Weight-matrix structured regularization provides optimal generalized least-squares estimate in diffuse optical tomography

Phaneendra K. Yalavarthy[a] and Brian W. Pogue
Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755

Hamid Dehghani
School of Physics, University of Exeter, Stocker Road, Exeter EX4 4QL, United Kingdom

Keith D. Paulsen
Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755

(Received 20 December 2006; revised 29 March 2007; accepted for publication 29 March 2007; published 17 May 2007)

Diffuse optical tomography (DOT) involves estimation of tissue optical properties using noninvasive boundary measurements. The image reconstruction procedure is a nonlinear, ill-posed, and ill-determined problem, so overcoming these difficulties requires regularization of the solution. While the methods developed for solving the DOT image reconstruction procedure have a long history, there is less direct evidence on the optimal regularization methods, or exploring a common theoretical framework for techniques which uses least-squares (LS) minimization. A generalized least-squares (GLS) method is discussed here, which takes into account the variances and covariances among the individual data points and optical properties in the image into a structured weight matrix. It is shown that most of the least-squares techniques applied in DOT can be considered as special cases of this more generalized LS approach. The performance of three minimization techniques using the same implementation scheme is compared using test problems with increasing noise level and increasing complexity within the imaging field. Techniques that use spatial-prior information as constraints can be also incorporated into the GLS formalism. It is also illustrated that inclusion of spatial priors reduces the image error by at least a factor of 2. The improvement of GLS minimization is even more apparent when the noise level in the data is high (as high as 10%), indicating that the benefits of this approach are important for reconstruction of data in a routine setting where the data variance can be known based upon the signal to noise properties of the instruments. © *2007 American Association of Physicists in Medicine*. [DOI: 10.1118/1.2733803]

Key words: near infrared, diffuse optical tomography, inverse problems, least-squares minimization

I. INTRODUCTION

Image reconstruction methods used in diffuse optical tomography (DOT) are mainly dependent on the type of data, the diffuse light model, and the number of available anatomical/spectral priors. There are numerous reconstruction techniques available in the literature depending on the application.[1–5] Yet despite the volume of work in this area there is no single investigation with a direct comparison of the least-squares (LS) minimization techniques using the same implementation scheme, especially in terms of data noise level and complexity in the test fields. Most of the comparisons in the literature have been in terms of implementation of minimization and convergence rates of one or two techniques at hand.[1–5] This work addresses this problem and compares minimization methods (more specifically different types of regularization) with the same implementation scheme for a direct quantitative comparison. Moreover, usage of weight matrices in the regularization which include the variance and covariance properties of data and image space are extensively explored here. A new covariance form borrowed from meteorological studies is introduced and proven to be effective for reconstructing highly noisy data in the generalized theoretical frame work.

Near infrared DOT involves reconstructing images of optical properties from transmission measurements using wavelengths from 650–1000 nm to interrogate tissue.[1,6–8] Optical absorption and scattering images obtained using multiple wavelengths can be used to estimate tissue hemoglobin, water concentration, scattering amplitude, and scattering power.[8] To overcome the inherent low-spatial resolution in DOT, there is a considerable interest in developing hybrid systems,[9–27] which use the spatial mapping of one system as the template for DOT. Image formation from the data collected by these (stand-alone/hybrid) systems involves solving an inversion problem. This article describes LS minimization techniques to solve the inverse problem and to quantitatively compare their performance in a systematic series of simulations. The inverse problem (image reconstruction procedure) in DOT is known to be a nonlinear, ill-posed, and ill-determined problem,[2] and to solve such a problem, a regularization term must be added to constrain the solution space in order to obtain a meaningful image. There are many

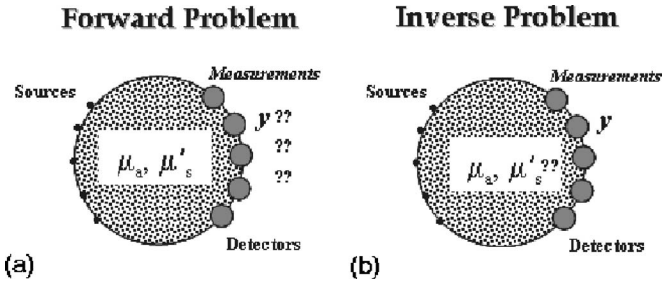

FIG. 1. An illustration of the forward and inverse problem in diffuse optical tomography is shown (see Ref. 64), where (a) the data y is estimated given values of μ_a and μ_s' and source/detector positions. In the inverse problem (b), the values of μ_a and μ_s' must be obtained given a set of measurements (y).

regularization methods available in the literature and this work focuses on the fact that most LS techniques presented in the literature can be encompassed within a generalized theoretical framework, which includes a regularization matrix that is based upon weights from the data and parameter variances. Note that Appendix A gives the terminology used in this work along with definitions of symbols.

Because of the interest in using spatial information derived from conventional imaging modalities in the DOT inverse problem, a number of methods have been presented in the literature.[9–27] These techniques were initially proposed by Barbour *et al.*[9] and Schweiger *et al.*[13] and used in to improve the quantitative outcome of reconstructed images. Ntziachristos *et al.*[14] used the magnetic resonance (MR) information to divide the imaging domain into tumor and nontumor regions to make the problem better posed. Li *et al.*[17] used an x-ray tomosynthesis volume to segment the breast into different subregions and used different regularization parameters depending on the size of the subregions. Recently Guven *et al.*[24] proposed a Bayesian frame work to include spatial prior information in an effective way which will not bias the image reconstruction problem to imperfect anatomical priors. Pogue and Paulsen,[10] Brooksby *et al.*[18,21,25] and Yalavarthy *et al.*[26] have extended these approaches for the use of anatomical prior information in which, depending on the connectivity and size of the subregion, the regularization term was scaled. Even though the effect of imperfect spatial prior information on the image reconstruction is a very active research area,[23,24,26] it was assumed here that the spatial priors were perfect. Other ongoing studies are examining this more complex issue.

II. DOT FORWARD PROBLEM

DOT involves solving a model (forward) and estimation (inverse) problem, sequentially as illustrated in Fig. 1. In this section, the forward problem is described, which involves generating the measurement data, for a given set of optical property estimates within the tissue, using a finite element solution to the diffuse transport equation.

Light propagation in a turbid elastic-scattering media, like tissue, is treated as "neutral-particle transport" rather than "wave propagation" and in the frequency domain, the diffusion equation is used, which is given by[2,28]

$$-\nabla \cdot D(\mathbf{r})\,\nabla \Phi(\mathbf{r},\omega) + [\mu_a(\mathbf{r}) + i\omega/c]\Phi(\mathbf{r},\omega) = Q_o(\mathbf{r},\omega), \tag{1}$$

where $\Phi(\mathbf{r},\omega)$ is the photon density at position $\mathbf{r}$ and the light modulation frequency is given by ω ($\omega=2\pi f$, in this work $f=100$ MHz). The isotropic source term is represented by $Q_o(\mathbf{r},\omega)$ and the speed of light in tissue by c, which is constant here. $\mu_a(\mathbf{r})$ is the optical absorption coefficient and $D(\mathbf{r})$ is the optical diffusion coefficient, which is defined as

$$D(\mathbf{r}) = \frac{1}{3[\mu_a(\mathbf{r}) + \mu_s'(\mathbf{r})]}, \tag{2}$$

where $\mu_s'(\mathbf{r})$ is the reduced scattering coefficient, which is defined as $\mu_s'=\mu_s(1-g)$. μ_s is the scattering coefficient and g is the anisotropy factor. A Robin (type-III) boundary condition is applied to model the refractive-index mismatch at the boundary.[29] The measured data for a frequency domain system are the amplitude and phase of the transmitted signal. If F is the forward model [finite element method (FEM) in here] which gives the fluence at every point, then the modeled data $G(\mu)$ can be obtained by sampling the forward model at the boundary given internal spatial distributions of optical properties and source-detector locations, where μ represents the parameters $\{\mu=[D(\mathbf{r});\mu_a(\mathbf{r})]\}$,

$$G(\mu) = S\{F(\mu)\}. \tag{3}$$

The details of the FEM formulation of the forward model are given in Refs. 30–32. The results presented are restricted to frequency-domain data, more specifically data (y) is the natural logarithm of the amplitude (A) and phase (θ) of the frequency-domain signal. Defining A and θ in terms of modeled data, $A=\sqrt{\mathrm{Re}\{G(\mu)\}^2+\mathrm{Im}\{G(\mu)\}^2}$ and $\theta=\tan^{-1}(\mathrm{Im}\{G(\mu)\}/\mathrm{Re}\{G(\mu)\})$. The Jacobian ($J$), which gives the rate of change of modeled data with respect to parameters, is calculated using the adjoint method.[30] Even though the actual parameters being estimated are $D(\mathbf{r})$ and $\mu_a(\mathbf{r})$, the results are presented in terms of $\mu_a(\mathbf{r})$ and $\mu_s'(\mathbf{r})$, which are spectroscopically more meaningful.

III. LEAST-SQUARES MINIMIZATION TECHNIQUES

This section outlines several different minimization schemes used in this work. These techniques are used to solve the inverse problem [Fig. 1(b)], which is achieved by minimizing the objective function (Ω) over the range of μ. Minimizing the objective function can be achieved by several different approaches. The most common approaches involve obtaining repeated solutions of the forward model and recomputation of the Jacobian (J) (and its inversion) at every iteration because of the nonlinear nature of the problem. There are also gradient-based optimization schemes available in the literature[33,34] to minimize the objective function which does not require an explicit inversion of the Hessian

matrix. In this work direct methods, known as full-Newton approaches,[2] are employed in minimization for all the regularization techniques used for a fair comparison. LS minimization has the effect of reducing high frequency noise, leading to smooth images of optical properties. Total variation methods and variants of this are used to obtain edge preservation in reconstructed images.[27,35] Solving the inverse problem using LS minimization can also be seen from a Bayesian prospective to obtain *maximum a posteriori* estimate.[24,36,37] A correlation between the Bayesian frame work and LS minimization techniques is given in Refs. 12, 38, and 39, but usage of the Bayesian frame work requires one to choose a particular noise model for both data and image space, which might not reflect the actual noise characteristics unless some prior information is available. Here, the emphasis is on LS minimization techniques with a focus on what the value of the regularization method can be. The LS methods are divided into two groups: (1) Without spatial priors and (2) with spatial priors.

