

# Diffuse Optical Tomography with a priori Anatomical Information

Murat Guven<sup>a</sup> Birsen Yazici<sup>a</sup> Xavier Intes<sup>b,c</sup> Britton Chance<sup>b</sup>

<sup>a</sup>Electrical and Computer Engineering Department, Drexel University, Philadelphia, USA

<sup>b</sup>Department of Biophysics and Biochemistry, University of Pennsylvania, Philadelphia, USA

<sup>c</sup>Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, USA

## ABSTRACT

Diffuse optical imaging is an emerging modality that uses Near Infrared (NIR) light to reveal structural and functional information of deep biological tissue. It provides contrast mechanisms for molecular, chemical, and anatomical imaging that is not available from other imaging modalities. Diffuse Optical Tomography (DOT) deals with 3D reconstruction of optical properties of tissue given the measurements and a forward model of photon propagation. DOT has inherently low spatial resolution due to diffuse nature of photons. In this work, we focus to improve the spatial resolution and the quantitative accuracy of DOT by using a priori anatomical information specific to unknown image. Such specific a priori information can be obtained from a secondary high-resolution imaging modality such as Magnetic Resonance (MR) or X-ray. Image reconstruction is formulated within a Bayesian framework to determine the spatial distribution of the absorption coefficients of the medium. A spatially varying a priori probability density function is designed based on the segmented anatomical information. Conjugate gradient method is utilized to solve the resulting optimization problem. Proposed method is evaluated using simulation and phantom measurements collected with a novel time-resolved optical imaging system. Results demonstrate that the proposed method leads to improved spatial resolution, quantitative accuracy and faster convergence than standard least squares approach.

**Keywords:** Diffuse Optical Tomography, inverse problem, anatomical a priori information, spatial resolution

## 1. INTRODUCTION

Diffuse Optical Tomography (DOT) has been the focus of extensive research due to its ability in revealing structural and functional information of biological tissue. Under certain conditions, the propagation of light in human tissue, such as breast, is adequately modeled by the diffusion equation.<sup>2,3</sup> DOT based on diffusion equation is a highly nonlinear problem with inherent low spatial resolution. In this work we formulate the inverse problem within a Bayesian framework with a spatially varying a priori probability density function extracted from a priori anatomical information. Proposed approach leads to a clinically applicable image reconstruction algorithm that can be easily incorporated to a combined imaging system, such as the novel MRI-NIR concurrent imaging system developed at the University of Pennsylvania Optical Imaging and Spectroscopy Laboratory.<sup>28</sup>

### 1.1. Related Work

Incorporation of structural information as the a priori information in inverse medical imaging problems has been suggested in Refs. 19-20. In Refs 15-16, MR images have been used to generate a finite element mesh to reconstruct a simulated rat cranium<sup>15</sup> and a complex head model.<sup>16</sup> Use of MR scans has been employed also for optical breast imaging<sup>17,18</sup> where the “reference medium” is obtained from anatomically accurate optical models of the tissue derived from MR images. In Ref. 28, it is suggested to combine pixels of the same tissue type based on MR prior information to reduce the number of unknowns. While this approach is computationally very viable, it is unable to resolve local variations within the same tissue type.

A Bayesian formulation of the DOT inverse problem has been suggested in Refs. 8-12 to incorporate a priori information. Introducing penalty functions<sup>13</sup> and uniform<sup>4-7</sup> or spatially varying regularization<sup>14</sup> terms in the inverse problem formulation is another way of incorporating a priori information. However, in these studies generic probability density functions or regularization terms rather than specific information about the unknown image have been used. Therefore, the improvements have been relatively limited.

## 1.2. Proposed Method

We have formulated the inverse problem within a Bayesian framework based on a linear forward model. The available high resolution anatomical image is segmented into sub-images that represent major tissue types. Prior probability density function of the image is formulated such that each sub-image is assigned a mean value that need not be equal to its actual optical value; and an unknown variance to allow local variations within sub-images. As a consequence, the overall formulation of the prior information becomes spatially varying, which is specific to the image of interest. The maximum a posteriori (MAP) estimate of the image is computed to determine the spatial distribution of the absorption coefficient of the medium. A conjugate gradient method with dynamically adjusted parameters is utilized to solve the resultant optimization problem. The proposed method is tested extensively using numerical simulations and real experimental data, which have been collected from a phantom using a time-resolved optical imaging system. Our results show that high resolution anatomical image obtained from a secondary modality can significantly improve the spatial resolution, the quantitative accuracy and the speed of DOT image reconstructions.

## 1.3. Organization of the Paper

The paper is organized as follows: Section 2.1 defines the forward model and section 2.2 provides the background on the formulation of the inverse problem. Section 2.3 and 2.4 formulate the data likelihood function and prior information, respectively. Section 2.5 formulates the inverse problem within a Bayesian framework. Section 2.6 describes the optimization procedure. Simulation and phantom experiments are demonstrated in section 3. Conclusion is presented in section 4.

# 2. THEORY

## 2.1. Forward Model

In order to characterize a forward model, we need to model the photon propagation in the scattering medium. In this work, we focus on the reconstruction of absorption coefficients, hence we keep scattering coefficient and consequently the diffusion coefficient constant throughout our calculations. As a result, the use of the following diffusion equation given in frequency domain is sufficient to define the forward model:

$$-\frac{i\omega}{c}\phi(r) - D\nabla^2\phi(r) + \mu_a(r)\phi(r) = A\delta(r_s) \quad (1)$$

$\phi(r)$  represents the spatially varying total field due to the point source  $A\delta(r_s)$  located at  $r = r_s$ .  $\omega$  denotes the frequency,  $c$  is the speed of light and  $i = \sqrt{-1}$ .  $D$  is the spatially invariant diffusion coefficient and  $\mu_a(r)$  stands for the spatially varying absorption coefficient.

The nonlinear relationship that is observed between  $\mu_a(r)$  and  $\phi(r)$  in equation (1) increases the computational burden in the reconstruction of the absorption coefficient distribution. In order to overcome this difficulty, we use a linear model, which is suitable for a medium with absorption coefficient values less than 0.5, such as human breast.<sup>1</sup>

Let  $\phi(r) = f(\mu_a(r))$  be the nonlinear function relating the spatial distribution of absorption coefficient  $\mu_a(r)$ , to the total field  $\phi(r)$ . A linear relationship between  $\phi(r)$  and  $\mu_a(r)$  can be obtained using the perturbation theory,<sup>21-23</sup> by expressing the spatially varying absorption coefficient as  $\mu_a(r) = \mu_{a0} + \delta\mu_a(r)$ , where  $\mu_{a0}$  is the spatially invariant background absorption coefficient and  $\delta\mu_a(r)$  is the spatially varying component.

