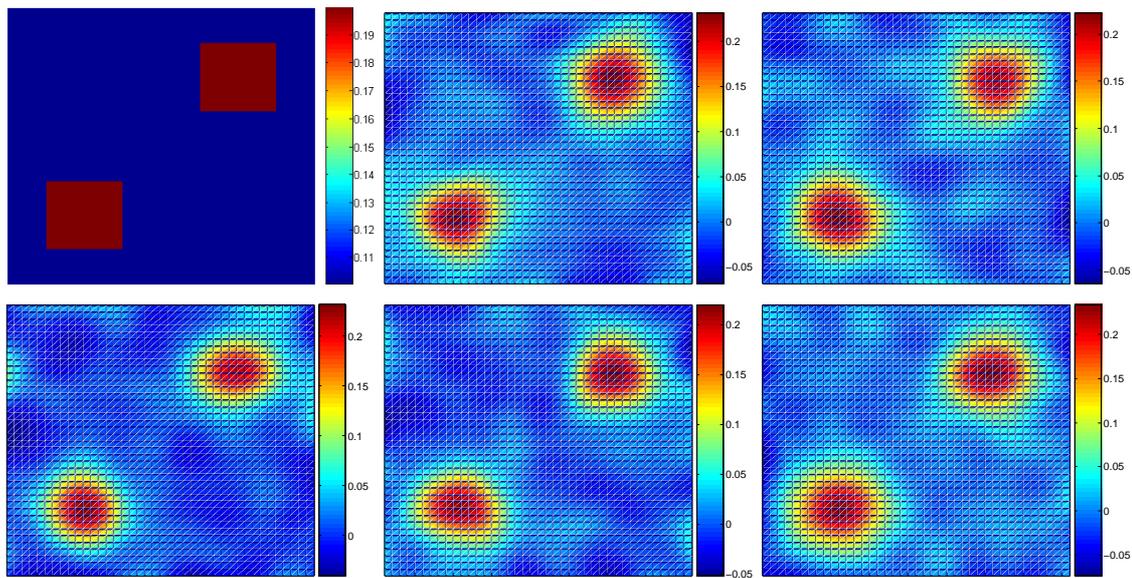


Figure 1: Reconstructions of the absorption coefficient in the two-dimensional domain $(0, 2) \times (0, 2)$. From top left to bottom right: real coefficient, reconstructions with Gauss-Newton, BFGS, Levenberg-Marquardt, Kaczmarz and augmented-Lagrangian methods. All reconstructions start from the same initial guess. Very small Tikhonov regularization (with the same parameter) has been applied to all reconstructions except for the Kaczmarz method where we simply applied a low-pass filter on the iteration the unknown in iteration (4.50).

augmented-Lagrangian type of method lies in the fact that it is easy to be implemented on parallel processors.

There is another special type of reconstruction methods that has not been presented in this paper: the Monte Carlo methods. Monte Carlo type of methods have been developed in [7, 62–64, 92]. Those methods are in general not as efficient as deterministic methods, but can be useful in special situations and are easier (than other deterministic methods) to implement when the domains of interest are of irregular shapes.

The reconstruction methods we discussed in this paper can be applied to almost any kinds of discretization on the radiative transport equation. In other words, the inversion methods are independent of how accurate the forward problem are discretized numerically, although the quality of the reconstructions will certainly depend on how the accurate the forward problems are solved. The focus on solving the forward problem in inverse transport applications is mostly on how to solve the problems fast enough. We will not be able to review results in the discretizations and solution of forward transport equations. We refer therefore, in a very subjective way and knowing that the list is by no means complete, interested readers to [4, 5, 51, 75, 76, 84, 87, 106].