A. Without spatial priors

1. Levenberg–Marquardt minimization

This approach is also known as a trust-region method[5,39] where experimental data is matched with modeled data iteratively.[40,41] The objective function for the DOT problem is defined as

$$\Omega = \{\|y - G(\mu)\|^2\}, \tag{4}$$

where $\mathbf{y}$ is the data and $G(\mu)$ is the modeled data. This equation is minimized by setting the first-order derivative equal to zero.

a. First-order condition. Minimizing Ω with respect to μ, which is achieved by setting $\partial\Omega/\partial\mu=0$,

$$\frac{\partial\Omega}{\partial\mu} = J^T\delta = 0, \tag{5}$$

where δ is the data-model misfit, $\delta=\mathbf{y}-G(\mu)$, J is the Jacobian, and T represents the matrix transpose operator.

b. Iterative update equation. Imagine a sequence of approximations to μ represented by μ_i, then using Taylor series on $G(\mu_i)$ and expanding around μ_{i-1} gives

$$G(\mu_i) = G(\mu_{i-1}) + J\Delta\mu_i + \ldots, \tag{6}$$

where $\Delta\mu_i=\mu_i-\mu_{i-1}$. Rewriting δ utilizing the first two terms of Eq. (6) (ignoring the rest, equivalently linearizing the problem) gives

$$\delta_i = y - G(\mu_i) = y - G(\mu_{i-1}) - J\Delta\mu_i = \delta_{i-1} - J\Delta\mu_i. \tag{7}$$

Rewriting Eq. (5) for the ith iteration

$$J^T\delta_i = 0. \tag{8}$$

Substituting Eq. (7) into Eq. (8) gives

$$J^T(\delta_{i-1} - J\Delta\mu_i) = 0. \tag{9}$$

Further simplification leads to the update equation

$$[\mathbf{J}^T\mathbf{J}]\Delta\mu_i = \mathbf{J}^T\delta_{i-1}. \tag{10}$$

When J^TJ is ill-conditioned, a diagonal term is added to stabilize the problem. In this case, the update equation becomes:

$$[\mathbf{J}^T\mathbf{J} + \alpha\mathbf{I}]\Delta\mu_i = \mathbf{J}^T\delta_{i-1}, \tag{11}$$

where $\Delta\mu_i$ is the update for the parameter in the ith step. Note that α monotonically decreases with iterations (always >0), and also that $\alpha\geqslant\|\delta\|^2$. The iterative method (or its modified version) is the commonly used minimization technique in DOT. It can be seen from Eqs. (10) and (11), when α becomes zero in Eq. (11) it becomes Eq. (10). It is also important to note that J^TJ is always symmetric, because $(J^TJ)^T=J^T(J^T)^T=J^TJ$. The advantage of using this method is in the simple choice of a regularization parameter (α). The limitations[41] of this method include:

- J^TJ must be positive definite.
- The initial guess (μ_0) should be close to the actual solution.
- The update equation [Eq. (11)] does not solve the first-order conditions unless $\alpha=0$.
- Since parameters are not involved in the minimization scheme, the inverse problem may be unstable.

Even though J^TJ is not positive definite in DOT, the Levenberg–Marquardt (LM) approach (or its modified version) has been used successfully in a number of instances.[2,6,7,28,42]

2. Tikhonov minimization

The generalized objective function[43,44] in the Tikhonov case includes parameters in the minimization function, which is defined as

$$\Omega = \{\|\mathbf{y} - G(\mu)\|^2 + \lambda\|L(\mu - \mu_0)\|^2\}, \tag{12}$$

where λ is the Tikhonov regularization parameter and L is a dimensionless regularization matrix (in this work). Here, μ_0 is the prior estimate of the optical properties, which in DOT has typically been obtained from calibrating the data.[45,46]

a. Choice of λ. Rewriting Eq. (12), normalizing both terms by their variances yields

$$\Omega = \left\{\frac{\|\mathbf{y} - G(\mu)\|^2}{(\sigma_y)^2} + \frac{\|L(\mu - \mu_0)\|^2}{(\sigma_{\mu-\mu_0})^2}\right\}, \tag{13}$$

where σ_y is the standard deviation in the data y and $\sigma_{\mu-\mu_0}$ is the standard deviation in the optical properties (or deviation from the prior estimate of optical properties). Note that the variance of data-model misfit $[\delta=\mathbf{y}-G(\mu)]$ is assumed from the data, i.e., $(\sigma_\delta)^2=(\sigma_y)^2+(\sigma_{G(\mu)})^2$ with $(\sigma_{G(\mu)})^2=0$ because synthetic data was used. Multiplying Eq. (13) by σ_y^2 and comparing the result with Eq. (12) leads to

$$\lambda = \frac{(\sigma_y)^2}{(\sigma_{\mu-\mu_0})^2}, \tag{14}$$

which shows that the Tikhonov regularization parameter (λ) should be equal to the square of the ratio of the standard deviation in data to the standard deviation of the parameters.

This is a subtle yet important point, especially since this parameter is rarely defined this way, and is most commonly derived empirically.

b. First-order condition. Minimizing Ω with respect to μ, which is achieved by setting $\partial\Omega/\partial\mu=0$,

$$\frac{\partial\Omega}{\partial\mu}=J^T\delta-\lambda L^TL(\mu-\mu_0)=0. \tag{15}$$

c. Update equation. Rewriting Eq. (15) for the ith iteration leads to

$$J^T\delta_i-\lambda L^TL(\mu_i-\mu_0)=0. \tag{16}$$

Substituting Eq. (7) into Eq. (16) results in

$$J^T(\delta_{i-1}-J\Delta\mu_i)-\lambda L^TL(\mu_{i-1}+\Delta\mu_i-\mu_0)=0. \tag{17}$$

Further simplification leads to the iterative update equation

$$[\mathbf{J}^T\mathbf{J}+\lambda\mathbf{L}^T\mathbf{L}]\Delta\mu_i=\mathbf{J}^T\delta_{i-1}-\lambda\mathbf{L}^T\mathbf{L}(\mu_{i-1}-\mu_0). \tag{18}$$

Note that L^TL is symmetric. The constraint on the choice of L is that it must be positive definite.[44] In the absence of spatial priors, a common choice for the form of L is the identity matrix (I), which leads to the update equation

$$[\mathbf{J}^T\mathbf{J}+\lambda\mathbf{I}]\Delta\mu_i=\mathbf{J}^T\delta_{i-1}-\lambda(\mu_{i-1}-\mu_0). \tag{19}$$

Refer to Appendix B for an analysis of the Tikhonov regularization in terms of singular values. This regularization method is particularly common for ill-posed problems. The advantage of the method, is that it includes parameters within the minimization scheme which can be selected to improve the stability of the solution. Its limitations are that:

- it requires a prior opinion about the noise characteristics of the parameter and data spaces (for λ) and
- it does not take into account the individual variances of the data points/parameters, nor their covariances.

However, the simplicity of the approach makes it attractive for use in ill-posed problems. When the dynamic range of the data is large (as in DOT), incorporation of the maximum variance in the data will cause the minimization to bias the solution to specific data points (e.g. near the boundaries at source-detector locations in DOT). To reduce the effect of bias, one can employ a generalized least squares (GLS) minimization scheme, described in the next section.

3. GLS minimization

Generalized least squares minimization schemes have been proposed in the context of Tikhonov minimization in the literature,[1,5,38] in which there is some ambiguity in choosing the regularization parameter (λ). In here, a direct inclusion of weight matrices (which are inverses of covariance matrices) in the minimization scheme was employed to explicitly remove the dependence of reconstructed image quality on the choice of regularization parameter. This type of choice leads to an objective function[47,48]

$$\Omega=\{[y-G(\mu)]^TW_\delta[y-G(\mu)]+(\mu-\mu_0)^TW_{\mu-\mu_0}(\mu-\mu_0)\}, \tag{20}$$

where W_δ is the weight matrix for data-model misfit (δ) with $W_\delta=[\mathrm{cov}(\delta)]^{-1}$ (Appendix A-4 of Ref. 47). $W_{\mu-\mu_0}$ is the weight matrix for optical properties $(\mu-\mu_0)$ with $W_{\mu-\mu_0}=[\mathrm{cov}(\mu-\mu_0)]^{-1}$ (Appendix A-4 of Ref. 47). Explicit forms for these weight matrices are discussed later. Since both are inverses of covariance matrices, they are symmetric and positive definite.

a. First-order condition. Minimizing Ω with respect to μ, which is achieved by setting $\partial\Omega/\partial\mu=0$ produces

$$\frac{\partial\Omega}{\partial\mu}=J^TW_\delta\delta-W_{\mu-\mu_0}(\mu-\mu_0)=0. \tag{21}$$

b. Update equation. Similar to the Tikhonov approach, linearizing the problem leads to the iterative update equation[48]

$$[\mathbf{J}^T\mathbf{W}_\delta\mathbf{J}+\mathbf{W}_{\mu-\mu_0}]\Delta\mu_i=\mathbf{J}^T\mathbf{W}_\delta\delta_{i-1}-\mathbf{W}_{\mu-\mu_0}(\mu_{i-1}-\mu_0). \tag{22}$$

4. Choice of W_δ

Since simulated data were used here, in the formation of the weight matrix (covariance matrix), it was assumed that the cov(δ) is due to measurement error only, which yields[47]

$$W_\delta=[\mathrm{cov}(\delta)]^{-1}=\{\mathrm{cov}[y-G(\mu)]\}^{-1}=[\mathrm{cov}(y)]^{-1}, \tag{23}$$

where cov represents the covariance operator. In the simulation, typically one generates the forward data and adds noise to it to form synthetic data

$$y=G(\mu)+\sigma_y\eta, \tag{24}$$

where η is a random number vector. Typically, a random number generator which follows a normal distribution with zero mean and unity variance is used. Here, σ_y is the standard deviation of the data, assuming the noise is totally uncorrelated (white noise) in which case, the covariance matrix becomes[47]

$$[\mathrm{cov}(y)]_{ij}=\begin{cases}0 & \text{if } i\neq j\\ (\sigma_y)_i^2 & \text{if } i=j.\end{cases} \tag{25}$$

