Optical Diffusion Tomography Using
Iterative Coordinate Descent Optimization in a
Bayesian Framework

Jong Chul Ye, Kevin J. Webb and Charles A. Bouman
School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN 47907

Rick P. Millane
Computational Science and Engineering Program
Purdue University, West Lafayette, IN 47907

April 2, 1999

Correspondence to: Kevin J. Webb
School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN 47907-1295
webb@ecn.purdue.edu
Tel.: 765-494-3373, Fax: 765-494-2706

OCIS Codes: 100.3010, 100.3190, 100.6950, 170.3010, 290.3200

To appear in JOSA-A
Abstract – Frequency-domain diffusion imaging uses the magnitude and phase of modulated light propagating through a highly scattering medium to reconstruct an image of the spatially dependent scattering and/or absorption coefficient in the medium. In this paper, the inversion algorithm is formulated in a Bayesian framework and an efficient optimization technique is presented for calculating the maximum a posteriori image. In this framework, the data are modeled as a complex Gaussian random vector with shot noise statistics, and the unknown image is modeled as a generalized Gaussian Markov random field. The shot noise statistics provide correct weighting for the measurement and the generalized Gaussian Markov random field prior enhances the reconstruction quality and retains edges in the reconstruction. A localized relaxation algorithm, the iterative coordinate descent algorithm, is employed as a computationally efficient optimization technique. Numerical results show that the Bayesian framework with the new optimization scheme out-performs conventional approaches in both speed and reconstruction quality.

Keywords – optical diffusion tomography, Bayesian image reconstruction, shot noise statistics, generalized Gaussian Markov random field, iterative coordinate descent algorithm
I Introduction

Optical diffusion tomography has generated considerable recent interest, and its potential for imaging in highly scattering media such as tissue, as an alternative to x-ray or ultrasonic tomography, has been demonstrated [1, 2, 3]. In this technique, with a red or near infra-red light source, the detected transmitted light is used to reconstruct the absorption and/or the scattering properties of the medium as a function of position. The low energy optical radiation presents significantly lower health risks than x-ray radiation. Also, suitable sources and detectors are relatively inexpensive, making such an instrument considerably less expensive than computed tomography (CT) and magnetic resonance imaging (MRI) systems. Furthermore, in an optical imaging application, a host of spectroscopic techniques can be applied. Given these desirable features, optical imaging has become a candidate for the screening of soft tissue tumors.

Two common approaches for optical diffusion tomography are a frequency domain method using an amplitude modulated optical source, where a coherent measurement is performed at the modulation frequency, and a time domain method using short optical pulses, where temporal gating methods are employed. We focus in this paper on the frequency domain method.

An accurate model for the propagation of photons through tissue can be obtained from transport theory [4]. This model ignores the optical phase, treating the photons as particles. A solution can be obtained by means of Monte-Carlo methods [5], which describes individual photon paths, or by means of the diffusion approximation [1]. While the Monte-Carlo method can model the photon path more accurately, the diffusion approximation is sufficiently accurate in highly scattering media such as tissue, and provides a computationally tractable forward model. Therefore, we use the diffusion equation as our forward model.

The inverse problem of reconstructing the absorption and/or scattering coefficients from measurements of scattered light is highly nonlinear because of the nonlinear coupling between the coefficients and the photon flux in the diffusion equation. To facilitate the computation of the unknown coefficients, several approaches attempt to locally linearize the inverse problem. For this class of problems, the Newton-Raphson (NR) method has been commonly used with the Levenberg-
Marquardt procedure based on a Taylor series expansion. A Levenberg-Marquardt method based on a variational formulation of the time domain diffusion equation has been applied to time domain data [6]. In this technique, the moments of the photon current at the detector locations are used as data in the inversion algorithm. In frequency domain methods, using sinusoidally modulated light, a similar algorithm has been developed using the magnitude and phase of the modulation envelope as data [7]. The forward frequency domain diffusion equation has been further approximated and formulated as an integral equation, and the Born approximation used to derive a linear integral equation [3]. In this approach, the equivalent wave number, which is a nonlinear function of absorption and scattering coefficients, is computed, and the unknown absorption and the scattering coefficients are recovered from the reconstructed equivalent wave number. These approaches use heuristic linearization to obtain the gradient necessary for the iterative solutions rather than rigorous optimization. Recently, we provided a detailed analysis of previous approaches (especially the integral equation approach), from a standard nonlinear optimization point of view, and showed that the conventional integral equation approach [3] does not utilize the correct Fréchet differential for the absorption and the scattering parameters [8, 9]. A new integral equation approach was derived and shown to have superior performance over previous methods [8]. Furthermore, the iterative method used in conventional approaches [6, 7, 3], which imposes a penalty on the $L_2$ norm of the update at each iteration, tends to over-smooth edges in the image or produce excessively noisy-images, depending on the control parameter value. This is because the $L_2$ penalty term for the new update is not a regularization in the Tikhonov sense [10], but a “trust region” constraint for a nonlinear least squares problem [11, 12].