We have applied Rytov approach<sup>21</sup> using a first order approximation to express the phase of the total field  $\phi(r, r_s)$  as:

$$\Phi(r, r_s) = \Phi_0(r, r_s) + \Phi_{sc}(r, r_s) \quad (2)$$

where the phase term  $\Phi_0(r, r_s)$  is due to the constant background absorption coefficient  $\mu_{a0}$  and the term  $\Phi_{sc}(r, r_s)$  is due to only the differential absorption  $\delta\mu_a(r)$  of the medium.

In the case of DOT, multiple source-detector pairs with multiple frequencies are used. The medium of interest is discretized into  $N$  pixels and the forward problem is expressed in terms of a system of linear equations that relates the diffuse Rytov phase  $\Phi_{sc}(r)$  to the spatially varying component  $\delta\mu_a(r)$  of the absorption coefficient distribution  $\mu_a(r)$ :

$$\begin{bmatrix} \Phi_{sc}^{f_1}(r_{d_1}, r_{s_1}) \\ \vdots \\ \Phi_{sc}^{f_1}(r_{d_m}, r_{s_m}) \\ \Phi_{sc}^{f_p}(r_{d_1}, r_{s_1}) \\ \vdots \\ \Phi_{sc}^{f_p}(r_{d_m}, r_{s_m}) \end{bmatrix} = \begin{bmatrix} W_{11}^{f_1} & \cdots & W_{1N}^{f_1} \\ \vdots & \ddots & \vdots \\ W_{m1}^{f_1} & \cdots & W_{mN}^{f_1} \\ W_{11}^{f_p} & \cdots & W_{1N}^{f_p} \\ \vdots & \ddots & \vdots \\ W_{m1}^{f_p} & \cdots & W_{mN}^{f_p} \end{bmatrix} \times \begin{bmatrix} \delta\mu_a(r_1) \\ \vdots \\ \delta\mu_a(r_k) \\ \delta\mu_a(r_{k+1}) \\ \vdots \\ \delta\mu_a(r_N) \end{bmatrix} \quad (3)$$

where  $\Phi_{sc}(r_{d_i}, r_{s_i})$  is the real part of the diffuse perturbative Rytov phase for the  $i^{th}$  source-detector pair;  $W_{ij}^{f_k}$  is the real value of the weight for the  $j^{th}$  pixel and the  $i^{th}$  source-detector pair at frequency  $f_k$  and  $\delta\mu_a(r_j)$  is the differential absorption coefficient of the  $j^{th}$  pixel. The relation above can be expressed equivalently as follows:

$$y = W \times x \quad (4)$$

where  $y$  is the measurement vector holding the perturbative Rytov phase  $\Phi_{sc}(r, r_s)$  for each source-detector pair,  $W$  denotes the linear forward model (weight matrix) which relates the differential absorption coefficient distribution  $x = [\delta\mu_a(r_1) \ \delta\mu_a(r_2) \ \cdots \ \delta\mu_a(r_N)]^T$  to the measurement vector  $y$ .

## 2.2. Inverse Problem

The inverse problem in imaging, namely the restoration and reconstruction of images can be defined as the general problem of estimating a 3-D field  $x(r)$  from some form of indirect observations related to this field. The problem can be stated as follows:

$$y = T\{x\} + \zeta \quad (5)$$

where  $y$  represents the measurements,  $x$  denotes the spatially varying 3-D field,  $\zeta$  is the additive noise in the measurement and  $T$  is an operator (linear or nonlinear) that relates the unknown field  $x$  to the available measurements  $y$ .

In this work we concentrate on 2D inverse DOT problems where the mathematical relationship between the measurements and the unknown field  $x$  is governed by a linear matrix equation as derived in the previous section. As a consequence, the operator denoted by  $T$  is replaced by the weight matrix  $W$ , as given in equation (4) and  $x$  denotes the unknown spatial distribution of the differential absorption coefficient  $\delta\mu_a$ . In other words,  $x$  represents the unknown image

$$x = [\delta\mu_a(r_1) \ \delta\mu_a(r_2) \ \cdots \ \delta\mu_a(r_N)]^T \quad (6)$$

where  $N$  is the total number of pixels/voxels.

The least squares (LS) solution for this problem is given as follows:

$$\hat{x}_{LS} = \arg \min_x \|y - Wx\|_2^2 \quad (7)$$

where  $\|\cdot\|_2$  denotes the  $L_2$  - norm. LS solution is equivalent to the maximum likelihood (ML) solution with Gaussian conditional probability density function with identity covariance matrix. However due to the ill-posed and/or underdetermined nature of the DOT problem, the LS solution is typically not robust. When the data is noisy, the LS solution will fit to the noise components as well. Since  $W$  is usually ill-conditioned in DOT applications, the noise components will be highly amplified in the LS solution, which is not a desired outcome.

The ill-posedness of the LS formulation can be resolved by incorporating a priori information of the image which in a way stabilizes the solution in the presence of noise. The basic idea is to constrain the solution such that the amplified noise effects are avoided. The Bayesian approach provides a natural framework to incorporate prior information. Maximum a posteriori (MAP) estimate of the image is given by:

$$\hat{x}_{MAP} = \arg \max_x \{\log p(x|y)\} \quad (8)$$

where  $p(x|y)$  is the conditional probability density function of  $x$  given the measurements  $y$ . Equivalently,

$$\hat{x}_{MAP} = \arg \max_x \{\log p(y|x) + \log p(x)\} \quad (9)$$

where  $p(y|x)$  is the data likelihood function and  $p(x)$  is the probability density function of the corresponding image. In the following two subsections, we shall discuss modelling the data likelihood and prior information.

### 2.3. Formulation of the Data Likelihood Function

The measurement vector  $y$  is formed as

$$y = [ y_{11}^{f_1} \quad y_{12}^{f_1} \quad \cdots \quad y_{1D}^{f_1} \quad y_{21}^{f_1} \quad \cdots \quad y_{SD}^{f_1} \quad y_{11}^{f_2} \quad \cdots \quad y_{SD}^{f_F} ]^T \quad (10)$$

where  $S$  is the number of sources,  $D$  is the number of detectors, and  $F$  is the number of frequencies associated with each source. The total number of measurements is then equal to  $P = S \times D \times F$ . Since we are interested in only the real part of the measurements,  $y_{ij}^{f_k}$  are all real.