The physics described by the radiative transport equation can be roughly classified into three regimes: the transport limit, the diffusion limit and the intermediate. The transport limit is the case when the underline scattering is very small so that photons can travel through the medium with little chance of getting scattered (i.e. change the

direction of traveling). In this case, inverse transport problems can be related to inverse (attenuated) Radon transform type of problems. There are analytical or half-analytical reconstruction methods. The resolution of the reconstruction is essentially controlled by sample resolution [49, 101], which can be very high. The diffusion limit is the case when absorption is sufficiently low and scattering sufficiently large. In this case, the transport process can be modeled macroscopically with the diffusion equation [41, 45], which in stationary case takes the form

$$\begin{aligned} -\nabla \cdot \mathcal{D} \nabla U + \Sigma_a(\mathbf{x})U(\mathbf{x}) &= 0, & \text{in } \Omega, \\ U + n\epsilon L_n \mathbf{v}(\mathbf{x}) \cdot \mathcal{D} \nabla U &= g(\mathbf{x}), & \text{on } \partial\Omega. \end{aligned} \quad (6.1)$$

Here $U(\mathbf{x})$ is the angularly-averaged photon flux at \mathbf{x} , an approximation of the quantity $\int_{S^{n-1}} u(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}$ in the transport equation. The diffusion coefficient $\mathcal{D}(\mathbf{x})$ is related to Σ_a and Σ_s . L_n and ϵ are known coefficients. Known results in [14] show that the inverse transport problem in diffusion regime is a severely ill-posed problem so that the resolution of the reconstruction is very low in practice. The reconstruction techniques we presented in the previous sections are mostly useful in the intermediate regime where we need to use the transport equation (instead of the diffusion equation) as the forward model but have no explicit reconstruction methods. Past numerical results show that even in regimes close to the diffusion regime, there are noticeable differences between reconstruction-based reconstructions and diffusion-based reconstructions [108]. This justifies somehow the use of transport equations in optical imaging.

In terms of future developments of numerical reconstruction methods for inverse transport problems, we believe the following aspects are very important.

The first aspect is to develop fast reconstruction methods to deal with problems with large data sets, about 10^6 larger than currently used [131]. It has been show that the use of those large data set can significantly improve the quality of the reconstructions. However, it is impossible to use those large data sets in the algorithms we have presented in this work since they are very slow. We believe that the combination of numerical methods with analytical or half analytical reconstruction methods, even for simple geometries, are important for future development in optical imaging with transport models [111].

The second aspect is to develop reconstruction methods that can efficiently utilize *a priori* information on the unknowns. Effective usage of *a priori* information can not only accelerate the reconstruction, but also improve the stability (thus the quality) of the reconstruction. The parameterization methods and the shape reconstruction methods we mentioned are clearly examples of such methods. One specific direction to go in this aspect is to follow the lines of the work in [53, 54].

The third aspect is to develop statistical methods for uncertainty quantification in inverse transport problems. In both optical tomography and optical molecular imaging, we assume that except for the objects we want to recover, all other parameters are known exactly. In practice, however, this is not true. For example, in BLT, we assume that the optical parameters are known and we only want to reconstruct the source term. However, the optical parameters are not known exactly since they come from a step of optical

tomography. We thus only know the parameters up to a certain accuracy. The uncertainty in the optical parameters will have an impact on the reconstruction of the source term. To characterize the uncertainty in the reconstruction of some parameters due to the uncertainty in other parameters, it is natural to introduce Bayesian type of statistical methods.

The fourth aspect is to develop methods that can utilize more efficiently time-dependent data. In all cases we have mentioned in this paper, time-dependent data are used by averaging information at different times; see for example, the linearization in (3.14) and the objective function (4.56). We know, however, photons reach the detector in early times carry different information about the medium than photons reach the detector in later times. Early photons are not scattered as much as later photons, so they carry mainly information on absorption property of the medium. Early time measurement would allow us to recover stably the absorption property. It would be of great interest to design methods that can efficiently utilize measurements in different time intervals.

Let us conclude the paper by the following remark. The inverse transport problems we considered in this paper are all inverse boundary value problems in the sense that the measurements for those problems (i.e., the data) are all taken on the boundary of the domain of interests. There are many other kinds of inverse boundary value problems that are of interests to practical applications. Many of the algorithms and ideas (such as parameterization and feature reconstruction) reviewed here can also be suitable for other inverse boundary value problems for differential equations.

Acknowledgments

The author would like to thank the anonymous referees for suggestions that helped us to improve the quality of this review paper. This work is partially supported by National Science Foundation (NSF) through grant DMS-0914825 and a faculty development award from the University of Texas at Austin. The author would like to thank Professor Guillaume Bal and Professor Andreas H. Hielscher (Columbia University) for stimulating discussions on inverse transport problems and applications in optical tomography and optical molecular imaging. The author is also grateful to Professor Simon Arridge (University College London) and Professor Hongkai Zhao (University of California at Irvine) for very useful discussions.

