SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Efficient reliable image reconstruction schemes for diffuse optical tomography

H. Egger M. Schlottbom

SFB-Report No. 2010–019 April 2010

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

Der Wissenschaftsfonds.

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Efficient reliable image reconstruction schemes for diffuse optical tomography

Herbert Egger* **Matthias Schlottbom†**

Abstract

This paper deals with the design and analysis of reliable and efficient numerical methods for the solution of inverse problems in diffuse optical tomography. The proposed fully discrete algorithms are based on iterative regularization methods, derived and analyzed on the continuous level, and their careful discretization by finite element methods. This guarantees convergence of the fully discrete algorithms under the same conditions as required on the continuous level, and allows to establish mesh independent reliability and performance. The derivatives and adjoints can be shown to be exact on the discrete level, which implies that the Gauß-Newton systems used in the reconstructions are real symmetric and positive definite, and the conjugate gradient method can be used for their efficient solution. We also present complexity estimates of the algorithms, and discuss a-posteriori error estimators for assessing the discretization error. The efficiency and robustness of the proposed methods is demonstrated in numerical tests.

1 Introduction

Optical diffusion tomography is a non-invasive imaging technique that utilizes near-infrared light to probe highly scattering media. Some typical applications are the monitoring of the oxygenation state of blood in the neonatal brain, or the detection of breast cancer, see e.g. [1, 14, 18, 19] and the references therein. In this paper, we consider the design and analysis of reliable and efficient fully discrete algorithms for image recontruction in diffuse optical tomography.

Due to the diffuse nature of light propagation in highly scattering media, the tomographic image reconstruction, i.e., the identification of the distribution of optical parameters from boundary measurements of light intensities, is a severely ill-posed and nonlinear inverse problem. The discretization of the governing equations, e.g., by a finite element method [30], turns the inverse problem into a large nonlinear ill-conditioned system of equations, and some regularization method has to be used for a stable solution. Several methods based on linearization or successive linearization have been applied successfully to the solution of the discretized inverse problem. We would like to mention in particular linear backprojection methods [6] or Newton-type iterations [28, 31], and refer to [1] for a comprehensive overview on discretization techniques as well as frequently used reconstruction algorithms.

While most of the work on numerical methods for optical diffusion tomography is concerned

* Institute for Mathematics and Scientific Computing, University Graz, Heinrichstraße 36, 8010 Graz, Austria.
† Aachen Institute for Advanced Study in Computational Engineering Science, RWTH Aachen University, Schinkelstraße 2, 52062 Aachen, Germany.

with the formulation of various discrete algorithms, and illustration of their good behaviour for certain test problems, we pursue a different goal here, namely we want to establish the reliability and efficiency of appropriate discrete reconstruction schemes. Such an approach inevitably requires a detailed analysis of the inverse problem and of appropriate regularization methods on the continuous level. This involves in particular the choice of appropriate topologies, which allow to establish continuity and differentiability of the forward operator. By a careful discretization, many properties derived on the continuous level can then be transfered to the discrete setting, which allows to establish convergence of fully discrete algorithms essentially under the same conditions that are required also on the continuous level. Moreover, the quality of reconstructions and the performance of the algorithms automatically becomes independent of the discretization level, as long as the discretization error is sufficiently small. In this way, the reliability and efficiency of the fully discrete algorithms can be ensured independent of the discretization level. This mesh-independent behaviour of our reconstruction algorithms will also be demonstrated in numerical tests.

The outline of the paper is as follows: In Section 2, we introduce the governing boundary value problem, and we define and analyze the forward operator, which maps the unknown parameters to the measurements. Section 3 then is concerned with the formulation of the inverse problem and its regularized solution by Tikhonov regularization. Two projected iterative reconstruction algorithms are presented in Section 4, and some convergence results are established as a theoretical backup. We also discuss the convergence of the iterations to stationary points and local minimizers of the Tikhonov functional, respectively. The remaining sections are concerned with the numerical realization: We first discuss in Section 5 the discretization of the governing boundary value problems by finite element methods, and show that important properties derived on the continuous level carry over verbatim to the discrete setting. This allows us to apply convergence results derived on the continuous level also for the discrete problems, and to obtain mesh-independent algorithms. Section 6 then summarizes some complexity estimates for the discrete algorithms. In Section 7, we remark on further implementational details, e.g., the estimation of the discretization error (which enters as part of the data noise), and the scaling of the measurements, which allows to improve the reconstructions considerably. Section 8 presents some results of numerical tests that demonstrate the reliability and mesh-independence of the algorithms, and we conclude with a short discussion of our results.

2 Physical model and the forward operator

The transport of intensity modulated light in highly scattering media is modelled by the diffusion approximation [1]

$$-\nabla \cdot (\kappa \nabla u) + (\mu + \iota k)u = 0 \qquad \text{in } \Omega, \tag{1}$$

$$\kappa \partial_n u + \rho u = \rho q \qquad \text{on } \partial\Omega. \tag{2}$$

Here, u is the complex amplitude of the photon density, κ and μ are the optical parameters characterizing the material under investigation, and k is the wavenumber, which depends on the modulation frequency ω. The Robin boundary condition (2) models a diffuse light source located at the boundary, and the parameter $\rho > 0$ is related to the refractive index of the material.

The following conditions on the domain and the coefficients are assumed to hold throughout this presentation.

Assumption 2.1. *(A1)* $\Omega \subset \mathbb{R}^m$, $m = 2, 3$, *is a bounded domain with* $C^{1,1}$ *boundary.*

(A2) The function κ *is uniformly positive and bounded, i.e., there exist positive constants* $\underline{\kappa}$, $\overline{\kappa}$ *such that* $\underline{\kappa} \le \kappa \le \overline{\kappa}$ *in* Ω, *and* $\overline{\kappa}/\underline{\kappa}$ *is sufficiently close to one.*

(A3) The function $\mu \in L^\infty(\Omega)$ *is non-negative and bounded from above, i.e., there exist non-negative constants* $\underline{\mu}$, $\overline{\mu}$ *such that* $\underline{\mu} \le \mu \le \overline{\mu}$ *in* Ω.

(A4) ρ *is a bounded and uniformly positive function on* $\partial\Omega$, *i.e., there exists positive constants* $\underline{\rho}$, $\overline{\rho}$ *such that* $\underline{\rho} \le \rho \le \overline{\rho}$ *on* $\partial\Omega$.

The complex version of the Lax-Milgram Theorem [10, Ch VII, Thm 1] ensures unique solvability of the boundary value problem (1)–(2).

Theorem 2.2. *Let (A1)–(A4) hold. Then the boundary value problem* (1)–(2) *has a unique (complex valued) solution* $u \in H^1(\Omega; \mathbb{C})$ *that satisfies* $\|u\|_{H^1(\Omega)} \le C\|q\|_{L^2(\partial\Omega)}$ *with a constant* C *depending only on the domain* Ω *and on the bounds for the coefficients.*

Remark 2.3. The smoothness of the boundary $\partial\Omega$ and the smallness condition on the ratio $\overline{\kappa}/\underline{\kappa}$ are not required for this statement, but for the differentiability results presented in the next section. Also, sources from the space $H^{-1/2}(\partial\Omega)$ could be allowed here, but L^2-regularity of the source terms is required for the results of the next section. Under the conditions of the theorem, the solution u can be shown to have some extra regularity, viz. $u \in W^{1,3}(\Omega)$, cf [12].

2.1 Measurements and the forward operator

The measurable quantity in optical tomography is the complex amplitude of the emitted light, which is given by [1]

$$m_D(u) = \int_{\partial\Omega} \eta_D(u - q)ds. \tag{3}$$

Here, $\eta_D \in L^\infty(\partial\Omega)$ is some given function modelling the characteristics of a detector. An optical experiment then consists of exciting Ω with a sequence of light sources q_j, $1 \le j \le s$ and recording the corresponding measurements $M_{ij} = m_i(u_j)$ with the detectors η_i, $1 \le i \le d$. A mathematical description of such an experiment is provided by a nonlinear forward operator

$$F : \mathcal{D}(F) \to \mathbb{C}^{d\times s}, \qquad (\kappa, \mu) \mapsto [M_{ij}], \qquad 1 \le i \le d, 1 \le j \le s, \tag{4}$$

which maps prescribed optical parameters to the corresponding virtual measurements. Using Theorem 2.2 and the trace theorem [10], the forward operator can be seen to be well-defined for all parameters (κ, μ) in the set

$$\mathcal{D}(F) := \{(\kappa, \mu) \text{ satisfying assumptions (A2)–(A3)}\} \tag{5}$$

of admissible parameters.

Remark 2.4. The image space of the forward operator was chosen to be finite dimensional here, which is the typical situation in experiments. The results stated below can however be generalized to the case of continuous measurements; e.g., one could similarly consider measurements $M \in [L^2(\partial\Omega, \mathbb{C})]^s$ modelling observation of the emitted light at the whole boundary but for finitely many excitations q_i, or full observation $M \in \mathcal{L}(L^2(\partial\Omega, \mathbb{C}), L^2(\partial\Omega, \mathbb{C}))$ of the Robin-to-Dirichlet map; for corresponding results we refer to [12].

2.2 Properties of the forward operator

The forward operator will be analyzed with respect to the following topologies: For the complex valued measurements we use the standard euclidean norm $\|M\|^2_{\mathbb{C}^{d\times s}} := \sum_{i,j} |M_{ij}|^2$, and the parameter space is equipped with the norm of $H^1(\Omega) \times L^2(\Omega)$.

Remark 2.5. The set $\mathcal{D}(F)$ is closed and convex with empty interior with respect to this topology, i.e., for any $(\hat\kappa, \hat\mu) \in \mathcal{D}(F)$ the ball $B_\varepsilon(\hat\kappa, \hat\mu) := \{(\kappa, \mu) : \|\kappa - \hat\kappa\|^2_{H^1} + \|\mu - \hat\mu\|^2_{L^2} < \varepsilon^2\}$ is not contained in $\mathcal{D}(F)$ for any $\varepsilon > 0$. This slightly complicates the statement and proof of some convergence results stated below, and also requires the use of appropriate iterative methods for the solution of the inverse problem. Choosing a stronger topology in the parameter space, e.g., that of $H^2(\Omega) \times L^2(\Omega)$, would facilitate the proof of Theorems 2.6–2.9, but also complicates the discretization, and affects source conditions and smoothness of (regularized) solutions. Choosing a too weak topology, on the other hand, would not allow to establish the required continuity and differentiability properties of the forward operator. Therefore, the topology of the parameter space has to be chosen with care, and we feel that $H^1(\Omega) \times L^2(\Omega)$ is just the appropriate setting for the problem under investigation; see also Remark 2.10.

In the following, we summarize some basic properties of the forward operator; for details and proofs we refer to [12].