Since synthetic data were used in this article, the weight matrix for the data (W_δ) becomes diagonal. In the experimental case, one needs to collect an ensemble of data sets from which a covariance matrix can be computed. In this case, "N" data sets need to be collected using the same phantom (different homogeneous phantoms need to be used for different signal levels), where N needs to be a large number. From this ensemble of $\{y\}$,

$$\{y\}=\bar{y}+\{\tilde{y}\}, \tag{26}$$

where $\bar{y}$ is the true or mean value of data and $\{\tilde{y}\}$ is perturbation due to noise. This leads to

$$[\mathrm{cov}(y)] = [\mathrm{cov}(\tilde{y})] = \overline{\{\tilde{y}\}\{\tilde{y}\}^T} = \frac{\sum_{i=1}^{N} \tilde{y}_i \tilde{y}_i^T}{N} \tag{27}$$

substituting Eq. (27) in Eq. (23) gives W_δ. Note that in Eq. (23), it was assumed the cov(δ) is due to measurement error, which is also true in the case of experimental data, as the data are calibrated to remove the offset and match the modeled data.[45,46]

5. Choices of $W_{\mu-\mu_0}$

Here, two forms were considered to highlight the versatility of the procedure, even though many other forms of the covariance matrix can exist.

a. Analytical covariance form. Borrowed from the meteorological studies, assuming the parameter field obeys the Helmholtz equation, an analytical form (for one-dimensional infinite domain case) for the covariance matrix is[47]

$$[\mathrm{cov}(\mu-\mu_0)]_{ij} = (\sigma_{\mu-\mu_0})^2 \left(1 + \frac{r_{ij}}{l}\right) e^{-r_{ij}/l}, \tag{28}$$

where r_{ij} is the separation distance between locations and l is the correlation length scale. $(\sigma_{\mu-\mu_0})^2$ is the expected variance for $\mu-\mu_0$. In this case, the weight matrix is constructed from $W_{\mu-\mu_0}=[\mathrm{cov}(\mu-\mu_0)]^{-1}$

b. Local Laplacian form. Here, the weight matrix is formed directly using a local Laplacian operator[5,49,50] between neighboring locations, where

$$W_{\mu-\mu_0} = [1/(\sigma_{\mu-\mu_0})^2] M^T M,$$

where M is the local Laplacian matrix, which is defined as

$$M_{ij} = \begin{cases} 0 & \text{if } i \text{ and } j \text{ are not neighbors} \\ -1 & \text{if } i \text{ and } j \text{ are neighbors} \\ \left|\sum_j M_{ij}\right| & \text{if } i = j \end{cases}, \tag{29}$$

where i and j represent the node numbers of the FEM mesh, which in turn become the indices of the local Laplacian matrix (M). The diagonal terms in M gives the total number of immediately connected nodes.

Computation of $W_{\mu-\mu_0}$ requires an estimate of variance of parameters $[(\sigma_{\mu-\mu_0})^2]$, as is the case for calculation of the Tikhonov regularization parameter [Eq. (14)]. The expected variance can be computed in many ways: the most common method for imaging problems is estimation from the literature. For example, the optical contrast between tumor and normal breast tissue is around 50%–400% (Refs. 51 and 52), which gives the expected standard deviation $(\sigma_{\mu-\mu_0})$ in the optical properties, and can be used to compute variance. The calibration of the experimental data is capable of giving a very good estimate of normal tissue optical properties.[45,46] Note that weight matrix containing the expected variance will not impose a hard constrain on the expected optical properties, but discourages update values ($\Delta\mu$) which are above these expected deviation in a given iteration.

The advantages of the GLS approach are that:

- It accounts for covariance among the parameters and data points.
- It also allows the individual data points/parameters to have different noise characteristics (variances).
- It constrains the problem through the weight matrices, to produce stable solutions.

The limitations of the procedure are:

- It requires prior knowledge about the noise characteristics of parameter and data spaces.
- The weight matrices may necessitate computation of the inverse of covariance matrices (increasing run time and memory requirements).
- It can generate unstable solutions when unreasonable constraints are inadvertently applied.

B. With spatial priors

Overall, the LS minimization schemes using spatial priors can be broadly classified into two approaches: (1) soft-priors and (2) hard-priors. The following two subsections will discuss these two approaches.

1. Soft-priors

In this approach, the regularization matrix L in the Tikhonov approach [Eq. (18)] encodes the spatial information.[21,26] Previous results have shown that using the spatial priors in this fashion do not bias the image estimate when the prior information is imperfect.[26] Typically, the conventional image is segmented into different regions depending on tissue type to generate the spatial constraints. The L matrix relates each nodal optical property in the numerical model to the other nodes in that region.[26] Two possible forms are indicated later.

a. Laplacian form[21]

$$L_{ij} = \begin{cases} 0 & \text{if } i \text{ and } j \text{ are not in the same region} \\ -1/N & \text{if } i \text{ and } j \text{ are in the same region} \\ 1 & \text{if } i = j \end{cases}, \tag{30}$$

where N is the number of sampling points (e.g., nodes in a FEM mesh) in that region.

b. Helmholtz form[26]

$$L_{ij} = \begin{cases} 0 & \text{if } i \text{ and } j \text{ are not in the same region} \\ -1/[N + (\kappa h)^2] & \text{if } i \text{ and } j \text{ are in the same region} \\ 1 & \text{if } i = j \end{cases}, \tag{31}$$

where N is the number of nodes in that region, $\kappa = 1/l$ with l being the covariance length and h is the distance between the nodes.

2. Hard-priors

In the hard-prior approach, also known as a parameter-reduction technique, the number of parameters to be esti-

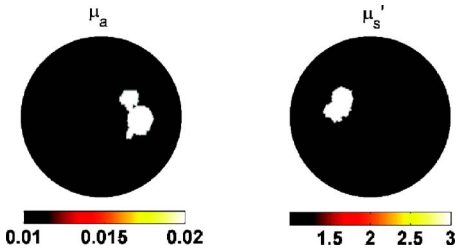

FIG. 2. The chosen optical property distribution/domain for the generation of synthetic data is shown. The diameter of the domain was 86 mm.

mated becomes the number of regions segmented from the other imaging modality (spatial priors).[26,53] Even though the number of parameters to be estimated reduces considerably (relative to soft-priors), the problem can still be ill-posed,[2] so a LM approach was used [Eq. (11)] in this case due to its simplicity. The main advantages of the method are:

- The problem is overdetermined, which also implies J^TJ is positive definite.
- It is computationally efficient.

The limitations include:

- The effect of error or uncertainty in the spatial priors can be amplified by the technique.
- The DOT problem may still be ill-posed (and ill-conditioned) after the constraints are added.[2]

3. Important notes about minimization schemes

There are additional important points about these minimization schemes.

- The weight matrices (W_δ and $W_{\mu-\mu_0}$) in the GLS scheme are computed before the iterative reconstruction procedure begins and are invariant during the iterative process. The same is true of the soft-priors L-matrix calculations.
- The first-order conditions [Eqs. (5), (15), and (21)] derived by minimizing the objective functions [Eqs. (4), (12), and (20)] in all minimization schemes appear on the right-hand side (rhs) of the update equations [Eqs. (11), (18), and (22)] which means that only when the rhs, has reached zero, the solution reached the global minimum.
- Computation of weight matrices, L matrices and the Tikhonov regularization parameter, requires a prior opinion about the variances of the parameters and data. Here, only the best prior estimates are used, which means that the actual variances of the parameter and data spaces are used in the reconstruction procedure. Variation from the best prior values can be examined also, to observe the effect of priors, but that work is beyond the scope of the present article.
- When spatial priors are used in this study (as well as in most studies), it is assumed that they are perfect. The effect of spatial prior uncertainty on the DOT inverse problem is discussed in Refs. 23, 24, and 26 and is the subject of ongoing study.
- The covariance lengths associated in the weight matrix [GLS-analytical covariance (AC) form, Eq. (28)], and the L matrix [Helmholtz form, Eq. (31)] calculations are chosen to be 10 and 5 mm, respectively. The effect of covariance length on the image reconstruction is discussed in Ref. 26.
- In the LM approach [Eq. (11)], the Jacobian is normalized by its optical properties. Also α was chosen initially to be 1 and it was reduced by a factor of $10^{0.25}$ at every iteration and multiplied by the maximum of the diagonal values of J^TJ. The normalization procedure is described in Ref. 54. Moreover, eight iterations were chosen for all the LM reconstructions, as it has been shown in the literature that after this iteration, error in the optical properties increases for this particular problem and algorithm.[55,56] This inherent instability can be attributed to the fact that J^TJ is not positive definite in DOT.
- For simplicity, all the reconstruction algorithms are tested only in the two-dimensional case. Comparison of three-dimensional reconstructions are left for future investigations.

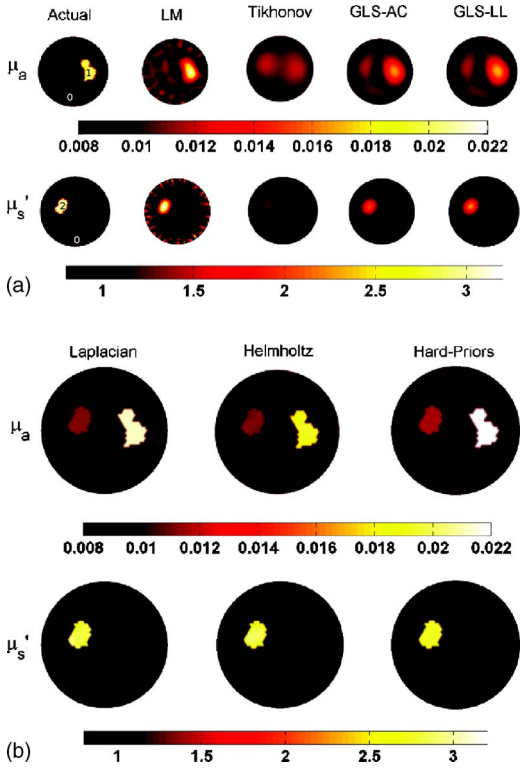

FIG. 3. Reconstruction results (top of the first row, abbreviations are given in Appendix A) are shown using noiseless data (bias calculations) (a) without spatial priors and (b) with spatial priors. The top row contains images of μ_a and bottom row shows μ_s' images.