References

- [1] G. S. Abdoulaev and A. H. Hielscher, Three-dimensional optical tomography with the equation of radiative transfer, *J. Electron. Imaging*, 12 (2003), pp. 594–601.
- [2] G. S. Abdoulaev, K. Ren, and A. H. Hielscher, Optical tomography as a PDE-constrained optimization problem, *Inverse Probl.*, 21 (2005), pp. 1507–1530.
- [3] B. D. Abramov and Y. V. Matveev, Some inverse problems for reactor point kinetics, *Transport Theory and Statistical Physics*, 37 (2008), pp. 327–343.

- [4] M. L. Adams and E. W. Larsen, Fast iterative methods for discrete-ordinates particle transport calculations, *Prog. Nucl. Energy*, 40 (2002), pp. 3–150.
- [5] V. Agoshkov, *Boundary Value Problems for the Transport Equations*, Birkhauser, Boston, 1998.
- [6] V. Antyufeev, Reconstruction of a phase function of the transport equation, *Inverse Probl.*, 18 (2002), pp. 125–137.
- [7] V. S. Antyufeev, *Monte Carlo Method for Solving Inverse Problems of Radiation Transfer*, VSP, Utrecht, Netherlands, 2000.
- [8] 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, 1991.
- [9] S. R. Arridge, Optical tomography in medical imaging, *Inverse Probl.*, 15 (1999), pp. R41–R93.
- [10] S. R. Arridge, O. Dorn, J. P. Kaipio, V. Kolehmainen, M. Schweiger, T. Tarvainen, M. Vauhkonen, and A. Zacharopoulos, Reconstruction of subdomain boundaries of piecewise constant coefficients of the radiative transfer equation from optical tomography data, *Inverse Problems*, 22 (2006), pp. 2175–2196.
- [11] S. R. Arridge and J. C. Hebden, Optical imaging in medicine: II. modelling and reconstruction, *Phys. Med. Biol.*, 42 (1997), pp. 841–853.
- [12] G. Bal, Inverse problems for homogeneous transport equations: I. The one-dimensional case, *Inverse Probl.*, 16 (2000), pp. 997–1011.
- [13] ———, Inverse problems for homogeneous transport equations: II. The multidimensional case, *Inverse Probl.*, 16 (2000), pp. 1013–1028.
- [14] ———, Inverse transport theory and applications, *Inverse Problems*, 25 (2009). 053001.
- [15] G. Bal, L. Carin, D. Liu, and K. Ren, Experimental validation of a transport-based imaging method in highly scattering environments, *Inverse Probl.*, 23 (2007), pp. 2527–2539.
- [16] G. Bal, I. Langmore, and F. Monard, Inverse transport with isotropic sources and angularly averaged measurements, *Inverse Problems and Imaging*, 2 (2008), pp. 23–42.
- [17] G. Bal and K. Ren, Atmospheric concentration profile reconstructions from radiation measurements, *Inverse Probl.*, 21 (2005), pp. 153–168.
- [18] ———, Transport-based imaging in random media, *SIAM J. Appl. Math.*, 68 (2008), pp. 1738–1762.
- [19] L. B. Barichello, R. D. M. Garcia, and C. E. Siewert, On inverse boundary-conditions problems in radiative transfer, *J. Quant. Spectrosc. Radiat. Transfer*, 57 (1997), pp. 405–410.
- [20] A. Belleni-Morante, An inverse problem for photon transport in interstellar clouds, *Transport Theory and Statistical Physics*, 32 (2003), pp. 73–91.
- [21] A. Belleni-Morante and F. Mugelli, Identification of the boundary surface of an interstellar cloud from a measurement of the photon far-field, *Math. Meth. Appl. Sci.*, 27 (2004), pp. 627–642.
- [22] A. Belleni-Morante and G. F. Roach, Gamma ray transport in the cardiac region: an inverse problem, *J. Math. Anal. Appl.*, 269 (2002), pp. 200–215.
- [23] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
- [24] E. Boman, *Radiotherapy Forward and Inverse Problem Applying Boltzmann Transport Equations*, PhD thesis, University of Kuopio, Finland, Kuopio, Finland, 2007.
- [25] W. Cai, M. Xu, and R. R. Alfano, Three-dimensional radiative transfer tomography for turbid media, *IEEE Journal of Selected Topics in Quantum Electronics*, 9 (2003), pp. 189–

