# Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography

**Andreas H. Hielscher**
**Sebastian Bartel**
State University of New York,
Downstate Medical Center,
Department of Pathology,
450 Clarkson Avenue, Box 25,
Brooklyn, New York 11203

**Abstract.** It is well known that the reconstruction problem in optical tomography is ill-posed. In other words, many different spatial distributions of optical properties inside the medium can lead to the same detector readings on the surface of the medium under consideration. Therefore, the choice of an appropriate method to overcome this problem is of crucial importance for any successful optical tomographic image reconstruction algorithm. In this work we approach the problem within a gradient-based iterative image reconstruction scheme. The image reconstruction is considered to be a minimization of an appropriately defined objective function. The objective function can be separated into a least-square-error term, which compares predicted and actual detector readings, and additional penalty terms that may contain *a priori* information about the system. For the efficient minimization of this objective function the gradient with respect to the spatial distribution of optical properties is calculated. Besides presenting the underlying concepts in our approach to overcome ill-posedness in optical tomography, we will show numerical results that demonstrate how prior knowledge, represented as penalty terms, can improve the reconstruction results. © *2001 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.1352753]

Keywords: optical tomography; image reconstruction; regularization; inverse problem.

Paper 42022 received Dec. 10, 1999; revised manuscript received Aug. 18, 2000; accepted for publication Dec. 12, 2000.

## 1 Introduction

Optical tomography (OT) is a fast growing field in which near-infrared light is used to image the distribution of optical properties inside the human body. Optical properties of interest are, for example, the absorption coefficient $\mu_a$, the reduced scattering coefficient $\mu_s'$, or the diffusion coefficient $D = c/(3\mu_a + 3\mu_s')$, where $c$ is the speed of light in the medium. The instrumentation for making light transmission measurements that are necessary for OT is nowadays widely available.[1–6] Furthermore, several algorithms that transform these measurements into useful cross-sectional images have matured to such a degree that first clinical trials are underway, especially in breast imaging.[7–30] However, a major difficulty in OT remains that the image reconstruction problem is ill-posed or underdetermined. In other words, there are many distributions of optical properties inside the medium under investigation that lead to the same set of detector readings on the surface of the medium.

Depending on the underlying structure and concepts of any particular reconstruction scheme, the problem of ill-posedness can be approached in different ways. Most of the currently employed reconstruction schemes fall in one of two classes, which we refer to as the linear-perturbation approach and the nonlinear-gradient method (see also Ref. 30). In both cases the goal is to minimize the difference between intensities measured on the boundary of the medium and some prediction for those measured values. Therefore the image reconstruction problem may be interpreted as an optimization problem in which an objective function is minimized. If we use the $\chi^2$ error norm to calculate the difference between measurements and predictions we can define the objective function as

$$\Phi(\boldsymbol{\zeta}) = \chi^2(\boldsymbol{\zeta}) \equiv \sum_s \sum_d \frac{(M_{s,d} - P_{s,d}(\boldsymbol{\zeta}))^2}{2\sigma_{s,d}^2}. \tag{1}$$

In this equation the parameter $\boldsymbol{\zeta}$ is a vector that contains the optical properties at all positions in the medium. If the image is discretized into $n$ pixels, and each pixel can vary in both absorption and scattering coefficient, the vector $\boldsymbol{\zeta}$ is of length $N = 2n$. $P_{s,d}(\boldsymbol{\zeta})$ is the predicted reading at detector location $d$, when light is injected at location $s$. The $P_{s,d}(\boldsymbol{\zeta})$ values are calculated with a forward model, e.g., by solving the diffusion equation for the given medium.[27] $M_{s,d}$ is the measured value at detector position $d$ given a source at $s$. The parameter $\sigma_{s,d}$ is a normalization constant.

In the linear perturbation approach, it is assumed that we have an estimate $\boldsymbol{\zeta}_0$ that is close to the true distribution $\boldsymbol{\zeta}$. In this case we can perform a Taylor expansion of $P_{s,d}$ around $\boldsymbol{\zeta}_0$. Neglecting nonlinear terms, we obtain

$$P_{s,d}(\boldsymbol{\zeta}) = P_{s,d}(\boldsymbol{\zeta}_0) + \partial P_{s,d}(\boldsymbol{\zeta})/\partial \boldsymbol{\zeta} \cdot (\boldsymbol{\zeta} - \boldsymbol{\zeta}_0)$$

$$= P_{s,d}(\boldsymbol{\zeta}_0) + \mathbf{J}_{s,d}\Delta\boldsymbol{\zeta}, \tag{2}$$

Andreas H. Hielscher is also at Department of Electrical and Computer Engineering, Polytechnic University, 5 MetroTech Center, Brooklyn NY 11201. Address all correspondence to Andreas H. Hielscher. E-mail: ahielscher@downstate.edu

where $\mathbf{J}_{s,d} = \partial P_{s,d}(\zeta)/\partial\zeta$ is a row in the Jacobian matrix $\mathbf{J}$, which is often also referred to as the weight-function matrix. If $Q$ is the number of source–detector pairs $(s,d)$ and $N$ the number of unknowns in the problem, then $\mathbf{J}$ is a $Q \times N$ matrix. The vector $\Delta\zeta = \zeta - \zeta_0$ is the difference in optical properties between the estimated and actual medium given by $\zeta$. Inserting Eq. (2) into Eq. (1) yields:

$$\Phi(\zeta) = \sum_s \sum_d \frac{(M_{s,d} - P_{s,d}(\zeta_0) - \mathbf{J}_{s,d}\Delta\zeta)^2}{2\sigma_{s,d}^2}. \quad (3)$$

Therefore, in this case, minimizing the functional in Eq. (1) is equivalent to solving the equation

$$\mathbf{J}\Delta\zeta = \mathbf{M} - \mathbf{P} \Leftrightarrow \mathbf{J}\mathbf{J}^T\Delta\zeta = \mathbf{J}^T(\mathbf{M} - \mathbf{P}), \quad (4)$$

where $\mathbf{M}$ and $\mathbf{P}$ are vectors that contain all $M_{s,d}$ and $P_{s,d}(\zeta_0)$ values, respectively. The image reconstruction problem is to find $\Delta\zeta$ and to determine the image given by $(\Delta\zeta + \zeta)$. The matrix $\mathbf{J}^T$ is the transpose of $\mathbf{J}$. Approaching the imaging problem this way, ill-posedness means that the quadratic matrix $\mathbf{J}\mathbf{J}^T$ is ill conditioned. Therefore, the determinant of the matrix $\mathbf{J}\mathbf{J}^T$ is almost zero and many different $\Delta\zeta$s solve Eq. (4). A standard way of overcoming this problem is to make $\mathbf{J}\mathbf{J}^T$ diagonally dominant.[17,30,31] Obviously, this is most easily accomplished by adding a diagonal matrix and solved

$$(\mathbf{J}\mathbf{J}^T + \lambda\mathbf{I})\Delta\zeta = \mathbf{J}^T(\mathbf{M} - \mathbf{P}), \quad (5)$$

where $\mathbf{I}$ is the identity matrix and $\lambda$ is usually referred to as the regularization parameter or hyperparameter. The goal is now to find a $\lambda$ that is large enough to avoid problems encountered with ill-conditioned matrixes, and at the same time small enough to not completely alter the basic relation defined in Eq. (1). For example, Jiang et al.,[18] Paulsen and Jiang,[17] and Pogue et al.[32] have derived such diagonal matrices for optical tomography.

