# Quantitative Photoacoustic Tomography

Hao Gao[1], Hongkai Zhao[2], and Stanley Osher[3]

[1] Department of Mathematics, University of California, Los Angeles, CA 90095, USA haog@math.ucla.edu
[2] Department of Mathematics, University of California, Irvine, CA 92697, USA zhao@math.uci.edu
[3] Department of Mathematics, University of California, Los Angeles, CA 90095, USA sjo@math.ucla.edu

**Summary.** This chapter focuses on quantitative photoacoustic tomography to recover optical maps from the deposited optical energy. After a brief overview of models, theories and algorithms, we provide an algorithm for large-scale 3D reconstructions, so-called gradient-based bound-constrained split Bregman method (GBSB).

## 1 Introduction

Photoacoustic tomography (PAT), a synergistic combination of ultrasound and optical imaging, has recently emerged as a potential imaging method to resolve optical contrasts with accurate quantification and high resolution [XW06, LW09, W09].

The first inverse problem in PAT concerns about the reconstruction of the deposited optical energy from the time-dependent boundary measurement of the acoustic pressure. Explicit inversion formulas exist for a large class of geometries of interest, when the problem is in free space, with constant sound speed, and without accounting of acoustic attenuation. The major efforts for this first step have been focused on the situations when any of aforementioned conditions is violated so that no explicit formulas exist. We refer the reader to other chapters in this volume and their references for the discussions of different topics in the first inverse problem in PAT.

The second inverse problem in PAT, so-called quantitative PAT (QPAT) consists of reconstructing optical maps, particularly the absorption coefficient or the chromophores, from the deposited optical energy that is recovered from the first step. In this chapter, we assume that the deposited optical energy is known and will focus on the simultaneous reconstruction of absorption coefficient and scattering coefficient. Please note that although the absorption map is usually of the major clinic interest, it is necessary to consider the

scattering map in reconstruction as well in order to accurately reconstruct the absorption map when the scattering coefficient is unknown.

There are mainly three methodologies for QPAT. First, using single optical illumination and assuming the scattering map is known, we can recover the absorption map [CAKB05, YWZJ07, BBVR08]; second, using multi-wavelength illuminations and assuming the spectral model of optical coefficients, both absorption and scattering maps can be obtained [CAB09]; third, using multiple optical illumination, so-called multiple-source QPAT (MS-QPAT), both absorption and scattering maps can be recovered [GZO10, BR10], for which the reconstruction uniqueness and stability estimates can be rigorously established [BU10]. MS-QPAT was also considered in [RN05, Z10] for the reconstruction of the absorption map. We will mainly focus on MS-QPAT in this chapter.

The major contribution of this chapter is to provide an algorithm for reconstructing optical maps in large-scale 3D QPAT, so-called GBSB, i.e., the gradient-based bound-constrained split Bregman method, based on the proposed method in [GZO10]. For the enhanced stability with respect to the initial guess or the noise, the simple bounds are imposed, and the solution is regularized with total variation (TV) norm [ROF92]. These two strategies are particularly important for scattering reconstruction, and thus for accurate absorption reconstruction. To be suitable for the large-scale computation, GBSB utilizes Quasi-Newton method as the inner loop with the computation of gradients rather than Jacobians, and split Bregman method as the outer loop that is particularly efficient for $L_1$-type optimization including TV-regularized problem.

The chapter is organized as follow. We first briefly overview existing models, theories, and algorithms of QPAT in Section 2; then discuss the ingredients of GBSB for the large-scale 3D QPAT, and present simulation results in Section 3; last end with a discussion section.

## 2 Overview of QPAT

### 2.1 Forward Models

The light migration in the photoacoustic imaging can be modeled by diffusion approximation (DA), a first-order phase approximation of radiative transport equation (RTE) in spherical harmonic bases [CZ67, A99]. DA so far is the most popular since its solutions can be solved practically and the solution methods are relatively simple [A99]. In contrast, although RTE is more accurate than DA, particularly in non-diffusive region, it hasn't been widely used mainly because it is challenging to solve RTE practically. However, the practical solvers of RTE on a daily laptop can be potentially available, advanced by state-of-art numerical algorithms [LM93, AL02, GZ09] and novel computer architectures [NVIDIA].

**Diffusion Approximation (DA)**

In DA, the photon density is simplified to be dependent only on the spatial variable $x$. On a bounded domain $\Omega$ with smooth boundary $\Gamma$, the DA with Robin boundary condition is

$$\begin{array}{c} -\nabla\cdot(D(\mathbf{x})\nabla\phi(\mathbf{x}))+\mu_a(\mathbf{x})\phi(\mathbf{x})=q(\mathbf{x}),\ \mathbf{x}\in\Omega,\\ 2\kappa D(\mathbf{x})(\mathbf{n}\cdot\nabla\phi(\mathbf{x}))+\phi(\mathbf{x})=q_b(\mathbf{x}),\ \mathbf{x}\in\Gamma, \end{array} \tag{1}$$

where $\mu_a$ is the absorption coefficient, $\mu_s$ the scattering coefficient , $\mu_s' = (1-g)\mu_s$ the reduced scattering coefficient with the anisotropic scattering factor $g$ (see (8)), $D$ so-called diffusion coefficient that can be defined by $D = 1/(3\mu_s')$, and $\kappa$ the constant for coupling the refraction index mismatch at the boundary, that can be usually determined through data fitting [A99]. Please note that DA is valid only in the diffusive regime, e.g., $\mu_a << \mu_s'$ and $\mathbf{x}$ is at least a few mean free paths away from the source. We refer the readers to [CZ67] for further discussions of DA and [A99] for DA based optical tomographic problems.

After discretization, DA (1) can be formulated as the linear system

$$A\Phi = Q. \tag{2}$$

The system matrix $A$, the source term $Q$ and the discretized density $\Phi$ in (2) are specific to discretization methods. For example, using finite element method (FEM) [J87] that is popular for DA solutions on $\Omega$ with irregular shapes, the photon density in piecewise-linear bases $\{\varphi_i, i \le N_p\}$ with $N_p$ nodes is

$$\phi(\mathbf{x}) = \sum_i \phi_i\varphi_i(\mathbf{x}) \text{ and } [\Phi]_i = \phi_i. \tag{3}$$

On the other hand, we discretize the optical coefficients in piecewise constants, i.e., $\{(\mu_{a,i}, \mu_{s,i}'), i \le N_t\}$ with $N_t$ elements in the mesh.

With this discretization, we have

$$[A]_{ij} = \textstyle\int_\Omega D(\nabla\varphi_i\cdot\nabla\varphi_j)d\mathbf{x} + \int_\Omega \mu_a\varphi_i\varphi_j d\mathbf{x} + \frac{1}{2\kappa}\int_\Gamma \varphi_i\varphi_j d\mathbf{x} \tag{4}$$

and

$$[Q]_j = \int_\Omega q\varphi_j d\mathbf{x} + \frac{1}{2\kappa}\int_\Gamma q_b\varphi_j d\mathbf{x}. \tag{5}$$

The available data in DA based QPAT have the form as the product of the absorption coefficient and the photon density, i.e., $\mu_a(\mathbf{x})\phi(\mathbf{x})$, which is assumed to be given here for all $x \in \Omega$ or can be reconstructed through acoustic inversion. In order to utilize the data in the discretized settings, we assume there are $N_d$ numerical detectors that are uniformly distributed across the domain $\Omega$. Although $N_d$ can be arbitrarily large, we will let it be equal to the number of freedom of the discretized $\mu_a$ to avoid redundancy, i.e., $N_d = N_t$. Let us also assume that there are $N_s$ optical illuminations in the

setting of MS-QPAT, each of which has a flux distribution $\Phi_i$ by solving (2) with the formula (4) and (5). Then the discrete mapping of the data can be thought of as the following linear functional with respect to $\mu_a$ and $\Phi$

$$F_j(\mu_a, \Phi_i) = \mu_{a,j} \sum_{j'=1}^{D} \alpha_{jj'} [\Phi_i]_{j'}, \ i \leq N_s, j \leq N_t, \tag{6}$$

where $\alpha_{jj'}$'s are the interpolation weights for the $j$th detector with $\sum_{j'} \alpha_{jj'} = 1$ and $D$, the degree of freedom of the used element (e.g., $D = 3$ for the triangular element).

**Radiative Transfer Equation (RTE)**

On a bounded domain $\Omega$ with angular space $\hat{S}$ (e.g., unit sphere in three dimensions (3D)), the steady-state RTE is

$$\hat{s}\cdot\nabla\phi(\mathbf{x},\hat{s})+\mu_t(\mathbf{x})\phi(\mathbf{x},\hat{s}) = \mu_a(\mathbf{x}) \oint_{\hat{S}} f(\hat{s},\hat{s}')\phi(\mathbf{x},\hat{s}')d\hat{s}'+q(\mathbf{x},\hat{s}), \ (\mathbf{x},\hat{s}) \in \Omega\times\hat{S}. \tag{7}$$

Now, the photon flux $\phi(\mathbf{x}, \hat{s})$ has not only spatial component, but also angular component, i.e., at position $\mathbf{x}$ in direction $\hat{s}$. The parameters in (7) are absorption coefficient $\mu_a$, scattering coefficient $\mu_s$, transport coefficient $\mu_t := \mu_a + \mu_s$ and the scattering kernel $f$. Very often $f$ is rotationally invariant, i.e., $f(\hat{s}, \hat{s}') = f(\hat{s} \cdot \hat{s}')$, and is normalized with $\oint f(\hat{s} \cdot \hat{s}')d\hat{s}' = 1$. For example, a popular $f$ for modeling photon transport in tissues is

$$f(\hat{s} \cdot \hat{s}') = \frac{1 - g^2}{4\pi(1 + g^2 - 2g\hat{s} \cdot \hat{s}')^{3/2}}, \tag{8}$$

so-called Henyey-Greenstein (H-G) function (3D version), where the anisotropy parameter $0 \leq g \leq 1$ measures the forward peaking of the scattering.