- 198.
- [26] K. M. Case, Inverse problem in transport theory, *Phys. Fluids*, 16 (1973), pp. 1607–1611.
 - [27] ———, Inverse problem in transport theory. II, *Phys. Fluids*, 18 (1975), pp. 927–930.
 - [28] ———, Inverse problem in transport theory. III, *Phys. Fluids*, 20 (1977), pp. 2031–2036.
 - [29] M. T. Chahine, Inverse problems in radiative transfer: Determination of atmospheric parameters, *J. Atmos. Sci.*, 27 (1970), pp. 960–967.
 - [30] A. J. Chaudhari, F. Darvas, J. R. Bading, R. A. Moats, P. S. Conti, D. J. Smith, S. R. Cherry, and R. M. Leahy, Hyperspectral and multispectral bioluminescence optical tomography for small animal imaging, *Phys. Med. Biol.*, 50 (2005), pp. 5421–5441.
 - [31] Y. Cheng, I. Gamba, and K. Ren, Recovering doping profiles in semiconductor devices with the Boltzmann-Poisson model. Preprint, 2010.
 - [32] S. R. Cherry, In vivo molecular imaging and genomic imaging: new challenges for imaging physics, *Phys. Med. Biol.*, 49 (2004), pp. R13–R48.
 - [33] M. Choulli and P. Stefanov, Inverse scattering and inverse boundary value problem for the linear Boltzmann equation, *Comm. Part. Diff. Eqn.*, 21 (1996), pp. 763–785.
 - [34] ———, Reconstruction of the coefficients of the stationary transport equation from boundary measurements, *Inverse Probl.*, 12 (1996), pp. L19–L23.
 - [35] ———, An inverse boundary value problem for the stationary transport equation, *Osaka J. Math.*, 36 (1998), pp. 87–104.
 - [36] A. X. Cong and G. Wang, Multispectral bioluminescence tomography: methodology and simulation, *Int. J. Biomed. Imag.*, 2006 (2006). 57614.
 - [37] W. Cong, K. Durairaj, L. Wang, and G. Wang, A Born-type approximation method for bioluminescence tomography, *Med. Phys.*, 33 (2006), pp. 679–686.
 - [38] C. H. Contag and M. H. Bachmann, Advances in in vivo bioluminescence imaging of gene expression, *Annu. Rev. Biomed. Eng.*, 4 (2002), pp. 235–260.
 - [39] A. Corlu, R. Choe, T. Durduran, K. Lee, M. Schweiger, S. R. Arridge, E. M. C. Hillman, and A. G. Yodh, Diffuse optical tomography with spectral constraints and wavelength optimisation, *Appl. Opt.*, 44 (2005), pp. 2082–2093.
 - [40] J. P. Culver, T. Durduran, D. Furuya, C. Cheung, J. H. Greenberg, and A. G. Yodh, Diffuse optical tomography of cerebral blood flow, oxygenation, and metabolism in rat during focal ischemia, *Journal of Cerebral Blood Flow & Metabolism*, 23 (2003), pp. 911–924.
 - [41] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology*, Vol VI, Springer-Verlag, Berlin, 1993.
 - [42] O. Dorn, A transport-backtransport method for optical tomography, *Inverse Probl.*, 14 (1998), pp. 1107–1130.
 - [43] ———, Scattering and absorption transport sensitivity functions for optical tomography, *Optics Express*, 7 (2000), pp. 492–506.
 - [44] ———, Shape reconstruction in scattering media with void using a transport model and level sets, *Canad. Appl. Math. Quart.*, 10 (2001), pp. 239–275.
 - [45] R. Elaloufi, R. Carminati, and J. Greffet, Time-dependent transport through scattering media: from radiative transfer to diffusion, *J. Opt. A: Pure Appl. Opt.*, 4 (2002), pp. S103–S108.
 - [46] 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. Am. A*, 5 (1988), pp. 366–373.
 - [47] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.
 - [48] F. Feng, P. Edström, and M. Gullikssone, Levenberg-Marquardt methods for parameter es-