More recently several groups (Saquib et al.,[24] Hielscher et al.,[25,27,28] Arridge and Schweiger,[26] Roy and Sevick,[29] and Ye et al.[33]) have developed so called gradient based iterative reconstruction (GIIR) schemes that do not solve Eqs. (4) or (5) to obtain an update of $\zeta$. Instead, starting from an initial guess $\zeta_0$ subsequent distributions $\zeta_{k+1}$ are obtained by calculating

$$\zeta_{k+1} = \zeta_k + \alpha\mathbf{A}g, \quad (6)$$

where $g = \partial\Phi(\zeta)/\partial\zeta$ is the gradient of the objective function [Eq. (1)] in column vector form of length $N$, $\mathbf{A}$ is an $N \times N$ matrix, and $\alpha$ is a real number representing the step size in the direction of the gradient. For the case of steepest gradient descent $\mathbf{A}$ equals the identity matrix, with only ones on the diagonal. The major advantage of GIIR algorithms over other currently employed algorithms is that no inversion of a full, ill-conditioned Jacobian matrix is necessary to obtain an update $\Delta\zeta$ of the optical properties in the medium. In GIIR schemes a Jacobian is calculated as part of the gradient calculation of the objective function. Once this gradient is found, a line minimization of the objective function along the direction of the gradient is performed to find the update for the optical properties. A more detailed description and discussion of GIIR algorithms can be found elsewhere.[24–30]

The question arises: what does ill-posedness mean using the GIIR approach? In the perturbation approach ill-posedness is identical to an ill-conditioned matrix $\mathbf{J}\mathbf{J}^T$. In the GIIR scheme, which interprets image reconstruction as a multidimensional minimization problem, ill-posedness means that a global minimum is not well defined. A global minimum is not well defined when either many minima with similar values exist or one minimum $\zeta_{\min}$, which is surrounded by many other $\zeta$, result in almost identical objective functions. Examples of such objective functions in optical tomography have been described, for example, by Arridge,[30] Schweiger and Arridge[34] and Hielscher et al.[35] In general, the reconstruction result will strongly depend on the initial guess $\zeta_0$. These phenomena are well known in the field of general optimization theory. A typical way to overcome this problem is to add penalty terms that provide additional constraints on the solution space. The penalty term may push the solution into the right area of the solution space and the minimization of the resulting objective function may provide a better-defined minimum. While the use of penalty terms has been studied for a variety of problems,[36–44] their use in GIIR schemes for optical tomography has not been explored.

The goal of this paper is to introduce and study the effects of various penalty functions on the reconstruction process in OT. In particular, we seek to derive penalty function from *a priori* knowledge about the systems under investigation. *A priori* knowledge is, in general, information that does not depend on the difference between predicted and measured data. For example, we know that optical properties are larger than zero. Furthermore in the near-infrared wavelength region $\mu_a < 10 \text{ cm}^{-1}$ and $\mu_s' < 100 \text{ cm}^{-1}$. In addition, it is often known how many different tissue types, and therefore how many different optical properties, are present in the interrogated medium. For example, in the brain we find cerebrospinal fluid, gray, and white matter. If magnetic resonance imaging data are available, one may even know the location of these tissues. The questions arise: how can this additional *a priori* knowledge be appropriately cast into a penalty function and can using this information improve image quality?

To illustrate the effects of penalty functions derived from *a priori* knowledge about the system, we will focus in this work on information regarding the composition of a given tissue volume. We consider the following cases: First, it is assumed that one knows that $n$ tissue types with $n$ different optical properties are present in the medium. However one neither knows the location nor their respective volume fraction of any given component in the tissue sample. Second, we will assume to have prior knowledge of all tissue types present and the volume percentage they occupy. However, just as in the first case, one does not know where the different tissue types are located inside the medium. This latter point corresponds to knowing the histogram of the medium to be reconstructed. We will derive the appropriate penalty terms for both cases. Numerical examples that demonstrate the effect of these penalty terms on the quality of the reconstructed images are shown and discussed. Furthermore, the relationship between penalty functions in gradient-based schemes and algebraic regularization mechanisms used in linear-perturbation schemes is discussed and put into context with our results.

## 2 Mathematical Background

### 2.1 Gradient-Based Iterative Image Reconstruction

Our gradient-based iterative image reconstruction scheme has three major components. **(1)** *Forward Model.* This model is a theory or algorithm that predicts a set of measured signals **U** given the positions $r_s$ of the light sources and the spatial distribution of optical properties $\zeta$. In this work we use as the governing equation for light propagation in tissue the time-dependent diffusion equation $\partial U/\partial t = \nabla(D\nabla U) - c\mu_a U + S$.[27] As optical parameters of interest we chose $\zeta = (c\mu_a(\mathbf{r}), D(\mathbf{r}))$. **(2)** *Analysis Scheme.* Here an objective function $\Phi$ is defined, which describes the difference between the measured **M** and predicted data **P**. An example is the least-square error norm, also called $\chi^2$ norm, described in Eq. (1). **(3)** *Updating Scheme.* Once the objective function is defined, the task becomes to minimize $\Phi$. This is accomplished in two substeps. First the gradient of the objective function $d\Phi(\zeta)/d\zeta$ is calculated by means of adjoint differentiation. Second, given the gradient an iterative line minimization in the direction of the gradient is performed. This step is refereed to as *inner iteration* and consists of several forward calculations in which the parameters $\zeta$ are varied. Once the minimum along the line is found, a new gradient is calculated at this minimum (*outer iteration*) and another line minimization is performed, now along a different direction in the $\zeta$ space. These steps are repeated until a distribution $\zeta$ is found for which $\Phi(\zeta)$ is smallest. A more detailed description of the GIIR algorithm used in this work can be found elsewhere.[27]

### 2.2 Penalty Functions

The incorporation of additional knowledge about the image to be reconstructed can be achieved by including a penalty term into the definition of the objective function. Instead of only trying to minimize the difference between the measurements and predictions, as is done if the objective function equals the $\chi^2$ error norm defined in Eq. (1), one can define the objective function as

$$\Phi(\zeta) = \chi^2(\zeta) + \omega \cdot \Pi(\zeta), \qquad (7)$$

where $\Pi(\zeta)$ denotes the penalty term and $\omega$ is the coupling parameter, also often referred to as the hyperparameter. The additional term $\omega\Pi(\zeta)$ is designed to provide a better-defined minimum in the solution space and push the gradient scheme toward more probable solutions. The latter is achieved by penalizing certain distributions $\zeta$ that are unlikely given our prior information of the system under investigation. Ideally, we expect $\Pi(\zeta)$ to constrain the $N$-dimensional minimum search to a small subspace of desirable solutions.

The penalty term should be continuously differentiable to be useful in GIIR schemes. Therefore the derivative of the penalty term with respect to the optical properties should not have any discontinuities. In general, smooth functions are much better suited for any gradient based minimum search. We will address these restrictions further when we derive the respective penalty functions.

### 2.2.1 *Tissue-Type Penalty Functions*

In this work we seek to provide penalty functions that are rooted in a particular kind of prior knowledge, namely the
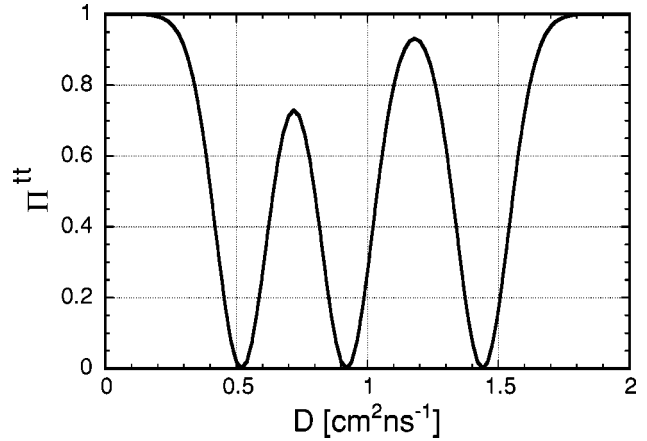


**Fig. 1** Example of a tissue-type penalty function $\Pi^{tt}$ [Eq. (8)] that assumes three different types of tissues ($k=3$), with $D_1=0.52$ cm$^2$ ns$^{-1}$, $D_2=0.92$ cm$^2$ ns$^{-1}$, and $D_3=1.44$ cm$^2$ ns$^{-1}$.

different types of tissues that are most likely to be found in the sample under investigation. The optical properties of these tissues are known within certain error margins. Areas of the reconstructed image that show any other than the expected optical parameters should therefore be penalized the more they differ from the expected values. Under these considerations we define the following penalty function:

$$\Pi^{tt}(\zeta) = \sum_{x \in S} \left( 1 - \sum_{k=1}^{K} \exp\left( -\frac{(a_k - \zeta_x)^2}{2\sigma_k^2} \right) \right). \qquad (8)$$