Given the normal $\hat{n}$ at the boundary $\Gamma$, let $\Gamma^+(\Gamma^-)$ represent $(\mathbf{x}, \hat{s}) \in \partial\Omega \times \hat{S}$ with $\hat{s} \cdot \hat{n} > 0 (\leq 0)$. Let $n_i(n_o)$ be the refraction index of the medium (environment). The reflection boundary condition is

$$\phi(\mathbf{x}, \hat{s}) = R(\mathbf{x}, \hat{s}, \hat{s}')\phi(\mathbf{x}, \hat{s}') + q_b(\mathbf{x}, \hat{s}), \quad (\mathbf{x}, \hat{s}) \in \Gamma^-, \ (\mathbf{x}, \hat{s}') \in \Gamma^+, \tag{9}$$

where $q_b$ is the boundary source, $R$ is the reflection energy ratio that can be computed through Fresnel formula [J99] and $\hat{s}'$ is the angle that reflects into $\hat{s}$ at the boundary $\mathbf{x}$. For the vacuum boundary condition that there is no refraction index mismatch, i.e., $n_i = n_o$, $R = 0$ in (9). Please note that the boundary condition for RTE is prescribed only for the incoming flux $\phi$ on $\Gamma^-$. We refer the readers to [CZ67] for the thorough treatment of RTE and [LM93] for its introductory numerical methods.

Similarly, RTE (7) along with the boundary condition (9) can be formally set up as the linear system (2), although (2) is rarely explicitly formulated

for RTE due to its memory requirement. For example, following [GZ09], we can discretize the angular variable $\hat{s}$ with FEM in piecewisely linear bases $\{\varphi^a_m, m \le M\}$ with $M$ angular directions and the spatial variable $\mathbf{x}$ with Discontinuous Galerkin method (DG) [CKS00] in piecewisely linear bases $\{\varphi^s_{ij}, i \le N_t, j \le D\}$ (e.g., $D = 4$ on tetrahedral meshes) with $N_t$ spatial elements. That is,

$$\phi(\mathbf{x}, \hat{s}) = \sum_{i,j,m} \phi_{ijm}\varphi^s_{ij}(\mathbf{x})\varphi^a_m(\hat{s}) \text{ and } [\boldsymbol{\Phi}]_{ijm} = \phi_{ijm}. \tag{10}$$

With this discretization, we have

$$\begin{aligned}
&[A]_{ijm,i'j'm'} = -\delta_{mm'}\delta_{ii'}\textstyle\int_{\Delta_{i'}} \varphi^s_{ij}(\hat{s}_{m'} \cdot \nabla\varphi^s_{i'j'})d\mathbf{x} \\
&+\delta_{mm'}\delta_{ii'}\textstyle\int_{\Gamma^+_{i'}} \varphi^s_{ij}\varphi^s_{i'j'}|\hat{s}_{m'} \cdot \hat{n}|d\mathbf{x} \\
&-\delta_{mm'}\delta_{T(i)i'}\textstyle\int_{\Gamma^-_{i'}} \varphi^s_{ij}\varphi^s_{ij'}|\hat{s}_{m'} \cdot \hat{n}|d\mathbf{x} - \delta_{R(m)m'}\delta_{ii'}\textstyle\int_{\Gamma^-} r_{mm'}\varphi^s_{ij}\varphi^s_{i'j'}|\hat{s}_{m'} \cdot \hat{n}|d\mathbf{x} \\
&+\delta_{mm'}\delta_{ii'}\mu_{t,i}\textstyle\int_{\Delta_{i'}} \varphi^s_{ij}\varphi^s_{i'j'}d\mathbf{x} + \delta_{ii'}\mu_{s,i}\textstyle\int_{\Delta_{i'}} w_{mm'}\varphi^s_{ij}\varphi^s_{i'j'}d\mathbf{x}
\end{aligned} \tag{11}$$

and

$$[Q]_{i'j'm'} = \int_{\Delta_{i'}} q\varphi^s_{i'j'}d\mathbf{x} + \int_{\Gamma^-} q_b\varphi^s_{i'j'}|\hat{s}_{m'} \cdot \hat{n}|d\mathbf{x}. \tag{12}$$

In (11), $\Delta_i$ denotes the spatial element, $\Gamma^+_i$ (resp. $\Gamma^-_i$) the internal edge or surface (excluding the domain boundary $\Gamma$) with the normal direction $\mathbf{n}$ satisfying $\hat{s}_m \cdot \mathbf{n} > 0$ (resp. $\hat{s}_m \cdot \mathbf{n} < 0$), $T(i)$ the upwind flux function that localizes the element that provides upwind fluxes to the element $i$, $R(m)$ the reflection function that localizes the direction from which the direction $m$ is reflected, $r_{mm'}$ the reflection weight into $m$ from the direction $m'$ that is discretized from $R$ in (9), the optical coefficients are assumed to be piecewisely constant, i.e., $\mu_{a,i}$ and $\mu_{t,i}$, the piecewisely linear scattering function $\tilde{f}$, i.e.,$\tilde{f} = \sum_m f_m\varphi^a_m$, and the angularly scattering weights $w_{mm'}$, i.e.,$w_{mm'} = \int_{\hat{S}} \tilde{f}\varphi^a_{m'}d\hat{s}$. We refer the readers to [GZ09] for details of solution algorithms and [GZ10] for error estimates.

Now, the available data in RTE based QPAT have the form as the product of the absorption coefficient and angularly-averaged fluxes, i.e., $\mu_a(\mathbf{x}) \int_{\hat{S}} \phi(\mathbf{x}, \hat{s})d\hat{s}$. Same as before, let us also assume that there are $N_s$ optical illuminations in the setting of MS-QPAT, each of which has a flux distribution $\boldsymbol{\Phi}_i$ by solving (2) with the formula (11) and (12). Then the discrete data for RTE can be formulated as follow

$$F_j(\mu_a, \boldsymbol{\Phi}_i) = \mu_{a,j} \sum_{j'=1}^{D} \alpha_{jj'}\Big(\sum_{m'=1}^{M} w_{m'}[\boldsymbol{\Phi}_i]_{jj'm'}\Big),\ i \le N_s, j \le N_t, \tag{13}$$

where $\alpha_{jj'}$'s again are the interpolation weights from the $j$th detector with $\sum_{j'} \alpha_{jj'} = 1$, and the angular weight $w_m = \int_{\hat{S}} \varphi^a_m d\hat{s}$.

### 2.2 Theory

In this section, we will discuss some established uniqueness and stability estimates mainly for MS-QPAT under single optical spectrum.

For QPAT based on DA, the reconstruction uniqueness and stability estimates were established for cases with at least two optical illuminations [BU10]. A non-uniqueness result was obtained in [BR10-2] for the case with single optical illumination, in which the available data are assumed to have the form $\mu_a\phi^2$ rather than $\mu_a\phi$ that comes from a different acousto-optic model [BS10]. Assuming the available data have the form $\gamma\mu_a\phi$ ($\gamma$: the Grüneisen coefficient), it was shown in [BR10] that only two out of the three coefficients $(\mu_a, D, \gamma)$ can be uniquely recovered even for an arbitrary number of illuminations.

The QPAT based on RTE was analyzed in [BJJ10]. The analysis assumes that the measurement operator can be decomposed into singular components, i.e., the ballistic component, the single scattering component and the multiple scattering component. With the ballistic component (the most singular component), $\mu_a$ and $\mu_s$ in (7) can be reconstructed; with the single scattering component, the anisotropy coefficient $g$ in H-G function (8) can be uniquely reconstructed. All the above parameters are obtained with Hölder-type stability. We refer the interested readers to [BJJ10] for further details. Please note that the multi-source setting was not considered in [BJJ10], since it is sufficient to recover $(\mu_a, \mu_s, g)$ assuming that each component can be well separated from the data. This assumption is appropriate when RTE is in the transport regime, which may limit its practical interest since in rare cases the regions of practical interest can be fully characterized by transport regime. That is RTE based QPAT with single optical illumination may not reconstruct $(\mu_a, \mu_s, g)$ due to the potential failure of extracting singular components in many practical cases. However, in this case, we should still be able to reconstruct at least $(\mu_a, \mu_s)$ in the setting of MS-QPAT assuming $g$ is known, since $(\mu_a, \mu_s')$ can be reconstructed for DA based MS-QPAT while the behavior of RTE is close to DA when the regime transits from being transport-like to diffusive-like. On the other hand, it is unclear whether $g$ can be recovered as well even in MS-QPAT in the hybrid regime of transport and diffusion.

#### DA Based QPAT

The discussions here follow the results in [BU10]. Slightly different from the aforementioned DA model (1), the considered model is the following DA with Dirichlet boundary condition on a bounded domain $X$ with the boundary $\Gamma$

$$\begin{gathered} -\nabla\cdot(D(\mathbf{x})\nabla\phi_i(\mathbf{x}))+\mu_a(\mathbf{x})\phi_i(\mathbf{x})=0,\ \mathbf{x}\in\Omega,\\ \phi_i(\mathbf{x})=q_i(\mathbf{x}),\ \mathbf{x}\in\Gamma, \end{gathered} \tag{14}$$

and the available data with $N_s$ optical illuminations are

$$F:=\{F_i=\mu_a(\mathbf{x})\phi_i(\mathbf{x}),\ i\le N_s\}. \tag{15}$$

The object here is to reconstruct $(\mu_a, D)$ from $F$, that is equivalent to reconstruct $(\mu_a, \mu_s')$ . Suppose that the optical coefficients $(\mu_a, D) \in V$ are sufficiently smooth, i.e.,

$$V = \{(\mu_a, D) \text{ such that } (\mu_a, \sqrt{D}) \in C^{k+1} \times Y, ||\mu_a||_{C^{k+1}} + ||\sqrt{D}||_Y \leq M\}, \tag{16}$$

where $k \geq 1$ and $Y$ is some subspace of $C^{k+2}$, the following uniqueness theorem (Theorem 1) with $N_s = 2$ and the stability estimate (Theorem 2) with $N_s = 2n$ ($n$: the dimension of the spatial domain $X$) can be obtained.