- timination problems in the radiative transfer equation, *Inverse Problems*, 23 (2007), pp. 879–891.
- [49] L. Florescu, J. S. J., and V. Markel, Single-scattering optical tomography, *Phys. Rev. E*, 79 (2009). 036607.
- [50] P. Gabriel, M. J. Christi, and G. L. Stephens, Calculation of Jacobians for inverse radiative transfer: An efficient hybrid method, *J. Quant. Spectrosc. Radiat. Transfer*, 97 (2006), pp. 209–227.
- [51] H. Gao and H. Zhao, A fast forward solver of radiative transfer equation, *Transport Theory and Statistical Physics*, 38 (2009), pp. 149–192.
- [52] ———, A multilevel and multigrid optical tomography based on radiative transfer equation, in *Proceedings of SPIE, Munich, Germany, 2009*.
- [53] ———, Multilevel bioluminescence tomography based on radiative transfer equation. part 1: l1 regularization. Preprint, 2009.
- [54] ———, Multilevel bioluminescence tomography based on radiative transfer equation. part 2: total variation and l1 data fidelity. Preprint, 2009.
- [55] A. P. Gibson, J. C. Hebden, and S. R. Arridge, Recent advances in diffuse optical imaging, *Phys. Med. Biol.*, 50 (2005), pp. R1–R43.
- [56] P. González-Rodríguez and A. D. Kim, Reflectance optical tomography in epithelial tissues, *Inverse Problems*, 25 (2009). 015001.
- [57] H. L. Graber, J. Chang, R. Aronson, and R. L. Barbour, A perturbation model for imaging in dense scattering media: Derivation and evaluation of imaging operators, in *Proc. SPIE, vol. IS11, Int. Soc. Opt. Eng., 1993*, pp. 121–143.
- [58] E. E. Graves, R. Weissleder, and V. Ntziachristos, Fluorescence molecular imaging of small animal tumor models, *Current Molecular Medicine*, 4 (2004), pp. 419–430.
- [59] X. Gu, K. Ren, J. Masciotti, and A. H. Hielscher, Parametric reconstruction method in optical tomography, in *Proceedings of the 28th IEEE EMBS Annual International Conference, 2006*.
- [60] M. Gurfinkel, S. Ke, X. X. Wen, C. Li, and E. M. Sevick-Muraca, Near-infrared fluorescence optical imaging and tomography, *Disease Markers*, 19 (2003), pp. 107–121.
- [61] W. Han and G. Wang, Theoretical and numerical analysis on multispectral bioluminescence, *IMA J. Appl. Math.*, 72 (2007), pp. 67–85.
- [62] C. Hayakawa and J. Spanier, Perturbation Monte Carlo methods for the solution of inverse problems, in *Monte Carlo and Quasi-Monte Carlo Methods*, H. Niederreiter, ed., Springer-Verlag, Berlin, 2004, pp. 227–241.
- [63] C. K. Hayakawa, J. Spanier, F. Bevilacqua, A. K. Dunn, J. S. You, B. J. Tromberg, and V. Venugopalan, Perturbation monte carlo methods to solve inverse photon migration problems in heterogeneous tissues, *Opt. Lett.*, 26 (2001), pp. 1335–1337.
- [64] C. K. Hayakawa, J. Spanier, and V. Venugopalan, Coupled forward-adjoint monte carlo simulations of radiative transport for the study of optical probe design in heterogeneous tissues, *SIAM J. Appl. Math.*, 68 (2007), pp. 253–270.
- [65] J. C. Hebden, S. R. Arridge, and D. T. Delpy, Optical imaging in medicine: I. experimental techniques, *Phys. Med. Biol.*, 42 (1997), pp. 825–840.
- [66] B. J. Hoenders, Existence of invisible nonscattering objects and nonradiating sources, *J. Opt. Soc. Am. A*, 14 (1997), pp. 262–266.
- [67] K. Ito and K. Kunisch, The augmented lagrangian method for parameter estimation in elliptic systems, *SIAM J. Control Optim.*, 28 (1990), pp. 137–157.
- [68] M. Jiang, T. Zhou, J. Cheng, W. Cong, and G. Wang, Development of bioluminescence