Here $K$ is the number of different tissue types $k$ in the system $S$. The parameter $a_k$ is the most likely optical property of tissue $k$, and $\zeta_x$ is the reconstructed optical property at pixel $x$. The parameter $1/\sigma_k$ can be interpreted as the confidence that $a_k$ is the exact value. For $1/\sigma_k \to \infty$ ($\Leftrightarrow \sigma_k \to 0$) the Gaussian becomes a delta function that is 1 only if $\zeta_x = a_k$ and zero otherwise. Therefore, only if $\zeta_x = a_k$ does the penalty term disappear. In this case Eq. (8) is not continuously differentiable. For any nonzero value of $\sigma_k$ there will be a range of $\zeta_x$ for which the penalty term changes smoothly from 1 to 0 and Eq. (8) becomes continuously differentiable. An example is shown in Figure 1, where $\Pi^{tt}$ is plotted for the case that three different tissue types with optical properties $D_1$, $D_2$, and $D_3$ are present and $\sigma_k = 0.1$ cm$^2$ ns$^{-1}$ for all three $D$ values. The derivative of this penalty function with respect to a pixel value $\zeta_x$ is easily calculated as

$$\frac{\partial \Pi^{tt}(\zeta)}{\partial \zeta_x} = -\sum_{k=1}^{K} \frac{a_k - \zeta_x}{\sigma_k^2} \exp\left( -\frac{(a_k - \zeta_x)^2}{2\sigma_k^2} \right). \qquad (9)$$

### 2.2.2 *Histogram Penalty Function*

If not only the various types of tissue in the sample are known but additionally their respective volume fraction, we are able to provide a most likely histogram $H^0$ to the reconstruction

scheme. In general, a histogram maps the $N$ pixel values of an image represented by $\zeta=(\zeta_1,\zeta_2,\ldots,\zeta_x,\ldots,\zeta_N)$ onto $L$ discrete intervals or bins of width $\Delta\zeta_l=\zeta_l-\zeta_{l-1}$ ($l=1,2,\ldots,L$). Therefore, the histogram $H(\zeta_l,\zeta)$ associated with an image $S$ or given set $\zeta$ of optical properties is defined by the following sum over all pixels $x$ of the image

$$H(\zeta_l,\zeta)\equiv\sum_{x\in S}\delta(\zeta_l,\zeta_x),\quad l=1,\ldots,L, \qquad (10)$$

where

$$\delta(\zeta_l,\zeta_x)\equiv\begin{cases}1 & \text{for } \zeta_{l-1}<\zeta_x\leqslant\zeta_l \\ 0 & \text{otherwise}\end{cases}. \qquad (11)$$

The sum in Eq. (10) sorts all pixel values $\zeta_x$ into $L$ bins of the histogram, since the $\delta$-function only contributes to the sum $H(\zeta_l,\zeta)$ if a pixel lies within the corresponding interval.

We can now define a penalty term $\Pi^{\text{Hist}}$, which evaluates the histogram $H(\zeta_l,\zeta)$ of a reconstructed image relative to the expected histogram $H^0(\zeta_l,\zeta)$. Here we choose the $\chi^2$ error norm between both functions and define the penalty term

$$\Pi^{\text{hist}}\equiv\chi^2_{H^0,H}=\sum_{l=1}^{L}\frac{(H^0(\zeta_l)-H(\zeta_l))^2}{H^0(\zeta_l)}, \qquad (12)$$

where we use the short notation $H(\zeta_l)$ for $H(\zeta_l,\zeta)$.

However, if the definition of the delta function in Eq. (11) is considered, we observe that Eq. (12) is not continuously differentiable, which makes it unsuitable for gradient based schemes. To overcome this problem we consider a representation of the delta function that uses a Gaussian

$$\delta(\zeta_l,\zeta_x)=\lim_{\sigma\to0}\sqrt{1/(2\pi\sigma^2)}\exp\left(-\frac{(\zeta_l-\zeta_x)^2}{2\sigma^2}\right). \quad (13)$$

With this we can calculate the derivative and obtain

$$\frac{\partial\Pi^{\text{hist}}}{\partial\zeta_x}=-\sum_{l=1}^{L}2\sqrt{1/(2\pi\sigma^2)}\frac{(H^0(\zeta_l)-H(\zeta_l))}{H^0(\zeta_l)}\frac{(\zeta_l-\zeta_x)}{\sigma^2}$$

$$\times\exp\left(-\frac{(\zeta_l-\zeta_x)^2}{2\sigma^2}\right), \qquad (14)$$

where we omitted the lim. By choosing an appropriate nonzero value for $\sigma$ and calculating the image histogram using Eq. (10) and the penalty term Eq. (12), we obtain a penalty term that is continuously differentiable and can be applied within the GIIR framework. Typically, we chose $\sigma$, so that the singular peaks overlapped (see Figure 2) to avoid *forbidden* intervals in the target histogram $H^0$.

Using Eq. (13) with a nonzero $\sigma$ to generate $H(\zeta_l)$ results in a smoothing operation on the histogram and is equivalent to convolving the exact histogram with a Gaussian of width $\sigma$. It may seem that the histogram thus obtained is blurred to an extent that spoils the goal of reconstructing only certain types of tissue. However, since we perform the same convolution on both, the histogram $H$ of the reconstructed image and the target histogram $H^0$ before they are compared [Eq. (12)] becomes minimal, if and only if the unconvolved functions are
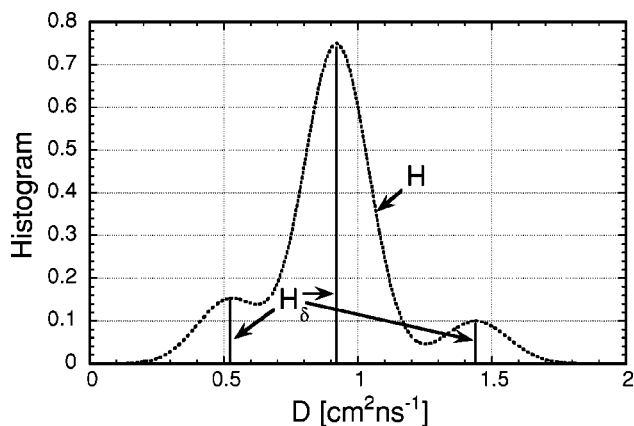


**Fig. 2** Ideal histogram $H_\delta$ for tissue that contains 15% tissue type 1 ($D_1=0.520$ cm$^2$ ns$^{-1}$), 75% tissue type 2 ($D_2=0.92$ cm$^2$ ns$^{-1}$), and 10% tissue type 3 ($D_3=1.44$ cm$^2$ ns$^{-1}$) and smoothed histogram $H$. The smoothed histogram results from convolution of $H_\delta$ with a Gaussian function of width $\sigma=0.15$ [Eq. (10) and Eq. (13)].

identical. Hence, let $(H^0_\delta,H_\delta)$ be accurate histograms [Eq. (10)] and $(H^0,H)$ be their convolution with Eq. (13) (with nonzero $\sigma$), then

$$H^0_\delta=H_\delta \;\Leftrightarrow\; H^0=H. \qquad (15)$$

Consequently, by minimizing the $\chi^2$ norm between two blurred histograms, we are at the same time minimizing the difference between the underlying sharp histograms. However, using the blurred histogram provides us with a continuously differentiable penalty term that is better suited for use in gradient based minimization schemes.