Photon detection can be modelled using shot noise statistics, which originates from Poisson statistics.<sup>8</sup> With a sufficiently large number of detected photons, the Poisson statistics can be approximated by a Gaussian distribution, uncorrelated between measurements and with a variance proportional to the measurement values. Consequently each measurement  $y_{ij}^{f_k}$  is assumed to be an independent real Gaussian random variable and the data likelihood is given by

$$p(y|x) = \frac{1}{K|\Lambda_\zeta(\lambda)|^{1/2}} \exp\left(-\frac{1}{2}\|y - Wx\|_{\Lambda_\zeta^{-1}(\lambda)}^2\right) \quad (11)$$

where  $\Lambda_\zeta(\lambda)$  is the covariance matrix of size  $P \times P$ ,  $K$  is the normalization constant and  $\|z\|_\Lambda^2 = z^T \Lambda z$ . Under the assumption of statistical independence,  $\Lambda_\zeta(\lambda)$  becomes a diagonal matrix:

$$\Lambda_\zeta(\lambda) = \lambda \Lambda_y = \begin{bmatrix} \lambda \sigma_{\zeta_1}^2 & 0 & 0 & \cdots & 0 \\ 0 & \lambda \sigma_{\zeta_2}^2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \lambda \sigma_{\zeta_p}^2 \end{bmatrix} \quad (12)$$

where  $\sigma_{\zeta_p}^2$  is equal to the corresponding measurement  $y_{ij}^{f_k}$ , hence  $diag(\Lambda_y) = y$  and the unknown term  $\lambda$  is related to the noise variance.

## 2.4. Formulation of the Prior Information

The available a priori anatomical information offers the opportunity to formulate a specific prior probability density function for the unknown image. Prior information comprising the smoothness and the optical boundaries provides additional advantages as compared to generic assumptions which can not take particular structural properties of the image of interest into account. In the case of DOT, such anatomical information can be obtained from concurrently or sequentially acquired Magnetic Resonance (MR) or X-ray images.

We assume that the optical and anatomical edges are the same. Edge information is utilized to decompose the unknown image  $x$  into  $M$  homogeneous sub-images having same prior probability density functions. Relying on the anatomical prior, each sub-image  $x_i$  is assigned a single mean value,  $\mu_i$  and a single variance  $\sigma_i^2$ , which is an unknown parameter. Under these assumptions

$$p(x_i) = \frac{1}{(2\pi\sigma_i^2)^{N_i/2}} \exp\left(-\frac{1}{2\sigma_i^2} \|x_i - \mu_i\|^2\right), \quad i = 1, 2, \dots, M \quad (13)$$

where  $M$  is the number of homogeneous regions and  $N_i$  is the number of pixels/voxels in the  $i^{th}$  sub-image.

As a result the probability density function of the whole image is modelled as follows:

$$p(x|\sigma) = \frac{1}{(2\pi)^{N/2} |\Lambda_x(\sigma)|^{1/2}} \exp\left(-\frac{1}{2} \|x - \mu\|_{\Lambda_x^{-1}(\sigma)}^2\right) \quad (14)$$

where  $\sigma$  is the vector holding the variances that belong to each sub-image

$$\sigma = [ \sigma_1^2 \quad \sigma_2^2 \quad \dots \quad \sigma_M^2 ] \quad (15)$$

and  $\Lambda_x(\sigma)$  is the covariance matrix of the image  $x$ . Further assumption of the uncorrelated sub-images result in an image whose pixels are all uncorrelated. Therefore the covariance matrix,  $\Lambda_x(\sigma)$  becomes a diagonal matrix

$$\Lambda_x(\sigma) = \begin{bmatrix} \sigma_1^2 I_{N_1 \times N_1} & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 I_{N_2 \times N_2} & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_M^2 I_{N_M \times N_M} \end{bmatrix} \quad (16)$$

where  $I_{N_i \times N_i}$  is an  $N_i \times N_i$  identity matrix and  $\mu$  is the vector holding the mean values assigned to each one of the pixels

$$\mu = \left[ \underbrace{\mu_1 \dots \mu_1}_{N_1} \quad \underbrace{\mu_2 \dots \mu_2}_{N_2} \quad \dots \quad \underbrace{\mu_M \dots \mu_M}_{N_M} \right]^T \quad (17)$$

## 2.5. Bayesian Formulation of the Inverse Problem

The MAP equation (9) can be rewritten with  $\lambda$  and  $\sigma$  unknown as follows:

$$\hat{x}_{MAP} = \arg \max_x \max_{\lambda} \max_{\sigma} \{ \log p(y|x, \lambda) + \log p(x|\sigma) \} \quad (18)$$

which is equivalent to minimization of the following cost function

$$\hat{x}_{MAP} = \arg \min_x \min_{\lambda} \min_{\sigma} \{ \log(|\Lambda_{\zeta}(\lambda)|) + \|y - Wx\|_{\Lambda_{\zeta}^{-1}(\lambda)}^2 + \log(|\Lambda_x(\sigma)|) + \|x - \mu\|_{\Lambda_x^{-1}(\sigma)}^2 \} \quad (19)$$

Once the estimate  $\hat{\lambda}$  of the parameter  $\lambda$  and the estimate  $\hat{\sigma}$  of the vector  $\sigma$  are computed, the MAP estimate  $\hat{x}_{MAP}$  takes the form

$$\hat{x}_{MAP} = \arg \min_x \{ \|y - Wx\|_{\Lambda_{\zeta}^{-1}(\hat{\lambda})}^2 + \|x - \mu\|_{\Lambda_x^{-1}(\hat{\sigma})}^2 \} \quad (20)$$

which is a quadratic function of  $x$ . Rewriting equation (20) in vector notations leads to

$$\hat{x}_{MAP} = \arg \min_x \{ [y - Wx]^T \Lambda_\zeta^{-1} [y - Wx] + [x - \mu]^T \Lambda_x^{-1} [x - \mu] \} \quad (21)$$

Taking the derivative of the above expression and setting it equal to zero yields an equation which is linear in  $x$ . Solution of this equation is the image estimate  $\hat{x}$ .

$$[W^T \Lambda_\zeta^{-1} W + \Lambda_x^{-1}] \hat{x}_{MAP} = W^T y - \Lambda_x^{-1} \mu \quad (22)$$

Note that, in regularization framework, equation (22) can be interpreted as a zero-order Tikhonov regularization.<sup>24</sup>

## 2.6. Optimization

Let  $\Psi(x, \lambda, \sigma)$  denote the cost function to be minimized as derived in equation (19)

$$\Psi(x, \lambda, \sigma) = \log(|\Lambda_\zeta(\lambda)|) + \|y - Wx\|_{\Lambda_\zeta^{-1}(\lambda)}^2 + \log(|\Lambda_x(\sigma)|) + \|x - \mu\|_{\Lambda_x^{-1}(\sigma)}^2 \quad (23)$$

A Conjugate Gradient (CG) method with dynamically adjusted parameters is applied to find a direction that minimizes the cost function presented above (Table 1). The  $\alpha$  parameter in the CG method is computed by the exact line search, as shown in table 1. We used Polak-Ribière-Polyak method<sup>25, 26</sup> to calculate  $\beta$ .

