Sparsity and level set regularization for diffuse optical tomography using a transport model in 2D

To cite this article: Kernel Prieto and Oliver Dorn 2017 Inverse Problems 33 014001

View the article online for updates and enhancements.

Related content
- Topical Review
  Oliver Dorn and Dominique Lesselier
- A nonlinear inversion method for 3D electromagnetic imaging using adjoint fields
  O Dorn, H Bertete-Aguirre, J G Berryman et al.
- History matching of petroleum reservoirs using a level set technique
  Oliver Dorn and Rossmary Villegas

Recent citations
- Sparsity regularization in inverse problems
  Bangti Jin et al.
Sparsity and level set regularization for diffuse optical tomography using a transport model in 2D

Kernel Prieto and Oliver Dorn

School of Mathematics, The University of Manchester, Oxford Road, M13 9PL, UK
E-mail: oliver.dorn@manchester.ac.uk

Received 14 December 2015, revised 22 September 2016
Accepted for publication 3 October 2016
Published 25 November 2016

Abstract
In this paper we address an inverse problem for the time-dependent linear transport equation (or radiative transfer equation) in 2D having in mind applications in diffuse optical tomography (DOT). We propose two new reconstruction algorithms which so far have not been applied to such a situation and compare their performances in certain practically relevant situations. The first of these reconstruction algorithms uses a sparsity promoting regularization scheme, whereas the second one uses a simultaneous level set reconstruction scheme for two parameters of the linear transport equation. We will also compare the results of both schemes with a third scheme which is a more traditional $L^2$-based Landweber–Kaczmarz scheme. We focus our attention on the DOT application of imaging the human head of a neonate where the simpler diffusion approximation is not well-suited for the inversion due to the presence of a clear layer beneath the skull which is filled with ‘low-scattering’ cerebrospinal fluid. This layer, even if its location and characteristics are known a priori, poses significant difficulties for most reconstruction schemes due to its ‘wave-guiding’ property which reduces sensitivity of the data to the interior regions. A further complication arises due to the necessity to reconstruct simultaneously two different parameters of the linear transport equation, the scattering and the absorption cross-section, from the same data set. A significant ‘cross-talk’ between these two parameters is usually expected. Our numerical experiments indicate that each of the three considered reconstruction schemes do have their merits and perform differently but reasonably well when the clear layer is a priori known. We also...
demonstrate the behavior of the three algorithms in the particular situation where the clear layer is unknown during the reconstruction.

Keywords: diffuse optical tomography, inverse transport problems, radiative transfer equation, level set regularization, sparsity promoting regularization, nonlinear Kaczmarsz

(Some figures may appear in colour only in the online journal)

1. Introduction

*Sparsity promoting regularization* has become an interesting alternative to the more classical pixel or voxel based $L^2$-minimization schemes (in the following simply called ‘$L^2$-based schemes’), the latter ones with or without standard $L^2$ Tikhonov style regularization terms. This has been demonstrated so far mainly for linear inverse problems [17, 31], but recently also nonlinear inverse problems have received increased attention by the scientific community [10–12, 16, 35, 43, 44, 55, 63]. Often (depending on the mathematical setup) the results of sparsity promoting regularization schemes are able to provide more compact supports of inclusions than standard $L^2$-based schemes can offer, in particular in situations where the probing fields are highly diffusive. This is a promising feature in situations where a priori information suggests that the objects of interest are in fact compactly embedded in some slightly varying inhomogeneous background of clearly different characteristics. In situations where sharp interfaces between regions of significantly different parameter values are expected, the so-called *level set technique for shape reconstruction* has been studied as another alternative to $L^2$-based schemes with very promising results [13, 26–28, 41, 53, 54, 66, 72]. In this paper we propose two such new reconstruction schemes, one based on sparsity and one based on level sets, for an inverse problem of the time-dependent linear transport equation having in mind situations that occur clinically in the imaging of the oxygen supply inside the heads of neonates. As we are proposing these two different reconstruction schemes for a given imaging scenario, it is also tempting to ‘compare’ their performances on particular situations, and to compare as well with more traditional $L^2$-based schemes. Therefore, in addition to introducing these two schemes, we will attempt to provide a ‘visual’ comparison by adding also a custom-made $L^2$-based inversion scheme described further below. Even though the performance of any inversion scheme is usually highly situation specific, we believe that such a comparison helps to understand better the underlying principles of these different approaches. The goal will not be to provide any form of ‘ranking’ between the schemes. We mention that interesting studies following a similar spirit but for a set of different inverse problems have been performed previously in [73].

As mentioned, we will concentrate on one specific application, namely DOT. This imaging modality uses near-infrared (NIR) light for investigating the domain of interest non-invasively. The NIR light is usually generated in a laser source and guided towards the boundary of the tissue, where it enters the region of interest. The distribution of outgoing light, after having passed through the head, over time and space provides then some information on the absorption and scattering behavior of the tissue inside the domain of interest. NIR light propagation inside human tissue resembles some form of random walk, where the density of a population of particles (photons) can be described by the linear transport equation in time domain. More details on this imaging modality can be found in [4, 6, 58].
We employ a linear transport model for this application which is very general and also finds important applications (with a very similar setup) in related fields such as photoacoustic tomography, fluorescence tomography or bioluminescence imaging \([7, 21, 58, 64, 67, 71]\). We deliberately keep the computational setup simple and generic, however without sacrificing important features of the underlying transport phenomena. We focus on situations where the usually employed diffusion approximation will not perform well due to the presence of so-called clear layers or due to other reasons. This is a typical situation which occurs when imaging the head of neonates using NIR light \([4, 6, 64]\). The impact of clear layers or regions on the inversion process has not yet been studied very extensively in the literature, but interesting work has been presented in \([8, 19, 24, 33, 40, 65]\), amongst others. The joint inversion of such transparent structures and the two-parameter optical profile from the same data set provides big challenges. Often this layer is either ignored in the synthetic modeling, or assumed known. We will investigate both of these approaches in our numerical experiments, but will not focus directly on the task of reconstructing clear layers by a tailor-made algorithm. Some reconstructions of clear layers are nevertheless provided as a side-product in our second set of numerical experiments.

2. Mathematical model

We will restrict ourselves to a computational setup in two spatial dimensions. In order to fix the mathematical setup, let the space \(X\) be defined by

\[
X := \Omega \times S^{n-1} \times [0, T],
\]

where \(\Omega\) is the domain of interest which is assumed to be a compact, convex subset in \(\mathbb{R}^n\) (here \(n = 2\)) with boundary \(\partial X = \partial \Omega \times S^{n-1} \times [0, T]\). The physical boundary \(\partial \Omega\) is assumed to be Lipschitz such that (ignoring uninteresting tangential directions) \(\partial X\) can be decomposed into the outflow part \(\partial X_+\), and the inflow part \(\partial X_-\), that is, \(\partial X = \partial X_+ \cup \partial X_-\), where \(\partial X_+ := \{(x, \theta, t) \in \partial X, \pm \theta \cdot \nu > 0\}\) and \(\nu\) is the outward normal direction at the given boundary point. The function \(u(x, \theta, t)\) for \((x, \theta, t) \in X\) describes the density of the photons travelling in the region \(\Omega\) at position \(x\) with direction \(\theta\) (and speed normalized to \(1 \text{ cm s}^{-1}\)) at time \(t\). Both, \(u\) and the source term \(q\) are correctly modeled as scalar quantities in \(L^1(X)\) and are assumed nonnegative \([14, 18]\). Then the one-speed time-dependent linear transport equation, in the framework of DOT often called radiative transfer equation (RTE), can be formulated as

\[
\frac{\partial u}{\partial t} + \theta \cdot \nabla u + (\mu(x) + b(x))u - b(x) \int_{S^{n-1}} \eta(\theta \cdot \theta') u(x, \theta', t) \mathrm{d}\theta' = q(x, \theta, t) \tag{1}
\]

inside \(X\) with initial condition

\[
u(x, \theta, 0) = 0 \quad \text{in} \quad \Omega \times S^{n-1} \tag{2}
\]

and boundary condition

\[
u(x, \theta, t) = 0 \quad \text{on} \quad \partial X_- \tag{3}
\]