- tomography, in *Developments in X-Ray Tomography*, U. Bonse, ed., vol. 6318 of Proc. SPIE, 2006.
- [69] A. Joshi, J. C. Rasmussen, E. M. Sevick-Muraca, T. A. Wareing, and J. McGhee, Radiative transport-based frequency-domain fluorescence tomography, *Phys. Med. Biol.*, 53 (2008), pp. 2069–2088.
 - [70] C. T. Kelley, *Iterative Methods for Optimization*, *Frontiers in Applied Mathematics*, Society of Industrial and Applied Mathematics, Philadelphia, 1999.
 - [71] M. E. Kilmer and E. de Sturler, Recycling subspace information for diffuse optical tomography, *SIAM J. Sci. Comput.*, 27 (2006), pp. 2140–2166.
 - [72] M. E. Kilmer, E. Miller, and C. Rappaport, QMR-based projection techniques for the solution of non-Hermitian systems with multiple right hand sides, *SIAM J. Sci. Comput.*, 23 (2001), pp. 761–780.
 - [73] A. D. Kim, Transport theory for light propagation in biological tissue, *J. Opt. Soc. Am. A*, 21 (2004), pp. 820–827.
 - [74] A. D. Kim, C. Hayakawa, and V. Venugopalan, Estimating tissue optical properties using the Born approximation of the transport equation, *Optics Lett.*, 31 (2006), pp. 1088–1090.
 - [75] A. D. Kim and M. Moscoso, Chebyshev spectral methods for radiative transfer, *SIAM J. Sci. Comput.*, 23 (2002), pp. 2075–2095.
 - [76] ———, Radiative transfer computations for optical beams, *J. Comput. Phys.*, 185 (2003), pp. 50–60.
 - [77] H. K. Kim and A. Charette, A sensitivity function-based conjugate gradient method for optical tomography with the frequency-domain equation of radiative transfer, *J. Quant. Spectrosc. Radiat. Transfer*, 104 (2007), pp. 24–39.
 - [78] H. K. Kim and A. H. Hielscher, A PDE-constrained SQP algorithm for optical tomography based on the frequency-domain equation of radiative transfer, *Inverse Problems*, 25 (2009). 015010.
 - [79] A. D. Klose and A. H. Hielscher, Fluorescence tomography with simulated data based on the equation of radiative transfer, *Opt. Lett.*, 28 (2003), pp. 1019–1021.
 - [80] ———, Quasi-Newton methods in optical tomographic image reconstruction, *Inverse Probl.*, 19 (2003), pp. 387–409.
 - [81] A. D. Klose, V. Ntziachristos, and A. H. Hielscher, The inverse source problem based on the radiative transfer equation in optical molecular imaging, *J. Comput. Phys.*, 202 (2005), pp. 323–345.
 - [82] E. W. Larsen, Solution of the inverse problem in multigroup transport theory, *J. Math. Phys.*, 22 (1981), pp. 158–160.
 - [83] ———, Solution of three-dimensional inverse transport problems, *Transport Theory and Statistical Physics*, 17 (1988), pp. 147–167.
 - [84] E. E. Lewis and W. F. Miller, *Computational Methods of Neutron Transport*, American Nuclear Society, La Grange Park, IL, 1993.
 - [85] 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, *Applied Optics*, 42 (2003), pp. 5181–5190.
 - [86] H.-Y. Li, A two-dimensional cylindrical inverse source problem in radiative transfer, *J. Quant. Spectrosc. Radiat. Transfer*, 69 (2001), pp. 403–414.
 - [87] G. I. Marchuk and V. I. Lebedev, *Numerical Methods in the Theory of Neutron Transport*, Harwood academic publishers, 1984.
 - [88] N. J. McCormick, Recent developments in inverse scattering transport methods, *Transport*