---

---

**Table 1.** Conjugate Gradient Algorithm

Initialize the image estimate to zero:  $\hat{x}^{(0)} = \vec{0}$   
Assign the mean value vector  $\mu$  its entries corresponding to each individual region  $x_i$   
Initialize the image variance:  $\hat{\sigma}_i^{(0)} = k \times \mu_i$ ,  $i = 1, 2, \dots, M$  and  $k > 0$   
Initialize the parameter  $\lambda$ :  $\hat{\lambda}^{(0)} = 1$   
Find gradient vector:  $\vec{g}^{(0)} = -\Omega(\hat{x}^{(0)}) = -\frac{\partial \Psi(x, \lambda, \sigma)}{\partial x} \Big|_{x=\hat{x}^{(0)}}$   
Set search direction:  $\vec{d}^{(0)} = \vec{g}^{(0)}$   
Define termination criterion  $\varepsilon$   
Set iteration counter:  $n = 0$   
**repeat**  
Find  $\alpha^{(n)}$  such that:  $\alpha^{(n)} = \arg \min_{\alpha > 0} \Psi(x^{(n)} + \alpha d^{(n)}, \hat{\lambda}, \hat{\sigma})$   
 $x^{(n+1)} = x^{(n)} + \alpha^{(n)} \vec{d}^{(n)}$   
Estimate  $\lambda$  using equation (25)  
Estimate  $\sigma$  using equation (27)  
 $\vec{g}^{(n+1)} = -\Omega(\hat{x}^{(n+1)})$   
 $\beta^{(n+1)} = \max\left(\frac{\vec{g}^{(n+1)T}(\vec{g}^{(n+1)} - \vec{g}^{(n)})}{\vec{g}^{(n)T} \vec{g}^{(n)}}, 0\right)$   
 $\vec{d}^{(n+1)} = \vec{g}^{(n+1)} + \beta^{(n+1)} \vec{d}^{(n)}$   
 $n = n + 1$   
**until**  $\|\alpha^{(n)} \Omega(\hat{x}^{(n)})\| < \varepsilon$

---

---

As the CG method is performed, the unknown parameter  $\lambda$  and the unknown vector of parameters  $\sigma$  are sequentially estimated. More specifically, at each iteration  $\lambda$  and  $\sigma$  are updated<sup>12</sup> as

$$\begin{aligned} \hat{\lambda} &= \arg \min_{\lambda} \Psi(\hat{x}, \lambda, \hat{\sigma}) \\ \hat{\sigma} &= \arg \min_{\sigma} \Psi(\hat{x}, \sigma, \hat{\lambda}) \end{aligned} \quad (24)$$

This is equivalent to computing the Maximum Likelihood (ML) estimate of the parameter  $\lambda$  and the vector  $\sigma$ . The minimization of the cost function with respect to the parameter  $\lambda$  given the updated estimates  $\hat{x}$  and  $\hat{\sigma}$  results

$$\hat{\lambda}_{ML} = \frac{\|y - W\hat{x}\|_{\Lambda_y^{-1}}^2}{P} \quad (25)$$

In order to find an estimate for the vector  $\sigma$ , we make use of the probability density function formulation for each individual sub-image with different mean value  $\mu_i$

$$p(x_i) = \frac{1}{(2\pi\sigma_i^2)^{N_i/2}} \exp\left(-\frac{1}{2\sigma_i^2}\|x_i - \mu_i\|^2\right) \quad (26)$$

hence the variance of each sub-image is calculated as

$$\hat{\sigma}_{i,ML}^2 = \frac{\|\hat{x}_i - \mu_i\|^2}{N_i}, \quad i = 1, 2, \dots, M \quad (27)$$

The estimates  $\hat{\lambda}$  and  $\hat{\sigma}$  are then used to update the image  $x$  according to the CG scheme. The CG algorithm converges to the solution after  $N$  iterations in principle. A general convergence estimate can be given as<sup>27</sup>

$$\|x^{(n)} - x^*\|_H \leq 2\|x^{(0)} - x^*\|_H \left[ \frac{\sqrt{\kappa(H)} - 1}{\sqrt{\kappa(H)} + 1} \right]^n \quad (28)$$

where  $x^*$  is the actual image and  $n$  is the number of iterations. The term  $\kappa(H)$  is called the condition number of  $H$  and  $H = W^T \Lambda_\zeta W + \Lambda_x^{-1}$ .

The CG algorithm will perform well if  $\kappa(H)$  is close to 1. Formulation of the inverse problem using a Bayesian framework with spatially varying prior information leads to a better conditioned system, as a result the convergence rate of the CG algorithm is improved with the incorporation of the prior information.

### 3. SIMULATION AND PHANTOM EXPERIMENTS

We perform two sets of experiments to demonstrate the performance of the proposed method. The first experiment set is purely numerical. We simulated optical measurements using a finite difference code for predetermined geometries (Fig. 1). The second set of experiments is collected from an optical phantom shown in Fig. 6 using a time resolved optical imaging system developed at University of Pennsylvania Optical Imaging and Spectroscopy Laboratory. We shall describe each set of experiments in the following subsections.