where we have chosen to model the incoming laser light as a source term \(q\) in (1) instead of an incoming boundary condition in (3). The coefficients \(\mu\) and \(b\) in (1) are the absorption and scattering coefficients of the medium, respectively. We denote by \(u_i, i = 0 \ldots, M - 1\), the solutions of (1)–(3) when applying the particular source terms.
where $\delta_x$, $\delta_y$, and $\delta_z$ are (if necessary suitable function space approximations to) the Dirac delta distributions corresponding to the position, angular and time variables, respectively. Here the locations $p_i \in \partial \Omega$ are assumed to be positioned at the boundary of the domain and $\theta_i$ is inward directed for modeling an incoming ultra-short laser pulse. When considering the RTE in 2D as in this paper, the domain $\Omega \subset \mathbb{R}^2$ and the direction vectors $\theta$ are in $S^1$. The RTE describes the balance of the number of photons in an infinitesimal volume over an infinitesimal time range [14]. The scattering function $\eta(\theta \cdot \theta')$, also known as the scattering kernel, is assumed here to depend only on the cosine of the scattering angle $\cos \psi = \theta \cdot \theta'$ formed between the incoming and outgoing directions $\theta'$ and $\theta$, respectively. In particular it is independent of the location $x$. This scattering function is assumed to be a priori known and normalized in the sense that

$$\int_{S^1} \eta(\theta \cdot \theta') d\theta' = 1,$$

which physically represents particle conservation in pure scattering events. Scattering functions are usually highly application dependent and often difficult to model correctly. A popular (since convenient) three-dimensional ($n = 3$) scattering model applied in DOT is described by the 3D Henyey–Greenstein (HG) scattering function

$$\eta(\cos \psi) = \frac{1 - g^2}{4\pi(1 + g^2 - 2g \cos \psi)^2},$$

When considering two-dimensional ($n = 2$) analogies for DOT, the choice of appropriate scattering functions is less well investigated and somehow arbitrary at this stage. One possible model for such 2D situations has been suggested in [39]. However, in this paper we follow a slightly different approach introduced in [21]. It projects the 3D version of (6) onto 2D maintaining some of its properties and at the same time guaranteeing particle conservation in 2D as it is requested by (5). The parameter $g$ is also here in the range of $-1 < g < 1$. Values of $g$ close to 1 indicate that scattering is primarily forward directed, whereas values close to $-1$ indicate that scattering is primarily backward directed. Typical values for $g$ in DOT obtained from animal tissues are $0.9 \leq g \leq 0.99$, which will be adopted as a reference for soft brain tissue in our numerical experiments. We mention that also some approximations to the linear transport equation have been specifically designed in the literature to take into account strong forward scattering [47, 51]. Notice also that the recently proposed Fournier–Forand scattering phase function provides another alternative to HG [38].

The outgoing flux across the boundary $\partial \Omega$ at receiver position $x_i$ and receiving time $t_i$ for given parameters $(\mu, b)$ is given by

$$G_i[\mu, b](x_i, t_i) = \int_{S^{n-1}} u_i(x_i, \theta, t_i) \nu(x_i) \cdot \theta \, d\theta \quad \text{on } \partial \Omega \times [0, T],$$

where $u_i$ is the solution of the RTE (1)–(3) for the source $q_i$, $i = 0, \ldots, M - 1$ and parameters $(\mu, b)$. Here $S^{n-1}$ denotes the subset of direction vectors $\theta \in S^{n-1}$ for which $\nu(x_i) \cdot \theta > 0$. When physically measured (or simulated) data $\hat{G}_i$ are available for each source position $q_i$, then we can define the corresponding residuals $\mathcal{R}_i[\mu, b]$ as

$$\mathcal{R}_i[\mu, b](x_i, t_i) = \hat{G}_i[\mu, b](x_i, t_i) - G_i[\mu, b],$$

describing the mismatch between predicted and measured data. If the measured data are noiseless, they can be considered as being associated with the ‘reference’ parameters $\tilde{\mu}, \tilde{b}$ for which the residuals become zero, that is, $\mathcal{R}_i[\tilde{\mu}, \tilde{b}(x_i, t_i) = 0$ for each source $q_i$. If noise
contained in the data, then this assumption is usually not true anymore, and the residual will only be approximately zero. In these situations usually some form of data misfit is minimized instead when practically solving the underlying inverse problem, and a suitable minimizer is deemed the solution of the inverse problem.

Let us in the following denote by $P$ the space of parameters $\mu$ and $b$, by $U$ the space of states $u$ and sources $q$, and by $Z$ the space of measurements $\mathcal{R}_i$ and (noiseless) residuals $\mathcal{R}_i$. In [14, 18] it is described that in many applications of the linear transport equation states and sources are naturally described as $L^1$-functions and parameters as $L^\infty$-functions due to their physical significance. Accordingly, in [20] it is shown that the residual operators $\mathcal{R}_i: \tilde{P} \rightarrow \tilde{Z}$ defined by (8) and (1)-(4) are Fréchet differentiable in the function spaces

$$\tilde{U} = \{ u \in L^1(X), (\partial_t + \theta \cdot \nabla)u \in L^1(X), \quad u(x, \theta, 0) = 0 \text{ on } \Omega \times S^{n-1}, u(x, \theta, t) = 0 \text{ on } \partial X \},$$

$$\tilde{P} = L^\infty(\Omega) \times L^\infty(\Omega), \quad \tilde{Z} = L^1(\partial \Omega \times [0, T]),$$

where these Banach spaces are equipped with appropriate norms. For more details on the inverse transport problem described in these spaces see [20, 21]. See also the related and more recent study in [70] which addresses a stationary transport problem.

## 3. Formulation of the inverse problem

Notice that in the following we will often denote the vector of unknowns in the inverse problem by the symbol $\sigma = (\mu, b)^T$. If only one of the parameters $\mu$ and $b$ is addressed and it does not matter which of both, we often denote it by $\kappa \in \{\mu, b\}$.

When deriving the different reconstruction schemes simultaneously in a consistent way it is more convenient to model all quantities as members of associated Hilbert spaces rather than using the natural Banach spaces $L^1$ or $L^\infty$ as outlined in the previous section. Therefore, we introduce now the function spaces

$$U = L^2(\Omega \times S^{n-1} \times [0, T]), \quad P = L^2(\Omega) \times L^2(\Omega), \quad Z = L^2(\partial \Omega \times [0, T]).$$

All these spaces are equipped with the standard inner products in $L^2$. The parameter-to-measurement operators $\mathcal{G}_i: P \rightarrow Z$ and the residual operators $\mathcal{R}_i: P \rightarrow Z$ are then mappings between Hilbert spaces described by (7), (8) and (1)-(4).

The classical formulation of the inverse problem in DOT can then be described as follows: Provided with the scattering function $\eta$ and the recorded data given by (7) for sources $q_i, i = 0, \ldots, M - 1$, find the absorption coefficient $\mu$ and the scattering coefficient $b$ (in the above function spaces) inside the domain $\Omega$ simultaneously.

One standard way to proceed now is to write down the least squares data misfit functional (often simply called ‘cost’) \n
$$\mathcal{J}(\sigma) = \mathcal{J}(\mu, b) = \sum_{i=0}^{M-1} \mathcal{J}_i(\mu, b); \quad \mathcal{J}_i(\sigma) = \mathcal{J}_i(\mu, b) = \frac{1}{2} \| \mathcal{R}_i[\mu, b] \|_Z^2,$$

which gives rise to the formal gradient

$$\nabla \mathcal{J}(\sigma) = \sum_{i=0}^{M-1} \nabla \mathcal{J}_i(\sigma). \quad (11)$$

When minimizing $\mathcal{J}(\sigma)$ with respect to $\sigma$, often iterative gradient-based techniques are employed which require the repeated evaluation of gradients given by (11). The so-called
adjoint scheme has proven to be a computationally viable way of doing this. We will use it in a form derived in [22] and adjust it to our situation as follows.

Let us denote the linearized residual operator by \( \mathcal{R}_{i}^{\cdot}[\mu, b] \) and its adjoint operator with respect to the function spaces \( P \) and \( Z \) by \( \mathcal{R}_{i}^{\cdot}[\sigma] = \mathcal{R}_{i}^{\cdot}[\mathcal{R}_{i}^{\cdot}[\sigma]] \). Let furthermore \( \xi_{i} = \mathcal{R}_{i}^{\cdot}[\sigma] \in Z, \sigma = (\mu, b)^{T} \in P \) and let \( z_{i} \) be the solution of the adjoint equation of the RTE

\[
-\frac{\partial z_{i}}{\partial t} - \theta \cdot \nabla z_{i} + (\mu(x) + b(x))z_{i}(x, \theta, t) - b(x) \int_{S^{n-1}} \eta(\theta \cdot \theta')z_{i}(x, \theta', t) d\theta' = 0
\]

in \( X \) with final condition

\[
z_{i}(x, \theta, T) = 0 \quad \text{on} \quad \Omega \times S^{n-1}
\]

and boundary condition

\[
z_{i}(x, \theta, t) = \xi_{i} \quad \text{on} \quad \partial X_{+}.
\]