- Theory and Statistical Physics, 13 (1984), pp. 15–28.
- [89] ———, Methods for solving inverse problems for radiation transport - An update, *Transport Theory and Statistical Physics*, 15 (1986), pp. 159–172.
- [90] ———, Inverse radiative transfer problems: A review, *Nuclear Science and Engineering*, 112 (1992), pp. 185–198.
- [91] ———, Analytic inverse radiative transfer equations for atmospheric and hydrologic optics, *J. Opt. Soc. Am. A*, 21 (2004), pp. 1009–1017.
- [92] G. A. Mikhailov, *Parametric Estimates by the Monte Carlo Method*, VSP, Utrecht, Netherlands, 1999.
- [93] F. Natterer, Algorithms in tomography, in *The State of Art of Numerical Analysis*, I. S. Duff and G. A. Watson, eds., Clarendon Press, 1997.
- [94] F. Natterer, Numerical methods in tomography, *Acta Numer.*, 8 (1999), pp. 107–141.
- [95] F. Natterer, *The Mathematical of Computerized Tomography*, Classics in Applied Mathematics, SIAM, Philadelphia, 2001.
- [96] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM Monographs on Mathematical Modelling and Computation, SIAM, Philadelphia, 2001.
- [97] U. Netz, J. Beuthan, H. Capius, H. Koch, A. D. Klose, and A. Hielscher, Imaging of rheumatoid arthritis in finger joints by sagittal optical tomography, *Medical Laser Application*, 16 (2001), pp. 306–310.
- [98] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.
- [99] V. Ntziachristos, Fluorescence molecular imaging, *Annu. Rev. Biomed. Eng.*, 8 (2006), pp. 1–33.
- [100] M. A. O’Leary, D. A. Boas, B. Chance, and A. G. Yodh, Experimental images of heterogeneous turbid media by frequency domain diffusion photon tomography, *Opt. Lett.*, 20 (1995), pp. 426–428.
- [101] A. N. Panchenko, Inverse source problem of radiative transfer: a special case of the attenuated Radon transform, *Inverse Problems*, 9 (1993), pp. 321–337.
- [102] H. M. Park and W. J. Lee, The solution of inverse radiation problems using an efficient computational technique, *J. Quant. Spectrosc. Radiat. Transfer*, 73 (2002), pp. 41–53.
- [103] H. M. Park and T. Y. Yoon, Solution of inverse radiation problems using the Karhunen-Loève Galerkin procedure, *J. Quant. Spectrosc. Radiat. Transfer*, 68 (2001), pp. 489–506.
- [104] V. Prapavat, W. Runge, J. Mans, A. Krause, J. Beuthan, and G. Müller, Development of a finger joint phantom for the optical simulation of early stages of rheumatoid arthritis, *Biomedizinische Technik*, 42 (1997), pp. 319–326.
- [105] H. Quan and Z. Guo, Fast 3-D optical imaging with transient fluorescence signals, *Optical Express*, 12 (2004), pp. 449–457.
- [106] K. Ren, G. S. Abdoulaev, G. Bal, and A. H. Hielscher, Algorithm for solving the equation of radiative transfer in the frequency domain, *Optics Lett.*, 29 (2004), pp. 578–580.
- [107] K. Ren, G. Bal, and A. H. Hielscher, Frequency domain optical tomography based on the equation of radiative transfer, *SIAM J. Sci. Comput.*, 28 (2006), pp. 1463–1489.
- [108] ———, Transport- and diffusion-based optical tomography in small domains: A comparative study, *Applied Optics*, 46 (2007), pp. 6669–6679.
- [109] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2nd ed., 2003.
- [110] R. Sanchez and N. J. McCormick, On the uniqueness of the inverse source problem for linear particle transport theory, *Transport Theory and Statistical Physics*, 37 (2008), pp. 236–263.
- [111] J. C. Schotland and V. A. Markel, Fourier-Laplace structure of the inverse scattering prob-