4. Special cases of GLS minimization

The update equation for the GLS scheme, Eq. (22), turns into the Tikhonov case [Eq. (18)] when $W_\delta = I$ and $W_{\mu-\mu_0}$

TABLE I. Mean and standard deviation of the reconstructed: (a) μ_a and (b) μ_s' values (in mm^{-1}) for different regions [labeled in first column of Fig. 3(a)] recovered with data having 0%, 5%, 10% noise for images shown in Figs. 3–5.

Method	Noise level	Region-0	Region-1	Region-2
Actual	—	0.01	0.02	0.01
LM	*0%*	0.0101±0.001	0.0172±0.0023	0.0105±0.0005
	5%	0.0102±0.0016	0.0125±0.0016	0.0123±0.0011
	10%	0.0103±0.0029	0.0132±0.0026	0.0118±0.0023
Tikhonov	*0%*	0.0102±0.0005	0.0117±0.0003	0.0117±0.0002
	5%	0.0102±0.0004	0.0114±0.0002	0.0112±0.0001
	10%	0.0102±0.0003	0.0108±0.0009	0.0107±0.0005
GLS-AC	*0%*	0.01±0.001	0.015±0.0011	0.0112±0.0003
	5%	0.0101±0.0014	0.0146±0.0012	0.0106±0.0004
	10%	0.0101±0.0013	0.0136±0.0009	0.0112±0.0008
GLS-LL	*0%*	0.01±0.001	0.0152±0.0012	0.0113±0.0003
	5%	0.0101±0.0016	0.0149±0.0015	0.0108±0.0006
	10%	0.0101±0.0016	0.0138±0.0009	0.0112±0.001
Laplacian	*0%*	0.0098±0.0001	0.0212±0.0001	0.0112±0.0001
	5%	0.0098±0.0002	0.0247±0.0001	0.0097±0.0001
	10%	0.0095±0.0001	0.0276±0.0002	0.0157±0.0128
Helmholtz	*0%*	0.0099±0.0001	0.019±0.0002	0.0111±0.0001
	5%	0.0099±0.0002	0.0193±0.0002	0.0099±0.0001
	10%	0.0098±0.0002	0.0174±0.0002	0.0136±0.0001
Hard-Priors	*0%*	0.0099	0.0218	0.0116
	5%	0.0098	0.0218	0.0131
	10%	0.0098	0.018	0.0166

Table 1(a)

Method	Noise level	Region-0	Region-1	Region-2
Actual	—	1.0	1.0	3.0
LM	*0%*	1.0356±0.2364	0.9995±0.0359	2.3758±0.5160
	5%	1.075±0.0357	1.0555±0.3254	1.8215±0.3144
	10%	1.2672±0.9086	1.3111±0.4128	1.7111±0.6112
Tikhonov	*0%*	1.0096±0.0397	1.1153±0.0260	1.1644±0.0251
	5%	1.0111±0.0004	1.0912±0.0189	1.0934±0.0104
	10%	1.0107±0.0216	1.0441±0.0062	1.0416±0.0035
GLS-AC	*0%*	1.0034±0.0688	1.0335±0.0199	1.6838±0.1961
	5%	1.0008±0.0916	1.0670±0.0362	1.6972±0.2037
	10%	0.9987±0.0831	1.0761±0.0343	1.3703±0.0773
GLS-LL	*0%*	1.0022±0.0693	1.03±0.0183	1.7801±0.2573
	5%	0.9998±0.1035	1.0567±0.0329	1.8502±0.3034
	10%	0.9981±0.0947	1.0839±0.0425	1.4271±0.0990
Laplacian	*0%*	0.9918±0.0155	0.9429±0.0015	2.8207±0.0491
	5%	0.9895±0.0202	0.8559±0.0036	3.6931±0.1551
	10%	1.0103±0.0124	0.7447±0.0011	1.9884±0.0096
Helmholtz	*0%*	0.9878±0.0154	1.0518±0.0018	2.7833±0.0854
	5%	0.9813±0.0199	1.1204±0.0081	3.4252±0.1947
	10%	0.9884±0.0121	1.2766±0.01	2.1761±0.0382
Hard-Priors	*0%*	0.9919	0.9266	2.7332
	5%	0.9874	1.0358	2.345
	10%	0.9854	1.3899	1.822

Table 1(b)

$=\lambda L^T L$. Moreover, if one assumes that $\Delta\mu=\mu-\mu_0$, which is equivalent to taking a single step in the iterative procedure, then Eq. (19) maps into Eq. (11) with $\alpha=2\lambda$. Hence, the LM technique can be viewed as a special case of the Tikhonov method, which itself is a special case of the GLS approach. It is important, however, to differentiate LM from the single-step Tikhonov approach because LM requires α to reach zero asymptotically with number of iterations, whereas in the Tikhonov scheme, λ is constant. Moreover, LM does not involve parameters in the objective function.

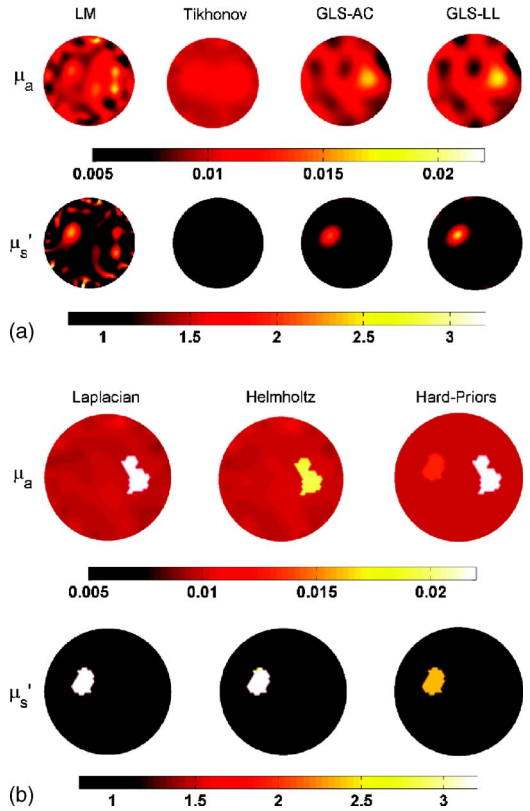

FIG. 4. Reconstruction results (top of the first row, abbreviations are given in Appendix A) are shown using 5% noisy data (a) without spatial priors and (b) with spatial priors. The top row gives images of μ_a and bottom row shows μ_s' images.

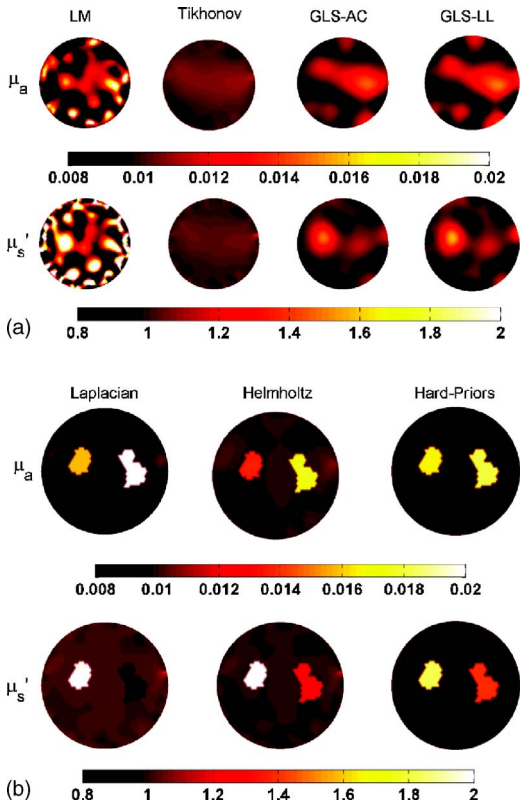

FIG. 5. Reconstruction results (top of the first row, abbreviations are given in Appendix A) are shown using 10% noisy data (a) without spatial priors and (b) with spatial priors. The top row gives images of μ_a and bottom row shows μ_s' images.

5. Stopping criterion

The importance of the stopping criterion in an iterative procedure cannot be ignored. The stopping criterion used in this work is based on the first-order conditions and data-model misfit, which in the limit ensures that the problem has reached the global minima. The iterative procedure is stopped when the $L2$ norm of the data-model misfit (δ) does not improve by more than $10^{-10}\%$ or the $L2$ norm of the first order conditions is less than $10^{-17}\%$. Beyond these values, the round-off error dominates. This stopping criterion is more robust because it involves first-order conditions as well.

IV. TEST PROBLEM

This section provides the details of the test problem considered here. The optical property distributions used for the synthetic data (y, noise added) generation are shown in Fig. 2. The diameter of the domain was 86 mm. The background optical properties were $\mu_a = 0.01\ \text{mm}^{-1}$ and $\mu_s' = 1.0\ \text{mm}^{-1}$. There were two irregular shaped targets, one in μ_a with a contrast of 2:1 to background and one in μ_s' with a contrast of 3:1 relative to the background. A mesh consisting of 4617 nodes (corresponding to 9040 linear triangular elements) was used for the generation of data. Sixteen light collection/delivery fibers were arranged equally spaced on the boundary of the circle, where one fiber was used as the source while all other fibers served as detectors in turn which produced a total of 240 measurements [that is 240 $\ln(A)$ data points and 240 θ data points]. The source was modeled as a Gaussian profile with a full width at half maximum of 3 mm to represent the light applied[57] and was placed at a depth of one transport scattering distance from the tissue boundary.[58] Noise levels of 1%, 3%, 5%, and 10% were added to the modeled data $\{[\ln(A);\theta]\}$ to form the experimental data (y). At the same time, the variances in the data were also computed to be used in the reconstruction algorithms.

