Article

Stochastic Proximal Gradient Algorithms for Multi-Source Quantitative Photoacoustic Tomography

Simon Rabanser [1], Lukas Neumann [2] and Markus Haltmeier [1],*

[1] Department of Mathematics, University of Innsbruck, Technikerstraße 13, A-6020 Innsbruck, Austria; simon.rabanser@uibk.ac.at
[2] Institute of Basic Sciences in Engineering Science, University of Innsbruck, Technikerstraße 13, A-6020 Innsbruck, Austria; lukas.neumann@uibk.ac.at
* Correspondence: markus.haltmeier@uibk.ac.at; Tel.: +43-512-507-53840

Abstract: The development of accurate and efficient image reconstruction algorithms is a central aspect of quantitative photoacoustic tomography (QPAT). In this paper, we address this issues for multi-source QPAT using the radiative transfer equation (RTE) as accurate model for light transport. The tissue parameters are jointly reconstructed from the acoustical data measured for each of the applied sources. We develop stochastic proximal gradient methods for multi-source QPAT, which are more efficient than standard proximal gradient methods in which a single iterative update has complexity proportional to the number applies sources. Additionally, we introduce a completely new formulation of QPAT as multilinear (MULL) inverse problem which avoids explicitly solving the RTE. The MULL formulation of QPAT is again addressed with stochastic proximal gradient methods. Numerical results for both approaches are presented. Besides the introduction of stochastic proximal gradient algorithms to QPAT, we consider the new MULL formulation of QPAT as main contribution of this paper.

Keywords: photoacoustic tomography; image reconstruction; radiative transfer equation; multilinear inverse problem; limited view; stochastic gradient method; limited data; Dykstra algorithm

1. Introduction

Photoacoustic tomography (PAT) is an emerging imaging modality, which combines the benefits of pure ultrasound imaging (high resolution) with those of pure optical tomography (high contrast); see [1,2]. The basic principle of PAT is as follows (see Figure 1): A semitransparent sample such as a part of a human patient is illuminated with short pulses of optical radiation. A fraction of the optical energy is absorbed inside the sample, which causes thermal heating, expansion, and a subsequent acoustic pressure wave depending on the interior absorbing structure of the sample. The acoustic pressure is measured outside of the sample and used to reconstruct an image of the interior.

One important reconstruction problem in PAT is recovering the initial pressure distribution (see, for example, [3–10]). The initial pressure distribution only provides qualitative information about the tissue-relevant parameters, as it is the product of the optical absorption coefficient and the spatially varying optical intensity, which again indirectly depends on the tissue parameters. Quantitative photoacoustic tomography (QPAT) addresses this issue and aims at quantitatively estimating the tissue parameters by supplementing the inversion of the acoustic wave equation with an inverse problem for light propagation (see, for example, [11–26]).
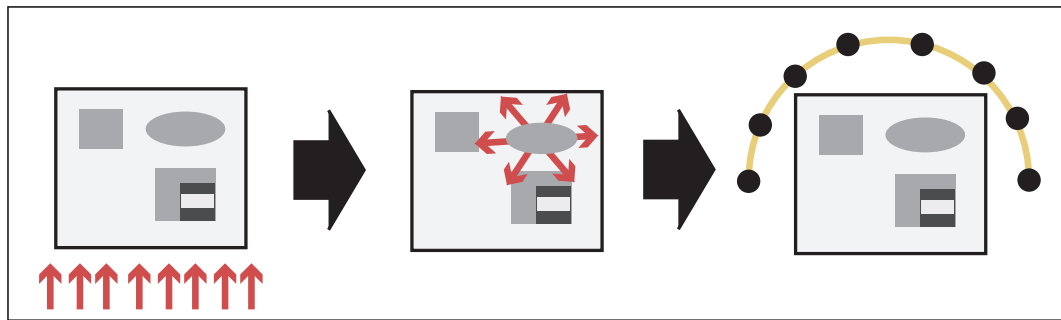
Figure 1. Basic principles of PAT. **Left**: the investigated object is illuminated with a short optical pulse; **Middle**: due to the thermoelastic effect, the absorbed light distribution induces an acoustic pressure wave depending on internal tissue properties; **Right**: the acoustic pressure wave is measured outside the object and used to reconstruct an image of the interior.

1.1. Multi-Source QPAT

In this paper, we consider image reconstruction in QPAT using multiple sources. We allow limited view measurements, where, for each illumination, partial data are collected only from a certain angular domain. For modeling the light transport, we use the radiative transfer equation (RTE), which is commonly considered as a very accurate model for light transport in tissue (see, for example, [27–30]). In particular, opposed to the diffusion approximation, the RTE allows for modeling directed optical radiation, which is required for a reasonable QPAT forward model. Additionally, it allows for including internal voids as regions of low scattering. As proposed in [18], we work with a single-stage reconstruction procedure for QPAT, where the optical parameters are reconstructed directly from the measured acoustical data. The image reconstruction problem of multi-source QPAT using N different sources can be formulated as a system of nonlinear equations (see, for example, [18,31])

$$\mathbf{F}_i(\mu) = v_i \quad \text{for } i = 1, \ldots, N. \tag{1}$$

Here, $\mathbf{F}_i$ is the operator that maps the unknown parameter pair $\mu = (\mu_a, \mu_s)$ consisting of the absorption coefficient $\mu_a \colon \Omega \to \mathbb{R}$ and the scattering coefficient $\mu_s \colon \Omega \to \mathbb{R}$ to the measured acoustic data v_i corresponding to the i-th source distribution (see Section 2 for precise definitions). There are two main classes of methods for solving the nonlinear inverse problem (1), namely, Tikhonov type regularization on the one and iterative regularization methods on the other hand [32–34]. Both approaches are based on rewriting (1) as a single equation $\mathbf{F}(\mu) = v$ with forward operator $\mathbf{F} = (\mathbf{F}_i)_{i=1}^N$ and data $v = (v_i)_{i=1}^N$. In Tikhonov regularization, one defines approximate solutions as minimizers of the penalized least squares functional $\frac{1}{2} \|\mathbf{F}(\mu) - v\|^2 + \lambda R(\mu)$. Here, $R(\,\cdot\,)$ is an appropriate regularization functional included to stabilize the inversion process and λ a regularization parameter that has to be carefully chosen depending on the data and the noise. In iterative regularization methods, stabilization is achieved via early stopping of iterative schemes. In such a situation, one usually applies iterative optimization techniques designed for minimizing the un-regularized least squares functional $\frac{1}{2} \|\mathbf{F}(\mu) - v\|^2$, and the iteration index plays the role of the regularization parameter.

Tikhonov type as well as iterative regularization methods can both be formulated as finding a solution of the optimization problem

$$\begin{cases} \min & \dfrac{1}{2} \sum_{i=1}^{N} \|\mathbf{F}_i(\mu) - v_i\|^2 + G(\mu), \\ \text{with} & \mu \in L^2(\Omega) \times L^2(\Omega). \end{cases} \tag{2}$$

In iterative regularization methods, one takes $G = \chi_{\mathbb{D}}$, the characteristic function of the domain of definition $\mathbb{D}$ of the forward operator (taking the value 0 in and the value ∞ outside of $\mathbb{D}$). In Tikhonov regularization, we take $G = \chi_{\mathbb{D}} + \lambda R$. Well established algorithms for solving Equation (2) are proximal gradient algorithms [35,36], which can be written in the form

$$\mu_{k+1} = \text{prox}_{s_k G}\left(\mu_k - s_k \sum_{i=1}^{N} \mathbf{F}_i'(\mu_k)^* \left(\mathbf{F}_i(\mu_k) - v_i\right)\right). \tag{3}$$

Here, $\text{prox}_{s_k G}$ is the proximity operator and s_k the positive step size; $\mathbf{F}_i'(\mu_k)$ denotes the derivative of the i-th forward operator evaluated at μ_k with $\mathbf{F}_i'(\mu_k)^*$ being its Hilbert space adjoint.

1.2. Stochastic Proximal Gradient Algorithms

Each iteration in the proximal gradient algorithm (3) can be numerically quite expensive, since it requires solving the forward and adjoint problems for all N equations in (1). In many cases, stochastic (proximal) gradient methods turn out to be more efficient since these methods only consider one of the equations in (1) per iteration. The stochastic proximal gradient method (see, for example, [37–42] and the references therein) for solving (2) is defined by

$$\mu_{k+1} = \text{prox}_{s_k G}\left(\mu_k - s_k \mathbf{F}_{i(k)}'(\mu_k)^* \left(\mathbf{F}_i(\mu_{i(k)}) - v_i\right)\right),$$

where $i(k) \in \{1, \ldots, N\}$ corresponds to one of the equations in (1) that is selected randomly for the update in the k-th iteration. In opposition to the standard proximal gradient method, this requires solving only one forward and one adjoint problem per iteration. Therefore, one iterative step is much cheaper for the stochastic gradient method than for the full gradient method. In the case of no regularization, $\lambda = 0$, the stochastic proximal gradient method reduces to the Kaczmarz method for inverse problems studied in [43–45].

The computationally most expensive task in the above methods is the numerical solution of the RTE. In this paper, we therefore additionally study a reformulation of the inverse problem of QPAT avoiding the computation of a solution of the RTE. For this purpose, the inverse problem is reformulated as multilinear inverse problem (35), where the RTE is added as a constraint instead of explicitly including its solution. The new formulation will be again addressed by Tikhonov regularization in combination with proximal stochastic gradient methods as discussed in Section 4.

Note that, in QPAT, it has often been assumed that the initial pressure distribution (corresponding to each illumination) is already recovered from acoustic measurements (see, for example, [13,16,23,25,26,46–49]). Research was focused on inverting the light propagation in tissues either modeled by the RTE or the diffusion approximation. In the case that acoustic measurements are only known on parts of the boundary, reconstruction of the initial pressure distribution is not possible in a stable manner. In order to obtain stable reconstruction results in [18], we propose a single-stage approach for QPAT, where the optical parameters are directly recovered from the acoustic boundary data. Throughout this paper, we will make use of this approach, which delivers stable results especially in the limited view situation. In opposition to [18], in this paper, we introduce (proximal) stochastic gradient methods, which effectively exploit the multi-illumination structure and turn out to be faster than the standard proximal gradient methods.

1.3. Outline

The remainder of this paper is organized as follows. In Section 2, we provide the mathematical model for QPAT (the forward problem) using the RTE. We allow multiple sources and partial acoustic measurements. We also recall known results for QPAT including differentiability of the forward problem. In Section 3, we address the inverse problem of QPAT using Tikhonov regularization and study the proximal stochastic gradient method for its solution. The new reformulation of the inverse

problem of QPAT as a multilinear inverse problem is presented in Section 4. For the solution of the proposed formulation, we again develop proximal gradient methods. Numerical results are presented in Section 5. The paper is concluded with a summary and outlook presented in Section 6.

2. The Forward Problem in QPAT

The image reconstruction problem of QPAT can be written as the system (1) of nonlinear equations, where the forward operators $\mathbf{F}_i$ map tissue relevant parameters to acoustic data sets recorded in specific regions outside the tissue. Precise formulations will be given in this section.

2.1. Mathematical Notation

We fix some mathematical notation that is used throughout this paper. We denote by $\Omega \subseteq \mathbb{R}^d$ a convex domain with piecewise smooth boundary modeling our domain of interest, with $d \in \{2, 3\}$ denoting the spatial dimension. In order to be able to impose appropriate boundary conditions for the RTE, it is convenient to split the set $\Gamma := \partial\Omega \times \mathbb{S}^{d-1}$ into inflow and outflow boundaries,

$$\Gamma_- := \left\{ (x, \theta) \in \partial\Omega \times \mathbb{S}^{d-1} \mid \nu(x) \cdot \theta \leq 0 \right\},$$

$$\Gamma_+ := \left\{ (x, \theta) \in \partial\Omega \times \mathbb{S}^{d-1} \mid \nu(x) \cdot \theta > 0 \right\},$$

with $\nu(x)$ denoting the outward pointing unit normal at $x \in \partial\Omega$ and $x \cdot y$ the standard inner product in $\mathbb{R}^d$. We write $B_R = \{x \in \mathbb{R}^n \mid \|x\| < R\}$ for the ball of radius R centered at the origin and suppose $B_R \supseteq \Omega$.