Then

\[
\nabla \mathcal{J}_{i}(\sigma) = \mathcal{R}_{i}^{\cdot}[\sigma]^{\#} \xi_{i} = \mathcal{R}_{i}^{\cdot}[\mathcal{R}_{i}^{\cdot}[\sigma]]
\]

is given by

\[
\mathcal{R}_{i}^{\cdot}[\sigma]^{\#} \xi_{i} = \left( \begin{array}{c} \mathcal{L}_{1}(u_{i}, z_{i}) \\ \mathcal{L}_{2}(u_{i}, z_{i}) \end{array} \right)
\]

with

\[
\mathcal{L}_{1}(u_{i}, z_{i}) := -\int_{0}^{T} \int_{S^{n-1}} u_{i}(x, \theta, t)z_{i}(x, \theta, t) d\theta dt,
\]

\[
\mathcal{L}_{2}(u_{i}, z_{i}) := \int_{0}^{T} \int_{S^{n-1}} \left( \int_{S^{n-1}} \eta(\theta \cdot \theta')u_{i}(x, \theta', t) d\theta' - u_{i}(x, \theta, t) \right)z_{i}(x, \theta, t) d\theta dt.
\]

Here \( \xi_{i} \) is understood to be applied uniformly in all directions \( \theta \) with \( \theta \cdot \nu > 0 \) in (14) and \( u_{i} \) solves (1)–(3) with source \( q_{i} \). The term \( \xi_{i} = \mathcal{R}_{i}^{\cdot}[\sigma] \) in (15) can be interpreted as an ‘adjoint’ source located at the detector positions having strength proportional to the residual values and giving rise to a back-propagating flux of virtual photons in reverse time. Due to this interpretation the reconstruction scheme considered here is often called ‘transport-backtransport scheme’ [22].

The above form of the adjoint to the linearized residual operator is mapping into standard \( L^{2} \)-spaces by construction. In the level set approach as well as in the sparsity regularization approach it will be convenient to have a slightly stronger form of the adjoint linearized residual operator available which maps into a Sobolev subspace of \( L^{2} \) as defined below. This will guarantee that the output of this operator has a sufficient degree of smoothness which helps stabilizing the mentioned reconstruction schemes.

In order to derive this alternative form, we follow [34]. Let us assume that the coefficients \( \mu \) and \( b \) are in the subset \( H^{1}(\Omega) \) of \( L^{2}(\Omega) \), where \( H^{1} \) stands for the Sobolev space defined by

\[
H^{1}(\Omega) = \{ \kappa \in L^{2}(\Omega), \partial_{\kappa}, \partial_{\kappa,\kappa} \in L^{2}(\Omega) \},
\]

and \( \partial_{\kappa}, \partial_{\kappa,\kappa} \) are the weak partial derivatives. This space is assumed to be equipped with the weighted inner product

\[
\langle p_{1}, p_{2} \rangle_{H^{1}} = \langle p_{1}, p_{2} \rangle_{L^{2}} + \gamma \langle \nabla p_{1}, \nabla p_{2} \rangle_{L^{2}},
\]

(19)

\( \gamma > 0 \) fixed, and the corresponding induced norm (see [1] for general background on such spaces). If we denote \( \mathcal{B} = H^{1}(\Omega) \times H^{1}(\Omega) \) and let \( \xi \in Z \) and \( \delta \sigma \) be a perturbation of
\[ \sigma = (\mu, b)^T \] in \( \hat{P} \), we define the smooth adjoint \( \mathcal{R}[\sigma]^* \xi \) by
\[
\langle \mathcal{R}[\sigma]^* \delta \sigma, \xi \rangle_Z = \langle \delta \sigma, \mathcal{R}[\sigma]^* \xi \rangle_P = \langle \delta \sigma, \mathcal{R}[\sigma]^* \xi \rangle_P,
\]
where \( P \) and \( Z \) are as defined in (9) and \( \langle \cdot \rangle_P \) refers to the inner product with respect to the space \( \hat{P} \) following (19) for each parameter. Then, the smooth gradient \( \nabla J^i \) is given by
\[
\langle \delta \sigma, \nabla J^i(\sigma) \rangle_P = \langle \delta \sigma, \nabla J(\sigma) \rangle_P.
\]
Integrating by parts on the left-hand side of (21) and using the initial and boundary conditions that correspond to the RTE and its adjoint, we see that \( \nabla J(\sigma) \) and \( \nabla J^i(\sigma) \) are related by
\[
(I - \gamma \Delta) \nabla J^i(\sigma) = \nabla J(\sigma) \quad \text{in} \quad \Omega,
\]
where \( I \) and \( \Delta \) in (22) are the Identity and the Laplacian operator, respectively, and a homogeneous Neumann boundary condition is assumed on \( \partial \Omega \). This relationship allows us to obtain the smooth gradient \( \nabla J^i(\sigma) \) from \( \nabla J(\sigma) \) by a simple post-processing (smoothing) step from the results obtained by the above described adjoint scheme. Due to this smoothing of the \( L^2 \) gradient for obtaining the \( H^1 \) gradient the latter one is called here ‘smoothed gradient’. In the literature variants of this scheme are often referred to as ‘Sobolev gradients’ [57].

4. A reconstruction approach using \( L^2 \)-based Landweber–Kaczmarz

Since our nonlinear problem is ill-posed, regularization techniques are typically applied for solving it. Iterative regularization techniques have been discussed extensively in the literature [46]. The focus of this paper is not on \( L^2 \)-regularization techniques, so we will keep this section very brief. One of the simplest approaches for iteratively regularized inversion of our nonlinear inverse problem is the Landweber method. In this approach we would simply write in step \( k \) of the iteration
\[
\sigma^{k+1} = \sigma^k - \tau \nabla J(\sigma^k), \quad \text{with} \quad \tau = \begin{bmatrix} \tau_\mu & 0 \\ 0 & \tau_\nu \end{bmatrix},
\]
where \( \tau \) is a relaxation parameter or fixed step-size and \( \nabla J \) is as defined in (11). This is practically calculated by obtaining the gradient directions \( \nabla J_i \) for each source following (15) and combining them as specified in (11). Convergence properties of such algorithms have been discussed in [37, 46]. In general the Landweber iteration technique is known to converge relatively slowly, but adding a Kaczmarz cycle over the index \( i \) in (23), (11) has proven to speed it up significantly in many situations [22, 36, 37, 52, 56]. Doing so we arrive at a technique which can also be considered a special case of a nonlinear variant of the algebraic reconstruction technique (ART) as pointed out in [56]. In more details, the Landweber–Kaczmarz (LK) method for solving the inverse problem proceeds in ‘cycles’ or ‘sweeps’ over the data subset corresponding to individual source positions (an idea going back to the original paper by Kaczmarz [45]) where each individual step can be written as
\[
\sigma^{k+1} = \sigma^k - \tau \nabla J_{[k]}(\sigma^k).
\]
Here we follow a simple sequential rule \([k] = (k \mod \text{mod} \ M) \in \{0, \ldots, M - 1\} \) and \( \nabla J_{[k]}(\sigma^k) \) is given by (15). The relaxation parameter \( \tau \) has the form as in the Landweber approach. \( M \) sequential steps complete a sweep of the algorithm. Notice that other non-sequential selection criteria for the individual subsets of data in (24) are possible as well, see for example [62].

We are aware that, instead of LK, higher order Newton, Gauss–Newton or Quasi–Newton type techniques certainly as well can be derived for solving the underlying \( L^2 \)-based
optimization task (see for example [30, 49]), but notice that the nonlinear version of the LK or ART scheme employed here has the very useful advantage that it is conceptually quite simple and therefore can be applied without major changes also to level set based reconstruction schemes or sparsity regularized reconstruction schemes. The LK method is also computationally inexpensive in terms of memory consumption compared with many other schemes. We will make heavy use of these advantages when addressing our other reconstruction techniques further below. Convergence properties of the LK technique in various applications have been discussed in [36, 37, 46, 56]. We also mention that no line search is implemented for this scheme in our numerical experiments. Small updates are used following a fixed step-size criterion.

Algorithm 1 describes the \( \text{LK-L}^2 \) reconstruction scheme for the reconstruction of absorption and scattering coefficients \( \mu, b \) in form of a Pseudo-code.