**Theorem 1.** *Let $N_s = 2$ and $\Omega$ be an open, bounded domain with $C^2$ boundary $\Gamma$. Assume $(\mu_a, D), (\tilde{\mu}_a, \tilde{D}) \in V$ and $D = \tilde{D}$ on $\Gamma$. Let $F$ and $\tilde{F}$ be the data for coefficients $(\mu_a, D)$ and $(\tilde{\mu}_a, \tilde{D})$ respectively. Then there is an open set of illuminations $(q_1, q_2) \in (C^{1,\alpha}(\Gamma))^2$ for some $\alpha > \frac{1}{2}$ such that if $F = \tilde{F}$, then $(\mu_a, D) = (\tilde{\mu}_a, \tilde{D})$.*

**Theorem 2.** *Let $N_s = 2n$, $k \geq 2$ and $\Omega$ be an arbitrary bounded domain with $C^{k+1}$ boundary $\Gamma$. Assume $(\mu_a, D), (\tilde{\mu}_a, \tilde{D}) \in V$ and $D = \tilde{D}$ on $\Gamma$. Let $F$ and $\tilde{F}$ be the data for coefficients $(\mu_a, D)$ and $(\tilde{\mu}_a, \tilde{D})$ respectively. Then there is an open set of illuminations $(q_1, \cdots, q_{2n}) \in (C^{k,\alpha}(\Gamma))^{2n}$ and a constant $C$ such that*

$$||\mu - \tilde{\mu}_a||_{C^k} + ||D - \tilde{D}||_{C^k} \leq C||F - \tilde{F}||_{(C^{k+1})^{2n}}. \tag{17}$$

Both theorems can be proved by first using the Liouville transform $u = \sqrt{D}\phi$ to transform the DA (14) into the Schrödinger equation

$$\begin{array}{c} \triangle u_i + \nu u_i = 0, \ \mathbf{x} \in \Omega, \\ u_i = \sqrt{D} q_i, \ \mathbf{x} \in \Gamma, \end{array} \tag{18}$$

with

$$\nu = -\frac{\triangle\sqrt{D}}{\sqrt{D}} - \frac{\mu_a}{D}, \tag{19}$$

and then proving the equivalent theorems by complex geometric optics (CGO) solutions for the Schrödinger equation (18) with the internal data $\{\mu u_i, i \leq N_s\}$, where $\mu$ is defined as

$$\mu = \frac{\mu_a}{\sqrt{D}}. \tag{20}$$

Once it is shown for the Schrödinger equation (18) that $(\mu, \nu)$ can be uniquely (resp. stably) reconstructed from $\{\mu u_i, i \leq N_s\}$, both theorems follow automatically for DA.

The assumption $N_s = 2n$ in Theorem 2 can be relaxed to $N_s = 2$ with additional geometric hypothesis on $X$. We refer the interested readers to [BU10] for this alternative stability estimate and the details of the proofs. Furthermore, the changes of variables (19) and (20) provide a constructive method for recovering $(\mu_a, D)$ as will further discussed next.

### 2.3 Algorithm

Given the data $Y = \{Y_{ij},\ i \le N_s, j \le N_t\}$, the objective of QPAT is to obtain the optical maps $X$ by minimizing the difference between the model prediction $F$ and the actual data $Y$

$$f(X) = \frac{1}{2}||F(X) - Y||_2^2 = \frac{1}{2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_t}(F_{ij}(X) - Y_{ij})^2. \tag{21}$$

For example, $X = \mu_a$ or $X = (\mu_a, \mu_s')$ in DA based QPAT; $X = \mu_a$ or $X = (\mu_a, \mu_s)$ in RTE based QPAT assuming $g$ is known. Please note that the simultaneous reconstruction of both optical maps based on DA is not unique in the conventional setting with single optical illumination under single wavelength. Similarly, in the RTE based QPAT, since it seems very unlikely to extract singular components in the practical cases that are not fully in the transport regime, it should be more appropriate to consider RTE based MS-QPAT as well in order to reconstruct both optical maps. Therefore, in the following we consider the setting of MS-QPAT with $N_s > 1$.

Here the model prediction $F$ is consistent with the previous definitions (6) or (13), i.e.,

$$F = \{F_{ij} := P_j^T \Phi_i,\ i \le N_s, j \le N_t\}, \tag{22}$$

where $P_j$ is the detection functional defined by (6) for DA or (13) for RTE that consists of $\mu_a$ and interpolation weights, and $\Phi_i$ is the photon density solution from the $i$th illumination that can be obtained from solving linear systems $A\Phi_i = Q_i$ with the formulas (4) and (5) for DA model, and the formulas (11) and (12) for RTE model.

#### Fixed-point Iteration

Assuming the knowledge of the scattering map, fixed-point iteration can be used to reconstruct $\mu_a$, in which $\mu_a$ and $\Phi$ are sequentially updated in iterations. That is formally the following with the initial guess $\mu_a^0$

$$\mu_a^{n+1} = \frac{Y}{\Phi(\mu_a^n)}, \tag{23}$$

where we implicitly assume $N_s = 1$, and the interpolation may be necessary for mapping $\Phi$ to the same discretized location of $\mu_a$, i.e., with the interpolation weights in (6) or (13). In this simple fixed-point iteration scheme (23), the next solution $\mu_a^{n+1}$ is directly updated each time accompanied by solving a linear system for $\Phi$ with respect to the last solution $\mu_a^n$.

Here the initial guess $\mu_a^0$ has to be considerably close to the true distribution in order for $\mu_a$ to converge. Overall, it is sensitive to the initial guess $\mu_a^0$, the insufficient knowledge of the scattering property, and the data noise.

**DA Based Non-iterative Methods**

With DA as forward model, QPAT can be solved in a non-iterative fashion. Before starting the discussion of non-iterative methods, it is important to notice that these methods can be non-iterative in the ideal setting, e.g., noise-free cases. In the practical setting, the non-iterative reconstruction can be problematic, for which the iterative regularization or optimization of the solution is usually required.

When the diffusion coefficient $D$ or the reduced scattering coefficient $\mu_s'$ is assumed to be known, one can simplify the DA (1) by inserting the data $F = \mu_a \phi$ directly, i.e.,

$$\begin{array}{c} -\nabla \cdot (D\nabla\phi) = -F, \ \in \Omega, \\ 2AD(\mathbf{n} \cdot \nabla\phi) + \phi = q, \ \mathbf{x} \in \Gamma. \end{array} \tag{24}$$

After solving (24) for $\phi$, $\mu_a$ can be simply computed by

$$\mu_a = \frac{F}{\phi}. \tag{25}$$

In the second case that $D$ is not known in $\Omega$, $(\mu_a, D)$ can still be recovered non-iteratively [BR10] assuming at least two optical illuminations i.e., $\{F_i = \mu_a \phi_i, \ i = 1, 2\}$, the exact $D$ in $\Gamma$ and the following DA with Dirichlet boundary condition

$$\begin{array}{c} -\nabla \cdot (D\nabla\phi_i) + \mu_a \phi_i = 0, \ \mathbf{x} \in \Omega, \\ \phi_i = q_i, \ \mathbf{x} \in \Gamma. \end{array} \tag{26}$$

From Dirichlet boundary condition and the data $F$, $\mu_a$ is also exactly known in $\Gamma$, i.e., $\mu_a = F_i/q_i$. The algorithm goes as follow.

By multiplying DA (14) for $\phi_1$ by $\phi_2$, DA (26) for $\phi_2$ by $\phi_1$, subtracting two equations, we have

$$-\nabla \cdot (D\phi_1^2 \nabla \frac{\phi_2}{\phi_1}) = 0. \tag{27}$$

Using Liouville variable $u = \sqrt{D}\phi_1$, we obtain the first-order transport equation with the variable $u^2$

$$\begin{array}{c} -\nabla \cdot (u^2 \nabla \frac{F_2}{F_1}) = 0, \ \mathbf{x} \in \Omega, \\ u^2 = Dq_1^2, \ \mathbf{x} \in \Gamma. \end{array} \tag{28}$$

After solving (28) for $u$, we can obtain $\nu$ (19) and $\mu$ (20) by formulas

$$\begin{array}{c} \mu = \frac{\mu_a}{\sqrt{D}} = \sqrt{\frac{(\mu_a \phi_1)^2}{D\phi_1^2}} = \frac{F_1}{u}, \\ \nu = -\frac{\triangle\sqrt{D}}{\sqrt{D}} - \frac{\mu_a}{D} = -\frac{\triangle(\sqrt{D}\phi_1)}{\sqrt{D}\phi_1} = -\frac{\triangle u}{u}, \end{array} \tag{29}$$

where DA (26) is used in deriving the second formula of (29).

Next observing that

$$\triangle\sqrt{D} + \nu\sqrt{D} = -\mu, \tag{30}$$

from (19) or the second formula of (29), we can compute $\sqrt{D}$ from $\nu$ and $\mu$ by solving the elliptic equation (30) with Dirichlet boundary condition, i.e., the exact values of $\sqrt{D}$ on $\Gamma$. Consequently, we can compute $D$, and $\mu_a = \mu\sqrt{D}$.

In summary, one can reconstruct $(\mu_a, D)$ from two measurements by solving a first-order transport equation (28) and an elliptic equation (30). Now we state one more assumption on $F$ in order for the solution to make sense. That is

$$\beta := F_1^2 \nabla \frac{F_2}{F_1} \tag{31}$$

is a vector in $W^{1,\infty}$ and $|\beta| \geq \alpha_0 > 0$.