### 2.2.3 *The Hyperparameter*

Each of the penalty functions has to be coupled to the error norm $\chi^2$ [Eq. (1)] with a hyperparameter. This parameter fixes the relative strength of the penalty term in the minimization scheme and describes the confidence one has that the additional information is correct. A good starting point is to choose $\omega$ in a way which ensures that the gradients of both the $\chi^2$ term and the penalty function are of similar magnitude. If we interpret the $\chi^2$ term and the $\omega\Pi$ term as potentials, the derivatives $\partial\chi^2/\partial\zeta$ and $\omega\partial\Pi/\partial\zeta$ can be interpreted as two forces that pull the pixel values in certain directions. By defining a new hyperparameter

$$\gamma=\left|\frac{\omega\partial\Pi}{\partial\zeta}\right|\bigg/\left|\frac{\partial\chi^2}{\partial\zeta}\right|, \qquad (16)$$

we can adjust the relative strength of these forces, originating from the $\chi^2$ potential and the $\Pi$ potential, respectively. If $\gamma=1$ then $\omega$ equals the ratio of the gradients and both forces are equally strong. If $\gamma>1$ the force due to penalty function $\omega\Pi$ is stronger than due to the $\chi^2$ term, and if $\gamma<1$ the $\chi^2$ term has a stronger influence. We will later see how the choice of $\gamma$ influences the reconstruction results.
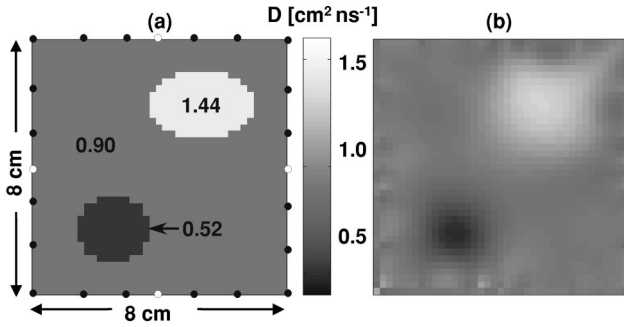
(a)    D [cm² ns⁻¹]    (b)

1.5

1.44

0.90

1.0

8 cm

0.5

← 0.52

8 cm

**Fig. 3** (a) Composition of two-dimensional example problem. An $8 \times 8$ cm domain is divided into a $40 \times 40$ grid. Source positions are indicated with white circles and detectors with black circles. (b) Reconstruction without penalty functions after 20 iterations. The initial guess for the first iteration is a homogeneous medium with $D=1$ cm$^{-2}$ ns$^{-1}$.

## 3  Results

### 3.1  Problem Setup

To test how assumptions about the medium can improve the image reconstruction we consider the following example. Given is an $8 \times 8$ cm medium that contains two objects (see Figure 3). The speed of light in this medium is considered to be constant $c = 22$ cm ns$^{-1}$. The optical properties of the background medium are given by $\mu_a = 0.1$ cm$^{-1}$, $\mu'_S = 8$ cm$^{-1}$ ($\Rightarrow D = c/(3\mu_a + 3\mu'_s) = 0.905$ cm$^2$ ns$^{-1}$). The optical properties of the inclusions are given by $\mu_a = 0.1$ cm$^{-1}$, $\mu'_S = 14.0$ cm$^{-1}$ ($\Rightarrow D = 0.520$ cm$^2$ ns$^{-1}$), and $\mu_a = 0.1$ cm$^{-1}$, $\mu'_S = 5.0$ cm$^{-1}$ ($\Rightarrow D = 1.438$ cm$^2$ ns$^{-1}$), respectively. The medium is discretized into a $40 \times 40$ mesh with $\Delta x = 0.2$ cm. Four sources, one centered at each side, surround the medium. For each source 20 detector readings are available, four on each side and one on each corner. These detector readings were simulated by using the time-dependent finite-difference forward model and adding Gaussian noise with a signal to noise ratio of 30 db to the result. The simulated detector readings each consists of 100 time-dependent fluence rates, with $\Delta t = 0.01$ ns. Therefore we have $4 \times 20 \times 100 = 8000$ data points. From these points the $\chi^2$ term is calculated as

$$\chi^2 = \sum_s \sum_d \sum_t \left( \frac{M^t_{s,d} - P^t_{s,d}(D,\mu_a)}{P^t_{s,d}(D,\mu_a)} \right)^2, \quad (17)$$

where $D$ and $\mu_a$ are vectors of length $40 \times 40 = 1600$, which contain the diffusion and absorption coefficients throughout the medium.

To quantify the quality of the image reconstruction, we furthermore define the relative image error as

$$I_E \equiv \frac{1}{N} \sum_n^N \sqrt{\frac{(D_n^{\text{target}} - D_n^{\text{wp}})^2}{(D_n^{\text{target}} - D_n^{\text{wop}})^2}}, \quad (18)$$

where $N$ is the number of pixels in the image, $D_n^{\text{target}}$ is the correct diffusion coefficient at pixel position $n$, $D_n^{\text{wop}}$ is the diffusion coefficient at pixel position $n$ reconstructed without using a penalty term, and $D_n^{\text{wp}}$ is the diffusion coefficient at pixel position $n$ reconstructed using a penalty term. Defining
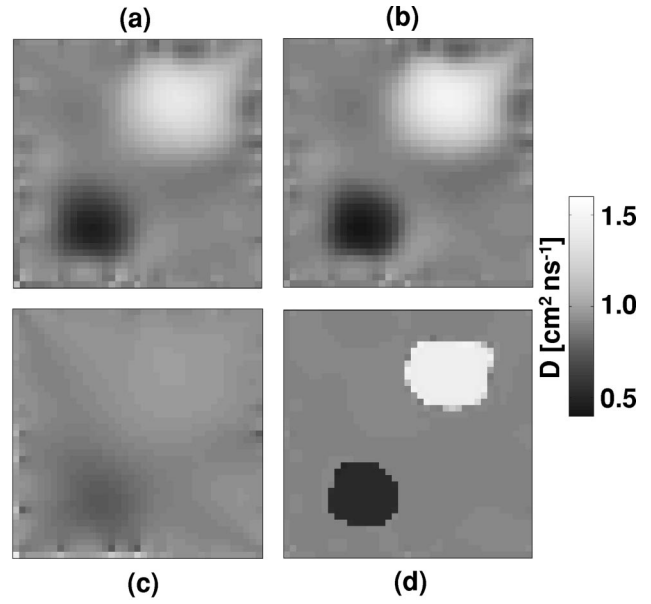


**Fig. 4** Image reconstruction with tissue-type penalty terms [Eq. (8)] after 20 iterations. The initial guess is a homogeneous medium with $D=1$ cm$^{-1}$: (a) $\gamma = 0.02$, (b) $\gamma = 0.08$, (c) $\gamma = 0.1$, and (d) $\gamma = 0.005$ $n^2$, where $n$ equals the number of iterations.

the image error this way allows us to quantify the improvement in image quality when using a penalty function as compared to not using a penalty function. A value of $I_E = 1$ indicates that the penalty term yields neither an improvement nor degradation in image quality.

### 3.2  Reconstruction Without Penalty Terms

We first use the GIIR algorithm to minimize the $\chi^2$ term, without any additional penalty terms. As the initial guess we chose $D = 1$ cm$^2$ ns$^{-1}$ and $\mu_a = 0.1$ cm$^{-1}$ for all points in the medium. Therefore, we only reconstruct the diffusion coefficient $D$ in this case, since the initial guess for $\mu_a$ equals the original value. The result of the reconstruction after 20 iterations is shown in Figure 3(b). The general features of the medium are recovered. The diffusion coefficient $D$ in the larger inclusion is increased while $D$ in the smaller inclusion is decreased. The minimal and maximal values of $D$ in the heterogeneities are $D = 0.50$ cm$^2$ ns$^{-1}$ and $D = 1.31$ cm$^2$ ns$^{-1}$, respectively. These values differ by approximately 10% from the original value. The sharp edges are not recovered and the inclusions appear blurred, as is typical for optical tomography. We found that as long as the initial guesses are within 20% of the average $D$ value of the image, the reconstruction results are very similar. When initial $D$ guesses are chosen that are outside the range of $D$ values present in the image strong artifacts start to appear.