Algorithm 1. \( \text{LK-L}^2 \)

\begin{itemize}
  \item[Input:] \( \sigma^0, \tau, I \) = number of sweeps.
  \item[Output:] \( K = IM \), \( \sigma^K \).
  \item[Step 1:] Perform \( I \) sweeps over the \( M \) source positions;
  \item[Step 2:] \( \forall k = 0, \ldots, IM - 1 \) do
    \begin{itemize}
      \item Compute the gradient \( \nabla \lambda_i(\sigma^k) \) according to (15);
      \item Calculate \( \sigma^{k+1} \) using (24) with fixed step-size \( \tau \).
      \item An optional stopping criterion for early termination;
    \end{itemize}
  \end{itemize}

5. Level set shape and contrast reconstruction

Various level set reconstruction schemes for nonlinear inverse problems have been reviewed in [13, 15, 26, 27, 41, 53, 72]. We adopt here a sufficiently general approach to be compared with \( L^2 \)-LK regularization on the one hand and sparsity-promoting regularization on the other hand. Originally the level set method has been developed by Osher and Sethian as a computational toolbox for tracking the propagation of interfaces in computational physics or image processing applications [59, 60, 69]. In the seminal paper [66] Santosa proposed to solve inverse problems for the shapes of unknown obstacles using such a level set method. This has been applied to DOT and Fluorescence Tomography using the time-dependent linear transport equation in [24, 25, 29] and to DOT using a diffusion approximation by [2, 68]. Later the reconstruction of internal parameter values has been attempted simultaneously with the corresponding shapes, see the above review papers for general applications and [5] for an example in DOT and [3] for an example from Fluorescence Tomography, both using the diffusion approximation. We will develop here a similar joint reconstruction algorithm for two parameters of the time-dependent linear transport equation for DOT. For brevity we leave out many practical details of the level set approach and refer for them to the given references.

In order to reconstruct the values and shapes for both, scattering and absorption coefficient, simultaneously in DOT, we introduce level set functions \( \phi, \psi \in P \) (or \( \phi, \psi \in \bar{P} \)) and write

\[
\mu(\phi, \mu_{int}) = \mu_{int}(1 - H(\phi)) + \mu_{ext}H(\phi), \quad b(\psi, b_{int}) = b_{int}(1 - H(\psi)) + b_{ext}H(\psi),
\]

\[\text{(25)}\]
where $H$ is the Heaviside function in 1D and $\mu_{\text{int}}$, $\mu_{\text{ext}}$, $b_{\text{int}}$, $b_{\text{ext}}$ can be either constants or functions, depending on the model used. In our numerical experiments we will assume that $\mu_{\text{int}}(x) = \mu$ and $b_{\text{int}}(x) = b$ are unknown constants, and $\mu_{\text{ext}}(x)$, $b_{\text{ext}}(x)$ are known functions. The level set functions $\phi(x)$ and $\psi(x)$ implicitly describe the unknown shapes, and therefore are as well unknowns of the inverse problem. We now want to construct an artificial evolution of the level set functions $\phi$ and $\psi$ and of the contrast parameters $\mu_{\text{int}}$ and $b_{\text{int}}$ that gradually reduces the cost (10) when plugged into the transport model via its associated shapes and contrast values. Since in this new model the parameters have the form (25), the expressions in (10) can now be written as (slightly abusing notation by using the same symbols throughout)

$$\mathcal{J}(\phi, \psi, \mu_{\text{int}}, b_{\text{int}}) = \frac{1}{2} \left\| R_{\text{int}}[\phi, \psi, \mu_{\text{int}}, b_{\text{int}}]\right\|_2^2$$

for $j = 0, \ldots, M - 1$. The descent direction for the functional $\mathcal{J}$ is obtained by taking the total derivative of the functional $\mathcal{J}$ with respect to the artificial evolution time $t$ (not to be confused with the physical time of the transport process), which, formally applying the chain rule, becomes

$$\frac{d\mathcal{J}}{dt} = \partial_\mu \mathcal{J} \frac{d\mu}{dt} + \partial_\phi \mathcal{J} \frac{d\phi}{dt} + \partial_\psi \mathcal{J} \frac{d\psi}{dt} + \partial_{\mu_{\text{int}}} \mathcal{J} \frac{d\mu_{\text{int}}}{dt} + \partial_{b_{\text{int}}} \mathcal{J} \frac{db_{\text{int}}}{dt}.$$ 

(27)

From our previous analysis we know that for a small variation $\delta \mu$ we have

$$\frac{\partial \mathcal{J}}{\partial \mu} \delta \mu = \int_\Omega \delta \mu R'_{\text{int}}[\mu, b]^* R_{\text{int}}[\mu, b] \, dx,$$

(28)

where $R'_{\text{int}}[\mu, b]^*$ is the adjoint of the (partial) Fréchet derivative of $R_{\text{int}}[\mu, b]$ with respect to the coefficient $\mu$ and is given by the corresponding component on the right-hand side in (16). A similar expression holds for the scattering parameter $b$. Using equation (25), one obtains furthermore

$$\frac{\partial \mu}{\partial \phi} = (\mu_{\text{ext}} - \mu_{\text{int}}) H'(\phi), \quad \frac{\partial \mu}{\partial \mu_{\text{int}}} = 1 - H(\phi),$$

$$\frac{\partial b}{\partial \psi} = (b_{\text{ext}} - b_{\text{int}}) H'(\psi), \quad \frac{\partial b}{\partial b_{\text{int}}} = 1 - H(\psi),$$

where $H'(\phi)$ and $H'(\psi)$ coincide with the 1D Dirac delta distributions (practically represented here by suitable function space approximations). Introducing the following short notation for the forcing terms

$$f_\phi(x, t) = \frac{d\phi}{dt}, \quad f_\psi(x, t) = \frac{d\psi}{dt}, \quad g_\mu(t) = \frac{d\mu_{\text{int}}}{dt}, \quad g_b(t) = \frac{db_{\text{int}}}{dt},$$

(29)

equation (27) obtains the form

$$\frac{d\mathcal{J}}{dt} = \int_\Omega f_\phi ((\mu_{\text{ext}} - \mu_{\text{int}}) H'(\phi) R'_{\text{int}}[\mu, b]^* R_{\text{int}}[\mu, b] \, dx$$

$$+ \int_\Omega f_\psi ((b_{\text{ext}} - b_{\text{int}}) H'(\psi) R'_{\text{int}}[\mu, b]^* R_{\text{int}}[\mu, b] \, dx$$

$$+ g_\mu \int_\Omega (1 - H(\phi)) R'_{\text{int}}[\mu, b]^* R_{\text{int}}[\mu, b] \, dx$$

$$+ g_b \int_\Omega (1 - H(\psi)) R'_{\text{int}}[\mu, b]^* R_{\text{int}}[\mu, b] \, dx,$$
where we have used that the terms $g_0$ and $g_b$ only depend on $t$ but not on $x$. Keeping in mind that $H'(\phi)$ and $H'(\psi)$ are formally positive, it is easily seen that in step $k$ of a level set based LK scheme analogous to (24) the individual forcing terms pointing into descent directions can be chosen as

$$f_{0,k} = - (\mu_{\text{ext}} - \mu_{\text{int}}) \mathcal{R}_{\{k\},b}[\mu^k, b^k] \mathcal{R}_{\{k\}}[\mu^k, b^k],$$

$$f_{0,k} = - (b_{\text{ext}} - b_{\text{int}}) \mathcal{R}_{\{k\},b}[\mu^k, b^k] \mathcal{R}_{\{k\}}[\mu^k, b^k],$$

and

$$g_{\mu,k} = - \int_{\mathcal{Y}_1} \mathcal{R}_{\{k\},b}^\prime[\mu^k, b^k] \mathcal{R}_{\{k\}}[\mu^k, b^k] \, dx,$$

$$g_{b,k} = - \int_{\mathcal{Y}_1} \mathcal{R}_{\{k\},b}^\prime[\mu^k, b^k] \mathcal{R}_{\{k\}}[\mu^k, b^k] \, dx,$$

where $\mathcal{Y}_1 = \{ x \in \Omega : \phi^k(x) \leq 0 \}$ and $\mathcal{Y}_2 = \{ x \in \Omega : \psi^k(x) \leq 0 \}$ and $[k] = (k \text{ modulo } M) \in \{0, \ldots, M-1\}$. Applying a forward Euler scheme for the artificial time evolution, in each step the two level set functions and the internal parameter values are updated according to the rules

$$\begin{pmatrix} \phi^{k+1} \\ \psi^{k+1} \end{pmatrix} = \begin{pmatrix} \phi^k + \tau \phi_{0,k} \\ \psi^k + \tau \phi_{0,k} \end{pmatrix}, \quad \begin{pmatrix} \mu^{k+1} \\ b^{k+1} \end{pmatrix} = \begin{pmatrix} \mu^k + \tau \mu_{\mu,k} \\ b^k + \tau b_{b,k} \end{pmatrix}.$$ \quad (33)

The initial guesses $\phi^0$, $\psi^0$ for the level set functions are chosen as so-called signed distance functions representing some arbitrary initial shapes. The initial guesses for the internal parameter values are $\mu_0^0 = \hat{\mu}_b + \epsilon_1$ and $b^0 = \hat{b}_b + \epsilon_2$ with some small contrast values $\epsilon_1$, $\epsilon_2 > 0$, where $\hat{\mu}_b$ and $\hat{b}_b$ are estimated or known background values.