By $L^2(\Omega)$ and $L^2(\Omega \times \mathbb{S}^{d-1})$, we denote the Hilbert spaces of square integrable functions on Ω and $\Omega \times \mathbb{S}^{d-1}$, respectively. By $L^2(\Gamma_-, |\nu \cdot \theta|)$, we denote the space of all $q_o \colon \Gamma_- \to \mathbb{R}$ for which $\|q_o\|_{L^2(\Gamma_-, |\nu \cdot \theta|)}^2 := \int_{\Gamma_-} |q_o(x, \theta)|^2 \, |\nu \cdot \theta| \, \mathrm{d}(x, \theta)$ is finite. We further write

$$\|\Phi\|_{\mathbb{W}}^2 := \|\Phi\|_{L^2(\Omega \times \mathbb{S}^{d-1})}^2 + \|\theta \cdot \nabla_x \Phi\|_{L^2(\Omega \times \mathbb{S}^{d-1})}^2 + \|\Phi|_{\Gamma_-}\|_{L^2(\Gamma_-, |\nu \cdot \theta|)}^2 \,,$$

$$\|v\|_{\mathbb{Y}}^2 := \int_0^\infty \int_{\partial B_R} |v(x, t)|^2 \, t \, \mathrm{d}x \mathrm{d}t \,,$$

and define

$$\mathbb{Q} := L^2(\Omega \times \mathbb{S}^{d-1}) \times L^2(\Gamma_-, |\nu \cdot \theta|) \,,$$
$$\mathbb{X} := L^2(\Omega) \times L^2(\Omega) \,,$$
$$\mathbb{W} := \{\Phi \colon \Omega \times \mathbb{S}^{d-1} \to \mathbb{R} \mid \|\Phi\|_{\mathbb{W}} < \infty\} \,,$$
$$\mathbb{Y} := \{v \colon \partial B_R \times (0, \infty) \to \mathbb{R} \mid \|v\|_{\mathbb{Y}} < \infty\} \,.$$

The inner products in $\mathbb{Q}, \mathbb{X}, \mathbb{W}, \mathbb{Y}$ will be denoted by $\langle \cdot, \cdot \rangle_{\mathbb{Q}}, \langle \cdot, \cdot \rangle_{\mathbb{X}}, \langle \cdot, \cdot \rangle_{\mathbb{W}}, \langle \cdot, \cdot \rangle_{\mathbb{Y}}$, respectively. The subspace of all $\Phi \in \mathbb{W}$ with $\Phi|_{\Gamma_-} = 0$ will be denoted by $\mathbb{W}_0$.

Elements in $\mathbb{X}$ will be written in the form $\mu = (\mu_a, \mu_s)$ and are the parameters we aim to determine. They are actually required to be contained in the convex subset

$$\mathbb{D}(\mathbf{T}) := \{\mu \in \mathbb{X} \mid 0 \leq \mu_a \leq \overline{\mu}_a, 0 \leq \mu_s \leq \overline{\mu}_s\} \,, \tag{4}$$

where $\overline{\mu}_a, \overline{\mu}_s > 0$. Elements in $\mathbb{Q}$ will be written in the form $q = (q_o, q_i)$ and model the optical sources. Elements in $\mathbb{W}$ describe the optical radiation, and elements in $\mathbb{Y}$ the measured acoustic data.

2.2. The Radiative Transfer Equation

To specify the forward operators, we require mathematical models for the light propagation, the conversion of optical into acoustic energy, and the propagation of the acoustic waves. These models will be presented in the rest of this section.

We model the optical radiation by a function $\Phi\colon \Omega \times \mathbb{S}^{d-1} \to \mathbb{R}$, where $\Phi(x,\theta)$ is the density of photons at position $x \in \Omega$ and propagating into direction $\theta \in \mathbb{S}^{d-1}$. The interaction of the photons with the background are described by absorption coefficient $\mu_a\colon \Omega \to \mathbb{R}$, the scattering coefficient $\mu_s\colon \Omega \to \mathbb{R}$, and the scattering operator $\mathbf{K}\colon \Phi \mapsto \mathbf{K}\Phi$, taking the form (see [27,30])

$$\forall\, (x,\theta) \in \Omega \times \mathbb{S}^{d-1}\colon \quad \mathbf{K}\Phi(x,\theta) = \int_{\mathbb{S}^{d-1}} k(\theta,\theta')\Phi(x,\theta')\mathrm{d}\theta'\,, \tag{5}$$

with scattering kernel $k\colon \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \to \mathbb{R}$. The absorption coefficient describes the ability of the background to absorb photons and the scattering coefficient describes the amount of photon scattering. The scattering kernel $k\,(\theta,\theta')$ describes the redistribution of velocity directions due to interaction of the photons with the background. From physical considerations, it is natural to assume k to be measurable, symmetric, nonnegative, and to satisfy $\int_{\mathbb{S}^{d-1}} k\,(\,\cdot\,,\theta')\,\mathrm{d}\theta' = 1$. In this article, we are concerned with the situation when the kernel is known a priori.

The photon density $\Phi(x,\theta)$ is supposed to satisfy the stationary radiative transfer equation (RTE),

$$(\theta \cdot \nabla_x + \mu_a + \mu_s(\mathbf{I} - \mathbf{K}))\,\Phi(x,\theta) = q_i(x,\theta) \quad \text{for } (x,\theta) \in \Omega \times \mathbb{S}^{d-1} \tag{6}$$

with boundary conditions

$$\Phi|_{\Gamma_-}(x,\theta) = q_o(x,\theta) \quad \text{for } (x,\theta) \in \Gamma_- \,. \tag{7}$$

Here, $q_i\colon \Omega \times \mathbb{S}^{d-1} \to \mathbb{R}$ denotes an internal photon source and $q_o\colon \Gamma_- \to \mathbb{R}$ a prescribed boundary source pattern. Note that PAT uses very short light pulses (below microseconds) and that light propagation happens on time scales much shorter than the scale of acoustic wave propagation. This justifies the use of the stationary case for the RTE; see [12] for a more complete discussion.

Theorem 1 (Well-posedness of the RTE). *For every $\mu \in \mathbb{D}(\mathbf{T})$ and $q \in \mathbb{Q}$, the stationary RTE (6) admits a unique solution $\Phi \in \mathbb{W}$. Moreover, there exists a constant C only depending on the parameters $\overline{\mu}_a, \overline{\mu}_s > 0$ (defining the domain $\mathbb{D}(\mathbf{T})$), such that*

$$\|\Phi\|_{\mathbb{W}} \le C\,(\|q_i\|_{L^2} + \|q_o\|_{L^2})\,. \tag{8}$$

Proof. See [50].　□

Definition 1 (Solution operator for the RTE). *The solution operator for the RTE is defined by*

$$\mathbf{T}\colon \mathbb{Q} \times \mathbb{D}(\mathbf{T}) \to \mathbb{W}\colon (q,\mu) \mapsto \mathbf{T}(q,\mu) := \Phi\,,$$

where Φ denotes the unique solution of (6).

Theorem 1 guarantees that the operator $\mathbf{T}$, mapping $(q,\mu) \in \mathbb{Q} \times \mathbb{D}(\mathbf{T})$ to the solution of the RTE, is well defined. Note that in the actual application $q = (q_i, q_o) \in \mathbb{Q}$ are prescribed sources, and $\mu = (\mu_a, \mu_s) \in \mathbb{D}(\mathbf{T})$ are the unknown parameters to be recovered.

2.3. Heating Operator

Due to the spatially varying absorption of photons, the tissue is locally heated and emits an acoustic pressure wave. The acoustic source is proportional to the amount of absorbed photons, the light intensity and the so-called Grüneisen parameter γ describing the efficiency of conversion

of optical to acoustical energy. We assume γ to be constant and after appropriate re-scaling we take $\gamma = 1$; for more details about the Grüneisen parameter, we refer to [17]. Therefore, the conversion of the optical energy into acoustic pressure wave is described by the heating operator defined as follows.

Definition 2 (Heating operator). *The heating operator is defined by*

$$\mathbf{H}\colon \mathbb{Q} \times \mathbb{D}(\mathbf{T}) \to L^2(\Omega) \tag{9}$$

$$(q, \mu) \mapsto \mu_a \int_{\mathbb{S}^{d-1}} \mathbf{T}(q, \mu)(\,\cdot\,, \theta) \mathrm{d}\theta\,.$$

If one introduces the averaging operator $\mathbf{A}\colon \mathbb{W} \mapsto L^2(\Omega)$ defined by $\mathbf{A}\Phi = \int_{\mathbb{S}^{d-1}} \Phi(\,\cdot\,, \theta) \mathrm{d}\theta$ one may write the heating operator in the form

$$\mathbf{H}(q, \mu) = \mu_a\,\mathbf{A} \circ \mathbf{T}(q, \mu) \quad \text{for } (q, \mu) \in \mathbb{Q} \times \mathbb{D}(\mathbf{T})\,.$$

Because $\mathbf{T}(q, \mu)$ models the photon density, $\mathbf{A} \circ \mathbf{T}(q, \mu)$ actually models the total light intensity. The heating operator is therefore given by the product of the absorption coefficient and the light intensity. The averaging operator $\mathbf{A}$ is well defined and bounded and therefore the heating operator is well defined as a mapping between $\mathbb{Q} \times \mathbb{D}(\mathbf{T})$ and $L^2(\Omega)$.

2.4. The Wave Equation

The local heating causes an acoustic pressure wave, where the initial pressure distribution p_0 is proportional to a fraction of the absorbed energy. Assuming constant speed of sound and after rescaling, the induced acoustic pressure $p\colon \mathbb{R}^d \times (0, \infty) \to \mathbb{R}$ satisfies the free-space wave equation:

$$\begin{cases} (\partial_t^2 - \Delta)p(x, t) = 0 & \text{for } (x, t) \in \mathbb{R}^d \times (0, \infty)\,, \\ p(x, 0) = p_0(x) & \text{for } x \in \mathbb{R}^d, \\ \partial_t p(x, 0) = 0 & \text{for } x \in \mathbb{R}^d\,. \end{cases} \tag{10}$$

Here, the function p_0 vanishes outside B_R, the ball of radius R, and acoustic data are collected on a subset of $\partial B_R \times (0, \infty)$ that we denote by $\Lambda \times (0, \infty)$. Recall that coupling of the RTE and the wave equation happens in such a way that the result of the heating operator $\mathbf{H}(q, \mu)$ acts as initial sound source p_0 depending on tissue parameters; see Definition 4. Standard existence and uniqueness theory for hyperbolic equations guarantees that, for any $p_0 \in H^1$, (10) has a unique solution $p \in H^1$, which continuously depends on p_0. Taking the trace results in loss of regularity by degree $1/2$. Therefore, $p_0 \mapsto p|_{\partial B_R \times (0, \infty)}$ is continuous between H^1 and $H^{1/2}$. The following Lemma implies the much stronger result that $p_0 \mapsto p|_{\partial B_R \times (0, \infty)}$ is actually an L^2-isometry.

Lemma 1. *Let $p_0 \in C^\infty(\mathbb{R}^n)$ have support in B_R and let p denote the solution of (10). Then,*

$$\int_{B_R} \int_0^\infty |p(x, t)|^2\,t\mathrm{d}t\,\mathrm{d}x = \frac{R}{2} \int_{B_R} |p_0(x)|^2\,\mathrm{d}x\,. \tag{11}$$

Proof. See [51] for d odd and [52] for d even. $\quad\square$

Definition 3. *We define the solution operator with full boundary data for the wave Equation (10) by*

$$\mathbf{U}\colon C^\infty(B_R) \subseteq L^2(B_R) \to \mathbb{Y}\colon p_0 \mapsto p|_{\partial B_R \times (0, \infty)}\,, \tag{12}$$

where p denotes the solution of (10).

According to Lemma 1, the operator **U** can be uniquely extended to a bounded linear operator defined on $L^2(B)$, denoted again by $\mathbf{U}\colon L^2(B_R) \to \mathbb{Y}$. The partial acoustic measurements made on $\Lambda \subseteq \partial B_R$ are then modeled by $\chi_{\Lambda \times (0,\infty)}\mathbf{U}p_0$.

2.5. Analysis of the Forward Problem in Multi-Source QPAT

We assume that we perform N individual experiments, where each experiment consists of separate optical sources and separate acoustic measurements. For the i-th experiment, we denote the source term by $q_i \in \mathbb{Q}$ and assume the acoustic measurements are made on $\Lambda_i \times (0, T_i) \subseteq \partial B_R \times (0, \infty)$.

Definition 4. *For any $i \in \{1, \dots, N\}$, we denote*

$$\mathbf{T}_i\colon \mathbb{D}(\mathbf{T}) \to \mathbb{W}\colon \mu \mapsto \mathbf{T}(q_i, \mu),$$
$$\mathbf{H}_i\colon \mathbb{D}(\mathbf{T}) \to L^2(\Omega)\colon \mu \mapsto \mu_a(\mathbf{A} \circ \mathbf{T}_i)(\mu),$$
$$\mathbf{U}_i\colon L^2(\Omega) \to \mathbb{Y}\colon p_0 \mapsto \chi_{\Lambda_i \times (0,T_i)}\mathbf{U}p_0,$$
$$\mathbf{F}_i\colon \mathbb{D}(\mathbf{T}) \to \mathbb{Y}\colon \mu \mapsto (\mathbf{U}_i \circ \mathbf{H}_i)(\mu).$$

Here, $\mathbf{T}_i$ denotes the i-th solution operator for the RTE, $\mathbf{H}_i$ the i-th heating operator, $\mathbf{U}_i$ the i-th partial solution operator for the wave equation, and $\mathbf{F}_i$ the i-th forward operator.

Recall that **T** stands for the solution operator for the RTE (6) given in Definition 1, $\mathbf{A}\Phi = \int_{\mathbb{S}^{d-1}} \Phi(\,\cdot\,, \theta)\mathrm{d}\theta$ is the averaging operator, and **U** the solution operator for the wave Equation (10); see Definition 3. The operator $\mathbf{T}_i$ models the photon transport and its solution (via the heating operator) acts as input for the solution of the wave equation and thereby couples the optical with the acoustical part.