Theorem 2.6. *The operator $F : \mathcal{D}(F) \to \mathbb{C}^{d\times s}$ is strongly continuous, i.e., if $\kappa_n \rightharpoonup \kappa$ (weakly) in $H^1(\Omega)$ and $\mu_n \rightharpoonup \mu$ (weakly) in $L^2(\Omega)$, then $F(\kappa_n, \mu_n) \to F(\kappa, \mu)$ (strongly) with respect to the topology of the measurement space.*

Remark 2.7. This result also holds for infinite dimensional observation spaces [12] equipped with the L^2-topology. The strong continuity implies the continuity and compactness of the forward operator, and guarantees weak lower semicontinuity of the least squares functional $\|F(\kappa, \mu) - M^\delta\|^2$ required in the proof of existence of minimizers for the regularized functional.

2.3 Derivative of the forward map

Under the regularity assumptions (A1)–(A4) on the domain and the coefficients, the forward operator can be shown to be differentiable.

Theorem 2.8. *For $j = 1, \dots, s$, let u_j be the solution of the forward system* (1)–(2) *for the source term q_j, and let $w_j \in H^1(\Omega; \mathbb{C})$ denote the (unique) solution of the* sensitivity equation

$$-\nabla \cdot (\kappa \nabla w_j) + (\mu + \iota k) w_j = \nabla \cdot (\hat\kappa \nabla u_j) - \hat\mu u_j \qquad \text{in } \Omega, \tag{6}$$

$$\kappa \partial_n w_j + \rho w_j = -\hat\kappa \partial_n u_j \qquad \text{on } \partial\Omega. \tag{7}$$

Then the directional derivative $\partial M := \partial_{(\hat\kappa, \hat\mu)} F(\kappa, \mu)$ – which is defined for all $(\hat\kappa, \hat\mu) \in H^1(\Omega) \times L^2(\Omega)$ such that $\kappa + t\hat\kappa > 0$ for $t > 0$ sufficiently small – is given by

$$[\partial M]_{ij} := \int_{\partial\Omega} \eta_i w_j ds, \qquad i = 1, \dots, d,\ j = 1, \dots, s.$$

Proof. The result follows directly from the linearity of the equations (1)–(2). □

Note that, according to Remark 2.5, the set of admissible directions $(\hat\kappa, \hat\mu)$ is only a dense subset of $H^1(\Omega) \times L^2(\Omega)$. By the uniform boundedness principle, we can however extend this definition to the whole space $H^1(\Omega) \times L^2(\Omega)$.

Theorem 2.9. *For any $(\kappa,\mu)\in\mathcal{D}(F)$, the mapping $(\hat\kappa,\hat\mu)\mapsto\partial_{(\hat\kappa,\hat\mu)}F(\kappa,\mu)$ can be extended to a bounded linear operator $F'(\kappa,\mu):H^1(\Omega)\times L^2(\Omega)\to\mathbb{C}^{d\times s}$, the derivative of F at (κ,μ). Moreover, there holds the following Lipschitz condition*

$$\|F'(\tilde\kappa,\tilde\mu)-F'(\kappa,\mu)\|\le L\big(\|\tilde\kappa-\kappa\|_{H^1(\Omega)}+\|\tilde\mu-\mu\|_{L^2(\Omega)}\big)$$

for some $L>0$ and all $(\tilde\kappa,\tilde\mu),(\kappa,\mu)\in\mathcal{D}(F)$.

Remark 2.10. The previous results rely on $W^{1,p}$ a-priori estimates for solutions of (1)–(2), which in turn require the smoothness of $\partial\Omega$, the closeness of $\underline{\kappa}$ to $\overline{\kappa}$, and some regularity of the sources, i.e., $q\in L^2(\partial\Omega)$ and not only $q\in H^{-1/2}(\partial\Omega)$. Moreover, we were not able to show Lipschitz continuity of the derivative operator for a weaker topology in the parameter space; in this sense the choice $H^1(\Omega)\times L^2(\Omega)$ seems to impose minimal regularity.

2.4 The adjoint of the derivative operator

For physical reasons, the parameters κ and μ are real valued. The derivative operator can however be defined formally also for complex variations $\hat\kappa\in H^1(\Omega;\mathbb{C})$, $\hat\mu\in L^2(\Omega;\mathbb{C})$. In the following, we will utilize this natural generalization, which is again denoted by the same symbol F', to define the adjoint of the derivative operator. The bracket $\langle\cdot,\cdot\rangle$ will be used to denote the complex scalar product with conjugation on the second element.

Theorem 2.11. *Let u_j, $j=1,\dots,s$ denote the solution of the forward system (1)–(2) for the source q_j, and let $v_i\in H^1(\Omega;\mathbb{C})$, $i=1,\dots,d$ be the solution of the* adjoint problem

$$-\nabla\cdot(\kappa\nabla v_i)+(\mu+\iota k)v_i=0 \qquad \text{in } \Omega, \tag{8}$$

$$\kappa\partial_n v_i+\rho v_i=\eta_i \qquad \text{on } \partial\Omega. \tag{9}$$

Moreover, define $F'(\kappa,\mu)^*:\mathbb{C}^{d\times s}\to H^1(\Omega;\mathbb{C})\times L^2(\Omega;\mathbb{C})$ by $F'(\kappa,\mu)^*r:=(h_\kappa,h_\mu)$ with

$$\tilde h_\kappa:=-\sum_{i,j}r_{ij}\nabla\overline{u}_j\nabla\overline{v}_i,\qquad h_\mu:=-\sum_{i,j}r_{ij}\overline{u}_j\overline{v}_i,$$

and

$$-\Delta h_\kappa+h_\kappa=\tilde h_\kappa \qquad \text{in } \Omega \tag{10}$$

$$\partial_n h_\kappa=0 \qquad \text{on } \partial\Omega. \tag{11}$$

Then $F'(\kappa,\mu)^*$ is the adjoint operator of $F'(\kappa,\mu)$, i.e., there holds

$$\langle F'(\kappa,\mu)[\hat\kappa,\hat\mu],r\rangle_{\mathbb{C}^{d\times s}}=\langle(\hat\kappa,\hat\mu),F'(\kappa,\mu)^*r\rangle_{H^1(\Omega;\mathbb{C})\times L^2(\Omega;\mathbb{C})} \tag{12}$$

for all $\hat\kappa\in H^1(\Omega;\mathbb{C})$, $\hat\mu\in L^2(\Omega;\mathbb{C})$, and $r\in\mathbb{C}^{d\times s}$.

Proof. Let w_j denote the solution of the sensitivity problem (6)–(7). Then

$$\begin{aligned}\big[F'(\kappa,\mu)[\hat\kappa,\hat\mu]\big]_{ij}&=\int_{\partial\Omega}\eta_i w_j ds=\int_\Omega\kappa\nabla v_i\nabla w_j+(\mu+\iota k)v_i w_j dx+\int_{\partial\Omega}\rho v_i w_j ds\\&=-\int_\Omega\hat\kappa\nabla v_i\nabla u_j+\hat\mu v_i u_j dx=-\langle\hat\kappa,\nabla\overline{u}_j\nabla\overline{v}_i\rangle_{L^2(\Omega;\mathbb{C})}-\langle\hat\mu,\overline{u}_j\overline{v}_i\rangle_{L^2(\Omega;\mathbb{C})}.\end{aligned}$$

The result then follows by multiplying with $\overline{r}_{ij}$, summing over i,j and using (10)–(11) to express the first term on the right hand side of the last equation as scalar product over $H^1(\Omega;\mathbb{C})$. □

Remark 2.12. Instead of the adjoint problems (8) one could as well use the solutions z_i of the formal adjoint equation

$$-\nabla \cdot (\kappa \nabla z_i) + (\mu - \iota k) z_i = 0,$$

with appropriate boundary conditions, and then exchange v_i by $\overline{z}_i$ in the definition of $F'(\kappa, \mu)^*$. From a computational point of view, we prefer (8), since it allows to re-use the system matrix (and its factorization) of the forward problem also for the solution of the adjoint problems; see Section 6. Alternatively, one might also define z_j as solutions of

$$\begin{aligned} -\nabla \cdot (\kappa \nabla z_j) + (\mu + \iota k) z_j &= 0 && \text{in } \Omega, \\ \kappa \partial_n z_j + \rho z_j &= \sum_i \overline{r}_{ij} \eta_i && \text{on } \partial\Omega, \end{aligned}$$

and replace the terms $\sum_{i,j} r_{ij} \nabla \overline{u}_j \nabla \overline{v}_i$ in the definition of $\tilde{h}_\kappa$ by the simple sums $\sum_j \nabla \overline{u}_j \nabla \overline{z}_j$. In this way, only s instead of d adjoint problems are required, i.e., the number of adjoint problems becomes independent of the number of detectors. This fact will be used below to obtain algorithms of optimal complexity.

3 The inverse problem and its regularization

The inverse problem of diffuse optical tomography is the following: Given measurements M^δ, find optical parameters $(\kappa, \mu) \in \mathcal{D}(F)$ such that

$$F(\kappa, \mu) = M^\delta, \qquad \|M - M^\delta\| \le \delta. \tag{13}$$

Here, M^δ denotes the perturbed measurements of the data $M = F(\kappa^\dagger, \mu^\dagger)$ corresponding to the true optical parameters $\kappa^\dagger, \mu^\dagger$, and δ is an upper bound for the data noise.

Remark 3.1. The inverse problem is ill-posed in the sense of Hadamard: for finitely many data, the problem is underdetermined; while for infinitely many data, F is compact with respect to any reasonable topology; see Remark 2.4 and 2.7. In the latter case, the solution of the inverse problem is unstable with respect to data noise. As a consequence, also the problems with finite dimensional data become increasingly unstable when the dimension of the measurement space is increased, and some regularization must be used to enforce stability of the reconstructions.

For the regularized solution of the inverse problem, we first consider Tikhonov regularization, for which convergence and convergence rates can be proven under relatively mild assumptions (that can actually be verified for our problem). Since our problems are nonlinear, some iterative algorithms have to be used in a numerical realization of Tikhonov's method. We consider two projected iterative methods for this purpose, which can be viewed as fixed point iterations for the first order optimality conditions for the Tikhonov functional. Although the analysis of these algorithms is formally independent of Tikhonov regularization, we think that the connection to Tikhonov's method provides some theoretical backup also for situations, which are not fully covered by the theory. Some results on the convergence of iterative methods towards minimizers of the Tikhonov functional can be found in [26, 27].

3.1 Tikhonov regularization

For $\alpha > 0$, a regularized solution $(\kappa_\alpha^\delta, \mu_\alpha^\delta)$ can be defined as minimizer of the Tikhonov functional [32, 13]

$$J_\alpha(\kappa,\mu) := \tfrac{1}{2}\|M^\delta - F(\kappa,\mu)\|^2_{\mathbb{C}^{d\times s}} + \tfrac{\alpha}{2}\left(\|\kappa-\kappa_0\|^2_{H^1(\Omega)} + \|\mu-\mu_0\|^2_{L^2(\Omega)}\right)$$

for $(\kappa,\mu) \in \mathcal{D}(F)$. Here, κ_0 and μ_0 denote some a-priori guesses for the unknown functions, and the regularization parameter α is assumed to be strictly positive. Under the conditions of Assumption 2.1, this method is well-defined.