### 3.3  Reconstruction with Tissue-Type Penalty Term

Next we consider how these results can be improved when the information is used that only three different tissues types with given optical properties are present in the medium [Eq. (8)]. However, no knowledge is assumed about where these three different regions are located, or how much of the image volume is occupied by a particular tissue type. Figures 4(a)–4(c)
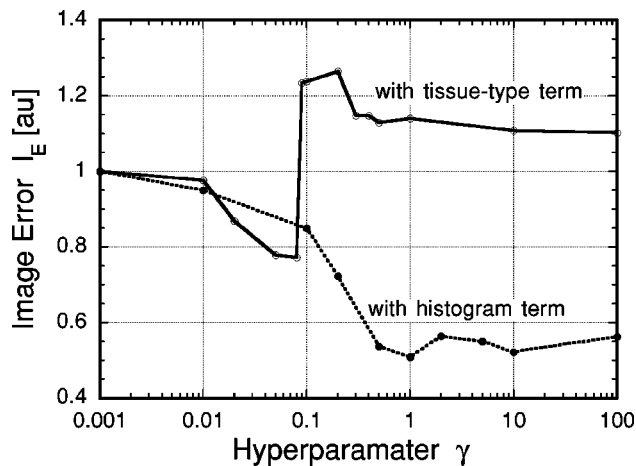
Fig. 5 Image error $I_E$ as function of the hyperparameter $\gamma$ for the tissue type prior $\Pi^{tt}$ (open circles) and histogram prior $\Pi^{hist}$ (filled circles). Both curves are normalized to 1 for the result of the reconstruction without penalty term ($\gamma=0$).
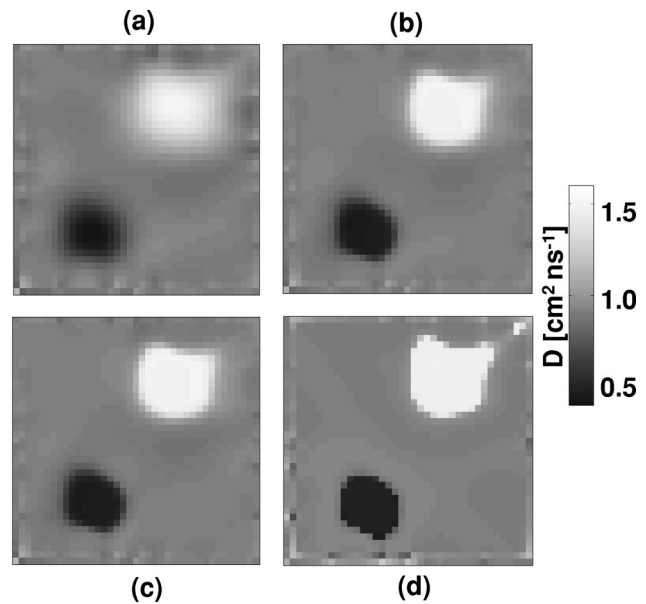


Fig. 6 Image reconstruction with histogram penalty terms [Eq. (12)] after 20 iterations. The initial guess is a homogeneous medium with $D=1$ cm$^{-1}$:(a) $\gamma=0.2$, (b) $\gamma=0.5$, (c) $\gamma=1.0$, and (d) $\gamma=10^6$.

show reconstructions obtained with the penalty term $\Pi^{tt}$ for three different parameters $\gamma$. In this example we chose $D_1 = 0.520$ cm$^2$ ns$^{-1}$, $D_2=0.92$ cm$^2$ ns$^{-1}$, and $D_3=1.44$ cm$^2$ ns$^{-1}$. Furthermore, we set the width of the Gaussians to $\sigma_k=0.1$ cm$^2$ ns$^{-1}$ [Eq. (8)]. The initial guess is a homogeneous medium with $D=1$ cm$^2$ ns$^{-1}$ and $\mu_a=0.1$ cm$^{-1}$. Reconstructions were terminated after 20 iterations. It can be seen that the penalty term has an effect on the image for $\gamma =0.02$ [Figure 4(a)]. Increasing $\gamma$ to 0.08 does produce better results [Figure 4(b)], however, a further increase to $\gamma=0.1$ yields reconstructions with almost constant $D$ values across the entire image [Figure 4(c)]. All pixel values are close to the background value $D=0.92$ cm$^2$ ns$^{-1}$. Figure 5 shows the dependence of the image error $I_E$ on the parameter $\gamma$. It can be seen that only for a small region of $0.01<\gamma<0.1$ does the image quality improve ($I_E<1$). For all other values $\gamma>0.1$ the image error is larger than without the penalty term ($I_E >1$).

Rather than keeping $\gamma$ fixed for all iterations, one can also change this parameter dynamically. Figure 4(d) shows the result of increasing $\gamma$ with the number of iterations $n$, so that $\gamma=0.005 \, n^2$. After 25 iterations both the location and $D$ values of the inclusions are correctly recovered. Deviations between the reconstruction and the original image are on the order of single pixels, and the image error is relatively small ($I_E=0.18$).

### 3.4 Reconstruction with Histogram Prior

Next, we consider the case where we have knowledge about the different tissue types in the medium and in addition know their respective volume fractions. Therefore, we assume to have knowledge about the histogram of the cross-sectional image of the tissue sample. Figures 6(a)–6(d) show reconstructed images using the histogram prior with increasing hyperparameter $\gamma$. As in previous cases, the initial guess is a homogeneous medium with $D=1$ cm$^2$ ns$^{-1}$ and $\mu_a=0.1$ cm$^{-1}$. Choosing $\gamma<0.1$ does not result in a significant improvement of the reconstruction and yields images similar to the unbiased case $\gamma=0$ [Figure 3(b)]. For $\gamma\geqslant0.2$ the two

objects become more localized and show plateaus of constant $D$ values. For $\gamma=0.5$ we observe increasingly sharp boundaries separating the heterogeneities from the background medium. Although their shape does not exactly match the original objects they reflect the correct size or rather volume fraction of these objects, as enforced by the histogram prior. Furthermore, the reconstructed $D$ values remain constant across most of the areas covered by the cubes, rather than showing a pronounced peak, as observed without the penalty term. This allows for a much better extraction of the absolute optical parameters from the reconstruction. As the strength of the prior is increased further, we observe only a weak dependence of the image quality on $\gamma$ [Figures 5, 6(c) and 6(d)]. For large values of $\gamma$, the reconstruction tends to exaggerate sharp edges and produces artifacts, while sacrificing the correct shape of the objects. As expected, the diffusion coefficient eventually assumes one of the three values imposed by the histogram. In Figure 6(d), the prior information is weighted $10^6$ stronger than the $\chi^2$ term, so that fitting the image's histogram to the prediction becomes the main objective. Nevertheless we still obtain reasonably good agreement with the original image.

### 4 Discussion

In general we found that the effectiveness of a penalty term depends as much on the right choice of the penalty term as on the right choice of the hyperparameter, which adjusts the influence of the penalty term during the reconstruction process. The results for the tissue-type penalty terms $\Pi^{tt}$ show that an improvement in the reconstruction result can only be obtained for a relatively small range of hyperparameters ($0.01<\gamma <0.1$). For values of $\gamma<0.01$, the additional penalty term has no effect, while for $\gamma>0.1$ the quality of the reconstruction is actually worse than without penalty function. Invoking the earlier discussed interpretation of the $\chi^2$ term and penalty term as potentials and there derivatives as forces, we see that

if $\gamma$ is chosen too strong the tissue-type penalty term traps all pixel values in the minimum closest to the initial guess ($D = 1$ cm$^2$ ns$^{-1}$, Figure 1). Forces due to the $\chi^2$ term are too weak to lift a pixel out of the valley centered around $D = 0.92$ cm$^2$ ns$^{-1}$.

Choosing the hyperparameter dynamically can improve the convergence of the GIIR scheme with $\Pi^{tt}$-type penalty terms, as shown in Figure 4(d). Giving only little weight to the penalty term in the early stages of the reconstruction process allows the algorithm to find areas with increased or decreased values of $D$ based on information from the $\chi^2$ term. Increasing the weight of the penalty term later in the reconstruction process leads to a sorting of each pixel in one of the three ''potential'' wells of the penalty term. However, even the use of a dynamically adjusted penalty term depends strongly on the appropriate choice of the initial value of $\gamma$ and the rate at which it is increased. One is forced to find these parameters empirically for each particular reconstruction problem. The same drawback was previously encountered for total variation minimization penalty terms (see Ref. 17).

The histogram penalty terms $\Pi^{hist}$ provides good reconstruction results for all $\gamma > 0.01$. Surprisingly even the very large $\gamma = 10^6$ provides a good reconstruction. The information contained in the histogram prevents the scheme from converging toward singular solutions, such as trapping all pixels at the same value of $D$. If more and more pixels assume only one of the expected values, the force acting on them diminishes and eventually becomes repulsive. Pixel values are pushed by penalty-term forces to alternative values to fit the overall histogram distribution. When the histogram of the reconstructed image approaches the ''true'' histogram the force acting on individual pixels diminishes and eventually becomes zero. At the global minimum of $\Pi^{hist}$ we may change single pixel values without increasing the penalty, by raising and lowering several pixels simultaneously. This is different in the scheme that uses the tissue-type penalty term $\Pi^{tt}$. To interchange two pixel values that are already located at the bottom of two different potential valleys (Figure 1), one always needs a force that pushes each pixel value over the potential hill in-between.