Next, we recall continuity and differentiability of the forward operators. For that purpose, we call $h \in \mathbb{X}$ a feasible direction at $\mu \in \mathbb{D}(\mathbf{T})$, if there exists some $\epsilon > 0$ with $\mu + \epsilon h \in \mathbb{D}(\mathbf{T})$.

Theorem 2 (Continuity and Differentiability).

1. *The operators $\mathbf{T}_i$, $\mathbf{F}_i$ and $\mathbf{H}_i$ are sequentially continuous and Lipschitz-continuous.*
2. *For every $\mu \in \mathbb{D}(\mathbf{T})$, the one-sided directional derivatives $\mathbf{T}_i'(\mu)(h)$, $\mathbf{F}_i'(\mu)(h)$ of $\mathbf{T}_i$, $\mathbf{F}_i$ at μ in any feasible direction h exist, and are given by*

$$\mathbf{T}_i'(\mu)(h) = \mathbf{T}\left(0, -(h_a + h_s - h_s\mathbf{K})\mathbf{T}(\mu), \mu\right), \tag{13}$$
$$\mathbf{F}_i'(\mu)(h) = \mathbf{U}_{\Lambda_i, T_i}\left(h_a\mathbf{A}\mathbf{T}_i(\mu) + \mu_a\mathbf{A}(\mathbf{T}_i'(\mu)(h))\right). \tag{14}$$

Proof. See [18]. □

Equations (13) and (14) define a bounded linear operator $\mathbf{F}_i'(\mu)\colon \mathbb{X} \to \mathbb{Y}$, which we call the derivative of $\mathbf{F}_i$ at $\mu \in \mathbb{D}(\mathbf{T})$. Numerical minimization schemes actually require the adjoint of $\mathbf{F}_i'(\mu)$, which we compute next.

Theorem 3 (Adjoint of $\mathbf{F}_i'(\mu)$). *Let $i \in \{1, \dots, N\}$ and $\mu \in \mathbb{D}(\mathbf{T})$. Furthermore, set $\Phi_i := \mathbf{T}_i(\mu)$ and let Φ_i^* denote the solution of the adjoint problem*

$$\left(-\theta \cdot \nabla_x + (\mu_a + \mu_s - \sigma\mathbf{K})\right)\Phi_i^* = -\mathbf{A}^*\left(\mu_a(\mathbf{U}_i^*v)\right) \tag{15}$$

with $\Phi_i^*|_{\Gamma_+} = 0$. Then, $\mathbf{F}_i'(\mu)^*\colon \mathbb{Y} \to \mathbb{X}$ is given by

$$\mathbf{F}_i'(\mu)^*v = \left[\begin{array}{c} \mathbf{A}(\Phi_i^*\Phi_i) + (\mathbf{A}\Phi_i)(\mathbf{U}_i^*v) \\ \mathbf{A}(\mathbf{I} - \mathbf{K})(\Phi_i^*\Phi_i) \end{array}\right]. \tag{16}$$

Proof. See [18]. □

Given data $v_1, \ldots, v_N \in \mathbb{Y}$, most numerical schemes for QPAT use gradients of the partial data-fidelity terms $F_I \colon \mathbb{D}(\mathbf{T}) \to \mathbb{R}$ for $I \subseteq \{1, \ldots, N\}$, where

$$F_I(\mu) = \sum_{i \in I} F_i(\mu) \quad \text{with} \quad F_i(\mu) := \frac{1}{2} \|\mathbf{F}_i(\mu) - v_i\|_{\mathbb{Y}}^2 \,. \tag{17}$$

By the chain rule, the gradient of F_I is given by $\nabla F_I(\mu) = \sum_{i \in I} \nabla F_i(\mu)$ with $\nabla F_i(\mu) = \mathbf{F}_i'(\mu)^* (\mathbf{F}_i(\mu) - v_i)$, where $\mathbf{F}_i'(\mu)^*$ can be computed by Theorem 3. Convergence of schemes such as the (stochastic) proximal gradient method considered in the following section require the Lipschitz continuity of ∇F_I, which will be shown in the following theorem.

Theorem 4 (Lipschitz continuity of ∇F_I). *For any data $v_1, \ldots, v_N \in \mathbb{Y}$ and any subset $I \subseteq \{1, \ldots, N\}$, the map $\mu \mapsto \nabla F_I(\mu)$ is Lipschitz-continuous.*

Proof. Without loss of generality, we assume $N = 1$, $I = \{1\}$ and write $v = v_1$, $F = F_{\{1\}}$, $\mathbf{T} = \mathbf{T}_1$, $\mathbf{U} = \mathbf{U}_1$, and $v(\mu) = \mathbf{F}(\mu) - v$. For any $\mu \in \mathbb{D}(\mathbf{T})$, let $\mathbf{T}^*(\mu)$ denote the solution of (15) with $v(\mu)$ in place of v. Then, for any $\mu, \tilde{\mu} \in \mathbb{D}(\mathbf{T})$,

$$\begin{aligned}
\|\nabla F(\mu) &- \nabla F(\tilde{\mu})\|_{\mathbb{X}}^2 \\
&= \|\mathbf{A}(\mathbf{T}^*(\mu)\mathbf{T}(\mu)) + (\mathbf{A}\mathbf{T}(\mu))(\mathbf{U}^*v(\mu)) - \mathbf{A}(\mathbf{T}^*(\tilde{\mu})\mathbf{T}(\tilde{\mu})) - (\mathbf{A}\mathbf{T}(\tilde{\mu}))(\mathbf{U}^*v(\tilde{\mu}))\|_{L^2(\Omega)}^2 \\
&\quad + \|\mathbf{A}(\mathbf{I} - \mathbf{K})(\mathbf{T}^*(\mu)\mathbf{T}(\mu)) - \mathbf{A}(\mathbf{I} - \mathbf{K})(\mathbf{T}^*(\tilde{\mu})\mathbf{T}(\tilde{\mu}))\|_{L^2(\Omega)}^2 \,.
\end{aligned} \tag{18}$$

For the second term in (18), we obtain

$$\begin{aligned}
\|\mathbf{A}(\mathbf{I} - \mathbf{K})&(\mathbf{T}^*(\mu)\mathbf{T}(\mu)) - \mathbf{A}(\mathbf{I} - \mathbf{K})(\mathbf{T}^*(\mu)\mathbf{T}(\tilde{\mu})) \\
&+ \mathbf{A}(\mathbf{I} - \mathbf{K})(\mathbf{T}^*(\mu)\mathbf{T}(\tilde{\mu})) - \mathbf{A}(\mathbf{I} - \mathbf{K})(\mathbf{T}^*(\tilde{\mu})\mathbf{T}(\tilde{\mu}))\|_{L^2(\Omega)} \\
&= \|\mathbf{A}(\mathbf{I} - \mathbf{K})(\mathbf{T}^*(\mu)[\mathbf{T}(\mu) - \mathbf{T}(\tilde{\mu})]) + \mathbf{A}(\mathbf{I} - \mathbf{K})(\mathbf{T}(\tilde{\mu})[\mathbf{T}^*(\mu) - \mathbf{T}^*(\tilde{\mu})])\|_{L^2(\Omega)} \\
&\leq c_1 \|\mathbf{A}(\mathbf{I} - \mathbf{K})\| \|\mathbf{T}^*(\mu)\|_{\mathbb{W}} L_{\mathbf{T}} \|\mu - \tilde{\mu}\|_{\mathbb{X}} + c_1 \|\mathbf{A}(\mathbf{I} - \mathbf{K})\| \|\mathbf{T}(\tilde{\mu})\|_{\mathbb{W}} L_{\mathbf{T}^*} \|\mu - \tilde{\mu}\|_{\mathbb{X}} \\
&\leq 2 c_1 \|\mathbf{A}(\mathbf{I} - \mathbf{K})\| \max(\|\mathbf{T}^*(\mu)\|_{\mathbb{W}}, \|\mathbf{T}(\tilde{\mu})\|_{\mathbb{W}}) \max(L_{\mathbf{T}^*}, L_{\mathbf{T}}) \|\mu - \tilde{\mu}\|_{\mathbb{X}} \,,
\end{aligned}$$

where $L_{\mathbf{T}^*}$ and $L_{\mathbf{T}}$ denote the Lipschitz constants of $\mathbf{T}^*$ and $\mathbf{T}$, and c_1 is a constant. The difference $\mathbf{A}(\mathbf{T}^*(\mu)\mathbf{T}(\mu)) - \mathbf{A}(\mathbf{T}^*(\tilde{\mu})\mathbf{T}(\tilde{\mu}))$ in the first term in (18) is estimated in a similar manner. Furthermore, we have

$$\begin{aligned}
\|(\mathbf{A}\mathbf{T}(\mu))&(\mathbf{U}^*v(\mu)) - (\mathbf{A}\mathbf{T}(\tilde{\mu}))(\mathbf{U}^*v(\mu)) + (\mathbf{A}\mathbf{T}(\tilde{\mu}))(\mathbf{U}^*v(\mu)) - (\mathbf{A}\mathbf{T}(\tilde{\mu}))(\mathbf{U}^*v(\tilde{\mu}))\|_{L^2(\Omega)} \\
&\leq \|(\mathbf{U}^*v(\mu))\mathbf{A}[\mathbf{T}(\mu) - \mathbf{T}(\tilde{\mu})]\|_{L^2(\Omega)} + \|\mathbf{A}\mathbf{T}(\tilde{\mu})(\mathbf{U}^*(v(\mu) - v(\tilde{\mu})))\|_{L^2(\Omega)} \,.
\end{aligned}$$

Noting that $\mathbf{A}$, $\mathbf{U}$ and $\mathbf{U}^*$ are linear and bounded, Theorem 2 and the computations above yield the Lipschitz continuity of ∇F. □

3. The Stochastic Proximal Gradient Method for QPAT

3.1. Formulation of the Inverse Problem

The inverse problem of multi-source QPAT consists in finding $\mu^\star \in \mathbb{X}$ from measured data

$$v_i = \mathbf{F}_i(\mu^\star) + z_i \quad \text{for } i = 1, \ldots, N \,. \tag{19}$$

Here, $\mu^\star = (\mu_a^\star, \mu_s^\star)$ are the unknowns to be estimated, z_i are the unknown error vectors, and $v_1, \ldots, v_N$ are the given noisy data. Using the notation

$$v := (v_1, \ldots, v_N) \in \mathbb{Y}^N,$$
$$\mathbf{F} := (\mathbf{F}_1, \ldots, \mathbf{F}_N) \colon \mathbb{D}(\mathbf{T}) \to \mathbb{Y}^N,$$

we can write (19) in the alternative form

$$\text{Estimate } \mu^* \in \mathbb{X} \text{ from } v = \mathbf{F}(\mu^*) + z. \tag{20}$$

Here, $z \in \mathbb{Y}^N$ denotes the error vector.

There are, at least, two different strategies to address such an inverse problem: Tikhonov type regularization on the one and iterative methods on the other hand. In this section, we give an overview of such methods. In particular, we describe proximal stochastic gradient methods (for minimizing the Tikhonov functional), which seem particularly well suited for multi-source QPAT but have not been investigated yet for that purpose.

3.2. Tikhonov Regularization in QPAT

In this section, we consider a quadratic Tikhonov regularization term for solving (19). Let

$$\mathbf{L} \colon \mathbb{D}(\mathbf{L}) \subseteq \mathbb{X} \to \mathbb{Z} \colon \mu \mapsto \mathbf{L}\mu$$

be a linear, densely defined, and possibly unbounded operator between $\mathbb{X}$ and another Hilbert space $(\mathbb{Z}, \langle \cdot, \cdot \rangle_{\mathbb{Z}})$ and set $\mathbb{D} := \mathbb{D} \cap \mathbb{D}(\mathbf{L})$. In this context, any element $\mu^+ \in \mathbb{D}$ with $\|\mathbf{L}\mu^+\| = \min\{\|\mathbf{L}\mu\| \mid \mathbf{F}(\mu) = v\}$ is called an $\|\mathbf{L}(\cdot)\|$-minimizing solution of $\mathbf{F}\mu = v$. Tikhonov regularization with regularization term $\frac{1}{2}\|\mathbf{L}\mu\|_{\mathbb{Z}}^2$ consists in computing a minimizer of the generalized Tikhonov functional $T_{v,\lambda} \colon \mathbb{X} \to \mathbb{R} \cup \{\infty\}$, defined by

$$T_{v,\lambda}(\mu) := \begin{cases} \frac{1}{2}\|\mathbf{F}(\mu) - v\|^2 + \frac{\lambda}{2}\|\mathbf{L}\mu\|^2, & \text{if } \mu \in \mathbb{D}, \\ \infty, & \text{otherwise}. \end{cases} \tag{21}$$

Here, $\lambda > 0$ denotes the regularization parameter that acts as a trade-off between the data fitting term and stability.

Theorem 5 (Well-posedness and convergence).