Theorem 3.2. *Let (A1)–(A4) hold. Then for any $\alpha > 0$ a minimizer $(\kappa_\alpha^\delta, \mu_\alpha^\delta) \in \mathcal{D}(F)$ of J_α exists.*

Proof. The domain $\mathcal{D}(F)$ is closed and convex, and by Theorem 2.6, F is weakly-continuous [12]. The existence of a minimizer now follows, since J_α is bounded from below (by zero), and coercive (by $\alpha > 0$). □

Under additional conditions on the solution, also quantitative convergence results can be derived.

Theorem 3.3. *Let $(\kappa^\dagger, \mu^\dagger) \in \mathcal{D}(F)$ denote a (κ_0,μ_0)-minimum norm solution for* (13)*, let Assumption 2.1 hold, and assume that an approximate source condition of the form*

$$(\kappa^\dagger,\mu^\dagger) = (\kappa_0,\mu_0) + \mathrm{Re}\big(F'(\kappa^\dagger,\mu^\dagger)^* w\big) + (r_\kappa, r_\mu) \tag{14}$$

holds for some $w \in \mathbb{C}^{d\times s}$ such that $L\|w\| < 1$, and some $(r_\kappa, r_\mu) \in H^1(\Omega)\times L^2(\Omega)$ with $\|r_\kappa\|^2_{H^1(\Omega)} + \|r_\mu\|^2_{L^2(\Omega)} \le \Delta^2$. Then for $\alpha \approx \delta + \Delta^2$ the following estimate

$$\|\kappa_\alpha^\delta - \kappa^\dagger\|_{H^1(\Omega)} + \|\mu_\alpha^\delta - \mu^\dagger\|_{L^2(\Omega)} = O(\sqrt{\delta+\Delta^2})$$

holds for any minimizer $(\kappa_\alpha^\delta, \mu_\alpha^\delta) \in \mathcal{D}(F)$ of the regularized Tikhonov functional.

Proof. For $\Delta = 0$, the result follows similarly to the proof of [13, Thm 10.4]; see also [12] for details concerning the present situation ($\mathcal{D}(F)$ has no interior points). Tikhonov regularization with approximate source conditions ($\Delta > 0$) has been considered in [22, 20]. □

Remark 3.4. Since we used a complex extension of the operator F' to define the adjoint, the elements $F'(\kappa,\mu)^* w$ are in general complex valued, and one has to take the real part in the definition of the source condition. Below we will see that the real part appears naturally in the derivation of the gradient of the Tikhonov functional.

Remark 3.5. We utilize approximate source conditions for the statement of the convergence results for the following reason: Since our problem is severly ill-posed, a strong source condition with $\Delta = 0$ would imply that the true solution has to be very smooth, such that the result cannot be applied in a realistic setting. On the other hand, convergence rates for weaker source conditions of Hölder-type can only be proven under additional nonlinearity conditions on the operator [20]; these can however not be verified for the problem under investigation. The use of the approximate source conditions allows us to derive similar convergence results also for iterative methods and to transfer source conditions and convergence results from the continuous to the discrete level, cf Remark 5.7.

Remark 3.6. Also the standard convergence result [13, Thm 10.3] holds: if M^{δ_n} denotes a sequence of perturbed data satisfying $\|M - M^{\delta_n}\| \leq \delta_n$ and $\alpha_n \sim \delta_n + \Delta_n^2 \to 0$, then at least a subsequence of minimizers $(\kappa_{\alpha_n}^{\delta_n}, \mu_{\alpha_n}^{\delta_n})$ converges to a solution of $F(\kappa, \mu) = M$, and any limit of such a subsequence is a solution of $F(\kappa, \mu) = M$.

Remark 3.7. Under additional assumptions on the parameters, and in case of full observation of the Robin-to-Neumann map, the inverse problem of diffuse optical tomography can be shown to have a unique solution, cf. e.g. [23, 15]; for the case of $\omega = 0$ see also [17], and [2] for an example of non-uniqueness. Assuming uniqueness, the convergence statement of Remark 3.6 simplifies, i.e., the minimizers $(\kappa_\alpha^\delta, \mu_\alpha^\delta)$ converge to the solution (κ, μ) as $\delta + \Delta^2 \to 0$.

3.2 Derivative and gradient of the Tikhonov functional

In the following section, we investigate iterative methods which aim at minimizing or at least finding stationary points of the Tikhonov functional. Since the formulation of these methods relies on the gradient of the Tikhonov function, let us shortly present explicit formulas for the derivative and the gradient.

Theorem 3.8. *The directional derivative $J'_\alpha(\kappa, \mu)[\hat\kappa, \hat\mu]$ of the Tikhonov functional at (κ, μ) in direction $(\hat\kappa, \hat\mu)$ is given by*

$$\begin{aligned} J'_\alpha(\kappa, \mu)[\hat\kappa, \hat\mu] = \operatorname{Re}\langle F'(\kappa, \mu)[\hat\kappa, \hat\mu], F(\kappa, \mu) - M^\delta\rangle_{\mathbb{C}^{d\times s}} \\ + \alpha\big(\langle \kappa - \kappa_0, \hat\kappa\rangle_{H^1(\Omega;\mathbb{R})} + \langle \mu - \mu_0, \hat\mu\rangle_{L^2(\Omega,\mathbb{R})}\big). \end{aligned}$$

Proof. Let us consider the first term in the Tikhonov functional which results from setting $\alpha = 0$. Then by standard arguments, we obtain

$$\begin{aligned} J'_0(\kappa, \mu)[\hat\kappa, \hat\mu] &= \tfrac{1}{2}\langle F'(\kappa, \mu)[\hat\kappa, \hat\mu], F(\kappa, \mu) - M^\delta\rangle_{\mathbb{C}^{d\times s}} + \tfrac{1}{2}\langle F(\kappa, \mu) - M^\delta, F'(\kappa, \mu)[\hat\kappa, \hat\mu]\rangle_{\mathbb{C}^{d\times s}} \\ &= \operatorname{Re}\langle F'(\kappa, \mu)[\hat\kappa, \hat\mu], F(\kappa, \mu) - M^\delta\rangle_{\mathbb{C}^{d\times s}}, \end{aligned}$$

The formulas for the remaining derivatives of the penalty terms follow similarly. □

The directional derivative is only defined for directions $(\hat\kappa, \hat\mu)$ such that $\kappa + t\hat\kappa$ is positive for $0 < t$ sufficiently small. The definition of the directional derivatives can however be extended by continuity to the whole parameter space.

Corollary 3.9. *The directional derivatives of the Tikhonov functional at $(\kappa, \mu) \in \mathcal{D}(F)$ define a densly defined linear operator that can be extended by continuity to a bounded linear functional $J'_\alpha(\kappa, \mu) : H^1(\Omega) \times L^2(\Omega) \to \mathbb{R}$, the derivative of the Tikhonov functional.*

The gradient of the Tikhonov functional is now defined via the Riesz isomorphism, i.e.,

$$\langle \nabla J_\alpha(\kappa, \mu), (\hat\kappa, \hat\mu)\rangle_{H^1(\Omega)\times L^2(\Omega)} = J'_\alpha(\kappa, \mu)[\hat\kappa, \hat\mu]$$

for all $\hat\kappa \in H^1(\Omega)$ and $\hat\mu \in L^2(\Omega)$.

Theorem 3.10. *Let h_k, h_μ be defined as in Theorem 2.11. Then the gradient of the Tikhonov functional is given by $\nabla J_\alpha(\kappa, \mu) := (g_\kappa, g_\mu)$ with $g_\kappa := \operatorname{Re}(h_\kappa) + \alpha(\kappa - \kappa_0)$ and $g_\mu := \operatorname{Re}(h_\mu) + \alpha(\mu - \mu_0)$.*

Proof. Using the identity (12), we have

$$\langle F'(\kappa, \mu)[\hat\kappa, \hat\mu], F(\kappa, \mu) - M^\delta\rangle_{\mathbb{C}^{d\times s}} = \langle (\hat\kappa, \hat\mu), F'(\kappa, \mu)^*(F(\kappa, \mu) - M^\delta)\rangle_{H^1(\Omega;\mathbb{C})\times L^2(\Omega;\mathbb{C})}.$$

The representation of the gradient follows from Theorem 3.8 together with the definition of the adjoint derivative given in Theorem 2.11. □

4 Iterative realization

In the following, we discuss iterative algorithms, which aim at solving the first order optimality conditions for minimizing the Tikhonov functional over the closed and convex set $\mathcal{D}(F)$. This means to find $(\kappa, \mu) \in \mathcal{D}(F)$ such that [21, Cor 1.2]

$$(\kappa, \mu) = \mathcal{P}_{\mathcal{D}(F)}\Big((\kappa, \mu) - \lambda \nabla J_\alpha(\kappa, \mu)\Big) \qquad \text{for any } \lambda > 0. \tag{15}$$

Here, $\mathcal{P}_{\mathcal{D}(F)} : H^1(\Omega) \times L^2(\Omega) \to \mathcal{D}(F)$ denotes the metric projector onto the convex set $\mathcal{D}(F)$.

Remark 4.1. The projector $\mathcal{P}_{\mathcal{D}(F)}$ is given by $\mathcal{P}_{\mathcal{D}(F)}(\kappa, \mu) = (\mathcal{P}_\kappa(\kappa), \mathcal{P}_\mu(\mu))$, where $\mathcal{P}_\kappa$, $\mathcal{P}_\mu$ are defined via the minimization problems

$$\begin{aligned} \|\mathcal{P}_\kappa(\kappa) - \kappa\|_{H^1(\Omega)} &= \inf_{(\tilde\kappa, \mu_0) \in \mathcal{D}(F)} \|\tilde\kappa - \kappa\|_{H^1(\Omega)} \\ \|\mathcal{P}_\mu(\mu) - \mu\|_{L^2(\Omega)} &= \inf_{(\kappa_0, \tilde\mu) \in \mathcal{D}(F)} \|\tilde\mu - \mu\|_{L^2(\Omega)}. \end{aligned}$$

On the continuous level, a solution of the second minimization problem can be given explicitly, namely $\mathcal{P}_\mu(\mu) = \min(\overline{\mu}, \max(\underline{\mu}, \mu))$. As we will see below, such a representation does in general not hold after discretization.