The weak dependence of the reconstruction on $\gamma$ and the fact that sensible results are still obtained as the penalty term becomes dominant, are important features of the histogram regularization. This makes it a very stable method to incorporate additional information into the reconstruction process. However, the full histogram information may not always be available, and only the tissue-type penalty term, which contain less information, may have to be used.

An additional effect of applying the histogram penalty function is a faster conversion of the reconstruction algorithm. Typically only 20 iterations are necessary to produce qualitatively correct images, compared to approximately 50 iterations without this prior information. As we increase $\gamma$ to values $\gg 1$, as little as 10 iterations suffice to produce the sharply contrasted results shown in Figure 6(d).

Given that the target histogram is known, one could be tempted to start with an image that has the right histogram. Therefore, rather than using a homogenous medium as the initial guess, one could chose a heterogeneous image as the initial guess, which has the correct volume fraction of certain tissue types. In Figures 7(a) and 7(b), a scrambled version of
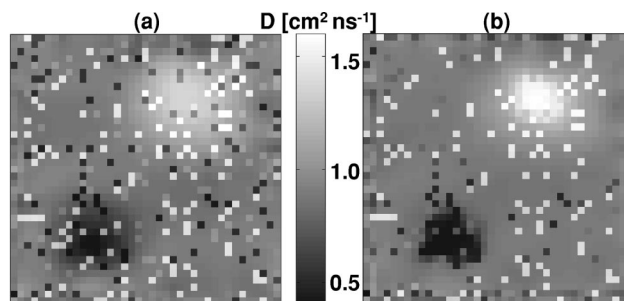


**Fig. 7** Image reconstruction starting from nonhomogeneous initial guess. Histogram of initial guess equals that of the target medium, however pixels with different $D$ are randomly placed: (a) reconstruction without penalty term; (b) result with histogram penalty term ($\gamma = 0.5$).

the exact image was used as input, so that during the first iteration $\Pi^{hist} = 0$ is already minimal. The hyperparameter $\gamma$ was set to 0 [Figure 7(a)] and 0.5 [Figure 7(b)], respectively. While for both cases the two objects are somewhat recovered, many of the initial single pixels at wrong positions remain, even when the histogram prior is switched on [Figure 7(b)]. Overall the image quality is much poorer than starting from a homogenous initial guess. In general, we found that starting with a homogeneous guess provides better images than starting with a wrong heterogeneous initial guess.

While the presented examples only show one type of medium, similar results have been obtained for a variety of differently structured media. The main results, namely that additional *a priori* information can be expressed in appropriate penalty terms, and that more information leads to better reconstructions and less sensitivity to a ''correct'' hyperparameter $\gamma$, is of general validity. Another example is shown in Figure 8 where instead of two compact elliptical inhomogeneities two curved lines with different diffusion coefficient cross a $4 \times 4$ cm medium. The medium is discretized into a $40 \times 40$ mesh with $\Delta x = 0.1$ cm. The optical properties of the background are given by $\mu_a = 0.1$ cm$^{-1}$, $\mu_S' = 8$ cm$^{-1}$, ($\Rightarrow D = 0.905$ cm$^2$ ns$^{-1}$), while the optical properties of the inclusions are given by $\mu_a = 0.2$ cm$^{-1}$, $\mu_S' = 13.0$ cm$^{-1}$ ($\Rightarrow D = 0.556$ cm$^2$ ns$^{-1}$), and $\mu_a = 0.05$ cm$^{-1}$, $\mu_S' = 4.5$ cm$^{-1}$ ($\Rightarrow D = 1.61$ cm$^2$ ns$^{-1}$), respectively. The initial guess for all reconstructions is $D = 1.0$ cm$^2$ ns$^{-1}$. Figure 8(a) shows the original image and Figures 8(b), 8(c), and 8(d) provide the reconstruction results after 40 iterations without a penalty term, with tissue-type penalty term, and histogram penalty term, respectively. Again we can see how penalty terms improve the image quality and that the use of more information leads to better results. The tissue-type penalty term only leads to improved images for $\gamma < 1.0$, while the histogram penalty term can be applied with a much wider range of $\gamma$ values. What type of penalty functions and choices of hyperparameters will be most suitable for various other applications such as optical tomographic imaging of the brain, breast, limb, joint, and other body parts remains to be determined for each case. Here we have limited ourselves to the derivation of the appropriate framework for such studies.
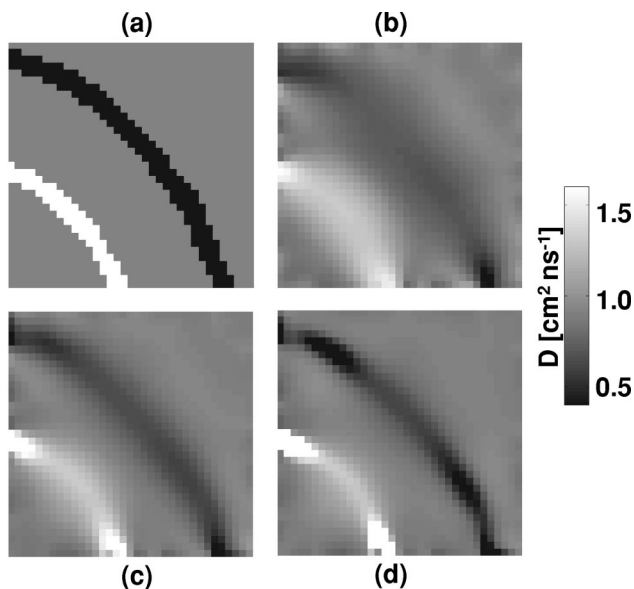
**(a)** **(b)**

**(c)** **(d)**

$D \ [\text{cm}^2 \ \text{ns}^{-1}]$

1.5

1.0

0.5

**Fig. 8** Example of reconstruction of two curved lines after 40 iterations. The initial guess is a homogeneous medium with $D=1$ cm² ns⁻¹: (a) original medium, (b) reconstruction without penalty term ($\gamma=0$), (c) reconstruction with tissue-type penalty term [Eq. (8), $\gamma=0.8$)], and (d) reconstruction with histogram penalty term [Eq. (12), $\gamma=0.8$].

Finally, it is interesting to compare the penalty term approach of gradient based schemes, with regularization schemes employed in the linear-perturbation approach (see Sec. I). It can be shown that the minimization of the functional in Eq. (7), which contains the $\chi^2$ term as well as a penalty term $\omega \cdot \Pi$, is equivalent to solving a matrix equation for $\Delta \zeta$ of the form[17,30,32]

$$(\mathbf{J}\mathbf{J}^T + \lambda \mathbf{R})\Delta \boldsymbol{\zeta} = \mathbf{J}^T(\mathbf{M} - \mathbf{P}(\boldsymbol{\zeta})) - \omega \Pi'(\boldsymbol{\zeta}), \quad (19)$$

where $\Pi'$ is the derivative of the penalty term with respect to the distribution of optical properties $\zeta(r)$. The matrix $\mathbf{R}$ is given by

$$\mathbf{R} \equiv \begin{vmatrix} \dfrac{\partial^2 \Pi(\boldsymbol{\zeta})}{\partial \zeta_1 \partial \zeta_1} & \cdots & \dfrac{\partial^2 \Pi(\boldsymbol{\zeta})}{\partial \zeta_1 \partial \zeta_j} & \cdots & \dfrac{\partial^2 \Pi(\boldsymbol{\zeta})}{\partial \zeta_1 \partial \zeta_N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 \Pi(\boldsymbol{\zeta})}{\partial \zeta_i \partial \zeta_1} & \cdots & \dfrac{\partial^2 \Pi(\boldsymbol{\zeta})}{\partial \zeta_i \partial \zeta_j} & \cdots & \dfrac{\partial^2 \Pi(\boldsymbol{\zeta})}{\partial \zeta_i \partial \zeta_N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 \Pi(\boldsymbol{\zeta})}{\partial \zeta_N \partial \zeta_1} & \cdots & \dfrac{\partial^2 \Pi(\boldsymbol{\zeta})}{\partial \zeta_N \partial \zeta_j} & \cdots & \dfrac{\partial^2 \Pi(\boldsymbol{\zeta})}{\partial \zeta_N \partial \zeta_N} \end{vmatrix}. \quad (20)$$

For the tissue-type penalty term we obtain

$$(R^{\text{tt}})_{ij} = \frac{\partial^2 \Pi^{\text{tt}}(\boldsymbol{\zeta})}{\partial \zeta_i \partial \zeta_j}$$

$$= \sum_{k=1}^{K} \{1/\sigma_k^2 - (a_k - \zeta_i)^2/\sigma_k^4\}$$

$$\times \exp\left(-\frac{(a_k - \zeta_i)^2}{2\sigma_k^2}\right) \quad \text{for} \quad i = j$$

$$= 0 \quad \text{for} \quad i \neq j. \quad (21)$$

Therefore, $(R^{\text{tt}})_{ij}$ has only entries on the diagonal, which means that no local coupling between pixels occurs. Adding this matrix to $\mathbf{J}\mathbf{J}^T$ always will make Eq. (19) diagonally dominant, if we choose the parameter $\lambda$ sufficiently large.