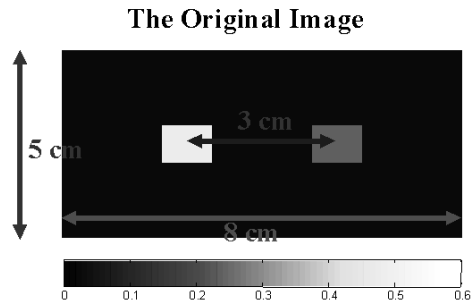
We evaluate the performance of image reconstructions in terms of convergence rate and mean square error given as

$$\epsilon(n) = \frac{1}{N} \sum_{i=1}^M \|\hat{x}_i^{(n)} - x_i^*\|^2 \quad (29)$$

where  $x_i^*$  is the actual  $\delta\mu_a$  value in sub-image  $x_i$ , and  $\hat{x}_i^{(n)}$  is the estimated image in sub-image  $x_i$  after  $n^{th}$  iteration.

#### 3.1. Simulation Experiment

The diffusion equation is solved using finite-difference method for the medium shown in Fig.1. The background absorption coefficient  $\mu_{a0}$  is equal to  $0.05 \text{ cm}^{-1}$  and the diffusion constant  $D \cong 1/3\mu'_s = 1/30 \text{ cm}$ . The  $1 \times 1 \text{ cm}^2$  square inclusions have the optical properties:  $\mu_a = 0.5 \text{ cm}^{-1}$  and  $\mu_a = 0.25 \text{ cm}^{-1}$  that correspond to  $\delta\mu_a$  values of  $0.45 \text{ cm}^{-1}$  for the heterogeneity on the left and  $0.2 \text{ cm}^{-1}$  for the heterogeneity on the right respectively. The scattering coefficient hence the diffusion coefficient of both heterogeneities are equal to those of the background. The medium is discretized into  $20 \times 40$  pixels each  $0.25 \text{ cm}$  by



**Figure 1.** The original image used in the simulation experiment

0.20 cm in length. 17 sources that operate at  $\omega = 150 \text{ kHz}$  are located along x-axis and 33 detectors are located on the opposite side of the medium, as a result the number of total measurements is equal to  $17 \times 33 = 561$ . The image is divided into three sub-images for the formulation of prior information, hence  $\mu = [\mu_1 \cdots \mu_1 \ \mu_2 \ \cdots \ \mu_2 \ \mu_3 \cdots \mu_3]^T$ , where  $\mu_1$  is the mean value of the background image,  $\mu_2$  is the mean value of the square heterogeneity on the left, and  $\mu_3$  is the mean value of the square heterogeneity on the right. Three main cases are examined in the reconstruction procedure and an LS solution is also presented for comparison.

### 3.1.1. Case 1

The reconstruction results are shown in Fig.2 for different values of the mean value vector  $\mu$  and for the LS solution. Results show that our technique achieves much better results when compared to LS solution. Deviations from the exact mean values in the prior information formulation do not lead to significant errors in the image reconstructions, which are superior when compared to the image obtained with LS solution.

Fig. 5(a) shows an “error vs. iteration number” plot for the same case. The plot depicts the superior convergence rate of the suggested technique over LS solution.

### 3.1.2. Case 2

For this case, we add zero-mean Gaussian noise to our measurement vector  $y$ , with Signal-to-Noise Ratio (SNR) values 14 dB and 10 dB respectively where we have defined SNR as

$$SNR = 10 \log_{10} \left( \frac{\text{Mean value of the measurements}}{\text{Standard deviation of the additive noise}} \right) \quad (30)$$

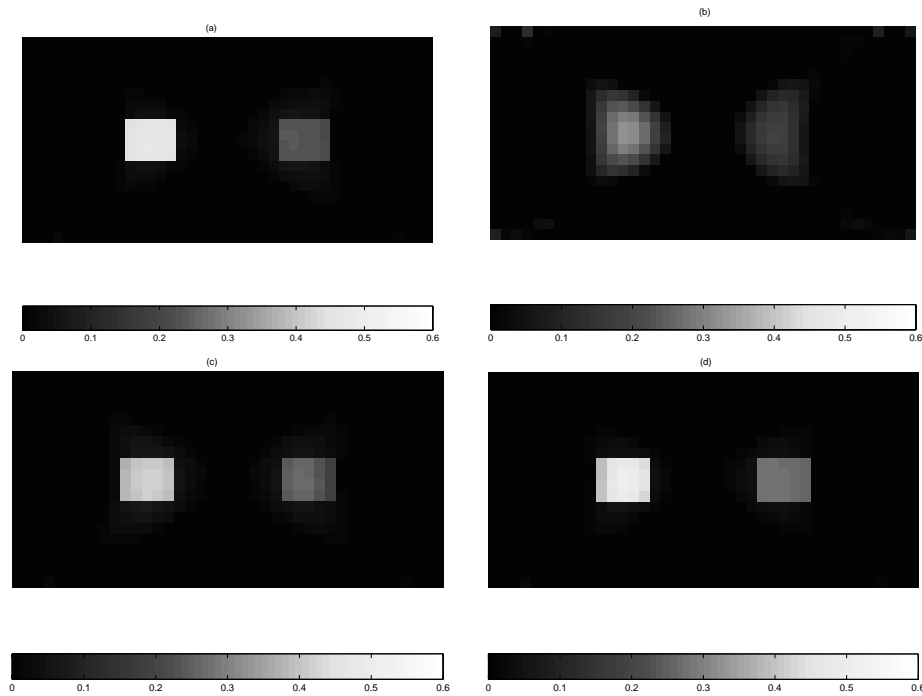
Application of our technique with the formulation of the prior information based on exact mean values results in superior results with a faster convergence rate compared to LS solution. The achieved performance is only slightly worse than the ideal case where there is no noise (Fig.3 and Fig.5(c)).

### 3.1.3. Case 3

For the last case; the prior information is shifted spatially by 1 pixel in both  $+y$  and  $+x$  directions. All of the results are displayed in the absence of noise for 2 different mean value vectors  $\mu$ , together with the LS solution and the reconstruction obtained by our technique under ideal conditions: spatially accurate prior information and exact mean values.

The images obtained with spatially shifted prior information shown in Fig.4(a)-(b) are poor in resolution, but still present quantitatively better results compared to LS solution. The solution converges to almost exact values within the pixels for which the defined prior information match spatially the exact anatomical map. The “error vs. iteration number” curves in Fig. 5(b) also display parallel results, where LS solution is slightly beaten. This relative performance decrease results from the fact that the quality evaluator given in equation (29) is defined for the whole image rather than the sub-images.





**Figure 2.** (a) The reconstruction with prior information formulation with  $\mu_1 = 0$ ,  $\mu_2 = 0.45$ ,  $\mu_3 = 0.2$ , (b) with LS solution, (c) The reconstruction with prior information formulation where  $\mu_1 = 0$ ,  $\mu_2 = 0.3$ ,  $\mu_3 = 0.12$ , (d)  $\mu_1 = 0$ ,  $\mu_2 = 0.55$ ,  $\mu_3 = 0.25$ .