The above derivations can be done analogously using the smooth adjoints $\mathcal{R}_{\{k\},b}^\prime[\mu, b]$ and $\mathcal{R}_{\{k\}}^\prime[\mu, b]$ in (28) following (22) with an appropriate weighting parameter $\gamma$ and with the obvious modifications in (30)-(32).

As step size criterion for the line search in the shape reconstructions we have prescribed a target value of pixels which maximally should change value in each step. A backtracking scheme is then applied (which does not need any forward calculations and is therefore computationally inexpensive and fast) to achieve these target values. In early sweeps the target values are chosen relatively large, and they are reduced during later iterations in order to control the speed of shape evolution. The update of the contrast values uses small but fixed step sizes.

The combined shape and contrast value reconstruction scheme for the absorption and scattering coefficients $\mu$, $b$ using level sets is summarized in algorithm 2 in form of a pseudo-code.

Algorithm 2. 1-K-level sets

Input: $\bar{\mu}^0$, $\bar{\psi}^0$, $\hat{\mu}^0$, $\hat{\psi}^0$, $I$ = number of sweeps.

Output: $K = IM$, $\phi^K$, $\psi^K$, $\mu^K$, $b^K$.

% Perform $I$ sweeps over the $M$ source positions;
for $k = 0, \ldots, IM - 1$ do

Compute the gradient $\nabla \mathcal{J}_{\Omega}(\mu^k, b^k)$ according to (15);
Compute the smooth gradient $\nabla \mathcal{J}_{\Omega}^\prime[\mu^k, b^k]$ following (22);
Do a backtracking line search on number of pixels that change value;
Calculate $\phi^{k+1}$, $\psi^{k+1}$, $\mu^{k+1}$, $b^{k+1}$ using (33);
An optional stopping criterion for early termination;
end
6. Sparsity promoting reconstruction

Sparsity regularization techniques for nonlinear inverse problems have been discussed in [11, 16, 35, 43, 44, 55], amongst others. In particular, [32, 42] describe a sparsity-promoting reconstruction scheme for the application of electrical impedance tomography (EIT). We follow this latter algorithmic and theoretical approach and apply the basic ideas to the new situation of DOT using a transport model. In contrast to the situation studied in [32, 42], in DOT two independent parameters need to be reconstructed simultaneously from the same data set, which requires some modifications of the algorithm. We mention that a diffusion approximation to the linear transport equation is used in [16] for a sparse two-parameter reconstruction problem in DOT, and in [61] for a single-parameter sparse reconstruction of the absorption parameter only.

Regularization usually amounts to making certain a priori assumptions on the parameter distribution to expect. In our sparsity promoting approach we assume that we are looking for anomalies occupying a compact shape of unknown topology and geometry embedded in a general inhomogeneous background. The overall volume of the support of the anomalies is assumed to be much smaller than the volume of the entire domain. The parameter values \( \sigma = (\mu, b)^T \) inside the anomalies are assumed to be significantly different from those values encountered in the background medium, such that the overall distribution can be written in the form

\[
\sigma = \sigma^0 + \delta \sigma,
\]

with some known background medium \( \sigma^0 = (\mu^0, b^0)^T \) into which the anomalies of interest \( \delta \sigma = (\delta \mu, \delta b)^T \) are embedded. Usually this background is ‘simple’ as for example a constant or smoothly varying low-contrast function. In our application of DOT we assume that it will be a constant background profile with an embedded clear layer, both of them being known but having a high contrast to each other. Our assumption is that sparsity promoting regularization schemes can be designed in a form which favor reconstructions of anomalies that have a small compact support embedded in such a background and making it this way easier to distinguish nearby compact anomalies [62]. Notice that this is only one way of applying sparsity to image reconstructions, and that the concept of sparsity can be generalized significantly when considering different bases or frames for the reconstruction [17]. We will not consider these generalizations here.

Let \( A \) be the admissible set of absorption and scattering coefficients given by

\[
A = \{ (\mu, b) \in H^1(\Omega) \times H^1(\Omega); c_0 \leq \mu \leq c_1, c_2 \leq b \leq c_3 \text{ a.e. } \Omega \},
\]

where \( c_0, c_1, c_2, c_3 \) are known positive constants. We introduce the following \( l_1 \) penalty terms

\[
\Psi_\mu(\delta \sigma) = \| \delta \mu \|_1 = \sum_{j \in \mathbb{N}} |(\delta \mu, \varphi_j)_{H^1(\Omega)}|,
\]

\[
\Psi_b(\delta \sigma) = \| \delta b \|_1 = \sum_{j \in \mathbb{N}} |(\delta b, \varphi_j)_{H^1(\Omega)}|,
\]

where \{ \varphi_j \} is an orthonormal basis or frame which can be a Fourier basis, or composed by wavelets, or alternatively (as in our case) formed by the characteristic functions of pixels or voxels of the domain \( \Omega \). We identify \( (\delta \mu, \varphi_j)_H, (\delta b, \varphi_j)_H \) as the expansion coefficients of \( \delta \mu \) and \( \delta b \) with respect to the basis \{ \varphi_j \}. Then we want to minimize the following cost

\[
\mathcal{F}(\sigma) = \mathcal{J}(\sigma) + \alpha_\mu \Psi_\mu(\delta \sigma) + \alpha_b \Psi_b(\delta \sigma) = \mathcal{J}(\sigma) + \alpha \cdot \Psi(\delta \sigma)
\]
over the set $A$, with $\mathcal{J}(\sigma)$ as in (10). We have written the $l_1$ penalty terms formally as an inner product $\alpha \cdot \Psi(\delta \sigma)$ between the row vector $\alpha = (\alpha_i, \alpha_b)$ and the column vector $\Psi(\delta \sigma) = (\Psi_i(\delta \sigma), \Psi_b(\delta \sigma))^T$. Notice that this penalty term $\alpha \cdot \Psi(\delta \sigma)$ is convex but not differentiable.

In our previous approaches, we anticipate also here that we will use the Landweber-Kaczmarz scheme for solving the problem (37). In this single-step approach we aim at minimizing in each step the functional

$$\mathcal{F}(\sigma) = \mathcal{J}(\sigma) + \alpha \cdot \Psi(\delta \sigma) \quad (38)$$

with $\sigma$ given as in (34) where the index $i$ refers to the source $q_i$. The publications [32, 42] propose a sparsity promoting regularization scheme for the nonlinear inverse problem EIT which uses a surrogate functional approach introduced earlier in [12, 17, 42]. In that algorithm, one real-valued parameter $\kappa(\lambda)$ (the conductivity) needs to be reconstructed from the given data. Our problem of DOT shares many similarities with EIT which motivates us to try an analogous approach for DOT. Using the surrogate functional approach with one real-valued parameter $\kappa$, the functional $\mathcal{F}(\kappa) = \mathcal{F}(\kappa^0 + \delta \kappa)$ can be locally approximated by the following surrogate functional

$$\mathcal{F}(\kappa^0 + \delta \kappa) \approx \mathcal{F}(\kappa^0) + (\delta \kappa - \delta \kappa^0, \nabla \mathcal{J}(\kappa^0))_{L^2(\Omega)} + \left(\tau^{-1}\frac{1}{2} ||\delta \kappa - \delta \kappa^0||^2_{L^2(\Omega)} + \alpha ||\delta \kappa||_1, \right.$$  

where $\tau$ is some step size to be discussed later. As shown in [32, 42], the direct application of the gradient $\nabla \mathcal{J}(\kappa^0)$ in this context might not retrieve accurate reconstructions because of its insufficient regularity properties. This is why we have agreed to use on $H^1(\Omega)$ the space of coefficients $\mu$, $b$, following a suggestion in [42]. This implies to use the Sobolev norm $|||H^1(\Omega)|||$ instead of $|||L^2(\Omega)||$ in the surrogate functional, which gives rise to the use of the smoothed gradients $\nabla \mathcal{J}^s(\kappa^0)$ instead of $\nabla \mathcal{J}(\kappa^0)$. The soft-thresholding approach is then proposed for minimizing this surrogate functional $\mathcal{F}$ by [12, 17, 63]

$$\delta \kappa^{k+1} = S_\lambda(\delta \kappa - \tau \nabla \mathcal{J}^s(\kappa^0)). \quad (39)$$

The soft-thresholding operator $S_\lambda$ is defined here for $\lambda > 0$ component-wise as

$$(S_\lambda(z))_j = \begin{cases} 0 & \text{if } |z_j| \leq \lambda, \\ \text{sign}(z_j)(|z_j| - \lambda), & \text{if } |z_j| > \lambda, \end{cases} \quad (40)$$

or, in a more compact way, as

$$(S_\lambda(z))_j = \text{sign}(z_j) \max\{|z_j| - \lambda, 0\}, \quad (41)$$

which is also called a shrinkage function. This operator is well-known to promote sparsity on $\delta \kappa$.