For approximating solutions of (15) by some iterative process, we consider projected iterations of the form

$$(\kappa_{n+1}, \mu_{n+1}) = \mathcal{P}_{\mathcal{D}(F)}\Big((\kappa_n, \mu_n) - \lambda_n C_n \nabla J_{\alpha_n}(\kappa_n, \mu_n)\Big), \qquad n \geq 0.$$

The choice $C_n = I$ yields the projected gradient method [8, 21], while the alternative choice $C_n = (F'(\kappa_n, \mu_n)^* F'(\kappa_n, \mu_n) + \alpha_n I)^{-1}$ amounts to a projected version of the Gauß-Newton method [4, 25].

Remark 4.2. Since in two or three space dimensions the spaces $H^1(\Omega)$ and $L^2(\Omega)$ are not continuously embedded in $L^\infty(\Omega)$, the elements $(\kappa_\lambda, \mu_\lambda) := (\kappa, \mu) + \lambda \nabla J_\alpha(\kappa, \mu)$ will in general not lie in $\mathcal{D}(F)$ for any $\lambda \neq 0$. Thus, the projection step in the iteration is required to guarantee that $\kappa_{n+1}, \mu_{n+1} \in \mathcal{D}(F)$ and $F(\kappa_{n+1}, \mu_{n+1})$ is well-defined; cf. [25] for a related problem and numerical tests.

In the following, we summarize some basic results concerning the convergence analysis of the projected gradient and the projected Gauß-Newton method. A full analysis of these methods is not in the scope of this paper, but we provide references for their proofs.

4.1 Projected gradient method

As a first method, let us consider the projected gradient method

$$(\kappa_{n+1}, \mu_{n+1}) = \mathcal{P}_{\mathcal{D}(F)}\Big((\kappa_n, \mu_n) - \lambda_n \nabla J_{\alpha_n}(\kappa_n, \mu_n)\Big), \qquad n \geq 0. \tag{16}$$

The basic convergence result for this iteration can be summarized as follows.

Theorem 4.3. *Let the sequence (κ_n, μ_n) be defined by (16), and assume that $\nabla J_{\alpha_n}(\kappa_n, \mu_n) \neq 0$. Then there exists an $\lambda_n > 0$ such that $J_{\alpha_n}(\kappa_{n+1}, \mu_{n+1}) < J_{\alpha_n}(\kappa_n, \mu_n)$.*

Proof. For a proof in the case that F is differentiable in a neighbourhood of $\mathcal{D}(F)$ see [8, 21]. The generalization to the present situation ($\mathcal{D}(F)$ has no interior points) is possible without changes in the proofs. □

Having established decrease in the Tikhonov functional, the convergence of subsequences to stationary points of the Tikhonov functional follows with the usual "sufficient decay" argument.

Theorem 4.4. *Let $\alpha_n \to \alpha$ from above, and λ_n be chosen by the Armijo rule [8, 21]. Then any weak cluster point of the sequence (κ_n, μ_n) satisfies* (15).

Proof. The proof for the case $\alpha_n = \alpha$ is classical, and can be found, e.g., in [8, 21]. The statement for the case $\alpha_n \to \alpha$ follows similarly. □

Remark 4.5. The convergence statement of Theorem 4.4 may seem unsatisfactory. However, since the Tikhonov functional is not convex, convergence of the iterates to local or global minimizers can in general not be expected without further assumptions, e.g. on the nonlinearity of the operator F or the "quality" of the solution; see also the convergence statement given in Remark 3.6.

The following result states that, under the conditions of Theorem 3.10, the iterates of the projected gradient method provide the same convergence rates that can be obtained by minimizing the Tikhonov functional. For this statement, we use the fact that the projected gradient method amounts to a projected version of a *modified Landweber iteration* proposed in [29].

Theorem 4.6. *Let (κ_n, μ_n) be defined by* (16) *with $\alpha_{n+1} \geq \alpha_n(1 - \alpha_n)$ and α_0 sufficiently small. Moreover, let $(\kappa^\dagger, \mu^\dagger)$ satisfy the source condition* (14)*, and assume that (κ_0, μ_0) is sufficiently close to $(\kappa^\dagger, \mu^\dagger)$. If the iteration is stopped as soon as $\alpha_N \leq (\delta + \Delta^2)$, then*

$$\|\kappa_N - \kappa^\dagger\|_{H^1(\Omega)} + \|\mu_N - \mu^\dagger\|_{L^2(\Omega)} = O(\sqrt{\delta + \Delta^2}).$$

Proof. The result follows from the non-expansiveness of the projection operator with similar arguments as in [29, Sec. 3]. A detailed proof can be found in [11]. A particular choice of parameters satisfying the assumption is given by $\alpha_n = (n+n_0)^{-1}$ with n_0 sufficiently large. □

Remark 4.7. Under the source condition (14), the iteration (16) can be shown to converge to a local minimum of the Tikhonov functional J_α, if $\alpha_n \to \alpha > 0$ appropriately. Moreover, this local minimizer of the Tikhonov functional satisfies the estimates of Theorem 3.3, and therefore, the iteration (16) can be considered to be a convergent minimization algorithm for the Tikhonov functional J_α. For convergence results based on similar arguments, see [26, 27].

4.2 Projected Gauß-Newton method

Gradient methods are known to converge slowly if the Hessian of the functional to be minimized is ill-conditioned. As an alternative, we therefore consider projected Gauß-Newton methods of the form

$$(\kappa_{n+1}, \mu_{n+1}) = \mathcal{P}_{\mathcal{D}(F)}\Big((\kappa_n, \mu_n) - \big([\mathrm{Re}(F'(\kappa,\mu)^* F'(\kappa,\mu) + \alpha_n I)]\big)^{-1} \nabla J_{\alpha_n}(\kappa_n, \mu_n)\Big). \quad (17)$$

Note that the operator $\mathrm{Re}(F'(\kappa,\mu)^* F'(\kappa,\mu) + \alpha I)$ is just the Gauß-Newton approximation for the Hessian of the Tikhonov functional.

Under the source condition of Theorem 3.10, a convergence rates result similar to that of Theorem 3.10 and 4.6 holds.

Theorem 4.8. *Let $0 < \alpha_{n+1} \leq \alpha_n$ be such such that $\inf_{n\geq 0} \frac{\alpha_{n+1}}{\alpha_n} =: r > 0$, and let the source condition of Theorem 3.10 hold. Moreover, assume that α_0 is sufficiently large, and (κ_0, μ_0) is sufficiently close to $(\kappa^\dagger, \mu^\dagger)$. Then the iterates of the projected Gauß-Newton iteration satisfy*

$$\|\kappa_n - \kappa^\dagger\|_{H^1(\Omega)} + \|\mu_n - \mu^\dagger\|_{L^2(\Omega)} \leq C\sqrt{\alpha_n},$$

for all n such that $\alpha_n \geq \hat{C}(\delta + \Delta^2)$ with $\hat{C}$ independent of δ and Δ. If the iteration is stopped as soon as $\alpha_N \leq \hat{C}(\delta + \Delta^2)$, then the estimate

$$\|\kappa_N - \kappa^\dagger\|_{H^1(\Omega)} + \|\mu_N - \mu^\dagger\|_{L^2(\Omega)} = O(\sqrt{\delta + \Delta^2})$$

holds.

Proof. The proof follows with slight modifications (since F is not Fréchet differentiable here) from the results in [4, Sec. 4]; see also [25] for the case $\Delta = 0$. □

Remark 4.9. Convergence results for the Gauß-Newton method without projection and $\Delta = 0$ are classical, and can be found, e.g., in [3, 24]. The generalization to the projected Gauß-Newton method requires the non-expansivness of the projection operator $\mathcal{P}_{\mathcal{D}(F)}$; in particular, a simple pointwise projection of the parameter κ is not suitable.

Remark 4.10. Similarly as for the projected gradient method, one can show that under the source condition (14) and for $\alpha_n \to \alpha > 0$, the iterates generated by the projected Gauß-Newton method converge to a local minimizer $(\kappa_\alpha^\delta, \mu_\alpha^\delta)$ of the Tikhonov functional. Since a choice $\alpha_n \to \alpha$ is allowed in Theorem 4.8, this local minimizer of the Tikhonov functional satisfies the estimate of the theorem and provides optimal convergence rates.

4.3 Summary of convergence results

The results of this section can be summarized as follows: Under the approximate source condition (14), convergence of the projected Gauß-Newton method and the projected gradient method can be shown, and order optimal convergence rate estimates hold. Without source condition, at least convergence of subsequences to critical points of the Tikhonov functional can be shown for the projected gradient method.

Remark 4.11. Under the source condition (14) and for $\alpha_n \to \alpha > 0$ appropriately, the Tikhonov functional can be shown to be locally convex around the iterates (κ_n, μ_n). Moreover, the iterations can be shown to converge locally to minimizers of the Tikhonov functional; cf [26, 27].

Remark 4.12. Some of the presented convergence results require the source condition already to ensure convergence of the methods. To the best of our knowledge, convergence of the projected gradient or the Gauß-Newton iteration with weaker or even without any source condition can be proven only under additional assumptions on the nonlinearity of the operator F, cf. [16, 24]. Such conditions could however not be verified for the problem considered here.

5 Discretization by finite elements

Let us now take a further step towards the computer solution of our inverse problem. For the discretization of the governing differential equations, we consider piecewise linear, continuous finite elements [9]. The same (but real valued) finite dimensional spaces are used for

discretization of the parameters κ and μ. For ease of notation, the discrete functions will be identified with their coordinate vectors throughout the text, and we use the symbol F_h to denote the discretized forward operator. We present in some detail the discretization of the forward operator, the derivative and the adjoint derivative operator, which are required for fully discrete solution algorithms. Additionally, we illustrate that a careful discretization allows to carry over important properties of these operators from the continuous to the discrete level, which will allow us to make statements about the performance of the fully discrete algorithms.

5.1 Discrete forward problem

After discretization, the boundary value problems (1)–(2) have the form of a linear system

$$[K(\kappa) + C(\mu + \iota k) + R]U = RQ. \tag{18}$$

The entries of the system matrices are defined by

$$K(\kappa)_{ij} = \int_\Omega \kappa \nabla\psi_i \nabla\psi_j dx, \qquad C(\mu)_{ij} = \int_\Omega \mu\psi_i\psi_j dx, \qquad R_{ij} = \int_{\partial\Omega} \rho\psi_i\psi_j ds.$$

Here and below, ψ_i denotes the finite element basis function corresponding to the ith vertex in the mesh, wich is assumed to be simplicial and shape regular [9]. The jth column of the source matrix Q stores the coordinate vector corresponding to the source q_j, and the jth column of the complex solution matrix U accordingly represents (a discretization of) the solutions of (1)–(2) for the source q_j. The measurement matrix can be computed by

$$M = D(U - Q), \qquad \text{where} \qquad D_{ij} = \int_{\partial\Omega} \eta_i\psi_j ds,$$

and the discretization of the forward operator is finally given by $F_h(\kappa, \mu) := M$.