### 3.2. Phantom Experiment

The proposed reconstruction algorithm is applied on the data obtained from a phantom experiment whose setup is shown in Fig.6. An intralipid layer of size  $12 \times 6.5 \text{ cm}^2$  with three circular holes of radius =  $1 \text{ cm}$  filled with blood solution constitute the optical medium. The system is a time-resolved system with 15 sources and 9 detectors. Experimental data is collected at frequencies ranging from 0 to  $238 \text{ MHz}$  with  $\Delta f = 14 \text{ MHz}$ . The blood solution has an approximate absorption coefficient value of  $0.09 \text{ cm}^{-1}$  at the wavelength that the source operates.

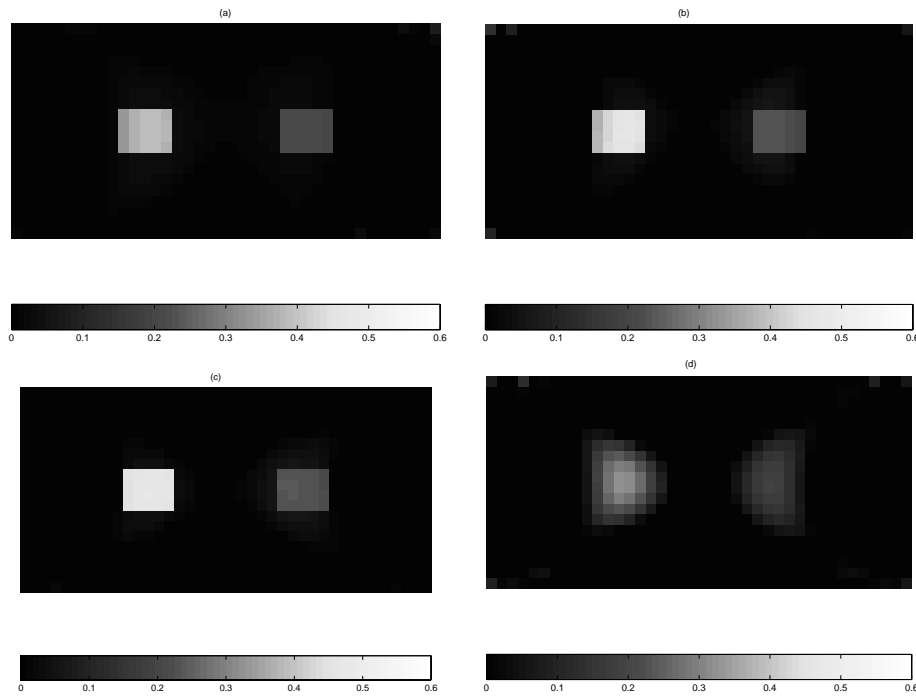
#### 3.2.1. Case 1

The medium is discretized into 48 pixels along x-axis and 26 pixels along y-axis. The image is divided into 3 sub-images with  $\mu = [\mu_1 \cdots \mu_1 \mu_2 \cdots \mu_2 \mu_3 \cdots \mu_3]^T$ , where  $\mu_1$  is the mean value of the background,  $\mu_2$  is the mean value assigned to the regions filled with blood. Due to the spatial mismatch between the available high-resolution prior information and low-resolution image, a transition region is defined with mean value  $\mu_3$ .

For this case  $\mu_1 = 0$ ,  $\mu_2 = 0.08$ , and  $\mu_3 = 0.06$ . The reconstruction results and the “error vs. iteration number” curve are shown in Fig. 7. Due to the resolution mismatch between the prior information and the image, artifacts, which are similar to the ones observed in case 3 of the simulation experiments, appear around the circular heterogeneities. Nevertheless, the algorithm converges to the exact values elsewhere.

#### 3.2.2. Case 2

For the last case the medium is discretized into 120 pixels along x-axis and 65 pixels along y-axis. The image is divided into 3 sub-images.



**Figure 3.** All prior information formulations are based on exact mean values (a) SNR = 10dB, (b) SNR = 14dB, (c) Reconstruction without noise , (d) LS solution without noise.

For this case the mean-value vector is again formed with mean values  $\mu_1 = 0$ ,  $\mu_2 = 0.08$ , and  $\mu_3 = 0.06$ . The reconstruction results and the “error vs. iteration number” curve are shown in Fig. 8. Since the transition region occupies a smaller area, there is a significant increase in the performance of the algorithm compared to case 1. The image has a better spatial resolution with an improved convergence rate.

#### 4. CONCLUSION

We have examined the potential utility of incorporating anatomical prior information within a Bayesian framework and demonstrated our method with simulation and phantom experiments. Results show that incorporating specific structural information improves the DOT image reconstruction in spatial resolution, quantitative accuracy and speed of convergence.

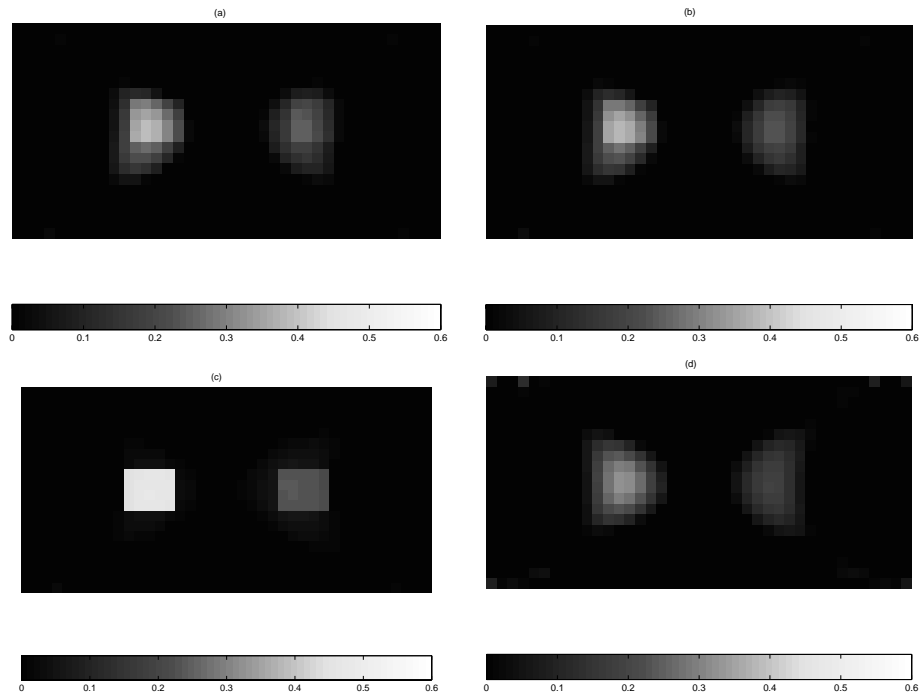
The results obtained in this work are encouraging for future study on combined NIR-MRI breast imaging, where the anatomical information will be extracted from the concurrent MR images.