Since in our DOT we have to reconstruct two parameters simultaneously, we extend this approach to

$$\delta \sigma^{k+1} = \left(\frac{\delta \mu}{\delta b}\right)^{k+1} = \begin{pmatrix} S_{\alpha \tau} & \left(\delta \mu - \tau_{\mu} \nabla \mathcal{J}^s_{\mu}(\mu^k, b^k)\right) \\ S_{\alpha \tau} & \left(\delta b - \tau_b \nabla \mathcal{J}^s_{b}(\mu^k, b^k)\right) \end{pmatrix}, \quad (42)$$

where the terms $\nabla \mathcal{J}^s_{\mu}(\mu^k, b^k)$ and $\nabla \mathcal{J}^s_{b}(\mu^k, b^k)$ represent the individual components of (16). Notice that also here we apply a Kaczmarz style approach with a simple sequential rule $i = [k + 1] = (k \text{ modulo } M) \in [0,...,M - 1]$.

The update of $\delta \sigma$ in (42) is composed of two stages. The first stage corresponds to a $L^2$-LK step as discussed in a previous section with step sizes $\tau_{\mu}$ and $\tau_b$ (and using thereby the $H^1$
gradient), whereas the second stage takes care of the promotion of sparsity using the shrinkage operator. We also have to consider the optimal choice of step-sizes. Similarly to [32, 42] the Barzilai–Borwein (BB) step size criterion is used here which was first proposed in [9] and later has been adopted to sparse reconstructions in [74]. It translates to our dual parameter problem as

\[
\tau_{\mu}^k = \frac{\langle \delta \mu^k - \delta \mu^{k-1}, \nabla J_{\mu,\sigma}^k(\mu^k, b^k) - \nabla J_{\mu,\sigma}^k(\mu^{k-1}, b^{k-1}) \rangle_{H'(\Omega)}}{||\delta \mu^k - \delta \mu^{k-1}||^2_{H'(\Omega)}},
\]

and

\[
\tau_{b}^k = \frac{\langle \delta b^k - \delta b^{k-1}, \nabla J_{b,\mu}^k(\mu^k, b^k) - \nabla J_{b,\mu}^k(\mu^{k-1}, b^{k-1}) \rangle_{H'(\Omega)}}{||\delta b^k - \delta b^{k-1}||^2_{H'(\Omega)}}.
\]

It has been noticed, however, that with the BB criterion the functional \( F_{\mu} \) does not automatically decrease monotonically, such that an additional line search is suggested for assuring overall convergence of the scheme. We adopt in our work a weak monotonicity condition as stopping criterion as described in [50]. To start with, we use the step size given by the BB rules (43), (44) as initial guess, and then reduce both, \( \tau_{\mu}^k \) and \( \tau_{b}^k \), simultaneously until the weak monotonicity condition is satisfied for both parameters, which reads for one single parameter \( \kappa \) as [50]

\[
F_{\mu}(\kappa_0) + S_{\mu,\sigma}(\delta \kappa^k - \tau^k \nabla J_{\mu,\sigma}^k(\kappa^k)) \leq \max_{k-m+1 \leq l \leq k} F_{\mu}(\kappa^l) - \tau^k \frac{\varepsilon}{2} ||S_{\mu,\sigma}(\delta \kappa^k - \tau^k \nabla J_{\mu,\sigma}^k(\kappa^k)) - \delta \kappa^k||^2_{H'}.
\]

Here \( m \in \mathbb{N} \) and \( \varepsilon \in \mathbb{R} \). In the dual-parameter setting, corresponding terms for \( \kappa = \mu \) and \( \kappa = b \) need to be combined such that this condition makes sense. We refer for more technical details to [50, 62] and to [16] for a related but slightly different approach. A standard monotonicity condition holds for \( m = 1 \). We have used the value \( m = 2 \) in our reconstructions. In our implementation, we use the step length given by the BB rule as the initial guess at each iteration and reduce it geometrically by a factor \( r = 0.6 \) until condition (45) holds. Also, the initial step length is constrained to \((\tau_{\min}, \tau_{\max})\). This line search iteration is terminated when either \( \tau^k \) falls below a pre-specified stopping value \( \tau_{\text{stop}} \) or when the iteration does not make any further progress. Similarly, the LK-sparsity iteration is terminated when either no further progress is made in reducing the cost or when a maximal number of iterations has been reached.

A Pseudo Code for the reconstruction scheme using the LK method with sparsity regularization is described in Algorithm 3.

**Algorithm 3.** LK-sparsity

\[\text{Input: } \sigma^0, \delta \sigma^0 = 0, \alpha, \tau_{\min}, \tau_{\max}, \tau_{\text{stop}}, m, I = \text{number of sweeps.}\]

\[\text{Output: } K = IM, \sigma^K.\]

\[\text{Perform } I \text{ sweeps over the } M \text{ source positions; for } k = 0, \ldots, IM - 1 \text{ do}\]

\[\text{Compute the gradient } \nabla J_{\mu,\sigma}^k(\sigma^k) \text{ according to (15);}\]

\[\text{Compute the smoothed gradient } \nabla J_{\mu,\sigma}^k(\sigma^k) \text{ following (22);}\]

\[\text{Reduce the step sizes } \tau_{\mu}^k \text{ and } \tau_{b}^k \text{ simultaneously, starting from (43), (44), until (45) is satisfied;}\]

\[\text{Calculate } \delta \sigma^{k+1} \text{ according to (42);}\]

\[\text{Update } \sigma^{k+1} = \sigma^k + \delta \sigma^{k+1};\]

\[\text{An optional stopping criterion for early termination; end}\]

13
7. Computational setup

We model an ‘idealized’ human head by a domain of size $5 \times 5$ cm$^2$. It is composed of a homogeneous background and an embedded clear layer which simulates the cerebrospinal fluid of the head. $\Omega$ is divided into a rectangular grid of $50 \times 50 = 2500$ pixels each having a size of $1 \times 1$ mm$^2$. We have considered two different possibilities of the clear regions. First, a clear layer of 2 mm thickness located 4 mm beneath the external boundary and enclosing the entire interior domain, and second a clear layer of 1 mm thickness located 5 mm beneath the external boundary. We have observed that qualitative results depend mainly on the presence (and sufficient knowledge) of the clear layers, but not so much on which of the two considered scenarios is selected. Therefore, for the sake of brevity, we demonstrate results of the first scenario (2 mm thickness) only in our first set of numerical experiments, and of the second one (1 mm thickness) only in the second set of numerical experiments. The other two possibilities have produced comparable results and are omitted here. The location and the internal values of the absorption and scattering coefficients of this clear layer are assumed to be known a priori in our first set of numerical experiments, whereas they are completely ignored in the second one. The total time $T$ of the measured photon flow considered in the forward or inverse problem using the RTE is 20 s, that is, our data are recorded in the interval of time $[0, 20]$ s. Recall that the speed of light is normalized in our transport model to $c = 1$ cm s$^{-1}$, such that in a real clinical experiment the measurement time frame would be much shorter.

Two absorbing and one scattering objects of compact support and moderate contrast to the background are embedded in the homogeneous part of the background medium whose precise positions and sizes can be seen in the figures below. The parameter values used are tabulated in table 1. These are in the range of biomedical applications [67]. As phase function parameter we have used $g = 0.9$ which indicates (as in the 3D version) a preferred forward scattering in each individual scattering event.

The discretization of the linear transport equation in 2D used in our numerical experiments closely follows the scheme described in details in [21, 22]. For the sake of brevity we refer interested readers to those publications for more details regarding this scheme. It uses an upwind finite-differences scheme for the spatial discretization, a forward Euler scheme for the temporal discretization, and a discrete ordinates approach for the discretization of the angular variable. As mentioned in section 2, the goal of this paper is mainly the demonstration of the general performance of different regularization schemes, in a proof-of-concept fashion, rather than aiming at describing a realistic 3D setup directly applicable to real data. However, we are very confident that the important ingredients of a realistic transport model for 3D diffuse optical tomography are well-represented in our 2D model, and that the use of a more sophisticated 3D implementation of a tailor-made transport model combined with our reconstruction algorithms as described here, even though being computationally far more expensive, would be directly applicable to the inversion of physically measured data.

<table>
<thead>
<tr>
<th></th>
<th>$\mu$ (cm$^{-1}$)</th>
<th>$b$ (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>0.1</td>
<td>100</td>
</tr>
<tr>
<td>Absorbing inclusions</td>
<td>0.5</td>
<td>100</td>
</tr>
<tr>
<td>Scattering inclusion</td>
<td>0.1</td>
<td>140</td>
</tr>
<tr>
<td>Clear layer</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 1. Optical parameters used in our numerical experiments.
Regarding DOT, we do not know of any reconstruction results published yet from real data that use a full 3D time-dependent transport model for the iterative reconstruction scheme in the presence of clear layers surrounding the brain. Some fundamental research is still necessary in order to reach that point.