5.2 Discrete derivative and adjoint operators

The derivative of the forward operator can be computed by $F_h'(\kappa, \mu)[\hat\kappa, \hat\mu] := DW$, where W is the complex solution matrix of the linear system

$$[K(\kappa) + C(\mu + \iota k) + R]W = -[K(\hat\kappa) + C(\hat\mu)]U. \tag{19}$$

This is the finite element discretization of the sensitivity problems (6)–(7), but the system for the derivative could be derived analogously also on the matrix level. As a consequence, we obtain exactness of the discrete derivative.

Theorem 5.1. *Let U and W denote the solutions of the discrete forward problem* (18) *and the discrete sensitivity system* (19)*. Then $F_h'(\kappa, \mu)[\hat\kappa, \hat\mu] := DW$ is the derivative of the discrete forward operator F_h at (κ, μ).*

Proof. The result follows directly from the linear dependence of the system matrices on the parameters, and the definition of U_i and W_i. □

Let us consider the adjoint operator $F_h'(\kappa, \mu)^*$ next. Similar to the continuous level, we define V as the solution of the adjoint system

$$[K(\kappa) + C(\mu + \iota k) + R]V = D^T. \tag{20}$$

The adjoint is then given by $F_h'(\kappa,\mu)^* r := (h_\kappa, h_\mu)$, with h_κ and h_μ defined

$$[K(1)+C(1)]h_\kappa = \Big(\sum_{i=1}^{d}\sum_{j=1}^{s} r_{ij} A(\overline{U}_j)\overline{V}_i\Big) \quad \text{and} \quad C(1)h_\mu = \Big(\sum_{i=1}^{d}\sum_{j=1}^{s} B(\overline{U}_j)\overline{V}_i\Big). \tag{21}$$

The matrices $A(u)$, $B(u)$ are given by

$$A(u)_{ij} := \int_\Omega \psi_i \nabla u \nabla \psi_j dx, \qquad \text{and} \qquad B(u)_{ij} := \int_\Omega \psi_i u \psi_j dx.$$

As above, V_i, U_j denote the ith respectively jth column of the solution matrices V and U.

This discretization of the adjoint operator is exact on the discrete level, which implies that important properties derived on the continuous level also hold on the discrete level.

Theorem 5.2. *Let W and V be the solutions of the discrete sensitivity and adjoint systems* (19) *and* (20)*, respectively. Then the discrete operators $F_h'(\kappa,\mu)[\hat\kappa,\hat\mu] := DW$ and $F_h'(\kappa,\mu)^* r := (h_\kappa, h_\mu)$ with h_κ, h_μ as above, are adjoint with respect to each other, i.e., there holds*

$$\langle F'(\kappa,\mu)[\hat\kappa,\hat\mu], r\rangle_{\mathbb{C}^{d\times s}} = \langle (\hat\kappa,\hat\mu), F'(\kappa,\mu)^* r\rangle_{H^1(\Omega;\mathbb{C})\times L^2(\Omega;\mathbb{C})}$$

for all $\hat\kappa, \hat\mu$ in the space of finite element functions and $r \in \mathbb{C}^{d\times s}$.

Proof. Let d_i denote the ith row of the measurement operator D, and W_j be the jth column of the solution to the sensitivity problem. Then by definition of V_i and W_j, we have

$$d_i W_j \overset{(20)}{=} V_i^T [K(\kappa) + C(\mu + ik) + R]^{-1} W_j \overset{(19)}{=} -V_i^T K(\hat\kappa) U_j - V_i^T C(\hat\mu) U_j = (*).$$

Using the definition of the matrices A and B, we further obtain

$$(*) = -V_i^T A(U_j)\hat\kappa - V_i^T B(U_j)\hat\mu.$$

The result then follows by multiplying with $\overline{r}_{ij}$, using the definition of the scalar products,

$$\langle \hat\kappa, h_\kappa\rangle_{H^1(\Omega;\mathbb{C})} := \overline{h}_\kappa^T [K(1)+C(1)]\hat\kappa \quad \text{and} \quad \langle \hat\mu, h_\mu\rangle_{L^2(\Omega;\mathbb{C})} := \overline{h}_\mu^T C(1)\hat\mu,$$

and summing over i and j. □

As a consequence of the exactness of the discrete adjoints, we obtain that the product $F_h'(\kappa,\mu)^* F_h'(\kappa,\mu)$ is a selfadjoint operator also on the discrete level.

Theorem 5.3. *Let $F_h'(\kappa,\mu)$, $F_h'(\kappa,\mu)^*$ denote the discrete derivative operator and its adjoint. Then the matrix $F_h'(\kappa,\mu)^* F_h'(\kappa,\mu)$ is positive semi definit (hermitian) with respect to the (complex) scalar product of the parameter space induced by $[K(1)+C(1)]$ and $C(1)$. Additionally, the matrix $\mathrm{Re}\Big(F_h'(\kappa,\mu)^* F_h'(\kappa,\mu)\Big)$ is real, symmetric and positive semi-definite with respect to this scalar product.*

Proof. Let $\hat\kappa$, $\hat\mu$ denote the vectors of degrees of freedom for the corresponding finite element functions. Then by the previous results, we have

$$\begin{aligned}&\langle (\hat\kappa,\hat\mu), F_h(\kappa,\mu)^* F_h'(\kappa,\mu)[\hat\kappa,\hat\mu]\rangle_{H^1(\Omega;\mathbb{C})\times L^2(\Omega;\mathbb{C})}\\ &\quad = \langle F_h'(\kappa,\mu)[\hat\kappa,\hat\mu], F_h'(\kappa,\mu)[\hat\kappa,\hat\mu]\rangle_{\mathbb{C}^{d\times s}} = \|F_h'(\kappa,\mu)[\hat\kappa,\hat\mu]\|^2_{\mathbb{C}^{d\times s}} \geq 0,\end{aligned}$$

which shows that positive semi-definiteness; the symmetry is obvious. □

5.3 Discretization of the Tikhonov functional

The discrete Tikhonov functional is now defined by

$$J_\alpha(\kappa,\mu) = \tfrac{1}{2}\|F_h(\kappa,\mu) - M^\delta\|^2_{\mathbb{C}^{d\times s}} + \tfrac{\alpha}{2}\big(\|\kappa-\kappa_0\|^2_{K(1)+C(1)} + \|\mu-\mu_0\|^2_{C(1)}\big),$$

where $\|u\|^2_K := u^T K u$ denotes the norm induced by the matrix K (for complex u the norm would be $(\overline{u}^T K u)^{1/2}$).

The gradient of the discrete Tikhonov functional, which is required for the iterative methods (16) and (17), can be computed like on the continuous level, i.e., $\nabla J_\alpha(\kappa,\mu) =: (g_\kappa, g_\mu)$ is given by

$$g_\kappa = \mathrm{Re}(h_\kappa) + \alpha(\kappa-\kappa_0), \qquad g_\mu = \mathrm{Re}(h_\mu) + \alpha(\mu-\mu_0),$$

with h_κ, h_μ as in (21).

Remark 5.4. The discrete gradient defined in this way is exactly the gradient of the discrete Tikhonov functional.

Remark 5.5. On the discrete level, the functions $\kappa - \lambda g_\kappa$ will be positive for $\lambda > 0$ small enough, so a descent in the Tikhonov functional can be achieved by descent iterations

$$(\kappa_{n+1},\mu_{n+1}) = (\kappa_n,\mu_n) - \lambda_n C_n \nabla J_\alpha(\kappa_n,\mu_n),$$

without projection, as long as C_n is symmetric positive definite, and $\lambda_n > 0$ is chosen small enough. The length of the admissible stepsize will however depend on the problem dimension (the meshsize), cf. Remark 4.1, and the projected algorithms have to be used in order to obtain mesh independent results. For numerical examples illustrating that the projection is in fact necessary, see [25].

5.4 The projection on the discrete level

In order to apply the iterative methods of the previous sections, one has to realize the projection operation in each iteration. On the discrete level, we have $\mathcal{P}_{\mathcal{D}(F)}(\kappa,\mu) = (\tilde\kappa,\tilde\mu)$ with $\tilde\kappa$ and $\tilde\mu$ defined by the quadratic programs

$$\begin{aligned} \|\tilde\kappa-\kappa\|^2_{K(1)+C(1)} &\to \min \qquad && \text{subject to} \qquad \underline{\kappa} \le \tilde\kappa \le \overline{\kappa}. \\ \|\tilde\mu-\mu\|^2_{C(1)} &\to \min \qquad && \text{subject to} \qquad 0 \le \tilde\mu \le \overline{\mu}. \end{aligned}$$

Remark 5.6. In contrast to the continuous level, the projection of μ cannot be characterized simply by $\tilde\mu = \min(\overline{\mu}, \max(\underline{\mu}, \mu))$ on the discrete level, since the mass matrix $C(1)$ is not diagonal. One could however preserve the "simple" projection , if the scalar product for the second parameter space is defined by a diagonal matrix, e.g., by the lumped mass matrix.

5.5 Remarks on the convergence of the fully discrete algorithms

The results derived in Section 3 on the continuous level also cover the discrete case, i.e., under the conditions of the convergence theorems, also the discrete analogues of the projected gradient and Gauß-Newton method converge. This can be seen as follows:

Remark 5.7. If an approximate source condition is satisfied on the continuous level, then also a discrete analogue holds. E.g., let $A : X \to Y$ be a linear, compact operator, and assume that $A_h : X_h \to Y_h$ is a discrete approximation (extended by 0 to $X_h^\perp$). Since A is compact, it can be approximated uniformly, so we can assume that $\|A - A_h\| \le h$. If $z = A^* w$ for some $w \in Y$, then

$$z = A_h^* w + (A^* - A_h^*)w,$$

and consequently z satisfies an approximate source condition with $\Delta = h\|w\|$. The projected solution $z_h = \mathcal{P}_{X_h} z$ satisfies a similar condition.

Remark 5.8. Assume that $F : D(F) \subset X \to Y$ is a continuous and compact operator with Lipschitz continuous derivative, and let F_h denote a discrete approximation to F such that $\|F(x) - F_h(x)\| \le h$ for all $x \in D(F)$. If $F(x) = y$ and $\|y^\delta - y\| \le \delta$, then by the triangle inequality $\|y^\delta - F_h(x)\| \le \delta + h$, i.e., the data y^δ are "admissible" for the discrete problem, but with noise level $\delta + h$. Some comments on how to incorporate the discretization error in the numerical algorithms will be made in Section 8.

Taking into account the discretization errors, we thus obtain algorithms that converge, independent of the discretization level, i.e., the number of iterations and the quality of reconstructions, do not deteriorate as the meshsize tends to zero. This *mesh independence* is an important property of the numerical algorithms used in actual computations.