#### ACKNOWLEDGMENTS

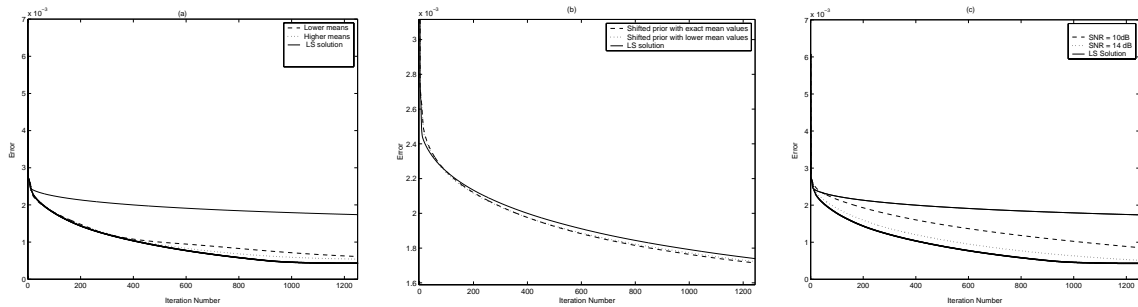
The authors are grateful to M. Holboke for developing the finite difference code. X.I. and B.C. acknowledge partial support from NIH CA 87046.

#### REFERENCES

1. D. A. Boas, “A fundamental limitation of linearized algorithms for diffuse optical tomography,” *Optics Express*, Vol.1, No. 13, pp.404-413, 1997.



**Figure 4.** (a) Reconstruction with spatially shifted prior information and exact mean values (b)with mean values  $\mu_1 = 0, \mu_2 = 0.3, \mu_3 = 0.12$ , (c) reconstruction with spatially accurate prior information and exact mean values, (d) LS solution.



**Figure 5.** Error vs. iteration number plots. Legend is given on the figures. Double-solid line, where available, denotes the error curve for the reconstruction with spatially accurate prior information formulation with exact mean values in the absence of noise. (a) Case 1, (b) Case 3, (c) Case 2.

2. D. A. Boas, M. A. O’Leary, B. Chance, and A. G. Yodh, “Detection and characterization of optical inhomogeneities with diffuse photon density waves: a signal-to-noise analysis,” *Applied Optics*, Vol. **36**, No. 1, pp.75-92 1997.
3. 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 Processing Magazine*, pp.57-75, 2001.
4. K. D. Paulsen and H. Jiang, “Enhanced frequency domain optical image reconstruction in tissues through total variation minimization,” *Appl. Opt.* **35**, pp. 3447-3458, 1996.
5. S. R. Arridge, “Forward and inverse problems in time resolved infrared imaging,” *Medical Optical Tomography*, G. Muller Ed., Vol. IS11, pp. 53-64, SPIE Optical Engineering, Bellingham, WA, 1993.

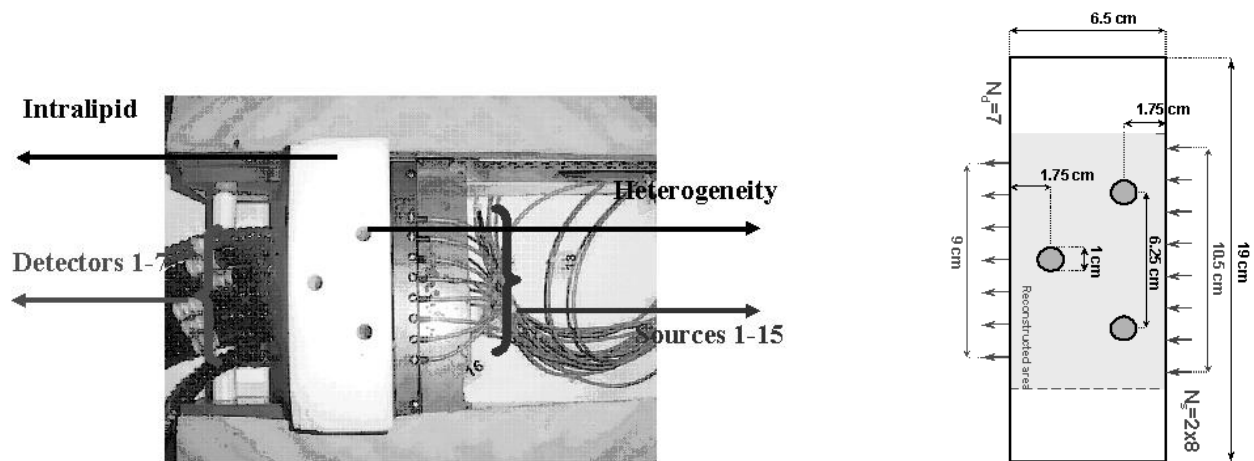


Figure 6. The setup and the phantom that are used in the experiment.