The angular variable $\theta$ was discretized into 12 direction vectors in a discrete ordinate fashion that are equidistantly distributed over the unit circle $S^1$. We use 16 point-like sources divided into four sets of four sources, each four sources being distributed equidistantly along one of the four straight boundary parts of $\partial \Omega$. These sources emit photons perpendicularly to the boundary into the domain. Receivers are positioned at each (discrete) boundary point along $\partial \Omega$. They record the outgoing flux given by (7). Notice that measurements in a small vicinity of the given source position are ignored since they are not expected to carry much information regarding the interior of the brain [23]. More precisely, only photons which arrive at the detectors at a spatial distance of more than 1 cm from the source positions and arriving between 2–20 s are considered as part of the data in the following numerical experiments.

In all the numerical simulations presented further below, the data are synthetically generated using the forward modeling scheme as described above, and the reconstructions as well use the same numerical scheme. Therefore, some form of ‘inverse crime’ is inherent in the presented results. A standard technique for dealing with this situation is to create numerical data on a different grid than used for the reconstruction. We partly follow this route and create all our numerical data on a grid which uses a different (twice as fine) temporal grid size than for the reconstruction. Then we add another 3% pointwise Gaussian noise to these synthetically generated data simulated on a mesh with the finer time-discretization. The noise is added as $\sigma^{\text{st}} = \sigma^{k} + \delta |\sigma^{k}| \varphi$, where $\varphi$ is the standard normal random variable and $\delta$ refers to the relative noise level. We do not use a different spatial grid when creating our synthetic data since it will affect strongly the modeling of the clear region which is assumed precisely known in the first set of our numerical experiments. This is important in order to distinguish the impact of different errors in the second set of numerical experiments where we try to evaluate the effect of knowledge of this clear layer on the reconstructions. In that part we will ignore the clear layer completely in the model for the reconstruction stage, even though it is present during the creation of the data. The performance of the three algorithms discussed here under such an erroneous assumption provides interesting insight into their general behavior against modeling errors, but also on the general impact of the presence of clear layers on the data.

During the above derivation of the three reconstruction techniques we chose not to specify in mathematical terms whether or not we know the clear layer \(a \text{ priori}\). Even though a known clear layer could be included in the theoretical derivation by modifying function spaces, we simply impose this additional constraint during the first set of numerical experiments by keeping the parameter values fixed inside the clear layer during the iterations using the correct values. During the second set of numerical experiments, on the contrary, we let initial background values be updated freely at these locations as requested by the algorithms.

Finally we mention that in all numerical experiments a positivity constraint is applied to the absorption and scattering parameters in (1) in order to avoid that they assume non-physical negative values.

8. Numerical results

Notice that for the \textit{cross-sections of the absorption parameter} displayed in the following figures a compact form is chosen for saving space. The left half of the plots shows a
horizontal cross-section at a vertical position 35 over the left half-plane of the domain, and the right half of the plots shows a horizontal cross-section at vertical position 23 over the right half-plane of the domain. This enables us to capture both absorbing inclusions in the same plot.

The cross-sections of the scattering parameter, on the other hand, are arranged as follows. Since only one scattering inclusion is present here, the plots are taken over the entire horizontal line at vertical position 40. As a positive side-effect this plot also allows us to monitor the cross-talk between the reconstruction of the scattering parameter and the reconstruction of the absorption parameter, since one of the absorbers is located as well on this cross-section.

Notice also that in all figures showing the initial guesses and the reconstructions the color scale has been adjusted in order to enhance visibility of the main features. Absolute values can be best seen in the cross sections shown in the plots, or by consulting the exact color scale shown next to each image.

8.1. Reconstruction with known clear layer

Figures 1 and 2 show the results of our first numerical experiment using $L^2$-LK as outlined in algorithm 1. We mention that a similar task as displayed in these two figures has been addressed already in [22], but without incorporating a clear layer in the modeling. Figures 3 and 4 show the results of our second numerical experiment using the newly proposed level set based reconstruction scheme described in algorithm 2. Figures 5 and 6 show the results of our third numerical experiment using the new sparsity promoting regularization scheme as outlined in algorithm 3. More details are given in the caption of each figure.
Figure 2. Simultaneous reconstruction of absorption and scattering coefficients $\mu$ and $b$ using $L^2$-LK as outlined in algorithm 1. Shown is the reconstruction of the scattering cross-section $b$. Top row from left to right: reference and initial scattering profile, evolution of $L^2$-residual norm. Bottom row from left to right: final reconstruction after 450 sweeps of the algorithm, cross-sections through reference and reconstructed scattering profiles.

Figure 3. Simultaneous reconstruction of shapes and contrast values of inclusions in the absorption and scattering coefficients $\mu$ and $b$ using a level set based reconstruction scheme algorithm 2. The top row shows from left to right the reconstruction of the absorption coefficient: initial guess (left), reconstruction after 4000 sweeps (center), reference absorption profile (right). The bottom row shows from left to right the reconstruction of the scattering coefficient: initial guess (left), reconstruction after 4000 sweeps (center), reference scattering profile (right).
8.2. Reconstruction with ‘unknown’ clear layer

As mentioned already before, we have performed a second set of numerical experiments where we do not assume to know the presence of a clear layer. Indeed, the clear layer is treated here as a ‘surprise’ and is not part of the model. The different reconstruction schemes are dealing in a different way with such a ‘surprise’ or model imperfection, as will become visible in the following numerical results. Notice that all performance parameters are kept the same here as in the first set of reconstructions, in order to facilitate a direct comparison. It is understood that, by changing and adjusting some or all of these performance parameters, better or different reconstructions might be achieved in this second set of numerical experiments.

Figures 7 and 8 show the results of our first numerical experiment using algorithm 1 in this new setup, figures 9 and 10 show the results of our second numerical experiment using algorithm 2, and figures 11 and 12 show the results of our third numerical experiment using algorithm 3. More details are given in the caption of each figure.

9. Discussion of numerical results

As mentioned, the first set of numerical experiments addresses the reconstruction of a few inclusions compactly embedded in a constant background with known clear layer.

Figures 1 and 2 show that the $L^2$-LK reconstruction just works as expected and as seen in many other applications. This scheme can be considered a reliable ‘workhorse’ which delivers
Figure 5. Simultaneous reconstruction of absorption and scattering coefficients $\mu$ and $b$ using sparsity promoting regularization as outlined in algorithm 3. Shown is the evolution of the absorption cross-section $\mu$. Top row from left to right: reference and initial absorption profile, evolution of $L^2$-residual. Bottom row from left to right: final reconstruction after 750 sweeps, combined cross-sections through reference and reconstructed absorption profiles.

Figure 6. Simultaneous reconstruction of absorption and scattering coefficients $\mu$ and $b$ using sparsity promoting regularization as outlined in algorithm 3. Shown is the evolution of the scattering cross-section $b$. Top row from left to right: reference and initial scattering profile, evolution of $L^2$-residual. Bottom row from left to right: final reconstruction after 750 sweeps, combined cross-sections through reference and reconstructed scattering profiles.
Figure 7. Simultaneous reconstruction of absorption and scattering coefficients $\mu$ and $b$ using $L^2$-LK as outlined in algorithm 1, but ignoring the presence of the clear layer in the reconstruction. Shown is the reconstruction of the absorption cross-section $\mu$. Top row from left to right: reference and initial absorption profile, evolution of $L^2$-residual norm. Bottom row from left to right: final reconstruction after 300 sweeps of the algorithm, combined cross-sections through reference and reconstructed absorption profiles.

Figure 8. Simultaneous reconstruction of absorption and scattering coefficients $\mu$ and $b$ using $L^2$-LK as outlined in algorithm 1, but ignoring the presence of the clear layer in the reconstruction. Shown is the reconstruction of the scattering cross-section $b$. Top row from left to right: reference and initial scattering profile, evolution of $L^2$-residual norm. Bottom row from left to right: final reconstruction after 300 sweeps of the algorithm, combined cross-sections through reference and reconstructed scattering profiles.
Figure 9. Simultaneous reconstruction of shapes and contrast values of inclusions in the absorption and scattering coefficients $\mu$ and $b$ using a level set based reconstruction scheme algorithm, but ignoring the presence of the clear layer in the reconstruction. The top row shows from left to right the reconstruction of the absorption coefficient: initial guess (left), reconstruction after 1000 sweeps (center), reference absorption profile (right). The bottom row shows from left to right the reconstruction of the scattering coefficient: initial guess (left), reconstruction after 1000 sweeps (center), reference scattering profile (right).