The actual reconstructions and forward modeled data computation were performed on different FEM meshes.[59] This mesh has the same diameter (86 mm) with 1785 FEM nodes, which corresponded to 3418 linear triangular elements.[58] The expected distribution of optical properties is given in Fig. 3(a) (first column). Background optical properties were used as initial estimates (μ_0) in the evaluation of reconstruction methods. The number of parameters to be estimated was 3570 (1785 in μ_a and 1785 in μ_s'). The number of data points available for reconstruction was 480 [240 of $\ln(A)$ and 240 of θ]. The dimension of J was 480×3570, W_δ was 480×480, and $W_{\mu-\mu_0}$ was 3570×3570. Optical property distributions were reconstructed from the data without noise (bias calculations) as well as with noise levels of 1%, 3%, 5%, and 10%. The reconstructions are repeated for the case of 3% noise in the data with increasing complexity (targets) in the optical property distributions.

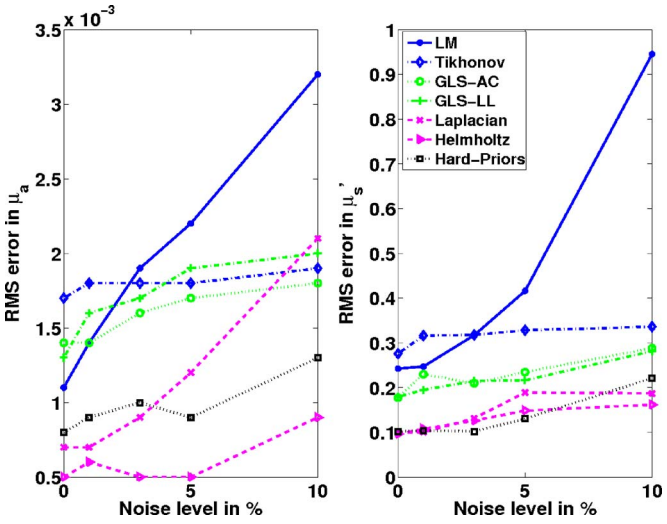

FIG. 6. A plot of the rms error in the estimated optical properties is shown as a function of increasing noise level for all reconstruction techniques.

V. RESULTS AND DISCUSSION

Initially all reconstruction techniques were executed on a data set without noise to estimate the bias. Note that for these calculations the variance was found between the data generated using meshes (described in Sec. IV) with 4617 nodes and 1785 nodes. The results without employing spatial prior information from the reconstruction techniques are given in Fig. 3(a). The first column shows the expected distribution for the 1785 node mesh used in the reconstruction and forward model calculations. The Tikhonov approach failed to recover the contrast. This was primarily due to the choice of λ, which was based on the maximum variance value, which biases the problem to data points that are above the average noise level. Since DOT is known to have a large dynamic range in the data (at least eight orders of magnitude[57]), this choice of λ deemphasize the data points that have low or intermediate variance values. The root-mean-square (rms) errors between the expected and reconstructed optical properties are plotted in Fig. 6. The mean and standard deviation in the reconstructed images for different regions [labeled in first column of Fig. 3(a)] using the reconstruction techniques discussed until now are given in Table I. In the case of no spatial priors, LM gives less bias in μ_a, where as GLS performs better in μ_s'. The bias calculations were repeated with spatial-priors and the reconstruction results are presented in Fig. 3(b). These rms errors in the optical properties are also plotted in Fig. 6. Surprisingly the soft-prior approach (Laplacian and Helmholtz) performed better than the hard-prior strategy. It can also be observed from Fig. 6 and Table I that the usage of spatial priors reduces the bias by at least a factor of 2.

Figure 4(a) shows reconstruction results using data with 5% noise in amplitude without employing spatial priors. Once again the Tikhonov approach fails to recover the contrast. The LM results are dominated by boundary artifacts. Figure 4(b) presents the results from the same data set when spatial priors were employed. Figures 5(a) and 5(b) show similar kinds of effort for the case of data with 10% noise. Note that, for the same choice of the regularization parameter (λ), Tikhonov minimization scheme with spatial priors yielded quantitatively more accurate results compared to without spatial priors case, indicating that the reconstructed image accuracy along with the quality largely depends on the prior information. The rms error in the reconstructed μ_a and μ_s' images are plotted in Fig. 6 as a function of increasing noise level. The rms error using the LM approach increases with increasing noise. GLS techniques perform very well even in the case of 10% noise [Figs. 5(a) and 6]. Among the GLS methods, usage of an analytical covariance form gives better results (≈13% less rms error) in μ_a and the local Laplacian form performs slightly better (≈3% less rms error) in μ_s'. In the case of employment of spatial-priors, it can clearly be seen [from Figs. 4(b), 5(b), and 6 and Table I] that hard-priors perform better in μ_s' reconstruction when the noise level is below 10%. Among the soft-prior results, for μ_a, the rms error linearly increases with increasing noise level in the Laplacian case (Fig. 6). In μ_s' reconstructions, the performance of Laplacian and Helmholtz are comparable, clearly Helmholtz performs slightly better (≈5%) when the noise level is above 3%. Interestingly, the Helmholtz regularization emerges with the lowest rms error in μ_a reconstruction. This is primarily because of the covariance length factor in the Helmholtz form of the regularization matrix [Eq. (31)], which ensures that the optical properties covary within that correlation length (in here it is 5 mm). The same explanation is true for the GLS-analytical covariance form [Eq. (28)], which performs better in μ_a estimation. It is also important to note that in the case of a limited number of wavelengths, Srinivasan *et al.*[60] have shown that 5% error in the optical property estimate (μ_a and μ_s') can lead to approximately 45% error in spectral properties (hemoglobin, water concentrations, oxygen saturation, and scattering estimates) of tissue. Any small improvement in the optical property estimates would be important for improvement in the utility of this type of imaging under practical conditions.

To emphasize the effects of complexity on the reconstruction procedures, a set of simulations were performed with an increasing number of targets. Each target was chosen to be circular with a diameter of 10 mm. The contrast to background optical properties was 2:1. The target locations and corresponding optical properties are shown in the first column of Fig. 7(a). The targets were also labeled from 1 to 4 (background is labeled as 0). The data used in this case have a noise level of 3%. A total of four different reconstructions were performed by adding each target at a time (from 1 to 4). The result of the four target case is shown in Fig. 7. Corresponding mean and standard deviation of the reconstructed optical properties for different regions [labeled in first column of Fig. 7(a)] are given in Table II. Figure 8 contains a plot of rms error in the reconstructed optical properties with increasing number of targets. The rms error increases with increasing number of targets for every reconstruction algorithm. Note that targets 3 and 4 were placed close to the center of the domain, where the sensitivity is low compared

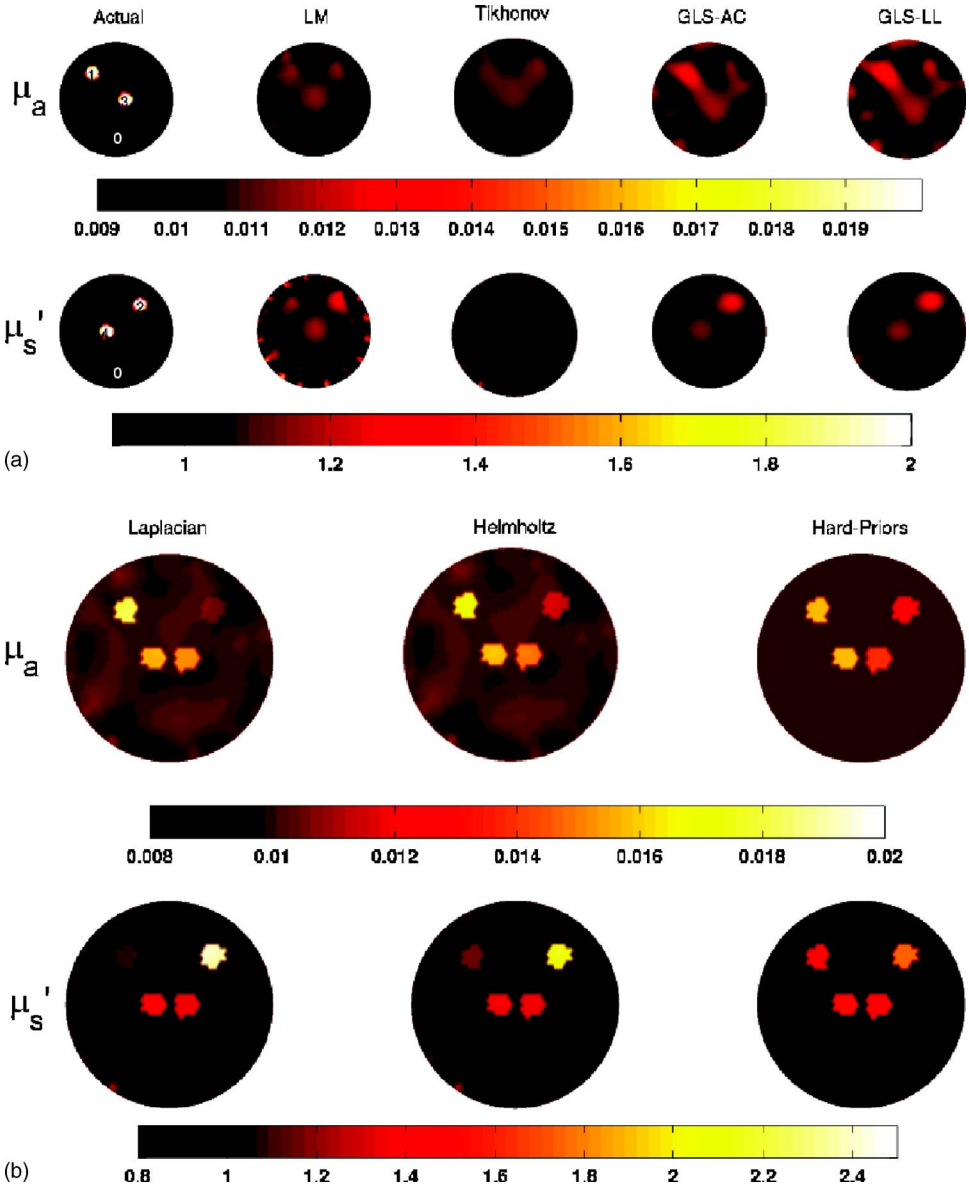

FIG. 7. Reconstruction results (top of the first row, abbreviations are given in Appendix A) are shown using 3% noisy data (a) without spatial priors and (b) with spatial priors for four targets in the tissue as shown. The top row gives images of μ_a and bottom row shows μ_s' images. The actual μ_a and μ_s' with target numbers are given in the first column of (a).

to the periphery.[58] Moreover, increasing the μ_a targets (from 1 to 2, target numbers 1 and 3), caused the rms error to increase by at least 30%. The same is true with the μ_s' targets. In the case of multiple targets, the Helmholtz type of regularization matrix resulted in the least error in both μ_a and μ_s'. Even though the hard-prior case performs very well in terms of lowest rms error for a single target, as the complexity (or number of parameters to be estimated) of the problem increases, it clearly performs poorer than most of the techniques presented (Fig. 8).