1. *For any $v \in \mathbb{Y}$ and any $\lambda > 0$, the Tikhonov functional $T_{\lambda,v}$ has at least one minimizer.*
2. *Let $v \in \text{ran}(\mathbf{F})$, $(\delta_m)_{m \in \mathbb{N}} \in (0, \infty)^{\mathbb{N}}$, $(v^m)_{m \in \mathbb{N}} \in \mathbb{Y}^{\mathbb{N}}$ with $\|v - v^m\| \leq \delta_m$. Suppose further that $(\lambda_m)_{m \in \mathbb{N}} \in (0, \infty)^{\mathbb{N}}$ satisfies $\lambda_m \to 0$ and $\delta_m^2 / \lambda_m \to 0$ as $m \to \infty$. Then:*

    ■ *Every sequence $(\mu^m)_{m \in \mathbb{N}}$ with $\mu^m \in \arg\min T_{v^m, \lambda_m}$ has a weakly converging subsequence.*
    ■ *The limit of every weakly convergent subsequence of $(\mu^m)_{m \in \mathbb{N}}$ is an $\|\mathbf{L}(\cdot)\|$-minimizing solution of $\mathbf{F}\mu = v$.*
    ■ *If the $\|\mathbf{L}(\cdot)\|$-minimizing solution of $\mathbf{F}\mu = v$ is unique and denoted by μ^+, then $(\mu^m) \rightharpoonup \mu^+$.*

Proof. See [18]. □

3.3. The Proximal Stochastic Gradient Algorithm for QPAT

Depending on the particular choice of $\mathbf{L}$, the Tikhonov functional (21) may be ill-conditioned. To address this issue in [18], we proposed the proximal gradient algorithm for minimizing (21), which is a very flexible algorithm for minimizing functionals of the form $F + G$, where F is smooth

and *G* is convex (see, for example, [35,36]). Here, we extend the approach to the proximal stochastic gradient algorithm. Additionally, we propose computing the proximal step using Dykstra's projection algorithm.

■ **Proximal gradient algorithm:** The proximal gradient algorithm is a splitting method that iteratively computes explicit gradient steps for *F* and implicit proximal steps for *G*. In our context, we take *F* as the data fidelity term and

$$G(\mu) = G_\lambda(\mu) := g_\lambda(\mu) + \chi_{\mathbb{D}}(\mu) := \frac{\lambda}{2} \|\mathbf{L}\mu\|^2 + \chi_{\mathbb{D}}(\mu),\tag{22}$$

where $\chi_{\mathbb{D}}$ is the characteristic function taking the value zero inside $\mathbb{D}$ and ∞ outside. The proximal gradient algorithm for minimizing the QPAT-Tikhonov functional (21) reads

$$\mu^{k+1} = \text{prox}_{s_k G_\lambda}\left(\mu^k - s_k \sum_{i=1}^N \nabla F_i(\mu^k)\right).\tag{23}$$

Here, $\text{prox}_{s_k G_\lambda} : \mathbb{X} \to \mathbb{D}$ denotes the proximal mapping corresponding to the functional $s_k G_\lambda$,

$$\text{prox}_{s_k G_\lambda}(x) = \arg\min\left\{\frac{1}{2}\|x - (\cdot)\|^2 + s_k G_\lambda\right\}.\tag{24}$$

Furthermore, $\nabla F_i(\mu^k)$ is the gradient of the *i*-th data fidelity term computed in Theorem 3.

■ **Dykstra's projection algorithm:** The constraint quadratic optimization problem (24) can efficiently be solved by a proximal variant of Dykstra's projection algorithm [35,36,53]. For that purpose, we write $s_k G_\lambda = \chi_{\mathbb{D}} + g$ with $g(x) := \frac{s_k \lambda}{2}\|\mathbf{L}x\|^2$. Setting $x_0 = \mu$, $p_0 = 0$ and $q_0 = 0$, Dykstra's projection algorithm for (24) reads, for $m \in \mathbb{N}$,

$$y_m = \text{prox}_g(x_m + p_m),\tag{25}$$
$$x_{m+1} = \mathbf{P}_{\mathbb{D}}(y_m + q_m),\tag{26}$$
$$p_{m+1} = x_m + p_m - y_m,\tag{27}$$
$$q_{m+1} = y_m + q_m - x_{m+1}.\tag{28}$$

Both proximal mapping in (25) and the projection in (26) can be computed explicitly. In fact, one readily verifies that

$$\text{prox}_g(x) = (\mathbf{I}_{\mathbb{X}} + s_k \lambda \, \mathbf{L}^* \mathbf{L})^{-1} x,\tag{29}$$
$$\mathbf{P}_{\mathbb{D}}(\mu) = \min\{\overline{\mu}, \max\{0, \mu\}\}.\tag{30}$$

Here, $\mathbf{I}_{\mathbb{X}}$ is the identity operator on $\mathbb{X}$ and $\mathbf{P}_{\mathbb{D}}$ the projection onto $\mathbb{D}$.

■ **Proximal stochastic gradient algorithm:** The methods described so far require in any iterative step the computation of the full gradient

$$\nabla F(\mu) = \sum_{i=1}^N \nabla F_i(\mu) \quad \text{with} \quad \nabla F_i(\mu) = \mathbf{F}'_i(\mu)^* \left(\mathbf{F}_i(\mu) - v_i\right).$$

The evaluation of each $\nabla F_i(\mu)$ requires the solution of the RTE and an adjoint problem and therefore is quite time-consuming. For multi-source QPAT, where $N > 1$, a significant acceleration may be obtained by a Kaczmarz strategy, where in each iterative step only one of the summands $\nabla F_i(\mu)$ is used. The resulting proximal stochastic gradient method for minimizing the Tikhonov functional (21) in QPAT reads

$$\mu^{k+1} = \text{prox}_{s_k G_\lambda}\left(\mu^k - s_k \nabla F_{i(k)}(\mu^k)\right),\tag{31}$$

where $i(k) \in \{1, \ldots, N\}$ is selected randomly for the update in the k-th iteration. Furthermore, $\text{prox}_{s_k G_\lambda}$ is the proximal mapping of $s_k G_\lambda$ that can be computed by Dykstra algorithm (25)–(28) and $\nabla F_i(\mu^k)$ is the gradient of the i-th data fidelity term that can be computed by Theorem 3.

One can also incorporate a block-iterative (or mini-batch) strategy in the stochastic gradient method, meaning that a small subset of $\{1, \ldots, N\}$ of equations is used per iteration instead of a single one. Such a variant could be especially useful in the case of a large number of different illumination patterns. For more details about stochastic gradient methods, see [37–42] and the references therein. Note that, in general, convergence of stochastic gradient methods requires asymptotically vanishing step size [42].

3.4. Iterative Regularization Methods

An alternative class of algorithms to address nonlinear inverse problems are iterative techniques. The most basic iterative method for solving the nonlinear inverse problem $v = \mathbf{F}(\mu)$ is the Landweber iteration. In the case that the domain of definition $\mathbb{D}$ is a proper subset, we have to combine the Landweber iteration with a projection step onto $\mathbb{D}$ as presented in this subsection. The projected Landweber iteration applied to multi-source QPAT reads

$$\mu^{k+1} = \mathbf{P}_\mathbb{D}\left(\mu^k - s_k \sum_{i=1}^{N} \nabla F_i(\mu^k)\right), \tag{32}$$

where ∇F_i is the gradient of F_i (see Equation (17)), and $\mathbf{P}_\mathbb{D}(\mu) = \min\{\overline{\mu}, \max\{0, \mu\}\}$ denotes the projection onto $\mathbb{D}$. In Tikhonov regularization, the regularity of solutions is enforced by an explicitly included penalty. In opposition to that, in iterative regularization methods, a stabilization effect is enforced by early stopping of the iteration. A common stopping rule is the discrepancy principle, where iteration is stopped at the smallest index $k \in \mathbb{N}$ satisfying $\|v - \mathbf{F}(\mu^k)\| \leq \tau\delta$, where δ is an estimate for the noise and $\tau \geq 1$. Formally, the projected Landweber iteration (32) arises as a special case of the proximal gradient iteration (23) for minimizing the Tikhonov functional, where the regularization parameter is taken as $\lambda = 0$ and where the proximal mapping (24) reduces to the orthogonal projection onto $\mathbb{D}$.

In a similar manner, one can also use a stochastic version of the projected Landweber iteration. Using the loping strategy of [43–45] in order to stabilize the iterative process, the resulting projected loping Landweber–Kaczmarz iteration reads

$$\mu^{k+1} = \mathbf{P}_\mathbb{D}\left(\mu^k - s_k \omega_k \nabla F_{i(k)}(\mu^k)\right), \tag{33}$$

$$\omega_k := \begin{cases} 1, & \|\mathbf{F}_{i(k)}(\mu^k) - v_{i(k)}\|_\mathbb{X} > \tau\delta_{i(k)}, \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

Here, $i(k) \in \{1, \ldots, N\}$ for any $k \in \mathbb{N}$ may be randomly selected, $\tau > 1$ is an appropriately chosen positive constant, $\nabla F_i(\cdot)$ is the gradient of the i-th data fidelity term computed in Theorem 3 and $\mathbf{P}_\mathbb{D}(\mu) = \min\{\overline{\mu}, \max\{0, \mu\}\}$ denotes the projection onto $\mathbb{D}$. The iteration (33), (34) terminates if $\|\mathbf{F}_i(\mu^k) - v_i\| \leq \tau\delta$ for all $i \in \{1, \ldots, N\}$. It is worth mentioning that, for noise free data, we have $\omega_k = 1$ for all k and, therefore, in this special situation, the iteration becomes $\mu^{k+1} = \mathbf{P}_\mathbb{D}(\mu^k - s_k \nabla F_{i(k)}(\mu^k))$, which formally arises from the proximal stochastic gradient method (31) with $\lambda = 0$. A convergence analysis of the loping Landweber–Kaczmarz method can be found in [43,44].

4. QPAT as Multilinear Inverse Problem

Since the RTE is time-consuming to solve, we are looking for a suitable reformulation of the inverse problem in multi-source QPAT avoiding computation of a solution of the RTE in each iterative step. In this paper, we propose to write (19) as a multilinear inverse problem, where we add the RTE as

a constraint instead of explicitly including its solution. The new formulation will again be addressed by Tikhonov regularization and proximal stochastic gradient methods.

4.1. Reformulation as Multilinear Inverse Problem

Recall the forward problem of QPAT governed by the RTE (6). With the abbreviation $\mathbf{M}(\mu) := \theta \cdot \nabla_x + \mu_a + \mu_s(\mathbf{I} - \mathbf{K})$, the RTE can be written in compact form $\mathbf{M}(\mu)\Phi = q$, where $\mu = (\mu_a, \mu_s)$ is the unknown parameter pair. In the case of exact data, the multi-source problem in QPAT (19) then can be reformulated as the problem of finding the tuple $\mathbf{z} := (\mu, (\Phi_i, H_i)_{i=1}^N) \in \mathbb{D} \times (\mathbb{W} \times L^2(\Omega))^N$ such that

$$\begin{aligned}
\mathbf{M}(\mu)\Phi_i &= q_i &\text{for } i = 1, \ldots, N, \\
H_i &= \mu_a \, \mathbf{A}\Phi_i &\text{for } i = 1, \ldots, N, \\
v_i &= \mathbf{U}(H_i) &\text{for } i = 1, \ldots, N.
\end{aligned} \tag{35}$$

Here, the index i indicates the i-th illumination, and $q_i \in \mathbb{Q}, \Phi_i \in \mathbb{W}, H_i \in L^2(\Omega), v_i \in \mathbb{Y}$ are the corresponding source, photon density, heating and acoustical data, respectively, and $\mathbf{A}\Phi_i = \int_{\mathbb{S}^{d-1}} \Phi_i(\,\cdot\,, \theta)d\theta$ is the averaging operator. We call (35) and resulting formulations below the multilinear (MULL) formulation of QPAT.

4.2. Application of Tikhonov Regularization

In the case that the data v_i are only known approximately, we use Tikhonov regularization for the stable solution of (35). For that purpose, we approximate (35) by the constrained optimization problem

$$\min_{(\mu,\Phi_i,H_i)_{i=1}^N} \left(\frac{1}{2} \sum_{i=1}^N \|v_i - \mathbf{U}(H_i)\|^2 + \frac{\lambda}{2} \|\mathbf{L}(\mu)\|^2 + \chi_{\mathbb{D}}(\mu) \right),$$

$$\text{s.t.} \begin{cases} \mathbf{M}(\mu)\Phi_i = q_i, \\ H_i = \mu_a \, \mathbf{A}\Phi_i \text{ for } i = 1, \ldots, N. \end{cases} \tag{36}$$

Here, the operator $\mathbf{L}\mu = (\mathbf{L}_a\mu_a, \mathbf{L}_s\mu_s)$ is possibly unbounded, $\frac{\lambda}{2}\|\mathbf{L}(\mu)\|^2$ is the regularization term and $\lambda > 0$ the regularization parameter. Note that (36) is equivalent to (21) and therefore the well-posedness and convergence results of Theorem 5 apply to (36) as well.