6 Complexity Estimates

In order to assess the efficiency of our numerical methods, let us derive some estimates for their computational complexity. The workload will naturally depend on the number of vertices N in the finite element mesh, on the number of sources s and detectors d, but also on the required accuracy, i.e., the noise level δ. In the following, we take a closer look on the individual steps required for the numerical solution.

(a) Assembling of the system matrix $S_n = [K(\kappa_n) + C(\mu_n + \iota k) + R]$**:** The assembling of any individual part of the system matrix has optimal, linear complexity and memory requirement $O(N)$. Note that the parts $K(\kappa_n)$, $C(\mu_n)$ which depend on the iterates are real valued, which can be used to accelerate the assembling process.

(b) Solution of the linear systems: The linear systems (18), (19), and (20) are complex symmetric, and can be solved with accuracy ε in $O(\log \varepsilon^{-1})$ iterations, if appropriate (e.g. multigrid preconditioned) solvers are used. (Since ε is bounded by the machine accuracy, one can assume that $O(\log \varepsilon^{-1}) = O(1)$). The total complexity for solving the forward system (18), or the sensitivity system (19) is therefore bounded by $O(sN)$ operations. Similarly, the solution of the adjoint equations (20) has complexity $O(dN)$.

Remark 6.1. If N is not too large, a sparse direct factorization of the system matrices $S_n = [K(\kappa_n) + C(\mu_n + \iota k) + R]$ can be used as alternative to the iterative solvers. The complexity for such a factorization is $O(N^{3/2})$ in two (or $O(N^2)$ in three) dimensions. Computing a solution by back substitution then requires $O(N \log N)$ (or $O(N^{3/2})$) operations. Note, that by our definition of the adjoint problems, the factorization of the system matrix of the forward problem can be used for solving the sensitivity and the adjoint problems as well.

Remark 6.2. The solutions U and V of the forward and adjoint systems have to be computed only once per Newton-iteration. They can then be used for the evaluation of the derivative and the adjoint derivative operator, see steps (d) and (e).

(c) Evaluation of the forward operator $F_h(\kappa,\mu)$: Since the detector matrix D is sparse (all detectors together cover at most the boundary of the domain), the evaluation of the forward operator requires at most $O(sN)$ operations, once the solution matrix U of the forward system (18) has been computed; see step (b).

(d) Application of the derivative operator $F_h'(\kappa,\mu)[\hat\kappa,\hat\mu]$: The jth column of the derivative is given by

$$\begin{aligned}\big(F_h'(\kappa,\mu)[\hat\kappa,\hat\mu]\big)_{\cdot,j} &= D[K(\kappa)+C(\mu+\iota k)+R]^{-1}[-K(\hat\kappa)U_j - C(\hat\mu)U_j]\\ &= V^T(-A^T(U_j)\hat\kappa - B^T(U_j)\hat\mu).\end{aligned}$$

The formula in the first line requires the assembly of the matrices $K(\hat\kappa)$ and $C(\hat\mu)$, the multiplication with s solutions U_j, the solution of s sensitivity problems, and the multiplication with D. The total complexity of this variant is thus $O(sN)$ operations. The formula of the second line avoids the solution of the sensitivity problems, but has a higher complexity $O(sdN)$, since each value $\big(F_h'(\kappa,\mu)[\hat\kappa,\hat\mu]\big)_{i,j}$ requires a scalar product of V_i with $-A^T(U_j)\hat\kappa - B^T(U_j)\hat\mu$.

(e) Application of the adjoint derivative $F_h'(\kappa,\mu)^*r$: Also the adjoint operator can be computed in two different ways. Let us consider the second component corresponding to the parameter μ. Then with the notation of Theorem 5.2, we have

$$C(1)h_\mu = -\sum_{i=1}^{d}\sum_{j=1}^{s} B(\overline{U}_j)\overline{V}_i r_{ij} = -\sum_{j=1}^{s} B(\overline{U}_j)\overline{Z}_j,$$

where we defined $Z_j := \sum_{i=1}^{d} V_i\overline{r}_{ij}$. The application of the adjoint operator thus requires $O(sdN)$ operations.

Alternatively, the function Z_j can be computed by solving the adjoint system

$$[K(\kappa)+C(\mu+\iota k)+R]Z = D^T\overline{r},$$

which requires the solution of s adjoint problems, and has $O(sN)$ complexity. Since D is sparse, the product $D^T\overline{r}$ requires at most $O(sN)$ operations. Consequently, the total work requirement for the application of the adjoint has $O(sN)$ complexity for this second variant.

(f) Gauß-Newton step: According to the definition of the derivative and its adjoint, we can then write

$$F_h'(\kappa,\mu)^* = G_h^{-1}F_h'(\kappa,\mu)^H, \qquad \text{where} \qquad G_h := \begin{bmatrix} K(1)+C(1) & 0\\ 0 & C(1)\end{bmatrix}$$

denotes the Gram matrix of the $H^1(\Omega)\times L^2(\Omega)$ scalar product of the parameter space. Here, we used that $F_h'(\kappa,\mu)$ and its adjoint can be represented by complex matrices, and denoted by S^H the hermitian of a complex matrix S. Multiplying the Gauß-Newton system with the Gram matrix G_h, we obtain the following equivalent definition of the Gauß-Newton step:

$$\begin{aligned}&[\mathrm{Re}(F_h'(\kappa,\mu)^H F_h'(\kappa,\mu)) + \alpha G_h](d\kappa,d\mu)\\ &\qquad = \mathrm{Re}F_h'(\kappa,\mu)^H(M^\delta - F_h(\kappa,\mu)) + \alpha G_h[(\kappa,\mu)-\kappa_0,\mu_0)].\end{aligned}$$

This is a real symmetric, and positive definite system (cf. Theorem 5.3), which can be solved efficiently with the conjugate gradient method. The number of iterations is bounded by the square root of the condition number. Thus, the number of iterations required to solve the Gauß-Newton system is of complexity $O(1/\sqrt{\alpha})$, and the overall complexity of determining the Gauß-Newton direction is $O(\alpha^{-1/2}sN)$, if the fast evaluation of the derivatives and adjoints is employed.

Remark 6.3. In [31], the GMRES method was employed for the solution of the Gauß-Newton system. As we showed, the system is in fact symmetric and positive definite, and can therefore be solved with the more efficient conjugate gradient method.

(g) Projection step: The projection step requires the solution of two quadratic minimization problems with simple box constraints, and in practice has $O(N)$ complexity.

Overall complexity: If α_n is chosen according to $\alpha_n = r^n\alpha_0$ with some $r < 1$ (see Theorem 4.8), the number of Newton iterations required to reach the stopping criterion depends logarithmically on $\alpha \sim \delta$. As outlined above, the computational effort of a single Gauß-Newton step is $O(\alpha_n^{-1/2}sN)$. Summing over all iterations, and assuming that the iteration is stopped when $\alpha_N \sim \delta$, the total workload can be estimated by $O(\delta^{-1/2}sN)$. The operations counts for the individual steps of the Gauß-Newton algorithm are summarized in Table 1.

procedure	# iterations	# operations
evaluation of the forward operator: $F_h(\kappa,\mu)^*$		$O(sN)$
application of the derivative: $F_h'(\kappa,\mu)[\hat\kappa,\hat\mu]$		$O(sN)$
application of the adjoint derivative: $F_h'(\kappa,\mu)r$		$O(sN)$
CG-Solve for Gauß-Newton system	$O(\alpha_n^{-1/2})$	$O(\alpha_n^{-1/2}sN)$
Gauß-Newton algorithm	$O(\log(\delta))$	$O(\delta^{-1/2}sN)$

Table 1: Computational complexity of the individual steps, iteration number estimates, and overall complexity of the Gauß-Newton algorithm.

Remark 6.4. The computational effort has optimal, linear increase with the meshsize N and the number of sources s. Moreover, the computational effort is independent of the number of detectors d, which is particularly advantageous in practice, where measurements may be recorded at high spatial resolution, e.g. by a CCD camera. One applicataion of the Gauß-Newton matrix has $O(sN)$ complexity, which is substantially lower that the algorithms proposed in [31], which require $O(sdN)$ operations.

Remark 6.5. The number of iterations in the conjugate gradient method for solving the Gauß-Newton system and the number of Gauß-Newton steps are independent of the level of discretization. This mesh-independent behaviour will be demonstrated also in our numerical tests.

7 Implementational details

Before we turn to numerical experiments, let us make some remarks on the proper scaling of the problem and on the choice of regularization parameters. Additionally, we would like to outline a strategy for estimating the discretization error, and give some guidelines for the design of appropriate computational meshes.

7.1 Scaling

The experimental setup, e.g., the position and sizes of detectors and sources, and the geometry imply a certain scaling of the measurements, which is independent of the solution (κ, μ) of the inverse problem. For example, the light intensity is much higher in the neighbourhood of a source than further away. Using a uniform weight for all measurments would thus discriminate measurements corresponding to source-detector pairs which are separated by some distance. Exactly these configurations, however, contain most of the information about the interior of the domain.

In order to incorporate geometric information about the measurement setup, we thus slightly modify the definition of the forward operator by weighting the individual measurements, i.e., we consider the weighted forward operator defined by

$$[F_w(\kappa, \mu)]_{ij} := W_{ij} D_i (U_j - Q_j).$$

Here $W \in \mathbb{R}^{d \times s}$ denotes a real valued weighting matrix with non-negative entries.

Remark 7.1. All results derived in the previous sections hold also for the weighted forward operator. The weighting can alternatively be incorporated in the definition of the norm in the measurement space, i.e., one can use the weighted norm $\|M\|^2_{\mathbb{C}^{d\times s}} := \sum_{i,j} W_{ij}^2 |M_{ij}|^2$ together with the original forward operator F. Note that the adjoint of the derivative operator with respect to this scalar product is given by $F'(\kappa, \mu)^{\#} r = F'(\kappa, \mu)^* [W_{ij}^2 r_{ij}]$, and consequently

$$F'(\kappa, \mu)^{\#} F'(\kappa, \mu) = F_w'(\kappa, \mu)^* F_w'(\kappa, \mu).$$

Therefore, weighting of the measurements is equivalent to a change of the scalar product in the measurement space.

Example 7.2. In order to take into account the measurement setup, we propose to use $W := 1/|F(\kappa_0, \mu_0)|$ or $W := 1/|M^\delta|$, where the division and absolute value have to be applied element wise. Such a weighting also introduces a natural scaling of the problem, i.e., the measurements will always be in the order of one.