For the histogram penalty function we obtain:

$$(R^{\text{hist}})_{ij} = \frac{\partial^2 \Pi^{\text{hist}}(\boldsymbol{\zeta})}{\partial \zeta_i \partial \zeta_j} = \begin{cases} A_{i,i} + B_{i,i} & \text{for} \quad i = j \\ A_{i,j} & \text{for} \quad i \neq j \end{cases}$$

with

$$A_{i,j} = \sum_{l=1}^{L} \frac{2\beta^2}{H^0(\zeta_l)} \frac{(\zeta_l - \zeta_i)(\zeta_l - \zeta_j)}{\sigma^4}$$

$$\times \exp\left(-\frac{(\zeta_l - \zeta_i)^2 + (\zeta_l - \zeta_j)^2}{2\sigma^2}\right)$$

$$B_{i,i} = \sum_{l=1}^{L} \frac{2\beta(H^0(\zeta_l) - H(\zeta_l))}{H^0(\zeta_l)} \left(\frac{1}{\sigma^2} - \frac{(\zeta_l - \zeta_i)^2}{\sigma^4}\right)$$

$$\times \exp\left(-\frac{(\zeta_l - \zeta_i)^2}{2\sigma^2}\right)$$

where $\beta = \sqrt{1/(2\pi\sigma^2)}$.

Therefore, $(R^{\text{hist}})_{ij}$ has nondiagonal entries, which means that coupling between pixels occurs. However, calculating $R^{\text{hist}}$ for the examples used in this work, we find that the values of diagonal elements are in general 3 orders of magnitude higher than the values of nondiagonal elements. Therefore, adding either the tissue-type or histogram penalty function to the $\chi^2$ term in a GIIR scheme, can also be interpreted as adding diagonally dominant matrices to the ill-conditioned matrix $\mathbf{J}\mathbf{J}^T$ obtained in the linear perturbation approach to the image reconstruction problem.

The advantage of the penalty-term approach in GIIR schemes over regularization schemes employed in linear perturbation methods seems to be that penalty terms have an immediate physical interpretation. Even though useful regularizers can obviously be derived without the concept of penalty terms as shown for example by Pogue et al.,[32] the concept of penalty terms may provide an intuitive conduit through which we can incorporate additional information into the reconstruction process. Since penalty terms can be considered as potentials that bias the $\chi^2$ potentials according to our prior knowledge about the system they may supply more insight into the underlying physical assumptions about the system to be reconstructed. Furthermore, matrix diagonality as de-

manded in the linear-perturbation approach to regularization, is clearly not the only factor to get reasonable reconstruction results. As shown in this work, the choice of a good hyperparameter is very critical, and much more work needs to be done to identify robust schemes for finding appropriate hyperparameters.

## 5   Summary

In this study we addressed the problem of ill-posedness of the image reconstruction problem in optical tomography. Ill-posedness is caused by the fact that different spatial distributions of optical properties inside the medium can lead to similar detector readings on the surface of the medium. The question arises as to how can one nevertheless find the correct distribution of optical properties inside the medium from measurements on the surface of the medium. In this work we approach the problem within a gradient-based iterative image reconstruction (GIIR) scheme. Using this scheme the image reconstruction problem is interpreted as an optimization problem in which an appropriately defined scalar objective function is minimized using the gradient of the objective function with respect to the distribution of optical properties. In this context ill-posedness is reflected by the absence of a unique global minimum. To overcome this problem we added penalty terms that are derived from *a priori* information about the system to the commonly used least-square-error term, which compares numerically predicted and actual detector readings.

Specifically we derived and tested so-called tissue-type-penalty terms and histogram-penalty terms. The tissue-type penalty term is derived from the knowledge that a fixed number of different tissues are present in the medium. The optical properties of these tissues are known within certain error margins. If a pixel in the reconstructed image takes on a value that is different from the expected value, a penalty is added to the objective function. The larger the difference the larger is the penalty. The histogram penalty term is derived from the knowledge of the exact histogram of the correct image. Therefore, in addition to the knowledge about the number of different tissue types, one also knows their respective volume fractions in the image.

Performing numerical studies we found that both types of penalty terms can improve the image quality as well as the convergence rate of the iterative image reconstruction process. However, a crucial variable is the coupling or hyperparameter that adjusts the relative strength of the penalty term with respect to the least-square-error term. The tissue-type penalty term only provides image improvement for a narrow range of hyperparameters. If the hyperparameter is not chosen correctly the addition of this penalty term can have a detrimental effect on the image quality. On the other hand, the histogram-penalty term is much less sensitive to the right choice of the hyperparameter and in almost all cases leads to improved reconstruction results.

Finally, we compared the penalty terms derived for the gradient-based iterative image reconstruction scheme, with regularization schemes employed in the linear-perturbation approach. We showed that both tissue-type and histogram-penalty terms lead to diagonally dominate matrixes in the linear-perturbation approach.

## References

1. K. Wells, J. C. Hebden, F. E. W. Schmidt, and D. T. Delpy, ''The UCL multichannel time-resolved system for optical tomography,'' in *Optical Tomography and Spectroscopy of Tissue*, B. Chance and R. R. Alfano, Eds., *Proc. SPIE* **2979**, 599–607 (1997).
2. H. Jess, H. Erdl, K. T. Moesta, S. Fantini, M. A. Franceschini, E. Gratton, and M. Kaschke, ''Intensity modulated breast imaging: Technology and clinical pilot study results,'' in *Advances in Optical Imaging and Photon Migration*, OSA Trends in Optics and Photonics, Vol. II, R. R. Alfano and J. G. Fujimoto, Eds., pp. 126–129, Optical Society of America, Washington, DC (1996).
3. J. P. Vanhouten, D. A. Benaron, S. Spilman, and D. K. Stevenson, ''Imaging brain injury using time-resolved near-infrared light scanning,'' *Pediatr. Res.* **39**, 470–476 (1996).
4. M. Miwa and Y. Ueda, ''Development of time-resolved spectroscopy system for quantitative noninvasive tissue measurement,'' in *Optical Tomography, Photon Migration, and Spectroscopy of Tissue and Model Media*, B. Chance and R. R. Alfano, Eds., *Proc. SPIE* **2389**, 142–149 (1995).
5. S. Fantini, M. A. Franceschini, J. S. Maier, S. A. Walker, B. Barbieri, and E. Gratton, ''Frequency domain multichannel optical detector for noninvasive tissue spectroscopy and oximetry,'' *Opt. Eng.* **34**, 32–42 (1995).
6. C. H. Schmitz, H. L. Graber, H. Luo, I. Arif, J. Hira, Y. Pei, A. Bluestone, S. Zhong, R. Andronica, I. Soller, N. Ramirez, S-L. S. Barbour, and R. L. Barbour, ''Instrumentation and calibration protocol for imaging dynamic features in dense-scattering media by optical tomography,'' *Appl. Opt.* **39**(34), 6466–6487 (2000).
7. M. A. Franceschini, K. T. Moesta, S. Fantini, G. Gaida, E. Gratton, H. Jess, W. W. Mantulin, M. Seeber, P. M. Schlag, and M. Kaschke, ''Frequency-domain techniques enhance optical mammography: initial clinical results,'' *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6468–6647 (1997).
8. S. R. Arridge, ''The Forward and inverse problems in time-resolved infrared imaging,'' in *Medical Optical Tomography*, G. Müller, Ed., *Proc. SPIE* **IS11**, 35–64 (1993).
9. S. R. Arridge, ''Photon-measurement density functions. Part I: Analytical forms,'' *Appl. Opt.* **24**, 7395–7409 (1995).
10. S. R. Arridge and M. Schweiger, ''Photon measurement density functions. Part 2: Finite-element-method calculation,'' *Appl. Opt.* **34**, 8026–8037 (1995).
11. M. Schweiger and S. R. Arridge, ''A system for solving the forward and inverse problems in optical spectroscopy and imaging,'' in *Advances in Optical Imaging and Photon Migrations*, OSA Trends in Optics and Photonics Series, R. R. Alfano and J. G. Fujimoto, Eds., Vol. 2, pp. 263–268, Optical Society of America, Washington, DC (1996).
12. R. L. Barbour, H. L. Graber, Y. Wang, J. H. Chang, and R. Aronson, ''A perturbation approach for optical diffusion tomography using continuous-wave and time-resolved data,'' in *Medical Optical Tomography*, G. Müller, Ed., *Proc. SPIE* **IS11**, 87–120 (1993).
13. H. L. Graber, J. Chang, R. Aronson, and R. L. Barbour, ''A perturbation model for imaging in dense scattering media: Derivation and Evaluation of Imaging Operators,'' in *Medical Optical Tomography*, G. Müller, Ed., *Proc. SPIE* **IS11**, 121–143 (1993).
14. R. L. Barbour, H. L. Graber, J. W. Chang, S. L. S. Barbour, P. C. Koo, and R. Aronson, ''MRI-guided optical tomography: Prospects