The constrained optimization problem (36) proposed in this paper can be addressed by various solution methods, for example using penalty methods or augmented Lagrangian techniques [54]. In this paper, we use a penalty approach for solving (36) where the constraints are included as penalty term. To simplify notation, we introduce the unconstraint functionals

$$\mathbf{J}^{(i)}(\mathbf{z}) := \frac{a_1}{2} \|\mathbf{M}(\mu)\Phi_i - q_i\|^2 + \frac{a_2}{2} \|\mu_a \, \mathbf{A}\Phi_i - H_i\|^2$$

$$+ \frac{a_3}{2} \|v_i - \mathbf{U}(H_i)\|^2 + \frac{\lambda}{2} \|\mathbf{L}(\mu)\|^2 + \chi_{\mathbb{D}}(\mu), \quad (37)$$

for certain parameters $a_1, a_2, a_3 > 0$ and $\mathbf{z}_i := (\mu, \Phi_i, H_i) \in \mathbb{Q} \times \mathbb{W} \times L^2(\Omega)$. The sum of the unconstraint functionals (37) over all illuminations will actually be minimized in our numerical implementations. For that purpose, we define

$$\mathbf{J}_1^{(i)}(\mathbf{z}) = \frac{1}{2} \|\mathbf{M}(\mu)\Phi_i - q_i\|^2,$$

$$\mathbf{J}_2^{(i)}(\mathbf{z}) = \frac{1}{2} \|\mu_a \, \mathbf{A}\Phi_i - H_i\|^2,$$

$$\mathbf{J}_3^{(i)}(\mathbf{z}) = \frac{1}{2} \|v_i - \mathbf{U}(H_i)\|^2,$$

$$\mathbf{J}_4^{(i)}(\mathbf{z}) = \frac{1}{2} \|\mathbf{L}(\mu)\|^2.$$

Then, we have $\mathbf{J}^{(i)}(\mathbf{z}) = \sum_{\ell=1}^{4} a_\ell \mathbf{J}_\ell^{(i)}(\mathbf{z}) + \chi_{\mathbb{D}}(\mu)$. For the approximate solution of (36), we minimize the unconstrained functional $\mathbf{J}(\mathbf{z}) = \sum_{i=1}^{4} \mathbf{J}^{(i)}(\mathbf{z})$, which can be written in the forms

$$\mathbf{J}(\mathbf{z}) = \sum_{i=1}^{N} \sum_{\ell=1}^{4} a_\ell \mathbf{J}_\ell^{(i)}(\mathbf{z}) + \chi_{\mathbb{D}}(\mu), \tag{38}$$

$$\mathbf{J}(\mathbf{z}) = \sum_{i=1}^{N} \sum_{\ell=1}^{3} a_\ell \mathbf{J}_\ell^{(i)}(\mathbf{z}) + \frac{\lambda}{2} \|\mathbf{L}(\mu)\|^2 + \chi_{\mathbb{D}}(\mu). \tag{39}$$

(Here and below, we also write $a_4 = \lambda$, if it simplifies notation.) The formulations (38) as well as (39) can be solved by various optimization techniques. In particular, as the functionals are given as the sum of simpler terms, the stochastic (proximal) gradient method is particularly appealing.

4.3. Solution of the MULL Formulation of QPAT Using Stochastic Gradient Methods

For solving QPAT in the novel MULL formulation (35), we use stochastic gradient methods similar to previous sections. For that purpose, we require the gradients (determining the steepest descent directions) of the individual functionals $\mathbf{J}_\ell^{(i)}(\mathbf{z}_i)$ with respect to $\mathbf{z}_i = (\mu, \Phi_i, H_i)$, which are given as

$$\nabla_{\mu_a} \mathbf{J}_1^{(i)}(\mathbf{z}) = \Phi_i(\mathbf{M}(\mu)\Phi_i - q_i), \tag{40}$$

$$\nabla_{\mu_s} \mathbf{J}_1^{(i)}(\mathbf{z}) = (\mathbf{I} - \mathbf{K})\Phi_i(\mathbf{M}(\mu)\Phi_i - q_i), \tag{41}$$

$$\nabla_{\Phi_i} \mathbf{J}_1^{(i)}(\mathbf{z}) = \mathbf{M}(\mu)(\mathbf{M}(\mu)\Phi_i - q_i), \tag{42}$$

$$\nabla_{\mu_a} \mathbf{J}_2^{(i)}(\mathbf{z}) = (\mathbf{A}\Phi_i)(\mu_a \mathbf{A}\Phi_i - H_i), \tag{43}$$

$$\nabla_{H_i} \mathbf{J}_2^{(i)}(\mathbf{z}) = -(\mu_a \mathbf{A}\Phi_i - H_i), \tag{44}$$

$$\nabla_{\Phi_i} \mathbf{J}_2^{(i)}(\mathbf{z}) = \mathbf{A}^*[\mu_a(\mu_a \mathbf{A}\Phi_i - H_i)], \tag{45}$$

$$\nabla_{H_i} \mathbf{J}_3^{(i)}(\mathbf{z}) = -\mathbf{U}^T(v_i - \mathbf{U}(H_i)), \tag{46}$$

$$\nabla_{\mu_a} \mathbf{J}_4^{(i)}(\mathbf{z}) = \mathbf{L}_a^* \mathbf{L}_a \mu_a, \tag{47}$$

$$\nabla_{\mu_s} \mathbf{J}_4^{(i)}(\mathbf{z}) = \mathbf{L}_s^* \mathbf{L}_s \mu_s. \tag{48}$$

(All other partial gradients are vanishing.) In the following, let N be the number of illuminations, write $\mathbf{z} = (\mu_a, \mu_s, (\Phi_i, H_i)_{i=1}^{N})$ and let $(s_k)_{k \in \mathbb{N}}$ be a sequence of step sizes. In this paper, we propose the following instances of the stochastic proximal gradient method for QPAT based on the multilinear formulation (35).

■ **MULL-projected stochastic gradient algorithm:** Here, we consider the form (38). For any iteration index $k \in \mathbb{N}$ choose $i(k) \in \{1, \dots, N\}$ and $\ell(k) \in \{1, \dots, 4\}$ and define the sequence of iterates $(\mathbf{z}^k)_{k \in \mathbb{N}}$ by

$$\mathbf{z}^{k+1} = (\mathbf{P}_{\mathbb{D}} \times \mathbf{I}) \left(\mathbf{z}^k - s_k \nabla \mathbf{J}_{\ell(k)}^{(i(k))}(\mathbf{z}^k) \right). \tag{49}$$

Here, the mapping $\mathbf{P}_{\mathbb{D}} \times \mathbf{I}$ is the proximal mapping corresponding to $\mathbf{z} \mapsto \chi_{\mathbb{D}}(\mu)$, which equals the projection $\mathbf{P}_{\mathbb{D}}$ in the μ component and equals the identity $\mathbf{I}$ in the other components.

■ **MULL-proximal stochastic gradient algorithm:** Here, we consider the form (39). For any iteration index $k \in \mathbb{N}$ choose $i(k) \in \{1, \dots, N\}$ and $\ell(k) \in \{1, \dots, 3\}$ and define sequence of iterates $(\mathbf{z}^k)_{k \in \mathbb{N}}$ by

$$\mathbf{z}^{k+1} = \text{prox}_{s_k G_\lambda} \left(\mathbf{z}^k - s_k \nabla \mathbf{J}_{\ell(k)}^{(i(k))}(\mathbf{z}^k) \right). \tag{50}$$

The second step implements the proximal mapping of $\mathbf{z} \mapsto s_k G_\lambda(\mu)$ with $G_\lambda(\mu) = \frac{\lambda}{2} \|\mathbf{L}(\mu)\|^2 + \chi_{\mathbb{D}}(\mu)$. As in the previous section, this can be computed with Dykstra's projection algorithm (25)–(28).

For better scaling, in our actual numerical implementation, we replace the scalar step sizes s_k by the adaptive step size rule

$$s_k^{i,\ell} := \arg\min\{\mathbf{z}^k - t\nabla\mathbf{J}_\ell^{(i)}(\mathbf{z}^k) \mid t \in \mathbb{R}\}. \tag{51}$$

Note that computing such step sizes does barely increase the computational time of the stochastic gradient method, since all involved calculations are anyhow necessary for computing the gradient for the iterative update. In opposition to that, calculating a similar adaptive step size for the algorithms proposed in Section 3 would require evaluation of the forward operators $\mathbf{F}_i$ and therefore would significantly increase the computation time. This might be seen as an additional advantage of the novel MULL formulation (35) and its regularized version (36).

5. Numerical Simulations

For the Tikhonov approach to multi-source QPAT, the radiative transfer equation is numerically solved by a streamline diffusion finite element method. Solving the RTE is required to evaluate the forward operator $\mathbf{F}$ and the gradient ∇F of the data fidelity term in every iterative step. For the alternative multilinear approach, these calculations are not necessary. However, the application of the transport operator to Φ has to be calculated for every update of $\mathbf{J}_1$. The simulations are performed on the square domain $\Omega = [-1\,\mathrm{cm}, 1\,\mathrm{cm}]^2$, where the absorption and the scattering coefficient are supported.

5.1. Numerical Solution of the RTE

Employing a finite element scheme, we derive the weak formulation of Equation (6) by integrating against a test function $w\colon \Omega \times \mathbb{S}^1 \to \mathbb{R}$ and replacing the exact solution Φ by a linear combination in the finite element space $\Phi^{(h)} = \sum_{i=1}^{N_h} c_i^{(h)} \psi_i^{(h)}(x,\theta)$ as in [18]. Here, the basis function $\psi_i^{(h)}(x,\theta)$ is the product of a basis function in space and a basis function in velocity. The spatial domain is triangulated uniformly with mesh size h and P_1-Lagrangian element function for the spatial and velocity domain. By choosing the test function $w(x,\theta) = \sum_{j=1}^{N_h} w_j(\psi_j(x,\theta) + D(x,\theta)\theta \cdot \nabla_x \psi_j(x,\theta))$ with streamline diffusion coefficient $D(x,\theta)$, we obtain

$$\int_\Omega \int_{\mathbb{S}^1} (D\theta \cdot \psi_i - \psi_i)\theta \cdot \nabla_x \psi_j \mathrm{d}\theta \mathrm{d}x + \int_{\Gamma_+} |\theta \cdot \nu| \mathrm{d}\sigma$$
$$+ \int_\Omega \int_{\mathbb{S}^1} (\mu_a + \mu_s - \mu_s\mathbf{K})(\psi_j + D\theta \cdot \nabla_x \psi_j)\psi_i \mathrm{d}\theta \mathrm{d}x = \int_{\Gamma_-} |\theta \cdot \nu| \psi_i \psi_j \mathrm{d}\sigma. \tag{52}$$

Equation (52) yields a system of linear equations $M^{(h)}c^{(h)} = b^{(h)}$, where evaluating the left-hand side of (52) provides the entries of $M^{(h)}$, the right-hand side gives the components of vector $b^{(h)}$. Note that the sparsity of matrix $M^{(h)}$ is low and solving the linear system for the Tikhonov approach is very time-consuming. On the other hand, the solution via the MULL formulation requires only a matrix vector multiplication, since in this case $\Phi^{(h)}$ is an independent variable. Thus, only the application to Φ has to be calculated and the transport equation does not need to be solved.

### 5.2. Test Scenario for Multiple Illumination

The sample is illuminated in orthogonal direction at the boundaries of $\Omega = [-1\,\mathrm{cm}, 1\,\mathrm{cm}]^2$. In our simulations, we use $N = 4$ homogenous illuminations and no internal sources. The illuminations are applied separately from each side (left, right, top and bottom) with acoustic data measured on a half circle on the same side as the illumination (see Figure 2). For the scattering kernel, we use the two-dimensional Henyey–Greenstein kernel,

$$k(\theta, \theta') := \frac{1}{2\pi} \frac{1 - g^2}{1 + g^2 - 2g\cos(\theta \cdot \theta')} \quad \text{for } \theta, \theta' \in \mathbb{S}^1,$$

where the anisotropy factor is chosen as $g = 0.5$ in all our experiments.
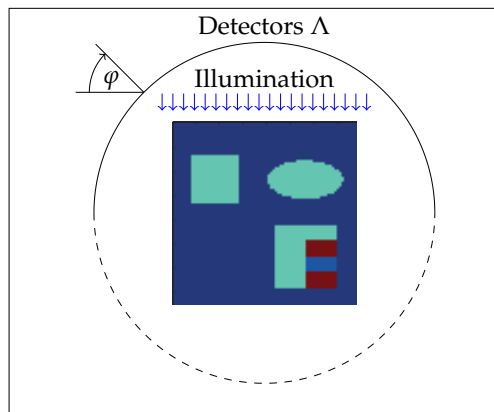
(**a**) Measurement setup.
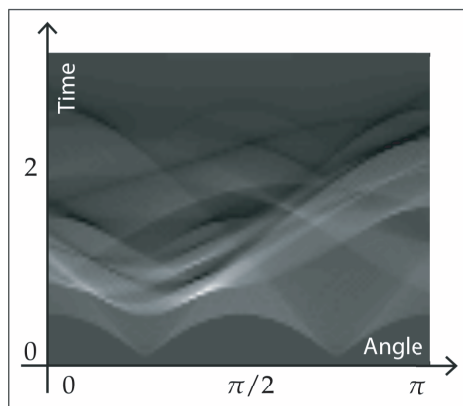
(**b**) Simulated pressure data.

Figure 2. (**a**) The phantom is defined on the square $\Omega = [-1\,\text{cm}, 1\,\text{cm}]^2$ and the acoustic pressure is measured on a semi-circle on the side of the illumination; (**b**) the simulated pressure correspond to the phantom and the illumination in (a) and are represented as gray scale density.