In implementation, since $\beta$ usually varies significantly over the domain, the transport equation (28) is solved in the normalized vector field $|\beta|/\mu^2$ instead. In the practical case with noise, the derivative of $F_2/F_1$ can become problematic. One way to deal with the noise is to formulate (28) as a linear system and iteratively solve it using some proper regularization.

This non-iterative reconstruction method can be potentially appealing for its non-iterative nature. However, we will not adopt this approach here in designing our solver for 3D large-scale QPAT, mainly because $(\mu_a, D)$ is assumed to be known on $\Gamma$. Another reason is that it seems that the practical performance is data-dependent, such as $\beta$ and its lower bound $\alpha_0$.

### Jacobian Based Method

In the Jacobian based method [GZO10], the QPAT is formulated as a least-square problem (21) with the forward model $F$ that is nonlinearly dependent on optical parameters $X$, and then the Jacobians of $F$ are iteratively computed so that the consequent subproblems are linear least square problems that can be solved by various least-square or convex optimization techniques [BV04, NW06, GZ10].

Jacobian based method is a natural way for solving (21) as an typical iterative linearization approach for nonlinear problems. That is the following with some initial guess $X^0$,

$$\begin{aligned} B^{n+1} &= Y + J(X^n)X^n - F(X^n), \\ X^{n+1} &= \text{argmin}_X ||J(X^n)X - B^{n+1}||_2^2 + R(X), \end{aligned} \tag{32}$$

where $J(X^n)$ is the Jacobian computed at $X^n$, and $R(X)$ the regularization term to regularize the solution that can be based on some priors of $X$, such as smoothness and sparsity. Another consequence for the image regularization is that the minimization problems of (32) are usually less illposed. This can be interpreted as a result of finding solutions in a more restrictive space with the global minimum.

The main advantages of Jacobian based method (32) are its simplicity and flexibility. For example, many state-of-art image processing techniques can be potentially applied here. But, for large-scale QPAT in 3D, the computation of Jacobian $J$ can be prohibitively slow, as will be explained next.

$J$ can be computed as follow. First we differentiate both sides of (2) with respect to $x_k$, each component of $X$,

$$A\frac{\partial \Phi_i}{\partial x_k} = -\frac{\partial A}{\partial x_k}\Phi_i \tag{33}$$

and use it to compute the Jacobian in the following way

$$\begin{aligned}
&[J]_{ij,k} = \tfrac{\partial F_{ij}}{\partial x_k} \\
&= \tfrac{\partial P_j^T}{\partial x_k}\Phi_i + P_j^T\tfrac{\partial \Phi_i}{\partial x_k} \\
&= \tfrac{\partial P_j^T}{\partial x_k}\Phi_i - P_j^T[A^{-1}(\tfrac{\partial A}{\partial x_k}\Phi_i)] \\
&= \tfrac{\partial P_j^T}{\partial x_k}\Phi_i - (P_j^T A^{-1})(\tfrac{\partial A}{\partial x_k}\Phi_i) \\
&= \tfrac{\partial P_j^T}{\partial x_k}\Phi_i - (\Psi_j^T)(\tfrac{\partial A}{\partial x_k}\Phi_i),
\end{aligned} \tag{34}$$

where $\Psi_j$ is so-called adjoint solution defined by

$$\Psi_j = (A^T)^{-1}P_j \text{ or } A^T\Psi_j = P_j. \tag{35}$$

As a result, this method for computing Jacobian is so-called the adjoint method, which is usually much more efficient than the direct method [A99]. But, it still requires the computation of linear systems (2) for $N_s + N_t$ times, that can be extremely time-consuming for 3D large-scale QPAT since $N_t$ can easily be around a million or more in 3D.

### Gradient Based Method

An alternative way without computing Jacobians is to consider (21) as a fully nonlinear function instead of the least-square problem [GZO10]. That is

$$X = \text{argmin}_X f(X) + R(X), \tag{36}$$

where $R(X)$ again is the regularization on the solution $X$.

Now we only need to compute the gradient of $f(X)$. The adjoint method for gradient computation goes as follow

$$\begin{aligned}
&[\partial f]_k = \textstyle\sum_{i,j}(F_{ij} - Y_{ij})\tfrac{\partial F_{ij}}{\partial x_k} \\
&= \textstyle\sum_{i,j}(F_{ij} - Y_{ij})\tfrac{\partial P_j^T}{\partial x_k}\Phi_i + \sum_{i,j}(F_{ij} - Y_{ij})P_j^T\tfrac{\partial \Phi_i}{\partial x_k} \\
&= \textstyle\sum_{i,j}(F_{ij} - Y_{ij})\tfrac{\partial P_j^T}{\partial x_k}\Phi_i - \sum_i S_i^T[A^{-1}(\tfrac{\partial A}{\partial x_k}\Phi_i)] \\
&= \textstyle\sum_{i,j}(F_{ij} - Y_{ij})\tfrac{\partial P_j^T}{\partial x_k}\Phi_i - \sum_i \Psi_i^T(\tfrac{\partial A}{\partial x_k}\Phi_i)
\end{aligned} \tag{37}$$

with the adjoint source

$$S_i = \sum_j (F_{ij} - Y_{ij}) P_j \tag{38}$$

and the adjoint solution

$$\Psi_i = (A^T)^{-1} S_i \text{ or } A^T \Psi_i = S_i. \tag{39}$$

Please note that the first term in (34) and (37) is the singular absorption term that is nonzero only for the absorption coefficient $\mu_a$, which reflects the fact that the sensitivity with respect to $\mu_a$ is stronger than with respect to the other variable. Consequently the reconstruction of $\mu_a$ should have a better resolution. On the other hand, the aforementioned fixed-point iteration actually corresponds to the truncated Jacobian involving only the first term for the absorption coefficient. Therefore, fixed-point iteration usually fails when the first term is no longer dominating or with considerable errors.

Now, for each gradient we only need to compute linear systems (2) for $2N_s$ times that are only a few. In contrast with the method based on Jacobians (34), the method based on gradients (37) requires much fewer computations of linear systems, and thus is feasible for 3D QPAT. The gradient based method has been developed for QPAT by [CAB09, GZO10]. It is also the base of GBSB algorithm for large-scale 3D QPAT that will be introduced next.

## 3 GBSB: an Algorithm for Large-scale 3D QPAT

In this section, we will aim at solving large-scale 3D QPAT, e.g., to recover both absorption map and scattering map in the setting of MS-QPAT. In the following, QPAT will be formulated as a bound-constrained nonlinear minimization problem with the solution regularized by TV norm, and then the development of the solution algorithm is based on the Split Bregman method [OBGXY05, GO09, COS09], namely, GBSB, gradient-based bound-constrained split Bregman method, which is an extension of our prior work [GZO10].

### 3.1 Formulation

For model simplicity, we adopt the following setting that is commonly assumed for QPAT: with DA as the forward model, the objective is to reconstruct the absorption coefficient $\mu_a$ and the reduced scattering coefficient $\mu_s'$ from the absorbed energy $\mu_a \phi$, which is assumed to be known (e.g., through acoustic inversion).

In the setting of MS-QPAT with $N_s$ optical illuminations, the available data $Y$ are

$$Y = \{Y_{ij} := F_j(\mu_a, \phi_i) + \epsilon_{ij},\ i \le N_s, j \le N_d\}, \tag{40}$$

where $\phi_i$ is the DA solution with the $i$th source, $F$ the measuring functional defined by (6) that is linearly dependent on $\mu_a$ and $\phi_i$, $N_d$ the number of the detectors, and $\epsilon_{ij}$ the data noise.

The parameters to be recovered are

$$X = (\mu_a, \mu_s'). \tag{41}$$

Assuming that the data noise $\epsilon$ obeys the Gaussian distribution, we formulate the data fidelity of QPAT as the nonlinear least-square summation

$$f(X) = \frac{1}{2}||F(X) - Y||_2^2 = \frac{1}{2}\sum_{i,j}[F_j(\mu_a, \phi_i(\mu_a, \mu_s')) - Y_{ij}]^2. \tag{42}$$

In this study, we specifically choose piecewise-constant discretization of $X$

$$X = \{(\mu_{a,i}, \mu_{s,i}'), i \le N_t\}, \tag{43}$$

and regularize it by TV norm

$$||MX||_1 = \sum_i |M_{i1}\mu_{a,i1} - M_{i2}\mu_{a,i2}| + |M_{i1}\mu_{s,i1}' - M_{i2}\mu_{s,i2}'|, i \le N_e. \tag{44}$$

Here $N_e$ is the total number of element edges in 2D or element surfaces in 3D that are inside the domain (excluding the domain boundary). It can be shown by TV coarea formula that $M_{i1}$ and $M_{i2}$ correspond to the edge length or the surface area [GZ10-2]. Please note that the triangulation is assumed to be conformal so that each internal edge or surface $i$ is shared by exactly two elements $i1$ and $i2$.

Consequently QPAT can be formulated as the following bound-constrained nonlinear optimization problem

$$X = \text{argmin}_X f(X) + \lambda||MX||_1, \text{ subject to } L \le X \le U. \tag{45}$$

Here $\lambda$ is the regularizing parameter, and $L$ (resp. $U$) is the lower (resp. upper) bound of $X$. Please note that (i) the use of TV regularization or simple bounds is for regularizing the illposed solution, particularly for $\mu_s'$; (ii) the enforced simple bounds are loose constraints for excluding some apparently undesired solutions (e.g., lower bound by zero and upper bound by a order of magnitude of the maximum), rather than tight *a priori* constraints targeting directly at the desired solutions.