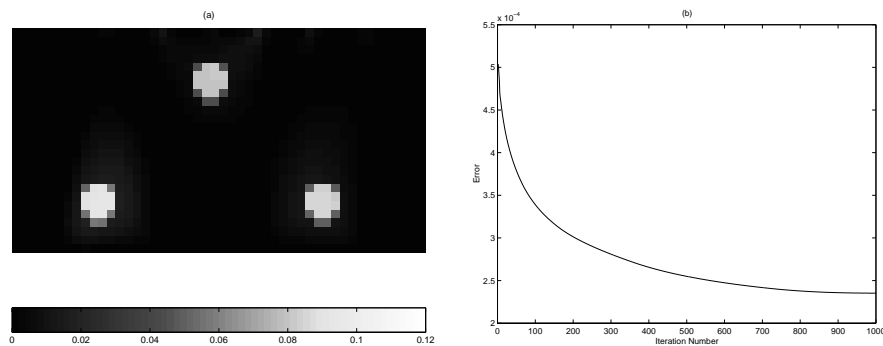
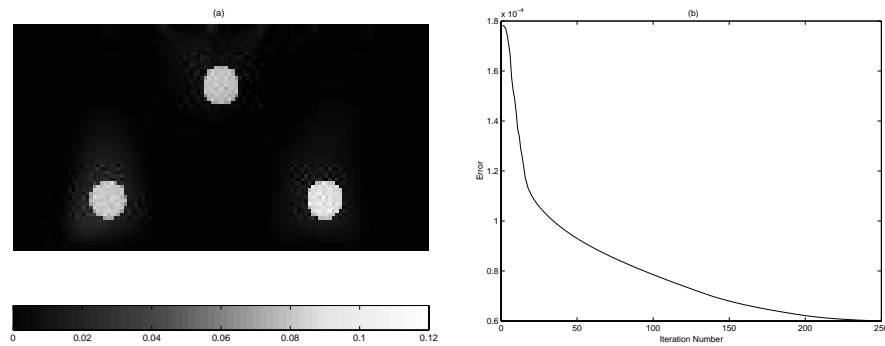


Figure 7. (a) Reconstructed image of pixel size  $26 \times 48$ , (b) The corresponding error vs. iteration number curve

6. H. Jiang, K. D. Paulsen, U. L. Osterberg, B. W. Pogue, and M. S. Patterson, "Optical image reconstruction using frequency- domain data: simulation and experiment," *J. Optical Society America A*, vol. **13**, no. 2, pp. 253-266, 1996.
7. Y. Yao, Y. Wang, Y. Pei, W. Zhu, and R. L. Barbour, "Frequency domain optical imaging of absorption and scattering distributions by a Born iterative method," *J. Optical Society America A*, vol. **14**, no. 1, pp. 325-342, 1997.
8. J. C. Ye, C. A. Bouman, K. J. Webb, and R. P. Millane, "Nonlinear Multigrid Algorithms for Bayesian Optical Diffusion Tomography," *IEEE Trans. on Image Processing*, pp. 909-922, vol. **10**, no. 6, 2001.
9. A. B. Milstein, S. Oh, J. S. Reynolds, K. J. Webb, and C. A. Bouman, "Three-dimensional Bayesian optical diffusion tomography with experimental data," *Optics Letters*, Vol. **27**, pp. 95-97, No. 2, 2002.
10. M. J. Eppstein, D. J. Hawrysz, A. Godavarty, and E. M. Sevick-Muraca, "Three dimensional near infrared fluorescence tomography with Bayesian methodologies for image reconstruction from sparse and noisy data sets," *Proc. Nat. Acad. Sci.*, **99**, 9619-9624, 2002.
11. M. Guven, B. Yazici, X. Intes, B. Chance, and Y. Zheng, "Recursive Least Squares Algorithm for Optical Diffusion Tomography," *Proceedings of the IEEE 28th Annual Northeast*, pp. 273-274, 2002.
12. S. Oh, A. B. Milstein, R. P. Millane, C. A. Bouman, and K. J. Webb, "Source-detector Calibration in Three-dimensional Bayesian Optical Diffusion Tomography", *J. Opt. Soc. Am. A*, Vol. **19**, No. 10, pp. 1983-1993, 2002.



**Figure 8.** (a) Reconstructed image of pixel size  $65 \times 120$ , (b) The corresponding error vs. iteration number curve

13. A. H. Hielscher, S. Bartel, "Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography," *Journal of Biomedical Optics*, Vol. **6**, pp.183-192, No. 2, 2001.
14. B. W. Pogue, T. O. McBride, J. Prewitt, U. L. Osterberg, and K. D. Paulsen, "Spatially Varying Regularization Improves Diffuse Optical Tomography," *Applied Optics*, Vol. **38**, No. 13, pp.2950-2961, 1999.
15. 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," *Optics Letters*, Vol. **23**, No. 21, pp. 1716-1718, 1998.
16. M. Schweiger and S. R. Arridge, "Optical tomographic reconstruction in a complex head model using a priori region boundary information," *Phys. Med. Biol.* **44**, pp. 2703-2721, 1999.
17. J. Chang, H. L. Graber, P. C. Koo, R. Aronson, S. L. Barbour, and R. L. Barbour, "Optical imaging of anatomical maps derived from magnetic resonance imaging using time-independent optical sources," *IEEE Trans. Med. Imaging* **16** pp.68-77, 1997.
18. R. L. Barbour, H. L. Graber, J. Chang, S-L S. Barbour, P. C. Koo and R. Aronson, "MRI-Guided optical tomography: Prospects and computation for a new imaging method," *IEEE Computational Science and Engineering*, pp.63-76, Winter 1995.
19. C-T Chen, X. Ouyang, W. H. Wong, X. Hu, V. E. Johnson, C. Ordonez, and C. E. Metz, "Sensor fusion in image reconstruction," *IEEE Trans. Nuclear Science* **38** pp.687-692, 1991.
20. G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Trans. Med. Imaging* **12** pp.670-680, 1993.
21. M. A. O'Leary, D. A. Boas, B. Chance, and A. G. Yodh, "Experimental images of heterogeneous turbid media by frequency-domain diffusing-photon tomography," *Opt. Lett.* **20**, pp.426-428, 1995.
22. S. R. Arridge, "Photon Measurement Density Functions. Part 1: Analytical Forms," *Appl. Opt.* **34**, pp. 7395-7409, 1995.
23. A. C. Kak, and M. Slaney, "*Principles of Computerized Tomographic Imaging*," pp. 208-218, IEEE, New York, 1988.
24. A. Tikhonov and V. Arsenin, "*Solutions of Ill-Posed Problems*," Winston and Sons, New York, 1977.
25. E. Polak, and G. Ribière, "Note sur la convergence de directions conjuguées", *Rev. Francaise Informat Recherche Opertionelle*, 3e annèe, 16 pp. 35-43, 1969.
26. B. T. Polyak, "The Conjugate Gradient Method in Extremem Problems," *USSR Comp. Math. Math. Phys.*, **9**, pp. 94-112, 1969.
27. C. T. Kelley, "*Iterative Methods for Optimization*," section 1.5.1 SIAM 1999.
28. V. Ntziachristos, A. G. Yodh, M. D. Schnall, and B. Chance, "MRI-guided Diffuse Optical Spectroscopy of malignant and benign breast legions," *Neoplasia*, Volume **4**, Issue 4, pp. 347-354, 2002.