For the simulated data, we choose a spatial mesh size $2/100$, in order to discretize the velocity direction the unit circle is divided in 64 subintervals. In order to avoid inverse crime, for the reconstruction, we use a different spatial mesh size $h = 2/80$ and use $N_\theta = 48$ velocity directions. Calculating the simulated data corresponds to evaluating the forward operators $\mathbf{F}_i$ with perpendicular boundary illumination constant along one side of the boundary square, $q_o(x, \theta) = \delta(\theta - \theta_i)\chi_i(x)1\,\text{mJ}\,\text{cm}^{-1}$, where δ is an approximation of the Dirac delta function and χ_i the indicator function of side i of Ω. In this way, we simulate data

$$v_i = \mathbf{U} \circ \mathbf{H}_i(\mu) + z_i^{\text{noise}} \quad \text{for } i = 1, \dots, 4.$$

Thereby, the heating operator is computed numerically by solving the RTE as described in Section 5.1. The wave operator $\mathbf{U}$ is evaluated by straightforward discretization of the well-known explicit formulas for (10) that can be found, for example, in [55,56]. In the following, we present results for exact data (where $z_i^{\text{noise}} = 0$) as well as for noisy data. For the noisy data case, we add 0.5% random noise to the simulated data, i.e., we take the maximum value of the simulated pressure and add white noise z_i^{noise}

with a standard deviation of 0.5% of that maximal value. The phantom, the setup and the simulated data for one of the four illuminations (top) are shown in Figures 2 and 3.
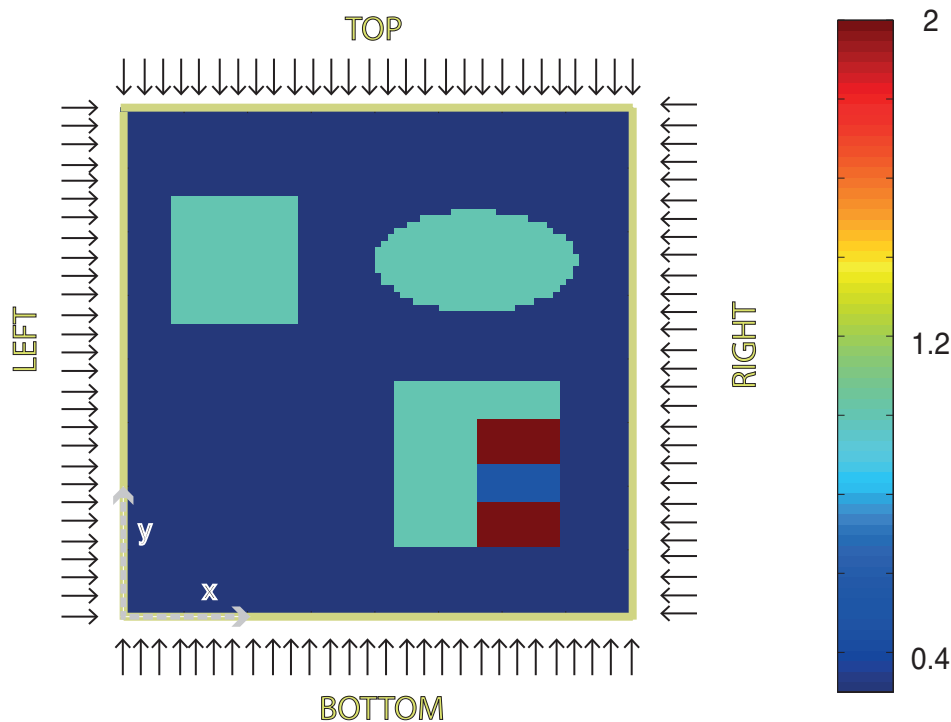

Figure 3. Absorption coefficient distribution of the tissue sample used for the numerical examples. Background absorption of the tissue is taken as $\mu_a = 0.3 \, \text{cm}^{-1}$, the blue obstacles have $\mu_a = 1 \, \text{cm}^{-1}$ and the red stripes $\mu_a = 2 \, \text{cm}^{-1}$. The area between the red stripes has absorption coefficient $\mu_a = 0.5 \, \text{cm}^{-1}$. The scattering coefficient is constant in the whole sample and chosen to be $\mu_s = 3 \, \text{cm}^{-1}$. Illuminations are applied consecutively from top, right, bottom and left. The corresponding boundary sources are given by $q_o(x, \theta) = \delta(\theta - \theta_i)\chi_i(x) 1 \, \text{mJ} \, \text{cm}^{-1}$. The x- and y-axis cover $[-1 \, \text{cm}, 1 \, \text{cm}]$.

5.3. Numerical Results

For regularizing the absorption and scattering coefficient, we make use of Laplace regularization and choose $\mathbf{L}_a = \Delta$ and $\mathbf{L}_s = 100\Delta$, respectively. We assume that the coefficient μ is known at the boundary of Ω and is therefore used as the starting value of our iterative schemes. Furthermore, we use the boundary value of μ for regularization; that is, we implement it in the Dykstra projection procedure (26) by iteratively projecting on the known boundary value. In the following, we discuss the methods that we have outlined in the previous section.

■ **Standard formulation of QPAT** (19)**:** We assume that the scattering coefficient is known and we restrict ourself to reconstructing the absorption coefficient. Then, the proximal gradient and proximal stochastic gradient algorithm, respectively, read

$$\mu_a^{k+1} = \text{prox}_{s_k G_\lambda}\left(\mu_a^k - \frac{s_k}{4}\sum_{i=1}^{4}\nabla F_i(\mu_a^k, \mu_s)\right), \tag{53}$$

$$\mu_a^{k+1} = \text{prox}_{s_k G_\lambda}\left(\mu_a^k - s_k \nabla F_{i(k)}(\mu_a^k, \mu_s)\right). \tag{54}$$

In contrast, to the full proximal gradient algorithm, the proximal stochastic gradient algorithm avoids evaluating the full gradient ∇F, but selects randomly an illumination number $i \in \{1, \ldots, 4\}$ for each iterative step. Because of formula (16), each iteration of the above procedures requires the calculation of the solution of the radiative transfer equation Φ as well of its adjoint Φ^*.

The top row in Figure 4 shows reconstruction result for the absorption coefficient using the original formulation with the proximal gradient method with $\lambda = 2 \times 10^{-8}$ and 10 iterative steps. The left picture shows the relative error $\|\mu_a - \mu_a^k\|/\|\mu_a\|$. Note that, in this case, solutions of the RTE and its adjoint have to be computed for four illuminations per iterative step. The reconstruction results in the bottom row in Figure 4 are obtained by the proximal stochastic gradient method with $\lambda = 2 \times 10^{-7}$. The regularization parameters λ have been selected empirically as a trade-off between stability and accuracy. The total number of iterations is taken as 30. In each iteration, a illumination pattern is chosen randomly and the computation of RTE and its adjoint is executed only for this single illumination. Therefore, the computational effort for the proximal stochastic gradient method is approximately 3/4 of the proximal gradient algorithm using full gradients. For the algorithms based on the standard formulation (19), calculating adaptive step sizes similar to (51) is time-consuming as this requires another evaluation of the forward operator $\mathbf{F}_i$ and therefore another solution of the RTE. Therefore, we simply use a constant step size rule; in our numerical experiments, it turned out that $s_k = 0.5$ is a suitable choice.
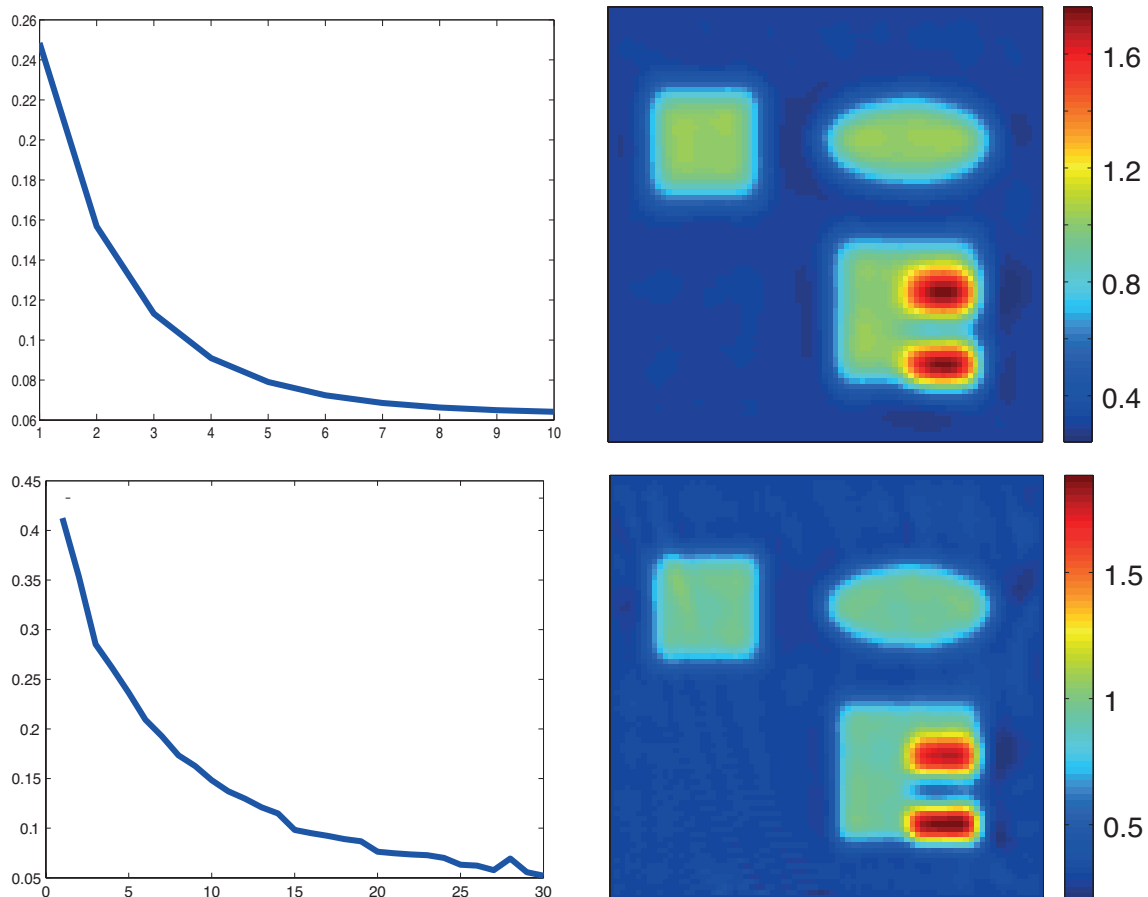

Figure 4. Reconstruction results based on standard formulation (19). **Top**: proximal gradient method; **Bottom**: proximal stochastic gradient method. The left images show the relative reconstruction errors of the reconstructed absorption coefficient as a function of the number of iterations, whereas the right pictures show the result after the final iteration. (The phantom is as described in Figure 3.)

■ **Novel MULL formulation of QPAT** (35)**:** The multilinear approach overcomes the problem of solving the RTE by minimizing (38) or (39). In both cases, one selects an arbitrary functional and performs a steepest descent step, resulting in an iterative scheme for the variables μ_a, μ_s, Φ and H. Recall that none of the partial gradients (40)–(48) requires solving the RTE (which is the most

time-consuming part for the standard formulation of QPAT). In each iterative step, we take a random illumination number $i \in \{1, \ldots, 4\}$ and a random functional number $\ell \in \{1, 2, 3\}$. The gradient step then consists of the update rule

$$
\begin{pmatrix} \mu \\ \Phi_i \\ H_i \end{pmatrix}^{k+1} = \begin{pmatrix} \mu \\ \Phi_i \\ H_i \end{pmatrix}^k + s_k \cdot \nabla \mathbf{J}_\ell ((\mu, \Phi_i, H_i)^k). \tag{55}
$$