### 3.2 Split Bregman Method

Since the simple bounds will be handled explicitly, now let us consider the unconstrained version of (45),

$$X = \text{argmin}_X f(X) + \lambda ||MX||_1. \tag{46}$$

The simple version of a typical approach for solving (46) is to iteratively solve

$$\begin{array}{l} X^{n+1} = \text{argmin}_X f(X) + \lambda_n ||MX||_1, \\ \lambda_{n+1} = \mu\lambda_n, 0 < \mu < 1. \end{array} \tag{47}$$

That is, the iteration begins with the regularized solution, that is from a less ill-conditioned system due to the regularization, and converges to the true solution as the regularization diminishes. A major difficulty of (47) is that the system can be too ill-conditioned to be solvable for small regularization parameters after some iterations.

To resolve this difficulty, we begin with the Bregman distance of TV norm

$$D(X, Y) = \lambda ||MX||_1 - \lambda ||MY||_1 - < V(Y), X - Y >, \tag{48}$$

where $V$ is a subgradient of the TV norm at $Y$.

Now instead of varying $\lambda_n$ as in the continuation strategy (47), we fix the regularizing parameter $\lambda$, replace the TV norm $\lambda||MX||_1$ by its Bregman distance $D(X, Y)$, and iteratively solve

$$X^{n+1} = \text{argmin}_X f(X) + D(X, X^n). \tag{49}$$

That is

$$\begin{array}{l} X^{n+1} = \text{argmin}_X f(X) + \lambda ||MX||_1 - V^n X, \\ V^{n+1} = V^n - \partial f(X^{n+1}). \end{array} \tag{50}$$

(50) is the precedent of Split Bregman Method that was proposed in [OBGXY05]. Please note that (50) avoids to solve ill-conditioned systems by fixing $\lambda$ while updating the Bregman distance. We refer the readers to [OBGXY05, COS09] for the well-definedness, convergence and several nice properties of the Bregman method when $f(X)$ is convex, and [BB09] for the analysis of more general $f(X)$ commonly occurring in inverse problems.

The similar Bregman method can be used to solve the non-differentiable $L_1$-type subproblems of (50)

$$X = \text{argmin}_X g^n(X) + \lambda ||MX||_1 \tag{51}$$

with $g^n(X) = f(X) - V^n X$.

The motivation comes from (i) the fact that the $L_1$ scalar minimization

$$\min \frac{1}{2}(x - y)^2 + \lambda |x| \tag{52}$$

has the explicit solution, so-called the shrinkage formula

$$x = T_\lambda(y) = \text{sgn}(y) \cdot \max(|y| - \lambda, 0), \tag{53}$$

where $\text{sgn}(y)$ denotes the sign of the scalar $y$;(ii) the split treatment of the differentiable data fidelity and the non-differentiable TV norm. Please note

that (53) also extends to the vector computation since the objective function of (51) is separable into the summation of (52).

The method is so-called Split Bregman Method [GO09, COS09] and it goes as follow. First we let the dummy variable $Z = MX$ so that TV becomes $||Z||_1$, for which the Shrinkage formula (53) can be applied later. Consequently, (51) is reformulated as an equality-constrained optimization problem

$$(X, Z) = \text{argmin}_{(X,Z)} g^n(X) + \lambda||Z||_1, \text{ subject to } MX = Z. \tag{54}$$

Now we enforce the equality constraints by quadratic penalties, however penalize the Bregman distance of $g^n(X) + \lambda||Z||_1$ iteratively rather than update the penalizing parameter $\mu$, i.e.,

$$\begin{aligned}
&(X^{m+1}, Z^{m+1}) = \text{argmin}_{(X,Z)} g^n(X) + \lambda||Z||_1 - V_x^m X - V_z^m Z + \tfrac{1}{2}\mu||MX - Z||_2^2,\\
&V_x^{m+1} = V_x^m + \mu M^T(Z^{m+1} - MX^{m+1}),\\
&V_z^{m+1} = V_z^m + \mu(MX^{m+1} - Z^{m+1}),
\end{aligned} \tag{55}$$

which can be simplified to

$$\begin{aligned}
&(X^{m+1}, Z^{m+1}) = \text{argmin}_{(X,Z)} g^n(X) + \lambda||Z||_1 + \tfrac{1}{2}\mu||MX - Z + W^m||_2^2,\\
&W^{m+1} = W^m + MX^{m+1} - Z^{m+1}.
\end{aligned} \tag{56}$$

Regarding the handling of equality constraints, Split Bregman Method (56) is similar to the augmented Lagrangian Method [H69, P69, B82, NW06], in which the Lagrangian multipliers are added to the object function and are iteratively estimated with the fixed quadratic penalty parameter $\mu$.

Then the first step of (56) can be solved by the iterative alternating optimization of $X$ and $Z$

$$\begin{aligned}
X^{k+1} &= \text{argmin}_X g^n(X) + \tfrac{1}{2}\mu||MX - Z^k + W^m||_2^2,\\
Z^{k+1} &= \text{argmin}_Z \lambda||Z||_1 + \tfrac{1}{2}\mu||MX^{k+1} - Z + W^m||_2^2,
\end{aligned} \tag{57}$$

where the second equation has the explicit solution by shrinkage formula (53)

$$Z^{k+1} = T_{\frac{\lambda}{\mu}}(MX^{k+1} + W^m). \tag{58}$$

Combining (50), (56), (57) and (58), Split Bregman Method is now summarized as

For $n = 1$ to $N$
  For $m = 1$ to $M$
    For $k = 1$ to $K$
      $X^{k+1} = \text{argmin}_X f(X) - V^n X + \frac{1}{2}\mu||MX - Z^k + W^m||_2^2$;
      $Z^{k+1} = T_{\frac{\lambda}{\mu}}(MX^{k+1} + W^m)$;
    End
    $W^{m+1} = W^m + MX^{m+1} - Z^{m+1}$;
  End

$V^{n+1} = V^n - \partial f(X^{n+1})$.
End

In particular, this Split Bregman loop with $M = K = 1$ has been proven to converge with certain numerical advantages [COS09] and will be adopted as the base of the GBSB algorithm. That is

$$\begin{aligned} &X^{n+1} = \text{argmin}_X f(X) + \tfrac{1}{2}\mu||MX - Z^n + W^n||_2^2 - V^n X, \\ &Z^{n+1} = T_{\frac{\lambda}{\mu}}(MX^{n+1} + W^n), \\ &W^{n+1} = W^n + MX^{n+1} - Z^{n+1}, \\ &V^{n+1} = V^n - \partial f(X^{n+1}). \end{aligned} \tag{59}$$

All steps in (59) except the first one are computationally cheap. Next we will turn to the discussion for solving this differentiable subproblem to update $X$.

### 3.3 L-BFGS: Quasi-Newton Approximation of the Hessian

Since the computation of Jacobians is extremely time-consuming, we will not formulate Jacobians from (21) according to the least-square form of $f(X)$, and solve the consequent optimization problems using least-square techniques. Instead, we rather treat (21) as a nonlinear objective function, and only compute its gradients (please see Section 2.3 for the details of gradient computation).

Specifically, we will adopt the Quasi-Newton method with adaptive updating of the Hessian by gradients (Section 3.3), compute the search direction while explicitly enforcing simple bounds (Section 3.4), and perform the line search satisfying Wolfe conditions and simple bounds (Section 3.5).

we are going to focus on solving the first subproblem of the Split Bregman loop (59). To simplify the notation, we let

$$g(X) = f(X) + \frac{1}{2}\mu||MX - Z^n + W^n||_2^2 - V^n X \tag{60}$$

and consider

$$X = \text{argmin}_X g(X). \tag{61}$$

Suppose that the current iterate is $X_k$, a quadratic approximation of (61) with $p = X - X_k$ is

$$Q_k(p) = g_k + \partial g_k^T p + \frac{1}{2} p^T H_k p, \tag{62}$$

where $g_k$, $\partial g_k$ and $H_k$ are the function value, the gradient and the Hessian at $X_k$ respectively. Consequently, from the optimal condition of (62),

$$p_k = -H_k^{-1} \partial g_k, \tag{63}$$

which can be used as the search direction at the $k$th iteration for (61).

For computational efficiency, instead of formulating $H_k$ explicitly and taking its inverse $H_k^{-1}$, we shall use a well-known Quasi-Newton method, namely, BFGS method [DS83, F87, NW06], to iteratively update $H_k^{-1}$, for which only $\partial g_k$ is required. The methodology goes as follow.

Let $p_k$ and $\alpha_k$ be the search direction and the step length at the current iterate $X_k$, i.e.,

$$X_{k+1} = X_k + \alpha_k p_k. \tag{64}$$

At the next iterate $X_{k+1}$, the new quadratic approximation $Q_{k+1}(p)$ should be consistent in the sense that

$$\begin{aligned} \partial Q_{k+1}(-\alpha_k p_k) &= \partial g_k, \\ \partial g_{k+1} + H_{k+1}(-\alpha_k p_k) &= \partial g_k. \end{aligned} \tag{65}$$

That is

$$s_k = H_{k+1}^{-1} y_k \tag{66}$$

with

$$s_k = x_{k+1} - x_k \text{ and } y_k = \partial g_{k+1} - \partial g_k. \tag{67}$$

To uniquely update $H_{k+1}^{-1}$ from $H_k^{-1}$, besides the symmetric requirement and the secant condition (66), we impose the additional condition that $H_{k+1}^{-1}$ is the closest to $H_k^{-1}$ in the weighted Frobenius norm $||\cdot||$ [NW06], i.e.,

$$\begin{aligned} &H_{k+1}^{-1} = \text{argmin}_{H^{-1}} ||H^{-1} - H_k^{-1}|| \\ &\text{subject to } H^{-1} = {H^{-1}}^T \text{ and } H^{-1} y_k = s_k. \end{aligned} \tag{68}$$

The unique solution of (68) is the well-known BFGS formula

$$H_{k+1}^{-1} = V_k^T H_k^{-1} V_k + \rho_k s_k s_k^T \tag{69}$$

with

$$\rho_k = \frac{1}{y_k^T s_k} \text{ and } V_k = I - \rho_k y_k s_k^T. \tag{70}$$

However, BFGS formula (69) is still not suitable for large-scale computation since $H_k^{-1}$ is usually dense and consequently can be prohibitive in terms of memory and speed. Therefore, we adopt the limited-memory version of BFGS, so-called L-BFGS [BNS94, NW06].