The artifacts due to poor regularization can be reduced by incorporation of a priori information using a Bayesian framework. In this framework, the maximum a posteriori (MAP) reconstruction is often computed by maximizing the posterior distribution. This Bayesian approach has been applied in many image restoration and reconstruction problems [13, 14]. More recently, Bayesian (or other regularization) methods have been applied to nonlinear inverse problems such as microwave imaging, impedance tomography, and optical imaging [15, 7, 16]. The individual approaches have
differed both in terms of the prior model (or stabilizing functional) used and the optimization algorithms employed to compute the MAP reconstruction. For example, Paulsen and Jiang [15] added a quadratic regularization term to their previous formulation [7] to stabilize the reconstruction. Each iteration of the optimization performed a linearization (similar to the Born approximation), followed by a full matrix inversion to solve the linearized problem. The computational complexity of this method is very high since $O(N^3)$ complex multiplications are required at each iteration, where $N$ is the number of image pixels. Saquib, Hanson and Cunningham proposed a more computationally efficient algorithm for the time domain diffusion problem in which each iteration alternates a linearization step with a single step of a conjugate gradient algorithm [16]. Arridge and Schweiger adapted the method of the reference [16] for the frequency domain diffusion tomography problem [17]. However, their method is computationally expensive because the line search used for each conjugate gradient update requires repeated evaluations of the forward model. Perhaps the research of Carfantan, Mohammad-Djafari and Idier is most closely related to ours, in that they used exact single-site updates for each pixel [18]. They observed that the single-site optimization had rapid convergence in terms of number of iterations. However, each iteration of this method is computationally expensive, requiring $O(N^2)$ complex multiplications. Finally, we note that previous Bayesian approaches have not incorporated the physics of the measurement noise into the Bayesian framework.

In this paper, we formulate the frequency domain optical diffusion inverse problem in a Bayesian framework and derive the maximum a posteriori (MAP) estimate for the reconstruction. Although the methodology we describe can in principle be applied in the general case of unknown absorption and scattering coefficients, for simplicity we focus on the estimation of the absorption coefficient under the assumption that the scattering coefficient is known.

Similar to the previous approach of Saquib, Hanson and Cunningham [16], we use the generalized Gaussian Markov random field (GGMRF) as a prior model of the unknown parameters. This results in stable and edge preserving regularization for the optical diffusion imaging problem. In addition, we incorporate a model for the detection statistics derived from the physics of the
measurement system. Since the dynamic range of the data is usually very large due to the source-detector geometry and strong attenuation in the medium, large intensity measurements may be over-weighted in the inversion procedure. Previously, heuristic scaling based on the time average of the measurements [17] has been used to ameliorate this problem. We address the scaling problem by deriving a model for the noise based on “shot noise” detection statistics [19]. This model provides a natural scaling for the data, which is based on the square root of the time average of the measurements. We believe that our approach is superior because it is based on the accuracy of the actual measurements and is extendible to a wide variety of physical measurement systems.

Another contribution of this work is the introduction of a new optimization technique that we call the iterative coordinate descent Born (ICD/Born) method. Each iteration of the ICD/Born method consists of a linearization step using the Born approximation, followed by a single pass of the ICD algorithm [20, 21]. Since the computational complexity of ICD/Born is $O(N)$, the method requires much less computation per iteration than the exact single pixel update algorithm [18], the Gaussian elimination technique for total variation minimization [15], and the conventional distorted Born iterative method (DBIM) [22].