and computation for a new imaging method,'' *IEEE Comput. Sci. Eng.* **2**, 63–77 (1995).

15. Y. Q. Yao, Y. Wang, Y. L. Pei, W. W. Zhu, and R. L. Barbour, ''Frequency-domain optical imaging of absorption and scattering distributions by Born iterative method,'' *J. Opt. Soc. Am. A* **14**, 325–342 (1997).

16. K. D. Paulsen and H. Jiang, ''Spatially varying optical property reconstruction using a finite element diffusion equation approximation,'' *Med. Phys.* **22**, 691–701 (1995).

17. K. D. Paulsen and H. Jiang, ''Enhanced frequency domain optical image reconstruction in tissues through total variation minimization,'' *Appl. Opt.* **35**, 3447–3458 (1996).

18. H. Jiang, K. D. Paulsen, and U. L. Österberg, ''Optical image reconstruction using DC data: simulations and experiments,'' *Phys. Med. Biol.* **41**, 1483–1498 (1996).

19. H. Jiang, K. D. Paulsen, U. L. Osterberg, B. W. Pogue, and M. S. Patterson, ''Simultaneous reconstruction of optical absorption and scattering maps in turbid media from near-infrared frequency-domain data,'' *Opt. Lett.* **20**, 2128–2130 (1995).

20. H. B. Jiang, K. D. Paulsen, U. L. Osterberg, B. W. Pogue, and M. S. Patterson, ''Optical image reconstruction using frequency domain data: Simulations and experiments,'' *J. Opt. Soc. Am. A* **13**, 253–266 (1996).

21. J. C. Schottland, J. C. Haselgrove, and J. S. Leigh, ''Photon Hitting Density,'' *Appl. Opt.* **32**, 448–453 (1993).

22. M. A. O'Leary, D. A. Boas, B. Chance, and A. G. Yodh, ''Experimental images of heterogeneous turbid media by frequency-domain diffusion-photon tomography,'' *Opt. Lett.* **20**, 426–428 (1995).

23. D. Y. Paithankar, A. U. Chen, B. W. Pogue, M. S. Patterson, and E. M. Sevick-Muraca, ''Imaging of fluorescent yield and lifetime from multiply scattered light reemitted from random media,'' *Appl. Opt.* **36**, 2260–2272 (1997).

24. S. S. Saquib, K. M. Hanson, and G. S. Cunningham, ''Model-based image reconstruction from time-resolved diffusion data,'' in *Medical Imaging: Image Processing*, *Proc. SPIE* **3034**, 369–380 (1997).

25. A. H. Hielscher, A. Klose, D. M. Catarious Jr., and K. M. Hanson, ''Tomographic imaging of biological tissue by time-resolved, model-based, iterative, image reconstruction,'' *OSA Trends in Optics and Photonics: Advances in Optical Imaging and Photon Migration II*, Vol. 21, R. R. Alfano and J. G. Fujimoto, Eds., Vol. 21, pp. 151–161, Optical Society of America, Washington, DC (1998).

26. S. R. Arridge and M. Schweiger, ''A gradient-based optimization scheme for optical tomography,'' *Opt. Express* **2**(6), 213–226 (1998).

27. A. H. Hielscher, A. D. Klose, and K. M. Hanson, ''Gradient-based iterative image reconstruction scheme for time-resolved optical tomography,'' *IEEE Trans. Med. Imaging* **18**, 262–271 (1999).

28. A. D. Klose and A. H. Hielscher, ''Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer,'' *Med. Phys.* **26**(8), 1698–1707 (1999).

29. 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**(10), 353–371 (1999).

30. S. R. Arridge, ''Optical tomography in medical imaging,'' *Inverse Probl.* **15**, R41–R93 (1999).

31. S. R. Arridge, ''Forward and inverse problems in time-resolved infrared imaging,'' in *Medical Optical Tomography*, G. Müller Ed., Vol. IS11, pp. 53–64, SPIE Optical Engineering, Bellingham, WA (1993).

32. B. W. Pogue, T. O. McBride, J. Prewitt, U. L. Osterberg, and K. D. Paulsen, ''Spatially variant regularization improves diffuse optical tomography,'' *Appl. Opt.* **38**, 2950–2961 (1999).

33. J. Chul Ye, K. J. Webb, C. A. Bouman, and R. P. Millane, ''Optical diffusion tomography by iterative-coordinate-descent optimization in a Bayesian framework,'' *J. Opt. Soc. Am. A* **16**, 2400–2412 (1999).

34. M. Schweiger and S. R. Arridge, ''Application of temporal filters to time resolved data in optical tomography,'' *Phys. Med. Biol.* **44**, 1699–1717 (1999).

35. A. H. Hielscher, A. D. Klose, and J. Beuthan, ''Evolution strategies for optical tomographic characterization of homogeneous media,'' *Opt. Express* **7**(13), 507–518 (2000). (http://www.opticsexpress.org/framestocv7n13.htm)

36. D. A. Pierre, *Optimization Theory with Applications*, Dover, New York, NY (1986).

37. D. H. Hristov and B. Gino Fallone, ''A continuous penalty function method for inverse treatment planning,'' *Med. Phys.* **25**(2), 208–224 (1998).

38. M. Ben-Daya and K. S. Al-Sultan, ''A new penalty function algorithm for convex quadratic programming,'' *Eur. J. Operational Res.* **101**(1), 155–164 (1997).

39. D. J. White, ''The maximization of a function over the efficient set via a penalty function approach,'' *Eur. J. Operational Res.* **94**(1), 143–154 (1996).

40. H. Xu, ''Stochastic Penalty Function Methods for Nonsmooth Constrained Minimization,'' *J. Optim. Theory Appl.* **88**(3), 709–725 (1996).

41. D. Vucina, ''Flow formulation FE metal-forming analysis with boundary friction via a penalty function,'' *J. Mater. Process. Technol.* **59**(3), 272–277 (1996).

42. B. C. Fabien, ''An extended penalty function approach to the numerical solution of constrained optimal control problems,'' *Opt. Control Appl. Methods* **17**(5), 341–355 (1996).

43. J. A. Snyman, N. Stander, and W. J. Roux, ''A dynamic penalty function method for the solution of structural optimization problems,'' *Appl. Math. Model.* **18**(8), 453–463 (1994).

44. K. Toma, ''Protein three-dimensional structure generation with an empirical hydrophobic penalty function,'' *J. Mol. Graphics* **11**(4), 222–232 (1993).