Dykstra's algorithm for smoothing the μ component is applied after each iterative step when $\ell \in \{1, 2\}$. Iteration (55) contains a gradient step for the RTE. Since one gradient step is not enough to obtain an appropriate approximation to the solution of the transport equation, we apply iteration (55) 40 times whenever $\ell = 1$ is chosen. In this situation, we apply the Dykstra iteration in the μ component after these 40 iteration steps, whereas the positivity projection is done in every step. Flowcharts of the stochastic gradient algorithms (standard and MULL formulations) are shown in Figure 5. For the projected stochastic gradient method, regularization of μ is done by incorporating the regularization functional $\mathbf{J}_4$ in the random choice of functionals; see (38). The positivity restriction is realized with the cut projection $\mathbf{P}_\mathbb{D}(\mu) = \max\{0, \mu\}$ applied after every iterative step.
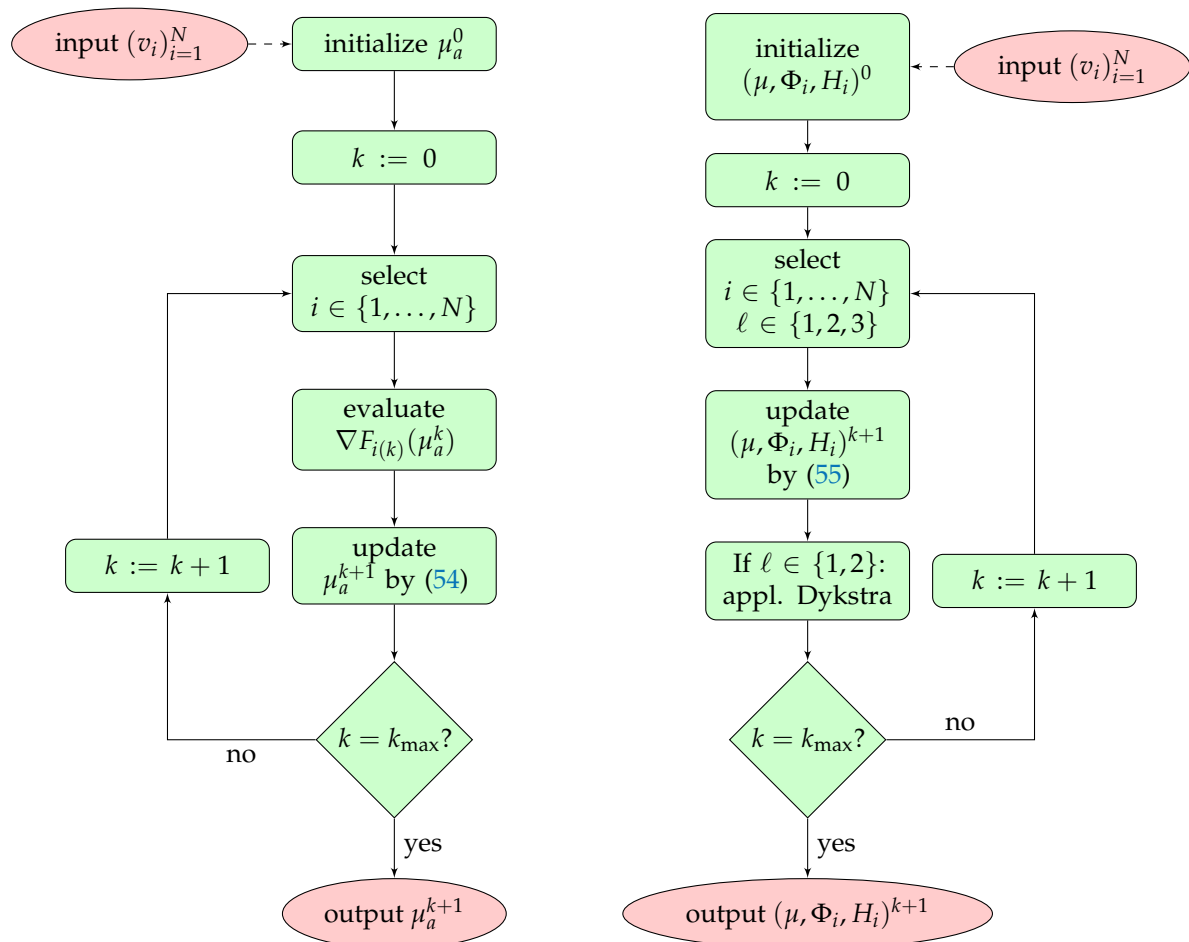

Figure 5. Flowcharts of stochastic gradient algorithms for QPAT proposed in this paper. **Left**: algorithm based on the standard formulation (19). **Right**: algorithm based on the novel MULL formulation (35). The update (54) using the standard formulation requires solving the forward RTE and the adjoint RTE, which is not required by (55) with the MULL formulation. Simulations are performed with $N = 4$.

Figure 6 shows reconstructions with the stochastic gradient methods for the novel MULL formulation of QPAT (35). The results in the top row are for the MULL-proximal stochastic gradient algorithm with $\lambda = 2 \times 10^{-8}$ and in the the bottom row results for the MULL-projected stochastic gradient method with $\lambda = 2 \times 10^{-8}$ are shown. In both cases, we used 1000 iterations.
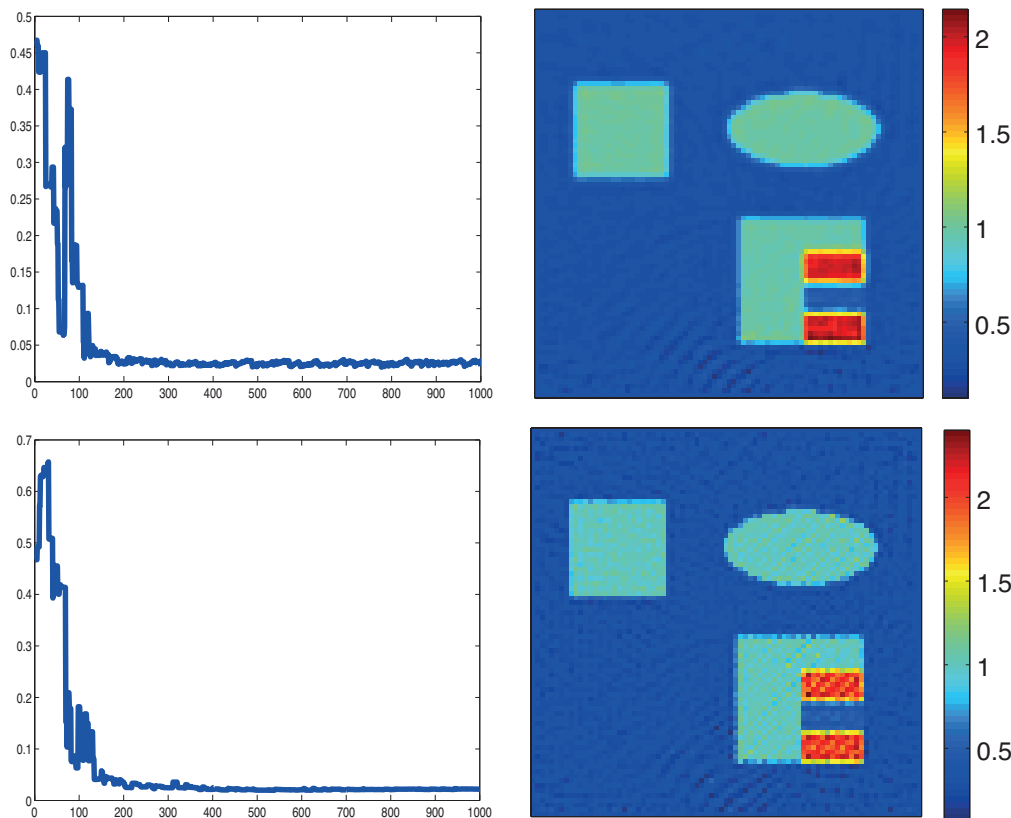

Figure 6. Reconstruction results based on the novel MULL formulation (35). **Top**: MULL-proximal stochastic gradient method based on the decomposition (38). **Bottom**: MULL-projected stochastic gradient method based on the decomposition (39). The left images show the relative reconstruction errors of the reconstructed absorption coefficient as a function of the number of iterations, whereas the right pictures show the results after the last iterations. (The phantom is as described in Figure 3.)

Remark 1. *Note that in the stochastic gradient methods for the novel MULL formulation of QPAT calculating the matrix vector product $M^{(h)} \cdot \Phi$ is the most costly part. In contrast, the standard formulation (19) requires the solution of the system $M^{(h)}c(h) = b^{(h)}$. Since the matrix $M^{(h)}$ is sparse only in its spatial domain, this is very time-consuming. On the other hand, the matrix $M^{(h)}$ (which is a discretization of $\theta \cdot \nabla_x + \mu_a + \mu_s(\mathbf{I} - \mathbf{K})$) has a simple dependence on the variables μ_a, μ_s. We therefore can compute the velocity entries of $M^{(h)}$ at the beginning of the iterative process to save computation time.*

The reconstruction times for the final reconstructions using all methods described above are shown in Table 1. For the standard formulation of QPAT, the reconstruction times seem to be in accordance with reported results using gradient or Newton-type methods for QPAT (see, for example, [48].) The methods based on the new MULL formulation (35) (after 1000 iterations) are faster than the methods based on the standard formulation (19) of QPAT (after 10, respectively, 40 iterations). From the relative reconstruction errors shown in Figures 4 and 6, one notices that, opposed to the methods based on (19), the methods using the MULL formulation could even be stopped much earlier while still obtaining a comparable reconstruction quality. We roughly estimate a speedup of a factor 10 using the novel MULL formulation instead of the standard formulation of QPAT.

Table 1. Reconstruction times for all methods. Recall that one iteration of the proximal stochastic gradient method is approximately four times cheaper than one iteration of the full proximal gradient method (both based on (19)). Further recall that one step in the methods based on the MULL formulation (35) is much less time consuming than for the methods based on (19); see Remark 1.

Algorithm	Model	Update	No. Iterations	Reconstruction Time
Proximal gradient	(19)	(23)	10	27.2 h
Proximal stochastic gradient	(19)	(31)	30	24.4 h
MULL-proximal stochastic gradient	(35)	(49)	1000	14.7 h
MULL-projected stochastic gradient	(35)	(50)	1000	11.9 h

In Figure 7, we show results for noisy data using the proximal gradient method based on the standard formulation (19) (top) and the proximal stochastic gradient method using the MULL formulation for QPAT (bottom). The regularization parameter is chosen as in the exact data case. Finally, in Figure 8, we show reconstruction results using only two consecutive illuminations applied from the top and from the left with noisy data. We use 10 iterations for the proximal gradient algorithm based on (19) (shown in left image in Figure 8) and 500 iterations for the stochastic gradient algorithms based on the MULL formulation (35) (shown in the right image in Figure 8).
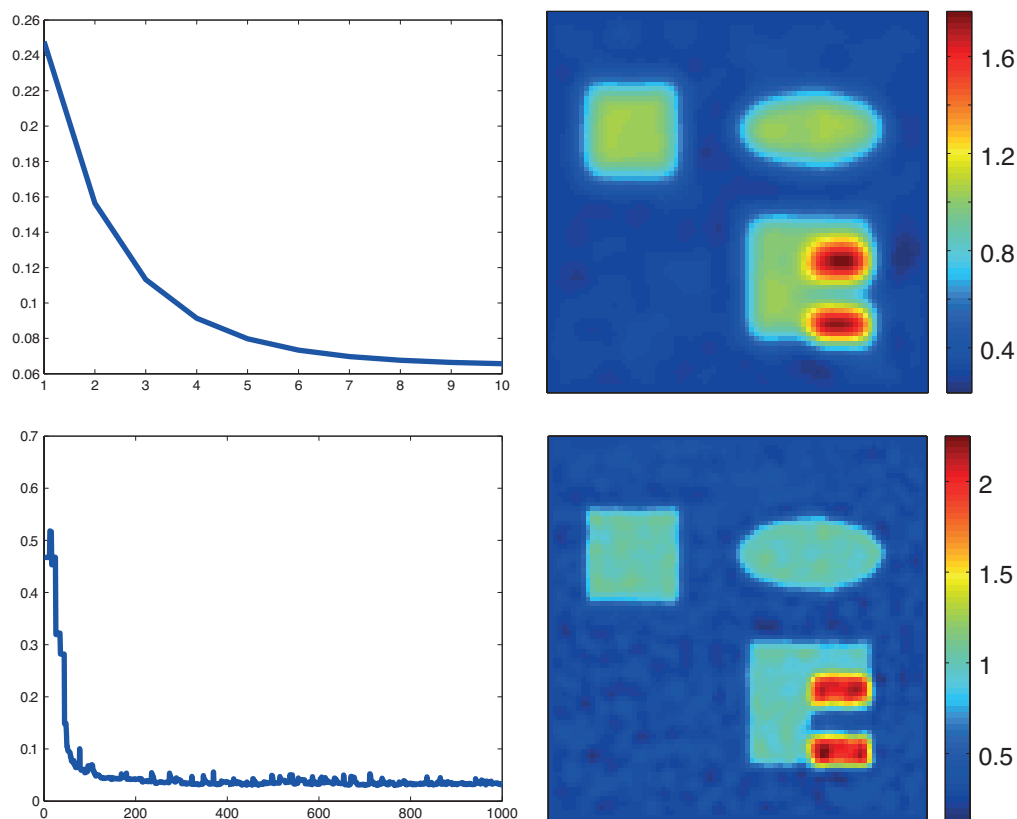

Figure 7. Reconstruction results from noisy data. **Top**: Proximal gradient method based on (19). **Bottom**: MULL-proximal stochastic gradient method. The left images show the relative reconstruction errors of the reconstructed absorption coefficient as a function of the number of iterations, whereas the right pictures show the results after the last iterations. (The phantom is as described in Figure 3.)
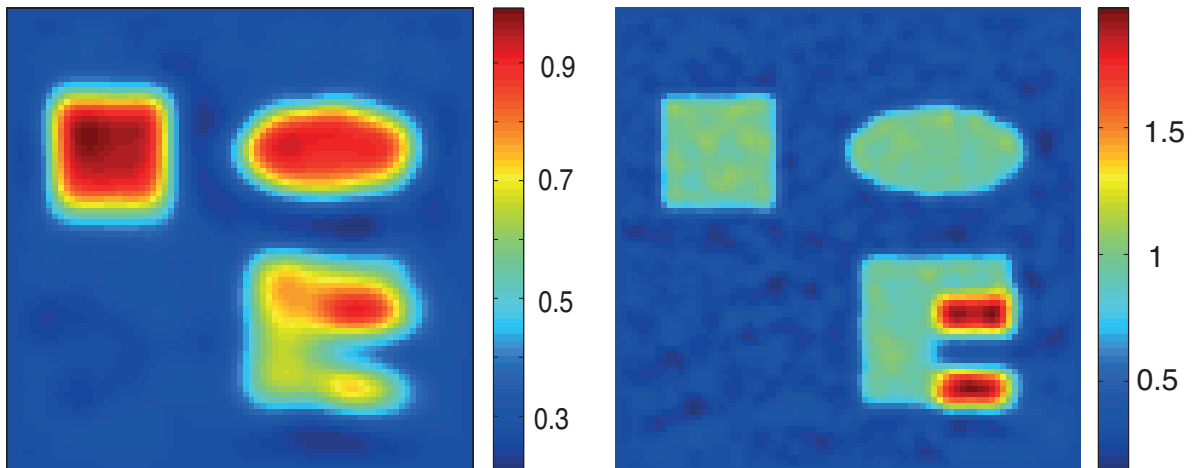
Figure 8. Reconstruction results from noisy data for two illuminations. **Left**: proximal gradient method based on (19) using 10 iterations. **Right**: MULL-proximal stochastic gradient method using only 500 iterations. The phantom is as described in Figure 3 and, for the reconstruction methods, we use two consecutive illuminations (from the top and from the left). The reconstruction time has been about 14 h for the method based on the standard formulation (19) and 3 h for the proposed MULL-proximal stochastic gradient method.