- lem for the radiative transport equation, *Inverse Problems in Imaging*, 1 (2007), pp. 181–188.
- [112] M. Schweiger, S. Arridge, , and D. Delpy, Application of the finite-element method for the forward and inverse models in optical tomography, *J. Math. Imaging Vision*, 3 (1993), pp. 263–283.
- [113] C. Sendra and M. A. Box, Retrieval of the phase function and scattering optical thickness of aerosols: a radiative perturbation theory application, *J. Quant. Spectrosc. Radiat. Transfer*, 64 (2000), pp. 499–515.
- [114] C. E. Siewert, Inverse solutions to radiative-transfer problems based on the binomial or the Henyey-Greenstein scattering law, *J. Quant. Spectrosc. Radiat. Transfer*, 72 (2002), pp. 827–835.
- [115] ———, Inverse solutions to radiative-transfer problems with partially transparent boundaries and diffuse reflection, *J. Quant. Spectrosc. Radiat. Transfer*, 72 (2002), pp. 299–313.
- [116] V. Y. Soloviev, C. D’Andrea, M. Brambilla, G. Valentini, R. B. Schulz, R. Cubeddu, and S. R. Arridge, Adjoint time domain method for fluorescent imaging in turbid media, *Applied Optics*, 47 (2008), pp. 2303–2311.
- [117] V. Y. Soloviev, K. B. Tahir, J. McGinty, D. S. Elson, M. A. A. Neil, P. M. W. French, and S. R. Arridge, Fluorescence lifetime imaging by using time gated data acquisition, *Applied Optics*, 46 (2007), pp. 7384–7391.
- [118] R. P. Souto, H. F. C. Velho, S. Stephany, and E. S. Chalhoub, Performance analysis of radiative transfer algorithms for inverse hydrologic optics in a parallel environment, *Transport Theory and Statistical Physics*, 33 (2004), pp. 449–468.
- [119] R. J. D. Spurr, T. P. Kurosu, and K. V. Chance, A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval, *J. Quant. Spectrosc. Radiat. Transfer*, 68 (2001), pp. 689–735.
- [120] P. Stefanov, *Inverse problems in transport theory. Inside Out: Inverse Problems and Applications, Vol (47)*, MSRI publications, 2003.
- [121] P. Stefanov and A. Tamasan, Uniqueness and non-uniqueness in inverse radiative transport. Submitted, 2008.
- [122] A. Tamasan, An inverse boundary value problem in two-dimensional transport, *Inverse Problems*, 18 (2002), pp. 209–219.
- [123] ———, *Optical tomography in weakly anisotropic scattering media*, in *Contemporary Mathematics*, AMS, Providence, RI, 2003.
- [124] T. Tarvainen, M. Vauhonen, and S. R. Arridge, Gauss-Newton reconstruction method for optical tomography using the finite element solution of the radiative transfer equation, *J. Quant. Spectrosc. Radiat. Transfer*, 109 (2008), pp. 2767–2778.
- [125] E. A. Ustinov, Adjoint sensitivity analysis of radiative transfer equation: temperature and gas mixing ratio weighting functions for remote sensing of scattering atmospheres in thermal IR, *J. Quant. Spectrosc. Radiat. Transfer*, 68 (2001), pp. 195–211.
- [126] ———, Adjoint sensitivity analysis of radiative transfer equation: 2. Applications to retrievals of temperature in scattering atmospheres in thermal IR, *J. Quant. Spectrosc. Radiat. Transfer*, 73 (2002), pp. 29–40.
- [127] A. P. Wang and S. Ueno, An inverse problem in a three-dimensional radiative transfer, *Astrophys. Space Sci.*, 155 (1989), pp. 105–111.
- [128] G. Wang, Y. Li, and M. Jiang, Uniqueness theorems in bioluminescence tomography, *Med. Phys.*, 31 (2004), pp. 2289–2299.
- [129] G. Wang, H. Shen, W. Cong, S. Zhao, and G. Wei, Temperature-modulated bioluminescence

- tomography, *Optical Express*, 14 (2006), pp. 2289–2299.
- [130] J.-N. Wang, Stability estimates of an inverse problem for the stationary transport equation, *Ann. Inst. Henri Poincaré*, 70 (1999), pp. 473–495.
- [131] Z. Wang, G. Panasyuk, V. Markel, and J. Schotland, Experimental demonstration of an analytic method for image reconstruction in optical tomography with large data sets, *Optics Lett.*, 30 (2005), pp. 3338–3340.
- [132] H.-C. Zhou, Y.-B. Hou, D.-L. Chen, and C.-G. Zheng, An inverse radiative transfer problem of simultaneously estimating profiles of temperature and radiative parameters from boundary intensity and temperature measurements, *J. Quant. Spectrosc. Radiat. Transfer*, 74 (2002), pp. 605–620.