Even though the choice of the Tikhonov regularization parameter (λ) given by Eq. (14) is the optimal, the other common way is to use *L*-curve analysis.[61] The *L* curve for DOT is shallow,[62] similar to the estimation problem in electrical impedance tomography, which poses a problem in selection of λ, and is shown to be unreliable in Ref. 59.

Table III gives the computational time per iteration for each of the reconstruction technique (in these two-dimensional cases) on Pentium IV (dual core) 2.8 GHz, 2GB RAM Linux work station. GLS schemes take little more computation time than the Tikhonov minimization, as expected hard priors took the least computation time.

Overall, the inclusion of spatial priors has an important positive effect. The errors in the estimated optical properties are also reduced by at least a factor of 2 with spatial information. The reconstructed images also contain the fine features extracted from conventional imaging modalities. Through the incorporation of the individual variability of the data points and optical parameters (GLS scheme), reconstruction performs better even when the noise level in the data is high. It is also important to note that, as mentioned before, iteration number 8 (which is the best result in terms of lowest rms error) is chosen for rms error calculations in LM approach, after this iteration, the solution becomes unstable. Whereas the rest of the approaches yield stable solutions (error in optical properties did not increase with increasing iterations). When the individual data point variances were not considered (Tikhonov approach), the reconstruction

TABLE II. Mean and standard deviation of the reconstructed: (a) μ_a and (b) μ_s' values (in mm^{-1}) for different regions [labeled in first column of Fig. 7(a)] recovered with data having 3% noise for images shown in Fig. 7.

Method	Region-0	Region-1	Region-2	Region-3	Region-4
Actual	0.01	0.02	0.01	0.02	0.01
LM	0.0101±0.0004	0.0113±0.0001	0.0112±0.0002	0.0111±0.0003	0.011±0.0002
Tikhonov	0.0102±0.0004	0.011±0.0001	0.0112±0.0001	0.0109±0.0001	0.011±0.0001
GLS-AC	0.0102±0.0009	0.0129±0.0003	0.0111±0.0003	0.0114±0.0003	0.0113±0.0003
GLS-LL	0.0102±0.0011	0.0133±0.0004	0.0115±0.0004	0.0113±0.0003	0.0113±0.0002
Laplacian	0.01±0.0002	0.0181±0.0001	0.0105±0.0001	0.0152±0.0001	0.0158±0.0001
Helmholtz	0.01±0.0002	0.0169±0.0001	0.0115±0.0001	0.0149±0.0001	0.0158±0.0001
Hard-Priors	0.01	0.0158	0.0126	0.014	0.0158

(a)

Method	Region-0	Region-1	Region-2	Region-3	Region-4
Actual	1.0	1.0	2.0	1.0	2.0
LM	1.0063±0.0986	1.1333±0.0027	1.24±0.0623	1.1191±0.0396	1.097±0.0366
Tikhonov	1.0051±0.0217	1.0341±0.0019	1.0575±0.0073	1.0321±0.0056	1.0329±0.0043
GLS-AC	0.9993±0.0489	0.9885±0.0139	1.2486±0.0447	1.021±0.0234	1.1184±0.0076
GLS-LL	0.9987±0.0553	0.9764±0.0127	1.2726±0.0596	1.0271±0.0262	1.1422±0.0105
Laplacian	0.9886±0.0163	1.0891±0.0023	2.3799±0.0242	1.3445±0.0043	1.4044±0.0036
Helmholtz	0.9899±0.0164	1.1499±0.0037	2.1122±0.0386	1.3382±0.0079	1.3521±0.0066
Hard-Priors	0.9856	1.3712	1.7319	1.4471	1.5255

(b)

algorithm may not have the ability to recover the contrast in the target. Moreover, simultaneous estimation of both absorption and scattering coefficients causes crosstalk between the two parameter estimates. Even with error-free spatial priors, as the complexity of the estimation problem (or number of targets) increased for a given noise level in the data, the parameter-reduction (hard-priors) technique failed to give the best estimates of the optical properties due to its LM implementation.

VI. CONCLUSIONS

The diffuse optical tomography inverse problem is often solved by Levenberg–Marquardt/modified Tikhonov minimization. A generalized approach for diffuse optical tomographic imaging which incorporates the expected variability of the data noise and magnitude of the optical parameter variation is presented as a structured weight-matrix regularization. It is also shown that Tikhonov minimization and the

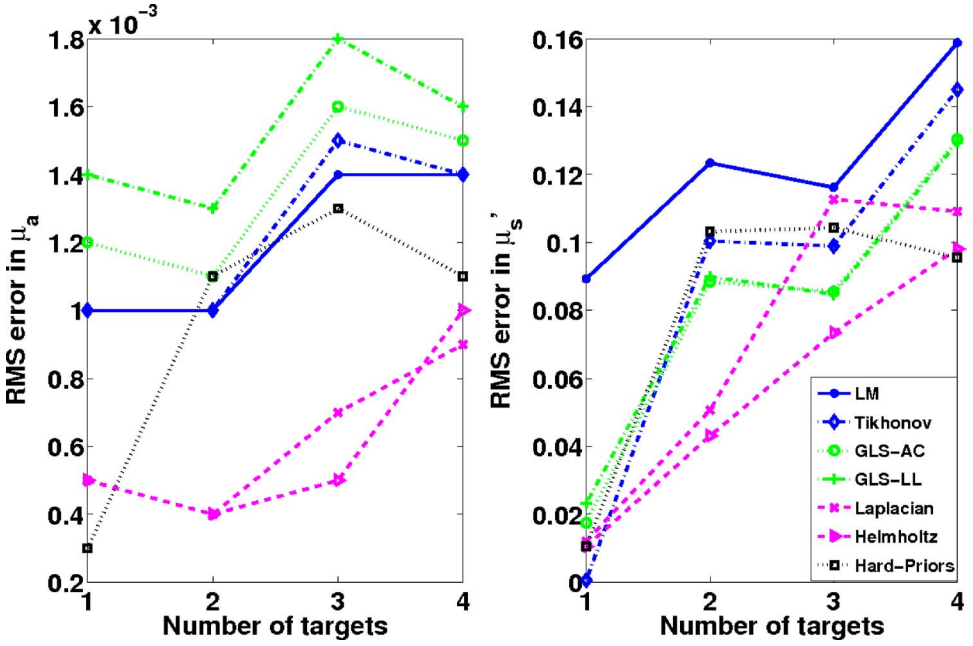

FIG. 8. Plot of the rms error in the estimated optical properties is shown for increasing number of targets with 3% noise in the data for all reconstruction techniques (legend of the figure). Abbreviations used for the techniques are given in Appendix A. The targets used are numbered in the images presented in Fig. 7(a).

TABLE III. Comparison of computation time per iteration for different reconstruction techniques on Pentium IV (dual core) 2.8 GHz, 2 GB RAM Linux work station. the abbreviations used for the reconstruction techniques are given in Appendix A.

Reconstruction method	Computation time per iteration
LM	17.92 Sec
Tikhonov	21.28 Sec
GLS-AC	23.39 Sec
GLS-LL	23.39 Sec
Laplacian	22.78 Sec
Helmholtz	22.78 Sec
Hard-Priors	10.73 Sec

Levenberg–Marquardt approach are special cases of this GLS minimization formalism. Weight matrices that are used in this reconstruction procedure, consisting of the variance and covariance among the data points and optical properties, penalize the solution to match the modeled data with the experimental data more appropriately. This framework can also be used to incorporate structural information, given by MR, computed tomography, or other imaging modalities when the two are acquired on the same tissue volume. Using a test problem, all of these techniques are studied in terms of the data noise level and test field complexity and a uniform comparison was made using the same implementation scheme for each minimization method. Even with highly noisy data, the GLS approach gives meaningful reconstruction results. It appears that the standard Levenberg–Marquardt approach may be unstable for the DOT problem. It is also shown that consideration of the individual variances of data points is the key for an estimation procedure to recover high optical contrast. Employing spatial information reduced the errors in the reconstruction results by at least a factor of 2. Parameter reduction using spatial priors can produce erroneous results when the noise level is high. The same is true for increasing numbers of targets. Future work includes investigating various approaches for incorporating spatial priors into the GLS scheme with experimental data sets. Moreover, a thorough examination of these techniques in three-dimensional case will be taken up as a future investigation. The computer algorithms and test data used in this article (along with some additional material) are given at this web page.[63]

ACKNOWLEDGMENTS

The authors are grateful to Professor Daniel R. Lynch for the useful discussions and valuable comments on this article. P.K.Y. acknowledges the DOD Breast Cancer predoctoral fellowship (BC050309). This work has been sponsored by the National Cancer Institute through Grant Nos. RO1CA78734, PO1CA80139, and DAMD17-03-1-0405.

APPENDIX A: TERMINOLOGY

DOT—diffuse optical tomography.

$\boldsymbol{\mu_a(\mathbf{r})}$—optical absorption coefficient of tissue.

$\boldsymbol{\mu_s'(\mathbf{r})}$—reduced (or transport) scattering coefficient of tissue.

$\boldsymbol{D(\mathbf{r})}$—optical diffusion coefficient of tissue$=1/3[\mu_a(\mathbf{r})+\mu_s'(\mathbf{r})]$.

$\boldsymbol{\mu}$—parameters (generalized) to estimate$=[D(\mathbf{r});\mu_a(\mathbf{r})]$.