Our numerical results indicate that the ICD/Born method together with the Bayesian framework yields accurate and fast reconstructions from synthetic data.

II Notation

The following notation is used in this paper.

$K$ = number of sources

$M$ = number of detectors

$P(=KM)$ = number of measurements

$N$ = number of image pixels

$\Omega; \partial \Omega$ = image domain; boundary of $\Omega$

$d_m$ = position vector of the $m$-th detector

$s_k$ = position vector of the $k$-th point source
\( r; r_i \) = position vector in \( \Omega \); position of the \( i \)-th voxel/pixel in \( \Omega \)

\( y_{km} \) = complex measurement at the \( m \)-th detector due to the \( k \)-th source

\( y \) = measurement vector, \( y = [y_{11}, y_{12}, \cdots, y_{1M}, y_{21}, \cdots, y_{KM}]^T \)

\( x \) = vector of unknown absorption coefficients, \( x = [\mu_a(r_1), \cdots, \mu_a(r_N)]^T \)

\( \psi_k(r, t) \) = time domain solution of diffusion equation (photon flux) at \( r \) due to the \( k \)-th source

\( \phi_k(r, \omega) \) = frequency domain solution of diffusion equation at \( r \) due to the \( k \)-th source

### III The Optical Diffusion Tomography Problem

In a highly scattering medium with low absorption, such as soft tissue in the 650-1300 nm wavelength range, the photon flux density is accurately modeled by the diffusion equation[1, 23, 24]. More specifically, let \( \psi_k(r, t) \) be the photon rate per unit area generated at time \( t \) and position \( r \in \Omega \) due to a modulated point source of light at position \( s_k \in \Omega \). Then, \( \psi_k(r, t) \) is given by the time domain diffusion equation as

\[
\frac{1}{c} \frac{\partial}{\partial t} \psi_k(r, t) - \nabla \cdot D(r) \nabla \psi_k(r, t) + \mu_a(r) \psi_k(r, t) = S(t) \delta(r - s_k),
\]

where \( c \) is the speed of light in the medium, \( S(t) \) is the time varying photon source density, and \( D(r) \) is the diffusion constant given by

\[
D(r) = \frac{1}{3(\mu_a(r) + \mu'_s(r))},
\]

where \( \mu_a(r) \) is the absorption coefficient, and \( \mu'_s(r) \) is the reduced scattering coefficient. The reduced scattering coefficient is defined by \( \mu'_s(r) = (1 - g)\mu_s(r) \) where \( \mu_s(r) \) is the scattering coefficient and \( g \) is the mean cosine of the scattering angle. Note that \( \psi_k(r, t) \) takes on real positive values since it corresponds to the number of photons which pass through a unit surface area per unit time.

Practical systems based on time domain measurements have been implemented [25, 1], but these systems tend to be expensive and noise sensitive. In order to circumvent these problems, we adopt a frequency domain approach to the optical diffusion problem [26, 7]. To do this, we assume that the light source is amplitude modulated at a fixed angular frequency \( \omega(\neq 0) \), so
that $S(t) = \text{Re}[1 + \beta \exp(j\omega t)]$, where $\beta$ is the modulation depth. At the detector, the complex modulation envelope is then measured by demodulating the in-phase and quadrature components of the measured sinusoidal signal $\psi_k(r, t)$. This technique allows low noise narrow-band heterodyne detection [19]. By taking the Fourier transform of (1), the partial differential equation that governs the complex modulation envelope $\phi_k(r, \omega)$ becomes

$$\nabla \cdot D(r) \nabla \phi_k(r, \omega) + (-\mu_a(r) + j\omega/c)\phi_k(r, \omega) = -\beta \delta(r - s_k).$$

(3)

In the frequency domain imaging approach, (3) is used as a forward model.

Throughout this paper, a two-dimensional (2-D) domain is considered, but the approach can be generalized to 3-D. Figure 1 illustrates the typical experimental scenario that we consider in this paper. The region to be imaged is denoted by $\Omega$ and is surrounded by $K$ point sources uniformly distributed around the 2-D boundary at positions $s_k \in \Omega$, and $M$ detectors interspersed between the sources at positions $d_m \in \Omega$. The reduced scattering coefficient $\mu'_s(r)$ is assumed to be known for all points $r \in \Omega$, but the absorption coefficient $\mu_a(r)$ in $\Omega$ is unknown. Our objective is then to determine the values of $\mu_a(r)$ from the measured values of $\phi_k(d_m, \omega)$. Note that the complex measurements of $\phi_k(d_m, \omega)$ must be made for each source and detector combination. The specific relationship between $\phi_k(r, \omega)$ and the physical measurements is described in Appendix I.