The motivation of L-BFGS comes from a recursive reformulation of (69)

$$\begin{aligned} H_k^{-1} &= (V_{k-1}^T \dots V_0^T) H_0^{-1} (V_0 \dots V_{k-1}) \\ &+ \rho_1 (V_{k-1}^T \dots V_1^T) s_1 s_1^T (V_1 \dots V_{k-1}) \\ &+ \dots + \rho_{k-1} s_{k-1} s_{k-1}^T. \end{aligned} \tag{71}$$

Now, for L-BFGS, we only save and use the most recent $m$ pairs of $(s, y)$ for updating $H_k^{-1}$, i.e.,

$$\begin{aligned} H_k^{-1} &= (V_{k-1}^T \dots V_{k-m}^T) H_{k,0}^{-1} (V_{k-m} \dots V_{k-1}) \\ &+ \rho_{k-m} (V_{k-1}^T \dots V_{k-m+1}^T) s_{k-m} s_{k-m}^T (V_{k-m+1} \dots V_{k-1}) \\ &+ \dots + \rho_{k-1} s_{k-1} s_{k-1}^T, \end{aligned} \tag{72}$$

where $H_{k,0}^{-1}$ is the initial guess of $H^{-1}$ at the $k$th iteration. An empirical effective choice is

$$H_{k,0}^{-1} = \gamma_k I \text{ with } \gamma_k = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}. \tag{73}$$

This L-BFGS recursive formula (72) allows an efficient computation of the search direction $p_k$ at the current iterate $X_k$, i.e.,

**L-BFGS Update:** $p_k$=LBFGS($\partial g_k$,$\{s_i, y_i, k-m \le i \le k-1\}$).
$q = \partial g_k$;
For $i = k-1$ to $k-m$
  $q = q - \alpha_i y_i$;
End
$r = H_{k,0}^{-1} q$;
For $i = k-m$ to $k-1$
  $r = r + s_i(\alpha_i - \rho_i y_i^T r)$;
End
$p_k = -r$.

We find that the L-BFGS (72) with around five truncated terms ($m = 5$) is generally sufficient in Q-PAT. We refer the readers to [DS83, F87, NW06] and the references therein for the general framework of Quasi-Newton methods, convergence analysis of BFGS (69) and L-BFGS (72), and other useful updates of Hessian.

### 3.4 Bound-constrained Search Direction

Now let us come back to the original constrained QPAT problem (45) and explicitly deal with its simple bounds. In this case, the Split Bregman loop (59) still works and instead we consider its first step as

$$X = \text{argmin}_X g(X) \text{ subject to } L \le X \le U \tag{74}$$

with $g(X)$ defined by (60).

In GBSB, we explicitly enforce bound constraints in the computation of search directions of (74) from the first-order necessary conditions (KKT conditions) [NW06]. The similar explicit formula for simple bounds was also implemented in the nonlinear optimization software LANCELOT [CGT92]. That is, assuming $p(X)$ is the exact decent direction of (61), it is not difficult to verify that the projected gradient $\hat{p}(X)$ onto the feasible region is exactly

**Projection of $p$ onto the Feasible Region:** $\hat{p}_k = \text{Proj}(p_k)$.

$$\hat{p}(x) = \begin{cases} -L(x), \text{ if } x + p(x) \leq L(x) \\ p(x), \text{ if } L(x) < x + p(x) < U(x),\ x \in X \\ -U(x), \text{ if } x + p(x) \geq U(x) \end{cases}. \tag{75}$$

(75) is the updating formula of search directions in the Quasi-Newton method. That is, the search direction $p(X)$ is first computed by L-BFGS (72), and then projected to $\hat{p}(X)$ via (75). It is $\hat{p}(X)$ that will be used in line search for the constrained QPAT (45). Please note that $\hat{p}(X)$ is not exact in our Quasi-Newton method since $p(X)$ is from L-BFGS in iterative quadratic approximations of $f$. Therefore, we will also enforce the bound constraints in the line search.

### 3.5 Line Search

Now we are going to complete the solution method for minimizing the nonlinear functional (74) as subproblems during Split Bregman iterations (59) for the QPAT model problem (45). The discussion concerns the choice of the step length $\alpha_k$ along the current search direction $\hat{p}_k$ so that new iterate $X_{k+1}$ can be updated by

$$X_{k+1} = X_k + \alpha_k \hat{p}_k. \tag{76}$$

In GBSB, we enforce the Wolfe Conditions [NW06] for selecting $\alpha_k$, i.e.,

$$\begin{aligned} &g(X_k + \alpha_k \hat{p}_k) \leq g(X) + c_1 \alpha_k \partial g^T(X_k) \hat{p}_k \\ &\partial g^T(X_k + \alpha_k \hat{p}_k) \hat{p}_k \geq c_2 \partial g^T(X_k) \hat{p}_k, \end{aligned} \tag{77}$$

where the first inequality guarantees the sufficient decrease of the objective function value, and the second inequality, so-called curvature condition, rules out unacceptably short steps. Please note that the curvature condition is necessary to guarantee the positive curvature, (i.e., $s_k^T y_k > 0$), which is implied by the secant equation (66). Otherwise,the performance of L-BFGS method may degrade due to the violation of the curvature condition.

Besides, we also enforce the bound constraints, i.e.,

$$L \leq X_k + \alpha_k \hat{p}_k \leq U. \tag{78}$$

As mentioned earlier, our projected gradient $\hat{p}$ is only an approximation, and it may not fulfill the simple bounds. As a result, (78) is necessary for $X_{k+1}$ to stay feasible.

Last, we implement the following Backtracking line search algorithm in GBSB, which is simple, yet empirically sufficient.

**Backtracking line search:** $\alpha_k = \text{Backtrack}(\hat{p}_k)$.
Choose $c_1 = 0.0001, c_2 = 0.9, \rho = 0.5, \alpha_k = 1$;

Do $\alpha_k = \rho\alpha_k$
Until $\alpha_k$ satisfies (77) and (78).

One observation from using this backtracking ling search algorithm is that $\alpha_k = 1$ is usually accepted after one or a few Quasi-Newton iterations (76).

### 3.6 GBSB Algorithm

Now, we can summarize our discussions so far on the GBSB algorithm for QPAT (45) as follow.

For $n = 1$ to $N$ (Bregman outer loop)

$g^n = f(X) - V^n X + \frac{1}{2}\mu||MX - Z^n + W^n||_2^2$;
$X_0^n = X^{n-1}$;
For $k = 1$ to $K$ (Quasi-Newton inner loop)
$p_k = \text{LBFGS}(\partial g_k^n, \{s_i^n, y_i^n, k - m \le i \le k - 1\})$;
$\hat{p}_k = \text{Proj}(p_k)$;
$\alpha_k = \text{Backtrack}(\hat{p}_k)$;
$X_{k+1}^n = X_k^n + \alpha_k \hat{p}_k$;
Break, if $||g_k^n|| \le \epsilon$;
End
$X^n = X_K^n$;

$Z^{n+1} = T_{\frac{\lambda}{\mu}}(MX^{n+1} + W^n)$;
$W^{n+1} = W^n + MX^{n+1} - Z^{n+1}$;
$V^{n+1} = V^n - \partial f(X^{n+1})$.
End

As discussed earlier, we fix the value of the regularizing parameter $\lambda$ to be a constant. Regarding the choice of this constant, any value of $\lambda$ other than a particularly small number is sufficient although the considerably large value of $\lambda$ may require extra iteration steps. Besides, $\mu = \lambda$ is recommended from our numerical experiences.

The stopping criterion for the Quasi-Newton inner loop of the GBSB algorithm is based on $\epsilon = \epsilon_i g_0^n$ with a small constant $\epsilon_0$, e.g., $\epsilon_i = 0.0001$. This $g_0^n$-dependent stopping criterion is motivated by the scale variation of $g^n$ in the the inner loop. On the other hand, the stopping criterion of Bregman outer loop is based on the difference between two consecutive iterative solutions, i.e., $||X^{n+1} - X^n|| < \epsilon_o||X^n||$,e.g., $\epsilon_i = 0.01$. Please note that in the case with considerably noisy data (40), the small $\epsilon_i$ is not recommended since $X^n$ may otherwise converge to the solution corresponding to the noisy data, which is not desirable.

We refer the readers to the previous sections for the details of individual steps of the GBSB algorithm, such as the algorithm parameters. Next we are

going to discuss on some practical factors that can be of great importance for the performance of GBSB algorithm in QPAT.

### 3.7 Data Scaling

The first scaling is with respect to the data fidelity term (21), which has considerably scale variation among different spatial elements. To balance the minimization of the inhomogeneous discrepancy between the model and the data, we can either consider the weighted data fidelity term with weights $Y_{ij}^{-1}$

$$f_1(X) = \frac{1}{2}\sum_{i,j}[Y_{ij}^{-1}F_j(\mu_a, \phi_i(\mu_a, \mu_s')) - 1]^2, \tag{79}$$

or the data fidelity term in logarithms

$$f_2(X) = \frac{1}{2}\sum_{i,j}[\log F_j(\mu_a, \phi_i(\mu_a, \mu_s')) - \log Y_{ij}]^2. \tag{80}$$

### 3.8 Parameter Scaling

The second scaling is with respect to the reconstruction variables (41), since $\mu_a$ and $\mu_s'$ usually differ from each other by 1 or 2 orders of magnitude. That is, now we consider the minimization problem with respect to the scaled parameters

$$X' = (\mu_a, r\mu_s'), \tag{81}$$

with $r$ as a scaling constant that can be set to the ratio of the mean of the initial absorption coefficient $\mu_a^0$ over the mean of the initial reduced scattering coefficient ${\mu'}_s^0$.