$\boldsymbol{\mu_0}$—prior value of the parameters (initial guess, generally obtained from prior calibration of data[45,46]).

$\boldsymbol{F(\mu)}$—forward model.

$\boldsymbol{G(\mu)}$—Modeled data (G—sampled forward model $=S\{F\}$).

$\boldsymbol{A}$—amplitude of the signal.

$\boldsymbol{\theta}$—phase of the signal.

$\boldsymbol{y}$—Measured data$=[\ln(A);\theta]$.

$||X||$—$L2$ norm of vector $X=\sqrt{\Sigma_{i=1}^{N}X_i^2}$.

$\boldsymbol{\delta}$—data-model misfit$=y-G(\mu)$.

$\boldsymbol{W_\delta}$—weight matrix for $\delta=[\mathrm{cov}(\delta)]^{-1}$ (Appendix A-4 of Ref. 47).

$\boldsymbol{W_{\mu-\mu_0}}$—weight matrix for $\mu-\mu_0=[\mathrm{cov}(\mu-\mu_0)]^{-1}$ (Appendix A-4 of Ref. 47).

$\boldsymbol{\lambda}$—Tikhonov regularization parameter.

$\boldsymbol{L}$—Tikhonov regularization matrix.

$\boldsymbol{I}$—identity matrix.

$\boldsymbol{\sigma^2}$—variance.

$\boldsymbol{J}$—Jacobian of the sampled forward model$=\partial G(\mu)/\partial\mu$.

$\boldsymbol{\Omega}$—objective function.

Error—true value-estimated value (prediction).

Bias—difference between the true optical property distribution and estimated optical properties in the case of model generated data (without adding the noise).

Ill-posed—Small changes in the data can cause large changes in the parameters.

Ill-conditioned—the condition number (ratio of largest singular value to smallest singular value) is large, which implies the inverse solution would not be unique.

Ill-determined—(or under-determined) the number of independent equations are smaller than number of unknowns.

Unstability—error gets amplified with iterations.

LM—Levenberg–Marquardt minimization (Sec. III A 1).

Tikhonov—Tikhonov minimization scheme without spatial-priors, $L=I$ (Sec. III A 2).

GLS-AC—generalized least squares minimization scheme (Sec. III A 3) with analytical covariance form for $W_{\mu-\mu_0}$ [Eq. (28)].

GLS-LL—generalized least squares minimization scheme (Sec. III A 3) with local Laplacian form for $W_{\mu-\mu_0}$ [Eq. (29)].

Laplacian—Tikhonov minimization scheme in the case of soft priors (Sec. III B 1) where L approximates Laplacian form, defined by Eq. (30).

Helmholtz—Tikhonov minimization scheme in the case of soft priors (Sec. III B 1) where L approximates Helmholtz form, defined by Eq. (31).

Hard-priors—parameter-reduction technique based on spatial priors (Sec. III B 2).

APPENDIX B: TIKHONOV REGULARIZATION-SINGULAR VALUES

It is interesting to examine Tikhonov regularization from the point of view of singular values. If one rewrites the update equation [Eq. (19)] as

$$[J^T J + \lambda I]\Delta\mu_i = J^T \delta_{i-1} + C, \tag{B1}$$

where $C=\lambda(\mu_{i-1}-\mu_0)$, as it is a constant vector for a chosen iteration i. Singular-value decomposition of J gives

$$J = USV^T, \tag{B2}$$

where U and V are orthonormal matrices containing the singular vectors of J, i.e., $U^TU=I$ and $V^TV=I$. S is a diagonal matrix containing the singular values (S_i) of J. Substituting this into update equation [Eq. (B1)] generates

$$(VS^TU^TUSV^T + \lambda I)\Delta\mu_i = VS^TU^T\delta_{i-1} + C. \tag{B3}$$

Using the orthonormal properties of U and left multiplying by V^T on both sides of Eq. (B3) yields

$$(V^TVS^TSV^T + \lambda V^T)\Delta\mu_i = V^TVS^TU^T\delta_{i-1} + V^TC. \tag{B4}$$

Now using the orthonormal properties of V and rearranging the terms leads to

$$(S^TS + \lambda I)V^T\Delta\mu_i = S^TU^T\delta_{i-1} + V^TC. \tag{B5}$$

Taking the inverse, left multiplying by V and simplifying the result gives

$$\Delta\mu_i = V(S^TS + \lambda I)^{-1}(S^TU^T\delta_{i-1} + V^TC). \tag{B6}$$

Writing Eq. (B7) in the form

$$\Delta\mu_i = VDP, \tag{B7}$$

where $P=(S^TU^T\delta_{i-1}+V^TC)$, a column vector, and D is a diagonal matrix which has the form

$$D_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ \dfrac{1}{S_i^2 + \lambda} & \text{if } i = j. \end{cases} \tag{B8}$$

Similar expressions hold for $L \neq I$ (Ref. 65) in Eq. (18). Considering the case $\lambda=0$, one can clearly see that for an ill-conditioned matrix J, implying some of the singular values are almost zero ($S_i \approx 0$), the inversion becomes unstable (some of the diagonal values of D become infinite). By using Tikhonov regularization, even when $S_i=0$, the inversion procedure is stabilized [Eq. (B8)]. The λ act as a filtering factor, giving the name Tikhonov filtering[65] for this procedure. Moreover, as this λ damps the amplification of the diagonal values of D for smaller values of S_i in Eq. (B8), this is also known as damped least squares minimization procedure.[65]

[a]Electronic mail: phaneendra.k.yalavarthy@dartmouth.edu

[1]A. Gibson, J. C. Hebden, and S. R. Arridge, "Recent advances in diffuse optical tomography," Phys. Med. Biol. **50**, R1–R43 (2005).
[2]S. R. Arridge, "Optical tomography in medical imaging," Inverse Probl. **15**, R41–R93 (1999).
[3]S. R. Arridge and J. C. Hebden, "Optical imaging in medicine: II. Modelling and reconstruction," Phys. Med. Biol. **42**, 841–853 (1997).
[4]S. R. Arridge and M. Schweiger, in "A general framework for iterative reconstruction algorithms in optical tomography, using a finite element method," *The IMA Volumes in Mathematics and Its Applications: Computational Radiology and Imaging, Therapy and Diagnostics*, edited by F. N. C. Börgers (Springer, New York, 1998), pp. 45–70.
[5]M. Schweiger, S. R. Arridge, and I. Nissila, "Gauss–Newton method for image reconstruction in diffuse optical tomography," Phys. Med. Biol. **50**, 2365–2386 (2005).
[6]D. A. Boas, D. H. Brooks, E. L. Miller, C. A. DiMarzio, M. Kilmer, R. J. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," IEEE Signal Process. Mag. **18**, 57–75 (2001).
[7]B. W. Pogue, M. Testorf, T. McBride, U. Osterberg, and K. D. Paulsen, "Instrumentation and design of a frequency-domain diffuse optical tomography imager for breast cancer detection," Opt. Express **1**, 391–403 (1997).
[8]S. Srinivasan, B. W. Pogue, S. Jiang, H. Dehghani, C. Kogel, S. Soho, J. J. Gibson, T. D. Tosteson, S. P. Poplack, and K. D. Paulsen, "Interpreting hemoglobin and water concentration, oxygen saturation and scattering measured *in vivo* by near-infrared breast tomography," Proc. Natl. Acad. Sci. U.S.A. **100**, 12349–12354 (2003).
[9]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).
[10]B. W. Pogue and K. D. Paulsen, "High-resolution near-infrared tomographic imaging simulations of the rat cranium by use of a priori magnetic resonance imaging structural information," Opt. Lett. **23**, 1716–1718 (1998).
[11]V. Ntziachristos, X. H. Ma, and B. Chance, "Time-correlated single photon counting imager for simultaneous magnetic resonance and near-infrared mammography," Rev. Sci. Instrum. **69**, 4221–4233 (1998).
[12]J. P. Kaipio, V. Kolehmanien, M. Vauhkonen, and E. Somersalo, "Inverse problems with structural prior information," Inverse Probl. **15**, 713–729 (1999).
[13]M. Schweiger and S. R. Arridge, "Optical tomographic reconstruction in a complex head model using a priori region boundary information," Phys. Med. Biol. **44**, 2703–2721 (1999).
[14]V. Ntziachristos, A. G. Yodh, M. Schnall, and B. Chance, "Concurrent MRI and diffuse optical tomography of breast after indocyanine green enhancement," Proc. Natl. Acad. Sci. U.S.A. **97**, 2767–2772 (2000).
[15]A. H. Hielscher and S. Bartel, "Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography," J. Biomed. Opt. **6**, 183–192 (2001).
[16]Q. Zhu, N. G. Chen, and S. C. Kurtzman, "Imaging tumor angiogenesis by use of combined near-infrared diffusive light and ultrasound," Opt. Lett. **28**, 337–339 (2003).
[17]A. Li, E. L. Miller, M. E. Kilmer, T. J. Brukilacchio, T. Chaves, J. Stott, Q. Zhang, T. Wu, M. Chorlton, R. H. Moore, D. B. Kopans, and D. A. Boas, "Tomographic optical breast imaging guided by three-dimensional mammography," Appl. Opt. **42**, 5181–5190 (2003).
[18]B. Brooksby, H. Dehghani, B. W. Pogue, and K. D. Paulsen, "Near infrared (NIR) tomography breast image reconstruction with *a priori* structural information from MRI: algorithm development for reconstructing heterogeneities," IEEE J. Sel. Top. Quantum Electron. **9**, 199–209 (2003).
[19]B. Brooksby, S. Jiang, C. Kogel, M. Doyley, H. Dehghani, J. B. Weaver, S. P. Poplack, B. W. Pogue, and K. D. Paulsen, "Magnetic resonance guided near infrared tomography of the breast," Rev. Sci. Instrum. **75**, 5262–5270 (2004).
[20]X. Intes, C. Maloux, M. Guven, B. Yazici, and B. Chance, "Diffuse optical tomography with physiological and spatial *a priori* constraints," Phys. Med. Biol. **49**, N155–N163 (2004).
[21]B. Brooksby, S. Jiang, H. Dehghani, B. W. Pogue, K. D. Paulsen, J. Weaver, C. Kogel, and S. P. Poplack, "Combining near infrared tomography and magnetic resonance imaging to study *in vivo* breast tissue: Implementation of a Laplacian-type regularization to incorporate magnetic resonance structure," J. Biomed. Opt. **10**, 051504:1–10 (2005).
[22]Q. Zhang, T. J. Brukilacchio, A. Li, J. J. Stott, T. Chaves, E. Hillman, T. Wu, M. Chorlton, E. Rafferty, R. H. Moore, D. B. Kopans, and D. A. Boas, "Coregistered tomographic x-ray and optical breast imaging: initial results," J. Biomed. Opt. **10**, 024033:1–024033:9 (2005).
[23]G. Boverman, E. L. Miller, A. Li, Q. Zhang, T. Chaves, D. H. Brooks, and D. Boas, "Quantitative spectroscopic diffuse optical tomography of the breast guided by imperfect a priori structural information," Phys. Med. Biol. **50**, 3941–3956 (2005).