The domain $\Omega$ is discretized into $N$ pixels, where the position of the $i$-th pixel is denoted by $r_i$ for $1 \leq i \leq N$. The set of unknown absorption coefficients is then denoted by the vector $x$ where

$$x = [\mu_a(r_1), \cdots, \mu_a(r_N)]^T.$$

(4)

Using this notation, we may express the forward model as a vector valued function $f(x)$. The function $f(x)$ then takes on the values of a $P = KM$ dimensional column vector with elements given by

$$f(x) = [f_1(x), f_2(x), \cdots, f_P(x)]^T = [\phi_1(d_1, \omega), \phi_1(d_2, \omega), \cdots, \phi_1(d_M, \omega), \phi_2(d_1, \omega), \cdots, \phi_K(d_M, \omega)]^T.$$

(5)

The elements of $f(x)$ then represent the “exact” values of the flux density for the assumed values of the absorption coefficient $x$. 

9
The measurements of the complex envelope $\phi_k(d_m, \omega)$ for source $k$ and detector $m$ are denoted by $y_{km}$. We also organize the measurements as a single column vector of length $P = KM$,

$$\mathbf{y} = [y_{11}, y_{12}, \ldots, y_{1M}, y_{21}, \ldots, y_{KM}]^T$$  \hspace{1cm} (6)

Note that there is a measurement corresponding to each source-detector pair.

The estimation of $\mathbf{x}$ from the measurement vector $\mathbf{y}$ is a classic example of an ill-posed inverse problem in which the solution is often underdetermined, non-unique, and noise sensitive. To address this problem, we formulate the solution in a Bayesian framework by computing the MAP estimate for $\mathbf{x}$ given $\mathbf{y}$. The MAP estimate, $\hat{\mathbf{x}}_{MAP}$, is given by

$$\hat{\mathbf{x}}_{MAP} = \arg \max_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y})$$

$$= \arg \max_{\mathbf{x}} \{ \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}) \},$$  \hspace{1cm} (8)

where $p(\mathbf{x}|\mathbf{y})$ is the posterior density of $\mathbf{x}$ given $\mathbf{y}$, $p(\mathbf{y}|\mathbf{x})$ is the conditional probability density of $\mathbf{y}$ given $\mathbf{x}$, and $p(\mathbf{x})$ is the prior density for the image. The density $p(\mathbf{y}|\mathbf{x})$ models the physical properties of the measurement system, while the prior density $p(\mathbf{x})$ models image characteristics such as smoothness that one would expect in the solution. In the following sections, we derive an expression $p(\mathbf{y}|\mathbf{x})$ based on our modeling assumptions, and we adopt a prior model $p(\mathbf{x})$ which enforces smoothness while allowing for abrupt changes in the estimate of $\mathbf{x}$.

## IV Measurement Model

In this section, we derive an expression for the distribution $p(\mathbf{y}|\mathbf{x})$ in terms of the photon flux density $\phi_k(d_m, \omega)$. The details of the model are derived in Appendix I, and are based on a shot noise model for the detected signal. In this model, the measurements are normally distributed with a mean equal to the exact (noiseless) measurement and a variance proportional to the exact measurement at a modulation frequency of zero (DC). The density function for a single datum is given by (Appendix I)

$$p(y_{km}|\mathbf{x}) = \frac{1}{2\pi \alpha |\phi_k(d_m, \omega)|} \exp \left[ -\frac{|y_{km} - \phi_k(d_m, \omega)|^2}{2 \alpha |\phi_k(d_m, \omega)|} \right],$$  \hspace{1cm} (9)
where \( \alpha \) is a constant determined by the modulation depth and the physical characteristics of the detector. We assume that the noise signals are independent for all source-detector pairs, so that the covariance matrix \( C \) for the data vector \( y \) is diagonal and given by

\[
C_{ii} = \alpha |\phi_k(d_m, \omega)| \approx \alpha |y_{km}|, \quad \text{where } i = M(k - 1) + m.
\]

To simplify notation, we define the diagonal matrix \( \Lambda \) by

\[
\Lambda = \frac{1}{2} C^{-1}.
\]

The data likelihood is then given by

\[
p(y|x) = \frac{1}{\pi^N |\Lambda|^{-1}} \exp \left[ -||y - f(x)||_\Lambda^2 \right]
\]

where \( ||z||_\Lambda^2 = z^H \Lambda z \), and \( H \) denotes the Hermitian transpose.