Consequently, the gradient of $f$ with respect to $X'$ is

$$\partial_{X'} f = (\partial_{\mu_a} f, r^{-1}\partial_{\mu_s'} f). \tag{82}$$

### 3.9 Initial Guess

Due to the nonlinear nature of QPAT and its ill-posedness besides the fact that the problem may have many local minimizers, in order to reconstruct the desired optical parameters, almost exclusively, any iterative algorithm requires an educational initial guess $X^0$. The extensive numerical tests on the stability of GBSB with respect to $X^0$ indicate that GBSB is quite stable even if the difference between $X^0$ and the underground truth is considerably large, which is ideal for the large-scale QPAT. Nonetheless, a reasonable $X^0$ is still required for GBSB to converge to the meaningful solution. An empirical strategy for $X^0$ is the model based fitting of the homogeneous optical background from the data, e.g.,

$$X^0 = \text{argmin}_X f(X). \tag{83}$$

Please note that here $X^0 = (\mu_a^0, \mu'^0_s)$ and is spatially independent so that (83) can be solved much more efficiently in comparison with the QPAT model problem (45).

### 3.10 Simulation Settings

We performed the GBSB algorithm for MS-QPAT with DA as forward model in both 2D and 3D.

In the 2D simulation, the phantom was based on the embedding of a 2D Shepp-Logan phantom into a circular domain of a 50mm diameter and the center (0,0). In addition, the absorption coefficient map (Fig. 1a) had another four 1.25mm-diameter circular inclusions centered at (-5,0), (-10,0), (-15,0) and (-20,0); the reduced scattering coefficient map (Fig. 1a) had another four 1.25mm-diameter circular inclusions centered at (5,0), (10,0), (15,0) and (20,0). To mimic the MS-QPAT, the phantom was illuminated respectively with the optical sources at the boundary located in each quadrant of the circular domain, and the total number of optical illuminations was four in 2D.

To alleviate the inverse crime in simulations, the data $Y$ (40) was generated based on a mesh with 17489 nodes and 34592 elements, and the reconstruction was performed on another independent mesh with 17137 nodes and 33888 elements. Then 1% Gaussian noise that is proportional to $Y$ was added to the data before the reconstruction. Please note that the phantom was first generated on a $128 \times 128$ grid, and then interpolated to the 1st triangular mesh for generating $Y$. Therefore, the reconstruction phantom was the latter one after the interpolation. In this case, the resolution of the reconstruction phantom can be regarded as $128 \times 128$ since the number of variables in the piecewise-constant discretization was sufficiently large. On the other hand, the optical parameters were first reconstructed on the 2nd triangular mesh, and then interpolated to the $128 \times 128$ cartesian grid for display, which corresponded to the $< 0.5mm$ resolution in the reconstruction phantom. On the other hand, as discussed earlier, the GBSB algorithm is stable with respect to the initial guess. The displayed reconstructed maps (Fig. 1b and 2b) were from a typical initial guess with $\mu_a^0 = 0.008$ and $\mu'^0_s = 0.8$, which did not overlap with any pixel value in the phantom (Fig. 1a and 2a).

Similarly, in the 3D simulation, the phantom was based on the embedding of a 3D Shepp-Logan phantom into a cylindrical domain of a 50mm diameter and a 50mm height and the center (0,0,0). In addition, the absorption coefficient map (Fig. 3a-c) had another four 2.5mm-diameter spherical inclusions centered at (-5,0,0), (-10,0,0), (-15,0,0) and (-20,0,0); the reduced scattering coefficient map (Fig. 4a-c) had another four 2.5mm-diameter spherical inclusions centered at (5,0,0), (10,0,0), (15,0,0) and (20,0,0). To mimic the MS-QPAT, the phantom was illuminated with the optical sources respectively

from the top flat surface, the bottom flat surface, and then at the boundary located in each quadrant of the side, and the total number of optical illuminations was six in 3D.

To alleviate the inverse crime in simulations, the data $Y$ was generated based on a mesh with 64356 nodes and 446823 elements, and the reconstruction was performed on another independent mesh with 56635 nodes and 348687 elements. Then 1% Gaussian noise that is proportional to $Y$ was added to the data before the reconstruction. Again the phantom was first generated on a $100 \times 100 \times 100$ grid, and then interpolated to the 1st tetrahedral mesh for generating $Y$. Please note that the resolution of the reconstruction phantom was actually less than $100 \times 100 \times 100$ due to the 1st tetrahedral mesh,i.e., $446823 < 100^3$. On the other hand, the optical parameters were first reconstructed on the 2nd tetrahedral mesh, and then interpolated to the $100 \times 100 \times 100$ cartesian grid for display, which corresponded to 0.5mm resolution in the reconstruction phantom. Again, the actual reconstructed resolution did not reach $100 \times 100 \times 100$ since the number of variables in the 2nd tetrahedral mesh was smaller than $100^3$ besides the fact that the reconstruction phantom did not even have the $100 \times 100 \times 100$ resolution. It was estimated that the reconstruction solution was approximately $70 \times 70 \times 70$. On the other hand, as discussed earlier, the GBSB algorithm is stable with respect to the initial guess. The displayed reconstructed maps (Fig. 3d-f and 4d-f) were from a typical initial guess with $\mu_a^0 = 0.008$ and $\mu_s'^0 = 0.8$, which did not overlap with any pixel value in the phantom (Fig. 3a-c and 4a-c).

### 3.11 Simulation Results

In this section, we present the simultaneously reconstructed absorption coefficient map $\mu_a$ and reduced scattering coefficient map $\mu_s'$ from the GBSB algorithm for MS-QPAT with DA as forward model.

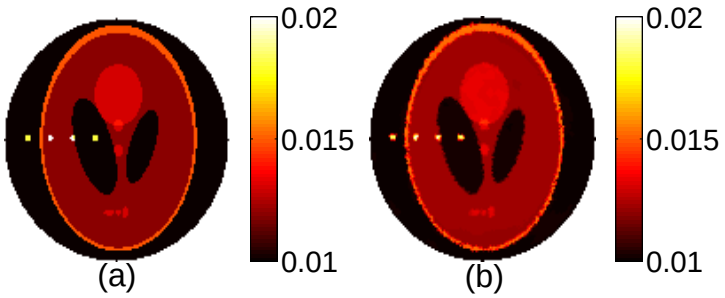


**Fig. 1.** 2D simultaneous reconstruction of absorption coefficient $\mu_a$ and reduced scattering coefficient $\mu_s'$ by the GBSB algorithm for MS-QPAT. (a) the true $\mu_a$ and (b) the reconstructed $\mu_a$. The displays are $128 \times 128$ with the display window [0.01 0.02]. The reconstruction data has 1% Gaussian noise.

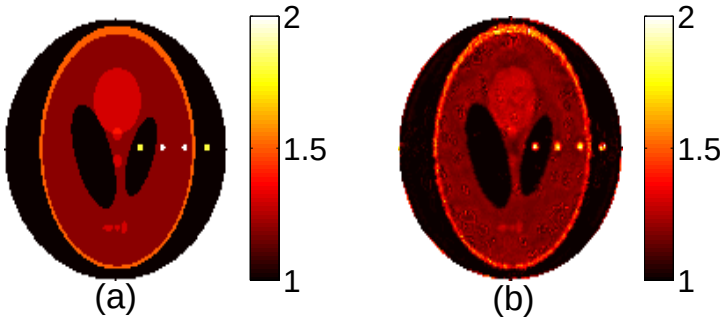


**Fig. 2.** 2D simultaneous reconstruction of absorption coefficient $\mu_a$ and reduced scattering coefficient $\mu_s'$ by the GBSB algorithm for MS-QPAT. (a) the true $\mu_s'$ and (b) the reconstructed $\mu_s'$. The displays are $128 \times 128$ with the display window [1 2].The reconstruction data has 1% Gaussian noise.

The 2D results are displayed with the $128 \times 128$ resolution in Fig. 1 and 2: Fig. 1 shows the true and the reconstructed $\mu_a$; Fig. 2 shows the true and the reconstructed $\mu_s'$. The phantom is 50mm in diameter, and therefore the image has $< 0.5mm$ resolution. In Fig. 1b, all Shepp-Logan inclusions and all four 1.25mm inclusions of the $\mu_a$ phantom (Fig. 1a) are successfully reconstructed. In Fig. 2b, major Shepp-Logan inclusions and all four 1.25mm inclusions of the $\mu_s'$ phantom (Fig. 1a) are successfully reconstructed while the small Shepp-Logan features are blurred and the background is relatively noisy. This reconstruction difference is due to the fact that the data are more sensitive to $\mu_a$ or the reconstruction is more ill-posed in $\mu_s'$.

The 3D results are displayed with the $100 \times 100 \times 100$ resolution in Fig. 3 and 4: Fig. 3 shows the true and the reconstructed $\mu_a$; Fig. 4 shows the true and the reconstructed $\mu_s'$. The phantom is 50mm in diameter and height, and therefore the image has $< 0.5mm$ resolution. However, as discussed earlier, the actual resolution is around $70 \times 70 \times 70$ due to the resolution of the used mesh. A better resolution should be achieved with finer meshes. In Fig. 3d-f, all Shepp-Logan inclusions and all four 2.5mm inclusions of the $\mu_a$ phantom (Fig. 3a-c) are successfully reconstructed. Again, in Fig. 4d-f, major Shepp-Logan inclusions and all four 2.5mm inclusions of the $\mu_s'$ phantom (Fig. 4a-c) are successfully reconstructed while the small Shepp-Logan features are blurred and the background is relatively noisy.