[24]M. Guven, B. Yazici, X. Intes, and B. Chance, "Diffuse optical tomography with *a priori* anatomical information," Phys. Med. Biol. **50**, 2837–2858 (2005).
[25]B. Brooksby, B. W. Pogue, S. Jiang, H. Dehghani, S. Srinivasan, C. Kogel, T. D. Tosteson, J. Weaver, S. P. Poplack, and K. D. Paulsen, "Imaging breast adipose and fibroglandular tissue molecular signatures using hybrid MRI-guided near-infrared spectral tomography," Proc. Natl. Acad. Sci. U.S.A. **103**, 8828–8833 (2006).
[26]P. K. Yalavarthy, H. Dehghani, B. W. Pogue, C. M. Carpenter, S. Jiang, and K. D. Paulsen, "Structural information within regularization matrices improves near infrared diffuse optical tomography" (unpublished).
[27]A. Douiri, M. Schweiger, J. Riley, and S. R. Arridge, "Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information," Meas. Sci. Technol. **18**, 87–95 (2007).
[28]H. Jiang, K. D. Paulsen, and U. 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).
[29]M. Schweiger, S. R. Arridge, M. Hiroaka, and D. T. Delpy, "The finite element model for the propagation of light in scattering media: Boundary and source conditions," Med. Phys. **22**, 1779–1792 (1995).
[30]S. R. Arridge and M. Schweiger, "Photon-measurement density functions. Part 2: Finite-element-method calculations," Appl. Opt. **34**, 8026–8037 (1995).
[31]H. Dehghani, B. W. Pogue, S. P. Poplack, and K. D. Paulsen, "Multiwavelength three-dimensional near-infrared tomography of the breast: Initial simulation, phantom, and clinical results," Appl. Opt. **42**, 135–145 (2003).
[32]H. Dehghani, B. Brooksby, K. Vishwanath, B. W. Pogue, and K. D. Paulsen, "The effects of internal refractive index variation in near infrared optical tomography: A finite element modeling approach," Phys. Med. Biol. **48**, 2713–2727 (2003).
[33]S. R. Arridge and M. Schweiger, "A gradient based optimization scheme for optical tomography," Opt. Express **2**, 213–226 (1998).
[34]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).
[35]K. D. Paulsen and H. Jiang, "Enhanced frequency domain optical image reconstruction in tissues through variation minimization," Appl. Opt. **35**, 3447–3458 (1996).
[36]J. C. Ye, C. A. Bouman, K. J. Webb, and R. P. Milane, "Nonlinear multigrid algorithms for Bayesian optical diffusion tomography," IEEE Trans. Image Process. **10**, 909–922 (2001).
[37]M. J. Eppstein, D. E. Dougherty, D. J. Hawysz, and E. M. Sevick-Muraca, "Three-dimensional Bayesian optical image reconstruction with domain decomposition," IEEE Trans. Med. Imaging **20**, 147–163 (2001).
[38]V. Kolehmainen, "Novel approaches to image reconstruction in diffusion tomography," Ph.D. thesis, Kuopio University, Finland, 2001.
[39]M. S. Zhdanov, *Geophysical Inverse Theory and Regularization Problems*, 1st ed. (Elsevier Science, New York, 2002).
[40]K. Levenberg, "A method for the solution of certain nonlinear problems in least squares," Proc. Soc. Exp. Stress Anal. **2**, 164–168 (1944).
[41]D. W. Marquardt, "An algorithm for least squares estimation of nonlinear parameters," J. Soc. Ind. Appl. Math. **11**, 431–441 (1963).
[42]M. Schweiger, S. R. Arridge, and D. T. Delpy, "Application of the finite element method for the forward and inverse models in optical tomography," J. Math. Imaging Vision **3**, 263–283 (1993).
[43]A. N. Tikhonov and V. A. Arsenin, *Solution of Ill-posed Problems* (Winston and Sons, Washington, 1977).
[44]A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM, 2004).
[45]B. W. Pogue, K. D. Paulsen, H. Kaufman, and C. Abele, "Calibration of near-infrared frequency-domain tissue spectroscopy for absolute absorption coefficient quantitation in neonatal head-simulating phantoms," J. Biomed. Opt. **5**, 185–193 (2000).
[46]S. Jiang, B. W. Pogue, T. O. McBride, M. M. Doyley, S. P. Poplack, and K. D. Paulsen, "Near-infrared breast tomography calibration with optoelastic tissue simulating phantoms," J. Electron. Imaging **12**, 613–620 (2003).
[47]D. R. Lynch, *Numerical Partial Differential Equations for Environmental Scientists and Engineers—A First Practical Course* (Springer, New York, 2005).
[48]P. K. Yalavarthy and D. R. Lynch, "Generalized least-squares minimization for diffuse optical tomography: Use of complex data," NML Lab Report, NML-06-9, Dartmouth College, 2006, http://www-nml.dartmouth.edu/Publications/internal_reports/NML-06-9/.
[49]B. Brandstter, K. Hollaus, H. Hutten, M. Mayer, R. Merwa, and H. Scharfetter, "Direct estimation of Cole parameters in multifrequency EIT using a regularized Gauss–Newton method," Physiol. Meas **24**, 437–448 (2003).
[50]N. Soni, K. D. Paulsen, H. Dehghani, and A. Hartov, "Finite element implementation of Maxwell's equations for image reconstruction in electrical impedance tomography," IEEE Trans. Med. Imaging **25**, 55–61 (2006).
[51]A. Cerussi, N. Shah, D. Hsiang, A. Durkin, J. Butler, and B. J. Tromberg, "*In vivo* absorption, scattering and physiologic properties of 58 malignant breast tumors determined by broadband diffuse optical spectroscopy," J. Biomed. Opt. **11**, 044005 (2006).
[52]V. Chernomordik, D. W. Hattery, D. Grosenick, H. Wabnitz, H. Rinneberg, K. T. Moesta, P. M. Schlag, and A. Gandjbakhche, "Quantification of optical properties of a breast tumor using random walk theory," J. Biomed. Opt. **7**, 80–87 (2002).
[53]H. Dehghani, B. W. Pogue, J. Shudong, B. Brooksby, and K. D. Paulsen, "Three-dimensional optical tomography: Resolution in small-object imaging," Appl. Opt. **42**, 3117–3126 (2003).
[54]S. Srinivasan, B. W. Pogue, H. Dehghani, S. Jiang, X. Song, and K. D. Paulsen, "Improved quantification of small objects in near-infrared diffuse optical tomography," J. Biomed. Opt. **9**, 1161–1171 (2004).
[55]B. W. Pogue, X. Song, T. D. Tosteson, T. O. McBride, S. Jiang and K. D. Paulsen, "Statistical analysis of non-linearly reconstructed near-infrared tomographic images: Part I—Theory and simulations," IEEE Trans. Med. Imaging **21**, 755–763 (2002).
[56]X. Song, B. W. Pogue, T. D. Tosteson, T. O. McBride, S. Jiang, and K. D. Paulsen, "Statistical analysis of non-linearly reconstructed near-infrared tomographic images: Part II—Experimental interpretation," IEEE Trans. Med. Imaging **21**, 764–772 (2002).
[57]T. O. Mcbride, B. W. Pogue, S. Jiang, U. L. Osterberg, and K. D. Paulsen, "A parallel-detection frequency-domain near-infrared tomography system for hemoglobin imaging of the breast *in vivo*," Rev. Sci. Instrum. **72**, 1817–1824 (2001).
[58]P. K. Yalavarthy, H. Dehghani, B. W. Pogue, and K. D. Paulsen, "Critical computational aspects of near infrared circular tomographic imaging: Analysis of measurement number, mesh resolution and reconstruction basis," Opt. Express **14**, 6113–6127 (2006).
[59]B. M. Graham and A. Adler, "Objective selection of hyperparameter for EIT," Physiol. Meas **27**, S65–S79 (2006).
[60]S. Srinivasan, B. W. Pogue, H. Dehghani, S. Jiang, X. Song, and K. D. Paulsen, "Effect of image reconstruction bias upon spectroscopy-based quantification of chromophores in near infrared tomography," Proc. OSA Biomedical Topical Meetings, OSA Technical Digest, WB3:1-3, Optical Society of America, Washington, DC, 2004.
[61]P. C. Hansen, "Analysis of discrete ill-posed problems by means of the *L*-curve," SIAM Rev. **34**, 561–581 (1992).
[62]R. J. Gaudette, D. H. Brooks, C. A. DiMarzio, M. E. Kilmer, E. L. Miller, T. Gaudette, and D. A. Boas, "A comparison study of linear reconstruction techniques for diffuse optical tomographic imaging of absorption coefficient," Phys. Med. Biol. **45**, 1051–1070 (2000).
[63]http://nir.thayer.dartmouth.edu/lscomparison.html.
[64]P. K. Yalavarthy, "Diffuse optical tomographic reconstruction in low-scattering tissue: Development of inversion algorithms based on Monte-Carlo simulation," M.Sc.(Engg.) thesis, Indian Institute of Science, Bangalore, India, 2004.
[65]M. Fedi, P. C. Hansen, and V. Paoletti, "Analysis of depth resolution in potential-field inversion," Geophysics **70**, A1–A11 (2005).