V The GGMRF Prior Model

In this section, we describe the prior model \( p(x) \) that we use for the absorption image \( x \). In many image reconstruction problems, the Markov random field (MRF) model [27] has proved useful in describing spatial correlations between neighboring pixels. MRFs have the property that the conditional distribution of a pixel given all other pixels is only a function of the pixel’s neighbors, i.e.,

\[
p(x_i|x_j, i \neq j) = p(x_i|x_{\partial i}),
\]

where \( \partial i \) denotes the pixels neighboring pixel \( i \). If the density function is constrained to be strictly positive, then a random field is a MRF if and only if its density function has the form of a Gibbs distribution (Hammersley-Clifford theorem [27]). A Gibbs distribution is any distribution with a density function that can be put in the form

\[
p(x) = \frac{1}{\sigma^N z(p)} \exp \left[ -\frac{1}{p} u(x/\sigma, p) \right],
\]
where $\sigma$ and $p$ are constants representing scale and shape parameters for the distribution, and $z(p)$ is a normalizing constant.

We further assume that the function $u(x/\sigma, p)$ has the form [28]

$$u(x/\sigma, p) = \sum_{(i,j)\in\mathcal{N}} b_{i-j} \rho \left( \frac{x_i - x_j}{\sigma}, p \right),$$

where $\mathcal{N}$ is the set of all neighboring pixel pairs, and $\rho(\cdot, \cdot)$ is a potential function which assigns a cost to differences between neighboring pixel values. A wide variety of functions for $\rho(\cdot, \cdot)$ have been used [29, 30, 13, 14]. However, we use here the generalized Gaussian MRF (GGMRF) model because it is both convex and scale invariant [30]. The convexity of the potential function of the GGMRF leads to continuous or stable MAP estimates, and the scale invariant property of the GGMRF potential functions eliminates the necessity of choosing an edge threshold which is often required for non-Gaussian potential functions [30]. For the GGMRF model, the density function for $x$ is given by

$$p(x) = \frac{1}{\sigma^N z(p)} \exp \left[ -\frac{1}{p\sigma^p} \sum_{(i,j)\in\mathcal{N}} b_{i-j} |x_i - x_j|^p \right], \quad 1 \leq p \leq 2.$$

Furthermore, since the absorption must be positive we also impose the constraint

$$x_i \geq 0, \quad i = 1, \ldots, N.$$

VI ICD/Born Optimization Technique

Referring to (8), (12), (16), and (17), the MAP estimate for $x$ is given by

$$\hat{x}_{MAP} = \arg \min_{x \geq 0} \left[ \|y - f(x)\|_A^2 + \frac{1}{p\sigma^p} \sum_{(i,j)\in\mathcal{N}} b_{i-j} |x_i - x_j|^p \right].$$

To compute the MAP reconstruction, we must perform the optimization (18). We choose to use the ICD algorithm [21] for a number of reasons. First, it has been shown that ICD updates work well with non-Gaussian prior models [21]. Second, the ICD algorithm is easily implemented with
A positivity constraint. On the other hand, a drawback of the conjugate gradient method is the
difficulty of incorporating positivity constraints [21].