6. Conclusions

In this paper, we developed efficient proximal stochastic gradient methods for image reconstruction in multi-source QPAT. We used the RTE as an accurate model for light transport and employed the single stage approach for QPAT introduced in [18]. One class of the proximal stochastic gradient methods has been developed based on the standard formulation for QPAT given in (19). Additionally, we developed another class using proximal stochastic gradient methods for the new MULL formulations (35) and (36) for QPAT. Besides proposing proximal stochastic gradient methods for QPAT, we also consider the formulations (35) and (36) as the main contributions of the present article. These new formulations avoid the time-consuming evaluation of the RTE at each iteration and allow for treating the QPAT problem as a constrained optimization problem, which enables the use of a variety of numerical algorithms. Here, we used a penalty approach in combination with stochastic gradient methods for the solution. Future work will be done in the direction of developing new algorithms based on (35) and (36). Additionally, we will investigate the use of different regularization terms in (36). Finally, the theoretical convergence analysis of proximal gradient algorithms and other iterative algorithms for solving (35) will be the subject of future research.

Author Contributions: Markus Haltmeier, Lukas Neumann and Simon Rabanser developed the reconstruction algorithms and the numerical implementation; Simon Rabanser performed the numerical experiments; Markus Haltmeier, Lukas Neumann and Simon Rabanser wrote the paper; Markus Haltmeier and Lukas Neumann supervised the project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Beard, P. Biomedical photoacoustic imaging. *Interface Focus* **2011**, *1*, 602–631.
2. Wang, L.V. Multiscale photoacoustic microscopy and computed tomography. *Nat. Photonics* **2009**, *3*, 503–509.

3. Agranovsky, M.; Kuchment, P.; Kunyansky, L. On reconstruction formulas and algorithms for the thermoacoustic tomography. In *Photoacoustic Imaging and Spectroscopy*; Wang, L.V., Ed.; CRC Press: Boca Raton, FL, USA, 2009; pp. 89–101.

4. Burgholzer, P.; Bauer-Marschallinger, J.; Grün, H.; Haltmeier, M.; Paltauf, G. Temporal back-projection algorithms for photoacoustic tomography with integrating line detectors. *Inverse Probl.* **2007**, *23*, S65–S80.

5. Haltmeier, M. Universal inversion formulas for recovering a function from spherical means. *SIAM J. Math. Anal.* **2014**, *46*, 214–232.

6. Haltmeier, M.; Nguyen, L.V. Analysis of iterative methods in photoacoustic tomography with variable sound speed. *SIAM J. Imaging Sci.* **2017**, *10*, 751–781.

7. Haltmeier, M.; Schuster, T.; Scherzer, O. Filtered backprojection for thermoacoustic computed tomography in spherical geometry. *Math. Methods Appl. Sci.* **2005**, *28*, 1919–1937.

8. Huang, C.; Wang, K.; Nie, L.; Wang, L.V.; Anastasio, M.A. Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media. *IEEE Trans. Med. Imaging* **2013**, *32*, 1097–1110.

9. Nguyen, L.V.; Kunyansky, L.A. A Dissipative Time Reversal Technique for Photoacoustic Tomography in a Cavity. *SIAM J. Imaging Sci.* **2016**, *9*, 748–769.

10. Rosenthal, A.; Ntziachristos, V.; Razansky, D. Acoustic Inversion in Optoacoustic Tomography: A Review. *Curr. Med. Imaging Rev.* **2013**, *9*, 318–336.

11. Ammari, H.; Bossy, E.; Jugnon, V.; Kang, H. Reconstruction of the Optical Absorption Coefficient of a Small Absorber from the Absorbed Energy Density. *SIAM J. Appl. Math.* **2011**, *71*, 676–693.

12. Bal, G.; Jollivet, A.; Jugnon, V. Inverse transport theory of photoacoustics. *Inverse Probl.* **2010**, *26*, 025011.

13. Bal, G.; Ren, K. Multi-source quantitative photoacoustic tomography in a diffusive regime. *Inverse Probl.* **2011**, *27*, 075003.

14. Chen, J.; Yang, Y. Quantitative photo-acoustic tomography with partial data. *Inverse Probl.* **2012**, *28*, 115014.

15. Cox, B.T.; Arridge, S.A.; Beard, P.C. Gradient-Based Quantitative Photoacoustic Image Reconstruction for Molecular Imaging. *Proc. SPIE* **2007**, *6437*, 64371T.

16. Cox, B.T.; Arridge, S.R.; Köstli, P.; Beard, P.C. Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method. *Appl. Opt.* **2006**, *45*, 1866–1875.

17. Cox, B.T.; Laufer, J.G.; Arridge, S.R.; Beard, P.C. Quantitative spectroscopic photoacoustic imaging: A review. *J. Biomed. Opt.* **2012**, *17*, 0612021.

18. Haltmeier, M.; Neumann, L.; Rabanser, S. Single-stage reconstruction algorithm for quantitative photoacoustic tomography. *Inverse Probl.* **2015**, *31*, 065005.

19. Haltmeier, M.; Neumann, L.; Nguyen, L.V.; Rabanser, S. Analysis of the Linearized Problem of Quantitative Photoacoustic Tomography. *arXiv* **2017**, arXiv:1702.04560.

20. Kruger, R.A.; Lui, P.; Fang, Y.R.; Appledorn, R.C. Photoacoustic Ultrasound (PAUS)–Reconstruction Tomography. *Med. Phys.* **1995**, *22*, 1605–1609.

21. Mamonov, A.V.; Ren, K. Quantitative photoacoustic imaging in radiative transport regime. *Commun. Math. Sci.* **2014**, *12*, 201–234.

22. Naetar, W.; Scherzer, O. Quantitative Photoacoustic Tomography with Piecewise Constant Material Parameters. *SIAM J. Imaging Sci.* **2014**, *7*, 1755–1774.

23. Ren, K.; Gao, H.; Zhao, H. A hybrid reconstruction method for quantitative PAT. *SIAM J. Imaging Sci.* **2013**, *6*, 32–55.

24. Rosenthal, A.; Razansky, D.; Ntziachristos, V. Fast Semi-Analytical Model-Based Acoustic Inversion for Quantitative Optoacoustic Tomography. *IEEE Trans. Med. Imaging* **2010**, *29*, 1275–1285.

25. Tarvainen, T.; Cox, B.T.; Kaipio, J.P.; Arridge, S.A. Reconstructing absorption and scattering distributions in quantitative photoacoustic tomography. *Inverse Probl.* **2012**, *28*, 084009.

26. Yao, L.; Sun, Y.; Jiang, H. Transport-based quantitative photoacoustic tomography: Simulations and experiments. *Phys. Med. Biol.* **2010**, *55*, 1917–1934.

27. Arridge, S.R. Optical tomography in medical imaging. *Inverse Probl.* **1999**, *15*, R41–R93.

28. Dautray, R.; Lions, J. *Mathematical Analysis and Numerical Methods for Science and Technology*; Springer: Berlin, Germany, 1993.

29. Egger, H.; Schlottbom, M. Numerical methods for parameter identification in stationary radiative transfer. *Comput. Optim. Appl.* **2015**, *62*, 67–83.

30. Kanschat, G. Solution of radiative transfer problems with finite elements. In *Numerical Methods in Multidimensional Radiative Transfer*; Springer: Berlin, Germany, 2009; pp. 49–98.

31. Gao, H.; Feng, J.; Song, L. Limited-view multi-source quantitative photoacoustic tomography. *Inverse Probl.* **2015**, *31*, 065004.

32. Engl, H.W.; Hanke, M.; Neubauer, A. *Regularization of Inverse Problems*; Springer Science & Business Media: Berlin, Germany, 1996.

33. Kaltenbacher, B.; Neubauer, A.; Scherzer, O. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*; Walter de Gruyter: Berlin, Geramny, 2008.

34. Scherzer, O.; Grasmair, M.; Grossauer, H.; Haltmeier, M.; Lenzen, F. *Variational Methods in Imaging*; Applied Mathematical Sciences; Springer: New York, NY, USA, 2009.

35. Combettes, P.L.; Wajs, V.R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **2005**, *4*, 1168–1200.

36. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*; Springer: Berlin, Germany, 2011.

37. Bertsekas, D.P. Incremental gradient, subgradient, and proximal methods for convex optimization: A survey. In *Optimization for Machine Learing*; Sra, S., Nowozin, S., Wright, S.J., Eds.; The MIT Press: London, UK, 2012.

38. Bertsekas, D.P. Incremental proximal methods for large scale convex optimization. *Math. Program.* **2011**, *129*, 163–195.

39. Xiao, L.; Zhang, T. A proximal stochastic gradient method with progressive variance reduction. *SIAM J. Optim.* **2014**, *24*, 2057–2075.

40. Duchi, J.; Singer, Y. Efficient online and batch learning using forward backward splitting. *J. Mach. Learn. Res.* **2009**, *10*, 2899–2934.

41. Li, H.; Haltmeier, M. The Averaged Kaczmarz Iteration for Solving Inverse Problems. *arXiv* **2017**, arXiv:1709.00742.

42. Pereyra, M.; Schniter, P.; Chouzenoux, E.; Pesquet, J.-C.; Tourneret, J.Y.; Hero, A.O.; McLaughlin, S. A survey of stochastic simulation and optimization methods in signal processing. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 224–241.

43. De Cezaro, A.; Haltmeier, M.; Leitão, A.; Scherzer, O. On steepest-descent-Kaczmarz methods for regularizing systems of nonlinear ill-posed equations. *Appl. Math. Comput.* **2008**, *202*, 596–607.

44. Haltmeier, M.; Leitao, A.; Scherzer, O. Kaczmarz methods for regularizing nonlinear ill-posed equations I: convergence analysis. *Inverse Probl. Imaging* **2007**, *1*, 289–298.

45. Haltmeier, M.; Kowar, R.; Leitao, A.; Scherzer, O. Kaczmarz methods for regularizing nonlinear ill-posed equations II: Applications. *Inverse Probl. Imaging* **2007**, *1*, 507–523.

46. Alberti, G.; Ammari, H. Disjoint sparsity for signal separation and applications to hybrid inverse problems in medical imaging. *Appl. Comput. Harmon. Anal.* **2017**, *42*, 319–349.

47. Ammari, H.; Garnier, J.; Kang, H.; Nguyen, L.; Seppecher, L. *Multi-Wave Medical Imaging, Mathematical Modelling & Imaging Reconstruction*; World Scientific Publishing: London, UK, 2017.

48. Saratoon, T.; Tarvainen, T.; Cox, B.T.; Arridge, S.R. A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation. *Inverse Probl.* **2013**, *29*, 075006.

49. Wang, C.; Zhou, T. On iterative algorithms for quantitative photoacoustic tomography in the radiative transport regime. *Inverse Probl.* **2017**, *33*, 115006.

50. Egger, H.; Schlottbom, M. Stationary radiative transfer with vanishing absorption. *Math. Models Methods Appl. Sci.* **2014**, *24*, 973–990.

51. Finch, D.; Patch, S.K.; Rakesh. Determining a function from its mean values over a family of spheres. *SIAM J. Math. Anal.* **2004**, *35*, 1213–1240.

52. Finch, D.; Haltmeier, M.; Rakesh. Inversion of spherical means and the wave equation in even dimensions. *SIAM J. Appl. Math.* **2007**, *68*, 392–412.

53. Combettes, P.L.; Pesquet, J.C. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*; Springer: New York, NY, USA, 2011; pp. 185–212.

54. Ito, K.; Kunisch, K. *Lagrange Multiplier Approach to Variational Problems and Applications*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 2008.

55. John, F. *Partial Differential Equations*, 4th ed.; Applied Mathematical Sciences; Springer: New York, NY, USA, 1982.

56. Evans, L.C. *Partial Differential Equations*; Graduate Studies in Mathematics; American Mathematical Society: Providence, RI, USA, 1998.