Figure 10. Simultaneous reconstruction of shapes and contrast values of inclusions in the absorption and scattering coefficients $\mu$ and $b$ using a level set based reconstruction scheme algorithm, but ignoring the presence of the clear layer in the reconstruction. Top row from left to right: evolution of absorption contrast value (left), evolution of scattering contrast value (center) and evolution of $L^2$-residual norm (right). Bottom row from left to right: cross section of the absorption parameter (left), cross section of the scattering parameter (right).
good and understandable results, but often over-smoothed and with low contrast when highly ill-posed problems (as this one) are addressed. General features of the inclusions such as their number and general location are recognizable at early iterations, but detailed information regarding specific size and contrast values are usually impossible to extract. A significant cross-talk with the absorption profile is visible in the reconstruction of the scattering coefficient in our general computational setup.

Figures 3 and 4 show that the combined level set approach, as proposed here, performs slightly better in estimating the sizes of the unknown inclusions and their contrast values than the $L^2$-LK scheme. The cross-talk is still visible in the scattering coefficient, but it might be easier in this shape representation to identify and discard absorption inclusions from the scattering reconstruction after comparing with the reconstruction of the absorption parameter. Notice that an increased extension of the reconstructed anomalies is expected when reconstructed contrast values are estimated too low [24]. A similar compensation effect can be achieved by the cross-talk itself, where one optical parameter can make up for an underestimated contrast in the other one. It is also visible that the level set technique, if started with an arbitrary initial guess, might take a large number of iterations in order to converge to the correct profiles, since it needs to ‘move’ the shapes from the initial location to the correct places. By doing so it often has to go over some ‘plateaus’ in the $L^2$-residual cost which makes it hard to establish efficient stopping criteria. An alternative, which has not been followed here, is to use early iterations of the $L^2$-LK scheme for extracting good initial shapes for this level set reconstruction, which should be able to reduce the number of level set

Figure 11. Simultaneous reconstruction of absorption and scattering coefficients $\mu$ and $b$ using Sparsity promoting regularization as outlined in algorithm 3, but ignoring the presence of the clear layer in the reconstruction. Shown is the evolution of the absorption cross-section $\mu$. Top row from left to right: reference and initial absorption profile, evolution of $L^2$-residual. Bottom row from left to right: final reconstruction after 750 sweeps, combined cross-sections through reference and reconstructed absorption profiles.
iterations to a small part of the current ones and avoid the occurrence of these plateaus in the evolution of the cost values [24].

Figures 5 and 6 show that the particular form of sparsity reconstruction as proposed here provides reconstructions which contain features somehow ‘inbetween’ the $L^2$-LK reconstruction and the level set reconstruction. The reconstructed inclusions have more compact shapes compared to the $L^2$-LK reconstructions, but they do not have the sharp interface to the background as in the level set reconstruction. The internal values are closer to the true values in some (mainly central) parts of the reconstructed inclusions here than in the level set based reconstructions, whereas at other parts (often close to the boundaries of the reconstructed inclusions) they are underestimated. Contrast values are significantly better though than in the $L^2$-LK scheme. The cross-talk in the scattering parameter with respect to the absorption profile seems to be further reduced here compared to both other schemes, but it is still visible.

The second set of reconstructions addresses the reconstruction of a few inclusions compactly embedded in a constant background with unknown (‘surprise’) clear layers.

Figures 7 and 8 show that the $L^2$-LK scheme reacts in a stable and reliable way on the changed situation and attempts to reconstruct the features which have the main impact on the data. In this situation these appear to be the missing clear layer. The embedded inclusions seem to become less important now, but reconstructions of parts of the absorption inclusions are still visible. It is seen that the $L^2$-LK scheme does not provide the level of resolution which is needed to correctly model the very thin clear layer in the reconstructions. Instead this layer is reconstructed as a broader band of very low absorption and scattering parameters along the correct regions.
Figures 9 and 10 show that the joint level set reconstruction similarly tries to reconstruct primarily the clear layer in the absorption coefficient. Resolution is much higher here than with the $L^2$-LK scheme. However, the scattering parameter does not perform any meaningful reconstruction at all. It can be argued that its shape seems to start traveling towards some of the inclusions, but does not really reach them. One disadvantage of the level set method here is that by construction it only can represent either the clear layers or the inclusions, but not both, since the values inside these different shapes are opposite to each other as seen from the background values. So, the algorithm needs to ‘make a choice’. We anticipate here that a generalized level set approach which dedicates one level set function to the clear layer and another different one to the embedded inclusions, in both the absorption and scattering coefficients, should perform much better. In the current setup, however, the clear layer seems to dominate the data and therefore is reconstructed quite precisely (compared to the other schemes) by the level set evolution in the absorption value.

Figures 11 and 12 show that our particular approach of sparsity promoting reconstruction seems to be well prepared to handle this new situation. A mixture is seen in the reconstructions of the absorption parameter between clear layer estimation and reconstruction of the inclusions. The reconstruction of the scattering parameter is dominated by the clear region.

10. Summary and conclusions

We have introduced two different reconstruction techniques for solving the inverse problem of a linear transport equation in 2D with applications in Diffuse Optical Tomography, and compared the performance of both schemes with each other and also with a third more traditional scheme. This third scheme follows mainly a strategy previously discussed in [22] and can nowadays be considered as part of a family of standard $L^2$-based regularization schemes for this application. The first new scheme proposed here can be considered being a generalization of a level set based scheme proposed in [24] for a much simpler situation. The second new regularization scheme follows closely the recently published work of [32, 42] which was considering a different imaging modality (EIT) and has been extended here to this new application. Using numerical experiments with simulated noisy data we have shown that all three strategies are able to invert these data simultaneously for an absorption and a scattering parameter profile, but with some significant differences in speed and accuracy when considering inclusions of well-defined compact support. Each of these schemes shows some advantages in certain situations, which might however become disadvantages in others. The high flexibility of the $L^2$-LK scheme comes with a lack of resolution and low contrast values in the reconstructions. The high precision of the level set technique in describing boundaries of shapes comes with a reduced flexibility when more than one type of inclusions is present. The sparsity promoting scheme performs quite well in our specific setup here, but it needs to be admitted that its underlying assumptions are well-suited to model our compactly embedded shapes. In situations where the unknown perturbations are more spread out it might not perform that well.

We also want to emphasize again that all three schemes described here do show great potential to be generalized and adjusted to better address specific situations which our basic setup might be unable to match well. Therefore, as mentioned in our introduction, the comparisons shown here need to be taken cum grano salis. Indeed, we believe that combining the various schemes to obtain new hybrid schemes might be a powerful route for dealing with a range of challenging inversion problems.
The numerical experiments presented here indicate some interesting routes for future research. Sparsity promoting regularization is a promising approach and still poses challenges for nonlinear inverse problems in particular when more than one parameters need to be reconstructed from the same data set. The same holds true for joint level set schemes. The avoidance of cross-talk is an important issue not only in DOT but in many other practically relevant multi-parameter imaging applications. Furthermore, the joint reconstruction of clear layers and unknown embedded inclusions in DOT still is a fairly open research topic. Needless to say that extensions to 3D are necessary when applying these concepts to more realistic setups, and eventually to physically obtained real data. This also entails more research on efficient forward modeling schemes for time-dependent linear transport models for situations in DOT. Extensions of the transport model to Fluorescence Tomography, Bioluminescence and Photoacoustic Tomography are also interesting routes to follow [29, 48, 58, 67].

Acknowledgments

The work of KP was funded by the Mexican Consejo Nacional de Ciencia y Tecnología (CONACYT) under the scholarship 213551. OD acknowledges support for this work from the RCUK Academic Fellowship scheme and from the EPSRC funded platform grant MAPLE EP/I01912X/1.

References

Inverse Problems 33 (2017) 014001
K Prieto and O Dorn


[31] Foucart S and Rauhut H 2013 A mathematical introduction to compressive sensing Applied and Numerical Harmonic Analysis (Birkhauser)


[41] Ito K, Kunisch K and Li Z 2001 Level-set function approach to an inverse interface problem Inverse Problems 17 1225–42


[50] Kluth T 2011 3d electrical impedance tomography with sparsity constraints: algorithm and implementation in application to the complete electrode model Master’s Thesis University of Bremen
[53] Leitão A and Scherzer O 2003 On the relation between constraint regularization, level sets, and shape optimization Inverse Problems 19 L1–1
[66] Santosa F 1996 A level-set approach for inverse problems involving obstacles ESAIM 1 17–33
[67] Saratono T, Tarvainen T, Cox B T and Arridge S R 2013 A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation Inverse Problems 29 075006
[70] Tang J, Han W and Han B 2013 A theoretical study for RTE-based parameter identification problems Inverse Problems 29 095002


[73] van den Doel K, Ascher U and Haber E 2013 The lost honour of l2-based regularization *Large Scale Inverse Problems* (Berlin: De Gruyter) pp 181–203