The ICD algorithm is implemented by sequentially updating each pixel of the image. After
every pixel has been updated, the procedure is repeated starting from the first pixel again. We
refer to a single update of every pixel in the image as a “scan”. The ICD algorithm therefore
consists of a number of scans until some convergence criterion is satisfied. Each scan consists of N
pixel updates. Each pixel update is chosen to minimize the MAP cost function, so that the update
\( \hat{x}_i \) of the absorption value of the \( i \)-th pixel is given by

\[
\hat{x}_i = \arg \min_{\tilde{x}_i \geq 0} \left[ \| y - f(\tilde{x}_i) \|_\Lambda^2 + \frac{1}{\rho \sigma^p} \sum_{j \in \mathcal{N}_i} b_{i-j} |\tilde{x}_i - x_j|^p \right], \tag{19}
\]

where \( \tilde{x}_i = [x_1, x_2, \ldots, x_{i-1}, \tilde{x}_i, x_{i+1}, \ldots, x_N]^T \) and \( \mathcal{N}_i \) is the set of pixels neighboring pixel \( i \). Note
that \( x_i \) is replaced by \( \hat{x}_i \) before the next pixel update. However, a direct approach [18] to the
update equation of (19) is very computationally expensive due to the highly nonlinear nature of
the forward model \( f(x) \). Furthermore, each evaluation of the function \( f(x) \) requires the solution of
the full partial differential equation (3) for each source.

The computational inefficiency is overcome by using the Born approximation at the beginning
of each scan, and call this approach the ICD/Born algorithm. We use the integer \( n \) to index the
scans of the algorithm, and \( x^n \) denotes the image after the \( n \)-th scan. At the beginning of the
\((n+1)\)-th scan, the approximation

\[
\| y - f(x) \|_\Lambda^2 \simeq \| y - f(x^n) - f'(x^n) \Delta x \|_\Lambda^2 \tag{20}
\]

is used where \( \Delta x = x - x^n \) and \( f'(x) \) represents the Fréchet derivative of \( f(x) \), which for the
discretized problem is the $P \times N$ complex matrix

$$f'(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_P(x)}{\partial x_1} & \cdots & \frac{\partial f_P(x)}{\partial x_N} \end{bmatrix}$$

and $f'(x^n)$ denotes the Fréchet derivative computed for the absorption parameter estimate $x^n$. We have shown previously [8] that the Fréchet differential for the $m$-th detector due to the $k$-th source, can be approximated by

$$\lim_{N \to \infty} \sum_{i=1}^{N} \frac{\partial \phi_k(d_m, \omega)}{\partial x_i} \Delta x_i = \int_{\Omega} dr \ g(d_m, r, \omega) \phi_k(r, \omega) \left\{ -1 + \frac{\mu_a^n(r)}{\mu_a(r)} + \mu_s(r) \right\} \Delta \mu_a(r)$$

where $\Delta \mu_a(r) = \mu_a(r) - \mu_a^n(r)$ denotes the change in $\mu_a$, and $g(d_m, r, \omega)$ is the Green’s function for the frequency domain diffusion equation (3). For the discretized domain, $\Delta x_i = \mu_a(r_i) - \mu_a^n(r_i)$

and the elements of the matrix $f'(x)$ in (21) are given by

$$\frac{\partial \phi_k(d_m, \omega)}{\partial x_i} = g(d_m, r_i, \omega) \phi_k(r_i, \omega) \left\{ -1 + \frac{\mu_a^n(r_i)}{\mu_a(r_i)} + \mu_s(r_i) \right\} A,$$

where $A$ is the pixel area. After the $n$-th scan, $f(x^n)$ and $f'(x^n)$ are calculated by computing $g(d_m, r, \omega)$ and $\phi_k(r, \omega)$ of (23) using a linear PDE solver for the diffusion equation (3) with the $n$-th estimate of absorption coefficient, $x^n$.

In an ICD scan, each pixel is updated in turn, and the new value $\hat{x}_i$ is given by

$$\hat{x}_i = \arg \min_{x_i \geq 0} \left[ ||\mathbf{y} - f(x^n) - [f'(x^n)]_{i\times}(\hat{x}_i - x^n_i)||_A^2 + \frac{1}{p\sigma^p} \sum_{j \in N_i} b_{i-j} |\hat{x}_i - x_j|^p \right],$$

where $[f'(x^n)]_{i\times}$ is the $i$-th column of the Fréchet derivative matrix. To compute the solution to (24), we express the first term as a quadratic function of $\hat{x}_i$ to obtain the expression

$$\hat{x}_i = \arg \min_{x_i \geq 0} \left[ \theta_1(\hat{x}_i - x_i^n) + \theta_2(\hat{x}_i - x_i^n)^2 + \frac{1}{p\sigma^p} \sum_{j \in N_i} b_{i-j} |\hat{x}_i - x_j|^p \right],$$
where $\theta_1$ and $\theta_2$ are given by