Remark 7.3. A logarithmic transformation $\widetilde{M} = \log M$ is frequently used in practice [1, 31]. The derivative of such a transformation is given by

$$[\partial \widetilde{M}]_{ij} = \Big[\big(\log(F(\kappa, \mu)) \big)' [\hat{\kappa}, \hat{\mu}] \Big]_{ij} = [F(\kappa, \mu)]_{ij}^{-1} \big[F'(\kappa, \mu)[\hat{\kappa}, \hat{\mu}] \big]_{ij},$$

i.e., the derivative $[\partial \widetilde{M}]$ of the transformed measurements is defined as a scaled version of the derivative ∂M of the original data. The scaling matrix is $1/M(\kappa, \mu)$ here, which is similar to the $1/|M_0|$ scaling proposed in Example 7.2. Therefore, this scaling can be expected to behave similar to the linear weighting. Note, however, that the logarithmic scaling introduces additional nonlinearity in the forward operator, and we do not see how to generalize our convergence analysis easily to this case.

Depending on the relative magnitudes of the parameters κ and μ, also a scaling of the regularization norm may be appropriate.

Example 7.4. In our numerical experiments, we use the scaled norm

$$\|(\kappa, \mu)\| = (\beta_\kappa \|\kappa\|^2_{H^1(\Omega)} + \beta_\mu \|\mu\|^2_{L^2(\Omega)})^{1/2},$$

for regularization with scaling parameters defined by $\beta_\kappa := \|\kappa_0\|^2_{H^1(\Omega)}$ and $\beta_\mu = \|\mu_0\|^2_{L^2(\Omega)}$, respectively. Since β_κ and β_μ are positive real numbers, the scaled norm is equivalent to the original norm, and our convergence analysis also covers this case.

7.2 Discretization error

For application of any regularization method, some estimate for the perturbations in the data is required. This includes measurement errors, but also the discretization error

$$\|F_h(\kappa,\mu) - M^\delta\| \le \|y - y^\delta\| + \|F_h(\kappa,\mu) - F(\kappa,\mu)\| \le \delta + \delta_h.$$

Once, bounds for δ and δ_h are available, we can apply the convergence theory derived on the continuous level also to the discrete setting; cf. Remarks 5.8 and 5.7. In the following, we assume that some bound δ for the measurement error is already available, and we consider the discretization error δ_h in more detail.

For our numerical tests, we utilize the following strategy: Assume that two meshes with meshsize h and $h/2$ are available (typically, the fine mesh is constructed by refinement from the coarse one). Then

$$\eta_h := \|F_{h/2}(\kappa,\mu) - F_h(\kappa,\mu)\|$$

may serve as an estimate for the discretization error. Under a simple saturation condition [5], this estimate can be shown to be reliable and efficient.

Lemma 7.5. *Assume that* $\|F(\kappa,\mu) - F_{h/2}(\kappa,\mu)\| \le \gamma \|F(\kappa,\mu) - F_h(\kappa,\mu)\|$ *for some* $0 < \gamma < 1$. *Then*

$$(1+\gamma)^{-1}\eta_h \le \|F(\kappa,\mu) - F_h(\kappa,\mu)\| \le (1-\gamma)^{-1}\eta_h.$$

Proof. The result follows directly from the triangle inequality. □

Remark 7.6. Let us shortly comment on other a-posteriori error estimates. The error in the ijth measurement can be expressed as the error in the linear functional

$$[F_h(\kappa,\mu)]_{ij} - [F(\kappa,\mu)]_{ij} = \int_{\partial\Omega} \eta_i (U_j - u_j).$$

This difference can further be estimated, e.g., by the method of dual-weighted-residuals [5] or by other goal-oriented or functional-type a-posteriori error estimators. A main design criterion of such estimators is, that they can be computed by local post-processing. The cheap evaluation is particularly important, if only one or a few variational problems have to be solved. In our applications, however, we have to solve the forward, sensitivity, and adjoint problems many times. Therefore, we feel that the additional effort of solving a forward problem on a finer mesh is acceptable.

Borrowing ideas from the dual weighted residual error estimators, one can show, that the discrete error can be decomposed into element contributions, i.e.,

$$|[F(\kappa,\mu)]_{ij} - [F_h(\kappa,\mu)]_{ij}| \le \sum_{T\subset\Omega} \eta_{ij}^T + \sum_{E\subset\partial\Omega} \eta_{ij}^E$$

where T and E denote elements respectively boundary elements of the mesh, and the local error contributions are defined by

$$\begin{aligned}\eta_{ij}^E &:= \|\rho\|_{\infty,E}\|U_{j,h} - U_{j,h/2}\|_E\|V_{i,h} - V_{i,h/2}\|_E \\ \eta_{ij}^T &:= \|\kappa\|_{\infty,T}\|\nabla(U_{j,h} - U_{j,h/2})\|_T\|V_{i,h} - V_{i,h/2}\|_T \\ &\qquad + \|\mu + \iota k\|_{\infty,T}\|U_{j,h} - U_{j,h/2}\|_T\|V_{i,h} - V_{i,h/2}\|_T.\end{aligned}$$

Remark 7.7. The quantities $\eta_h^T := \sum_{i,j}\eta_{ij}^T$ and $\eta_h^E := \sum_{i,j}\eta_{ij}^E$ can be used for adaptive mesh design [5]. Note that the solutions u_j and v_i of the forward and adjoint problems decay fast away from the sources q_j respectively the detectors η_i. This a-priori argument already implies, that the meshes should be refined towards the boundary, and we utilize such graded meshes in our numerical experiments.

7.3 Choice of parameters

According to the convergence results of Section 3, and Remarks 5.7 and 5.8, the regularization parameter should be chosen in the order of the data error, i.e., $\alpha \sim \delta + \delta_h$ to obtain optimal convergence rates. For the execution of the iterative reconstruction methods, one has to specify additional parameters, e.g., n_0 for the projected gradient method, or α_0 as well as r for the projected Gauß-Newton iteration.

Remark 7.8. The constants involved in the choice of parameters depend on the properties of the operator F, and could in principle be estimated a-priori. In practice, one can assume that the noise level δ and the discretization error δ_h are of the same order for similar experiments, so we propose to select appropriate parameters, which work well for a series of test problems, and then utilize these parameter sets also for the actual computations. Due to the mesh-independence of our algorithms, the parameters only depend on the problem, and not on the discretization; so the same sets of parameters can be used for various discretizations and measurement configurations. This will be illustrated with numerical test in the next section.

Remark 7.9. Instead of the a-priori stopping rules utilized in Theroems 3.10, 4.6, and 4.8, one might as-well use a-posteriori stopping rules, e.g., the Lepskij principle [7] or a discrepancy principle. To the best of our knowledge, a complete convergence analysis for the latter case is only available under additional assumptions on the nonlinearity of the forward operator, which we could not verify for the problem considered in this paper.

8 Numerical tests

In this section, we report on numerical results obtained with the projected Gauß-Newton and the projected gradient method. As a test case, we consider the following benchmark problem defined in [31]: Ω is a two-dimensional circular domain with 25mm radius, and the optical parameters κ, μ are depicted in Figure 1, cf [31]. The parameter in the boundary conditions is set to $\rho = 1/3.25$, and the modulation frequency is $f = 150\text{MHz}$ ($\omega = 2\pi f$). As initial guess we use in our tests the background values of the true parameters, i.e., $\kappa_0 \equiv 1.4815\text{mm}$ and $\mu_0 \equiv 0.025\text{mm}^{-1}$.

At the boundary of the domain we put $s = 32$ equally distributed sources, and one of $d = 32$ detectors is placed between each pair of neighbouring source locations. The distribution of

sources and detectors is displayed in Figure 1, together with one of the adaptive meshes used for our computations; cf. Rmeark 7.7. For our numerical experiments, we use simulated data, which are generated on a fine mesh with 556.033 vertices. To avoid inverse crimes, the reconstructions are performed on coarser meshes, and the data are additionally perturbed by random noise.

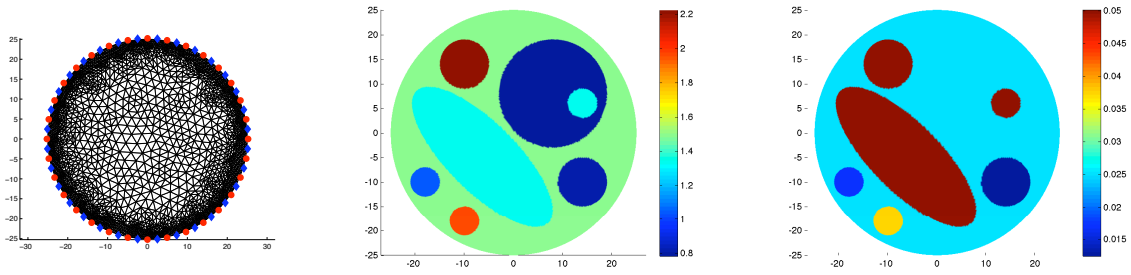

Figure 1: Mesh with $N = 2293$ vertices (left) together with source (red) and detector arrangement (blue); optical parameters used for the reconstructions: diffusion coefficient κ (middle) and absorption coefficient μ (right).

8.1 Discretization errors

As outlined in Section 5.5, discretization errors have to be taken into account in order to guarantee convergence of the fully discrete schemes. For our simulations, we utilize the strategy outlined in Section 7.2. To verify the efficiency and reliability of the error estimator $\eta_h := \|F_{h/2}(\kappa, \mu) - F_h(\kappa, \mu)\|$, we computed the error estimates and the true error (i.e., the error of the finite element solution at meshsize h to the one obtained at the finest discretization level). The results are listed in Table 2.

N	η_h	disc. error
606	4.91e-01	5.13e-1
2293	6.95e-03	7.02e-3
8913	1.18e-03	1.30e-3
35137	2.12e-04	3.48e-4

Table 2: Estimated error η_h and true discretization error (i.e. error of the finite element solution at meshsize h, with the one obtained on a mesh with 556.033 vertices) vs. the number of unknowns.

As predicted by the theory, the estimates are very close to the true discretization errors on all meshes. The numerical tests also reveal that a mesh with at least 2.293 vertices should be used in order to appropriately treat data with 1% data noise, i.e., such that $\delta_h < \delta$ and the discretization error can be neglected.

8.2 Choice of parameters

In order to select parameters α_n for the iterative reconstruction methods, we ran several tests with different data sets, and determined the following parameter choice rule, which was used

in all our numerical tests: $\alpha_0 = 10^{-5}$, and $\alpha_N(\delta) = 10^{-4}\delta^{3/2}$. Similar parameters α_N were obtained, when the iterations were stopped by the discrepancy principle

$$r_n := \|F(\kappa_n, \mu_n) - M^\delta\|_{\mathbb{C}^{d\times s}} \leq \tau\delta$$

with $\tau = 2$. The parameters according to our a-priori choice are compared with those obtained with the discrepancy principle in Table 3.

δ	$\alpha_N^{(disc)}$	$\alpha_N^{(apriori)}$
10e-1	5.00e-6	3.16e-6
10e-2	2.50e-7	1.00e-7
10e-3	1.95e-8	3.16e-9
10e-4	3.05e-10	1.00e-10
10e-5	2.38e-12	3.16e-12

Table 3: Regularization parameters α_N obtained with the discrepancy principle and $\tau = 2$ on a mesh with $N = 2.293$ vertices (left), and a-priori choice $\alpha_N = 10^{-4}\delta^{3/2}$ used for the numerical tests below (right). In order to obtain results for very small noise levels, data generated on the same mesh were used here.