## 4 Discussion

The proposed GBSB algorithm can be easily extended for multi-wavelength QPAT [CAB09] and RTE based QPAT. In the latter case, it is interesting to investigate whether it is possible to reconstruct $(\mu_a, \mu_s, g)$ in the hybrid regime of transport and diffusion. Another interesting question is to study the direct

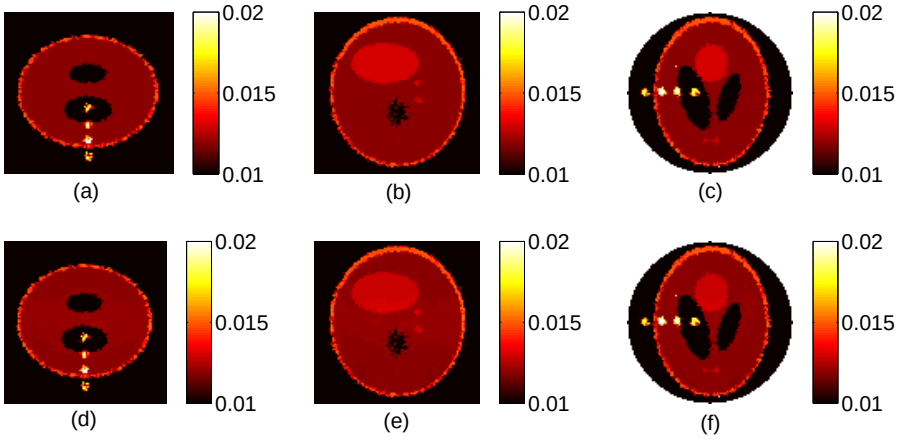


**Fig. 3.** 3D simultaneous reconstruction of absorption coefficient $\mu_a$ and reduced scattering coefficient $\mu_s'$ by the GBSB algorithm for MS-QPAT. (a)-(c) the sagittal, coronal, and transverse plane of the true $\mu_a$ and (d)-(f) the sagittal, coronal, and transverse plane of the reconstructed $\mu_a$. The displays are $100 \times 100 \times 100$ with the display window [0.01 0.02]. The reconstruction data has 1% Gaussian noise.

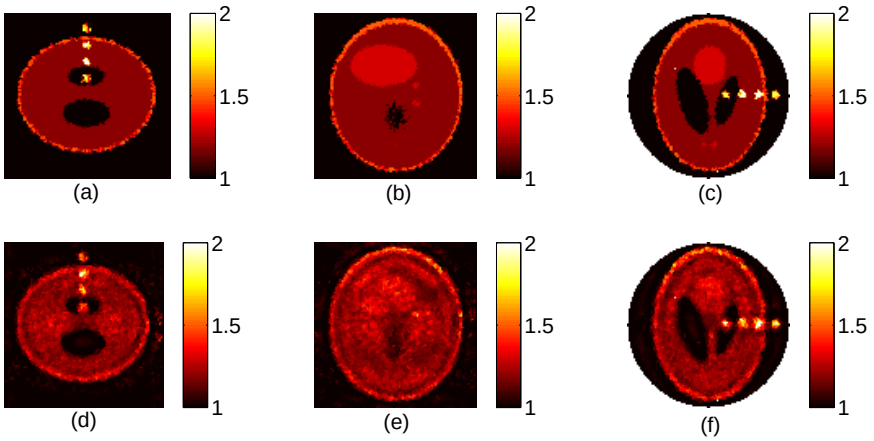


**Fig. 4.** 3D simultaneous reconstruction of absorption coefficient $\mu_a$ and reduced scattering coefficient $\mu_s'$ by the GBSB algorithm for MS-QPAT. (a)-(c) the sagittal, coronal, and transverse plane of the true $\mu_s'$ and (d)-(f) the sagittal, coronal, and transverse plane of the reconstructed $\mu_s'$. The displays are $100 \times 100 \times 100$ with the display window [1 2]. The reconstruction data has 1% Gaussian noise.

reconstruction of optical maps from boundary acoustic measurement, since GBSB can also be extended to this case and the forward/adjoint problem can be defined similarly once a practical acoustic model is available.

## References

[A99] Arridge, S.R.: Optical tomography in medical imaging. Inverse Problems, **15**, R41–R93 (1999)

[AL02] Adams, M.L., Larsen, E.W.: Fast iterative methods for discrete-ordinates particle transport calculations. Progress in Nuclear Energy, **40**, 3–159 (2002)

[B82] Bertsekas, D.R.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York (1982)

[BB09] Bachmayr, M., Burger, M.: Iterative total variation schemes for nonlinear inverse problems. Inverse Problems, **25**, 105004 (2009)

[BBVR08] Banerjee, B., Bagchi, S., Vasu, R.M., Roy, D.: Quantitative photoacoustic tomography from boundary pressure measurements: noniterative recovery of optical absorption coefficient from the reconstructed absorbed energy map. J. Opt. Soc. Am. A, **25**, 2347–2356 (2008)

[BJJ10] Bal, G., Jollivet, A., Jugnon, V.: Inverse transport theory of photoacoustics. Inverse Problems, **26**, 025011 (2010)

[BNS94] Byrd, R.H., Nocedal, J., Schnabel, R.B.: Representations of quasi-Newton matrices and their use in limited-memory methods. Mathematical Programming, Series A, **63**, 129–156 (1994)

[BR10] Bal, G., Ren, K.: Multiple-source quantitative photoacoustic tomography. Preprint (2010)

[BR10-2] Bal, G., Ren, K.: Non-uniqueness result for a hybrid inverse problem. Preprint (2010)

[BS10] Bal, G., Schotland, J.C.: Inverse scattering and acousto-optic imaging. Phys. Rev. Letters, **104**, 043902 (2010)

[BU10] Bal, G., Uhlmann, G.: Inverse diffusion theory of photoacoustics. Inverse Problems, **26**, 085010 (2010)

[BV04] Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, New York (2004)

[CAB09] Cox, B.T., Arridge, S.R., Beard, P.C.: Estimating chromophore distributions from multiwavelength photoacoustic images. J. Opt. Soc. Am. A, **26**, 443–455 (2009)

[CAKB05] Cox, B.T., Arridge, S.R., Köstli, K., Beard, P.: Quantitative photoacoustic imaging: fitting a model of light transport to the initial pressure distribution. Proc. SPIE, **5697**, 49–55 (2005)

[CGT92] Conn, A.R., Gould, N.I.M., Toint, P.L.: LANCELOT: A FORTRAN Package for Large-scale Nonlinear Optimization (Release A). Springer-Verlag, New York (1992)

[CKS00] Cockburn., B., Karniadakis, G.E., Shu, C.W.: Discontinuous Galerkin Methods: Theory, Computation and Applications. Springer, New York (2000)

[COS09] Cai, J.-F., Osher, S., Shen, Z.: Split Bregman methods and frame based image restoration. SIAM Multiscale Model. Simul., **8**, 337–369 (2009)

[CZ67] Case, K.M., Zweifel, P.F.: Linear Transport Theory. Addison-Wesley, Massachusetts (1967)
[DS83] Dennis J.E., Schnabel, R.B.: Numerical Methods for Unconstrained Optimization and Nolinear Equations. Prentice-Hall, Englewood Cliffs, NJ (1983)
[F87] Fletcher, R.: Practical Methods of Optimization. John Wiley & Sons, New York (1987)
[GO09] Goldstein, T., Osher, S.: The split Bregman method for $l_1$ regularized problems. SIAM J. Imaging Sci., **2**, 323–343 (2009)
[GZ09] Gao, H., Zhao, H.K.: A fast forward solver of radiative transfer equation. Transport Theory and Statistical Physics, **38**, 149–192 (2009)
[GZ10] Gao, H., Zhao, H.K.: Analysis of a forward solver of radiative transfer equation. Preprint (2010)
[GZ10-2] Gao, H. Zhao, H.: Multilevel bioluminescence tomography based on radiative transfer equation Part 2: total variation and l1 data fidelity. Optics Express, **18**, 2894–2912 (2010)
[GZO10] Gao, H., Zhao, H.K., Osher, S.: Bregman methods in quantitative photoacoustic tomography. CAM Report, **10-42**, (2010)
[H69] Hestenes, M.R.: Multiplier and gradient methods. Journal of Optimization Theory and Applications, **4**, 303–320 (1969)
[J87] Johnson, C.: Numerical Solution of Partial Differential Equations by the Finite Element Method. Cambridge University Press, New York (1987)
[J99] Jackson, J.D.: Classical Electrodynamics. Wiley, New York (1999)
[LM93] Lewis, E.E., Miller, W.F.: Computational Methods of Neutron Transport. ANS Inc., La Grange Park, Illinois (1993)
[LW09] Li, C.H., Wang, L.V.: Photoacoustic tomography and sensing in biomedicine. Physics in Medicine and Biology, **54**, R59–R97 (2009)
[NVIDIA] NVIDIA: NVIDIA CUDA Compute Unified Device Architecture, Programming Guide version 2.2.(2009)
[NW06] Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (2006)
[OBGXY05] Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterated regularization method for total variation based image restoration. SIAM Multiscale Model. Simul., **4**, 460–489 (2005)
[P69] Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed) Optimization. Academic Press, New York (1969)
[RN05] Ripoll, J., V. Ntziachristos, V.: Quantitative point source photoacoustic inversion formulas for scattering and absorbing media. Phys. Rev. E, **71**, 031912 (2005)
[ROF92] Rudin, L. Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. J. Phys. D, **60**, 259–268 (1992)
[W09] Wang, L.V.: Multiscale photoacoustic microscopy and computed tomography. Nat. Photonics, **3**, 503–509 (2009)
[XW06] Xu, M., Wang, L.V.: Photoacoustic imaging in biomedicine. Review of Scientific Instruments, **77**, 041101 (2006)
[YWZJ07] Yin, L., Wang, Q., Zhang, Q., Jiang, H.: Tomographic imaging of absolute optical absorption coefficient in turbid media using combined photoacoustic and diffusing light measurements. Opt. Lett., **32**, 2556–2558 (2007)
[Z10] Zemp, R.J.: Quantitative photoacoustic tomography with multiple optical sources. Applied Optics, **49**, 3566–3572 (2010)