For the Gauß-Newton iteration, we then set $\alpha_n = \alpha^0 r^n$ with $r = 0.5$, and for the projected gradient method, we chose $\alpha_n = \alpha_0(n+1)^{-1}$.

8.3 Mesh independent performance

In order to illustrate the mesh-independent performance of the fully discrete iterative reconstruction methods, we performed the following numerical tests: We used data generated on the finest mesh with 556.033 vertices and additionally added 1% random noise. The reconstruction errors, which were obtained in our computations are listed in Table 4.

N	r_N	e_N/e_0	GN it.	CG it.	time in s
2293	9.75e-3	0.53	7	405	116
8913	9.60e-3	0.56	7	292	369
35137	9.78e-3	0.58	7	219	1288

Table 4: Reconstruction errors $e_N = e(\kappa_N, \mu_N) = (\|\kappa_N - \kappa^\dagger\|^2_{L^2(\Omega)} + \|\mu_N - \mu^\dagger\|^2_{L^2(\Omega)})^{1/2}$ obtained with the projected Gauß-Newton method (17) with $\alpha_n = 2^{-n} \times 10^{-5}$ and $\alpha_N = 10^{-7}$ (cf. Table 3). Also displayed are the weighted residuals $\|(F_h(\kappa_N, \mu_N) - M^\delta)/|F_h(\kappa_0, \mu_0)|\|$, the number of Gauß-Newton iterations, and the total number of conjugate gradient iterations required for solving the Gauß-Newton systems.

As predicted by our theory, the reconstruction errors as well as the computational effort do not deteriorate as the meshsize is decreased.

For comparison with the complexity estimates derived in Section 6, we list in Table 5 the number of conjugate gradient iterations required for the solution of the Gauß-Newton systems in dependence of the regularization parameter α_n.

The increase of the iteration numbers has exactly the predicted behaviour. Note that most of the work is always performed in the final iteration, i.e., if $\alpha_n = \alpha_0 r^n$ decays exponentially,

n	0	2	4	6	8	10	rate
CG steps	12	28	67	125	228	376	-0.50
α_n	1.00e-5	2.50e-6	6.25e-7	1.56e-7	3.92e-8	9.77e-9	1.00

Table 5: Number of conjugate gradient iterations required for the solution of the Gauß-Newton systems in dependence of α_n obtained on mesh with 2.293 vertices.

then the total number of conjugate gradient iterations is a small multiple of the iterations required for the last Newton step; see also Table 1.

As a final test, let us compare the performance of the Gauß-Newton method to that of the projected gradient method. The results of these computations are listed in Table 6.

δ	$\alpha_{\min}$	e_N/e_0	it	time	e_N/e_0	it	time
1e-1	3.16e-6	0.75	2(34)	13	0.80	4	5
1e-2	1.00e-7	0.50	7(404)	111	0.60	100	55
1e-3	3.16e-9	0.35	12(2013)	550	0.40	3163	1768

Table 6: Comparison of the performance of the Gauß-Newton method and the projected gradient method on a mesh with 2.293 vertices.

Note that for relatively large noise levels ($\delta \geq 1\%$), the projected gradient method is a viable alternative to the algorithmically more complicated Gauß-Newton iteration. The Gauß-Newton methods seems to deliver slightly better reconstructions. Let us mention that for our choice of stopping parameters, the residuals obtained with the Gauß-Newton method is somewhat smaller than that of the modified Landweber iteration with the same parameter α_N. If the discrepancy principle is used for determining the stopping index, both methods yield very similar results.

9 Discussion

In this paper, we derived and analyzed fully discrete reconstruction methods for optical diffusion tomography. While several steps, e.g., the choice of topologies, the analysis of mapping properties of the forward operator, depend on the particular application, the general framework is applicable to a wide class of inverse problems.

Our approach for devising reliable discrete schemes is based on a careful analysis of the inverse problem on the continuous level. In view of the Banach-Steinhaus theorem, such an analysis is inevitable for the construction of robust and mesh-independent algorithms. The proof of continuity and differentiability of the forward operators always depends on the choice of appropriate topologies. Additionally, some constraints (e.g. the positivity of solutions) may have to be imposed in order to guarantee well-definedness of the reconstruction algorithms already on the continuous level. All these considerations for the continuous problem have to be respected also in discrete setting.

As reconstruction algorithms, we considered projected versions of iterative regularization methods. Similar reconstruction methods (without projection and different regularizationn norms)have been discussed already in [31, 28]. We think, that our investigations contribute to

the assessment of these methods in several ways: (i) Based on the analysis of the continuous problem, we included the projection step into the formulation of the algorithms; (ii) we established convergence results on the continuous level, which together with an appropriate discretization also guarantee convergence of the fully discrete algorithms; (iii) we showed the exactness of the discrete derivatives and adjoints, which gave rise to real symmetric and positive definite Gauß-Newton systems. These could be solved efficiently by conjugate gradient methods; (iv) we provided complexity estimates for the fully discrete schemes; (v) we proposed implementations of the derivative and the adjoint operator, which are of optimal complexity $O(sN)$, i.e., independent of the number of detectors. Our algorithms are therefore more efficient than those relying on the assembly of the Jacobian presented in [31].

Although our reconstruction algorithms have optimal complexity and work reasonably fast, we did not aim at optimizing the implementation for a specific computer architecture. In fact, the solutions corresponding to different source terms or detectors could easily be computed on multiple processors, which would be a natural starting point for parallelization. The use of modern computer architectures, e.g., multicore systems or GPUs, would provide an additional possibility for further acceleration of the reconstruction process.

Acknowledgments

Financial support by the Deutsche Forschungsgemeinschaft (German Research Association) through grant GSC 111 is gratefully acknowledged.

References

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

[2] S. R. Arridge and W. R. B. Lionheart. Nonuniqueness in diffusion-based optical tomography. *Optics Letters*, 23:882-884, 1998.

[3] A. Bakushinsky. The problem of the convergence of the iteratively regularized Gauss-Newton method. *Comput. Math. Math. Phys.*, 32:1353–1359, 1992.

[4] A. B. Bakushinsky and M. Yu Kokurin. *Iterative Methods for Approximate Solution of Inverse Problems*. Springer, Dordrecht, 2004.

[5] W. Bangerth and R. Rannacher. *Adaptive Finite Element Methods for Differential Equations (Lectures in Mathematics Eth Zurich)*, Birkhäuser, 2003.

[6] D. C. Barber and A. D. Seagar. Fast reconstruction of resistive images. *Clin. Phys. P. A*, 8:47–54, 1987.

[7] F. Bauer and T. Hohage. A Lepskij-type stopping rule for regularized Newton methods. *Inverse Problems*, 21:1975–1991, 2005.

[8] D. P. Bertsekas. On the Goldstein-Levitin-Poljak gradient projection method. *IEEE Trans. Automat. Control*, AC-21:174–184, 1976.

[9] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, New York, Oxford, 1978.

[10] R. Dautray and J. L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology, Functional and Variational Methods*, volume 2. Springer, Berlin, 1993.

[11] H. Egger. On the convergence analysis of modified Landweber iteration for nonlinear inverse problems. Technical report, SFB-2010-017, University Graz, 2010.
[12] H. Egger and M. Schlottbom. Analysis and Regularization of Problems in Diffuse Optical Tomography. Technical Report, AICES-2009-23, RWTH Aachen University, 2009.
[13] H. W. Engl and M. Hanke and A. Neubauer. *Regularization of Inverse Problems.* Kluwer, Dordrecht, 1996.
[14] A. P. Gibson, J. C. Hebden, and S. R. Arridge. Recent advances in diffuse optical imaging. *Physics in Medicine and Biology*, 50(4):R1–R43, 2005.
[15] N. I. Grinberg. Local uniqueness for the inverse boundary problem for the two dimensional diffusion equation. *Eur. J. Appl. Math.*, 11:473–489, 200.
[16] M. Hanke and A. Neubauer and O. Scherzer. A convergence analysis of Landweber iteration for nonlinear ill-posed problems. *Numer. Math.*, 72:21–35.
[17] B. Harrach. On uniqueness in diffuse optical tomography. *Inverse Problems*, 25:055010, 2009.
[18] J. C. Hebden and S. R. Arridge and D. T. Delpy. Optical imaging in medicine: I. Experimental techniques. *Physics in Medicine and Biology*, 42:825–840, 1997.
[19] J. C. Hebden and S. R. Arridge and D. T. Delpy. Optical imaging in medicine: II. Modelling and reconstruction. *Physics in Medicine and Biology*, 42:841–853, 1997.
[20] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems – chances and limitations. *Inverse Problems*, 25:035003, 2009.
[21] M. Hinze and R. Pinnau and M. Ulbrich and S. Ulbrich. *Optimization with PDE Constraints.* Springer, 2009.
[22] B. Hofmann. Approximate source conditions in Tikhonov–Phillips regularization and consequences for inverse problems with multiplication operators. *Math. Meth. Appl. Sci.*, 29:351–71, 2006.
[23] V. Isakov. *Inverse Problems for Partial Differential Equations, 2nd ed.*. Springer, 2006.
[24] B. Kaltenbacher. Some Newton-type methods for the regularization of nonlinear ill-posed problems. *Inverse Problems*, 13:729–753, 1997.
[25] B. Kaltenbacher and A. Neubauer. Convergence of projected iterative regularization methods for nonlinear problems with smooth solutions. *Inverse Problems*, 22:1105–1119, 2006.
[26] R. Ramlau. A steepest descent algorithm for the global minimization of the Tikhonov functional. *Inverse Problems*, 18:381–405, 2002.
[27] R. Ramlau. TIGRA – an iterative algorithm for regularizing nonlinear ill-posed problems. *Inverse Problems*, 19: 433–465, 2003.
[28] R. Roy and E. M. Sevick-Muraca. Truncated Newton's optimization scheme for absorption and fluorescence optical tomography: Part I. Theory and formulation. *Opt. Express*, 4:353–371, 1999.
[29] O. Scherzer. A modified Landweber iteration for solving parameter estimation problems. *Appl. Math. Optim.*, 38:25–68, 1998.
[30] M. Schweiger and S. R. Arridge and D. T. Delpy. Application of the finite-element method for the forward and inverse models in optical tomography. *J. Math. Imag. Vis.*, 3:263–283, 1993.
[31] M. Schweiger and S. R. Arridge and I. Nissilä. Gauss–Newton method of image reconstruction in diffuse optical tomography. *Phys. Med. Biol.*, 50: 2365–2386, 2005.
[32] A. N. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems.* Wiley, New York, 1977.