$$
\begin{align*}
\theta_1 &= -2Re \left( [f'(x^n)_{*i}]^H \Lambda \epsilon(i) \right) \\
\theta_2 &= 2[f'(x^n)_{*i}]^H \Lambda f'(x^n)_{*i},
\end{align*}
$$

and the error vector $\epsilon(\cdot)$ for the first pixel is

$$
\epsilon(1) = y - f(x^n),
$$

and for subsequent pixels is updated as

$$
\epsilon(i + 1) = \epsilon(i) - [f'(x^n)]_{*i}(\hat{x}_i - x^n_i).
$$

Solution of equation (25) requires minimization of a one-dimensional function. We achieve this by solving for the root of the derivative of the expression in the square brackets in (25), i.e.,

$$
\theta_1 + \theta_2(\hat{x}_i - x^n_i) + \frac{1}{\sigma^p} \sum_{j \in N_i} b_{i-j} |\hat{x}_i - x_j|^{p-1} \text{sign}(\hat{x}_i - x_j) = 0.
$$

This root-finding procedure is done using a half-interval search [31] because the function in (29) is monotone decreasing [21]. The lower and upper bounds of the update $\hat{x}_i$ in (25) are first computed from the observation that [21]

$$
\min \left\{ x^n_i - \frac{\theta_1}{\theta_2}, \ x_j \in N_i \right\} \leq \hat{x}_i \leq \max \left\{ x^n_i - \frac{\theta_1}{\theta_2}, \ x_j \in N_i \right\}.
$$

Then, these bounds are used as end points for initiating a half interval search. Since the half-interval search has guaranteed exponential convergence, one can either choose to terminate after a fixed number of iterations or after a fixed tolerance is reached [31]. Figure 2 summarizes the complete procedure used to implement the ICD/Born algorithm.

VII Computational Complexity

In this section, we compare the computational cost of our ICD/Born method with the conventional DBIM. This is done by counting the number of complex multiplications (referred as cflops) required
for one update of the whole image. Recall that one complete iteration or scan of the ICD/Born method implies a single update of each pixel in the image formed by the unknown absorption coefficients. Table 1 summarizes the computational complexity analysis, and Table 2 gives comparisons in the number of complex multiplications for two typical cases.

We first analyze the ICD/Born algorithm. The computational cost for evaluating an element of the Fréchet derivative $f'(x^n)$ in (21) consists of calculating the Green’s function of (3), the flux $\phi_k(r_i, \omega)$, and evaluating (23). Evaluation of $\phi_k(r_i, \omega)$ requires $K$ forward solutions of (3), one solution for each source location. The evaluation of the Green’s function in (23) implicitly involves placing a source at each grid point $r_i$ and computing the flux at each detector point $d_m$, which requires $N$ forward solutions. This computational cost can be dramatically reduced by using reciprocity [32] of $\phi_k(r_i, \omega)$ in (3) so that

$$g(d_m, r_i, \omega) = g(r_i, d_m, \omega).$$

Hence, we can place a source at each detector point, thereby requiring only $M \ll N$ forward solutions. Furthermore, to compute the Green’s function and flux for the new parameter $x^{n+1}$, we can use the corresponding values from the previous iteration as initial guesses in an iterative solver. Since the ICD/Born algorithm typically produces a small change $(x^{n+1} - x^n)$, this approach gives rapid convergence by reducing the number of iterations required for the forward solver. Using a standard five point Laplacian discrete approximation, each forward iteration of (3) by an iterative linear solver (e.g. SOR [33]) requires $5N$ complex multiplications. Therefore, the total multiplications required for the Green’s function and flux update is $5(K + M)LN$, where $L$ is the number of iterations required for the linear solver. In addition, $(2MK + 2)N$ multiplications are required to fill the Fréchet derivative matrix in (21).

The solution of (29) is usually computationally inexpensive since the neighborhood $\mathcal{N}_i$ typically contains only a few pixels. Therefore, the computation is dominated by the calculation of $\theta_1$ and $\theta_2$ in (26). Since the number of columns of $f'(x^n)$ is $MK$ and $\Lambda$ is diagonal, the number of multiplications required to compute an individual value of $\theta_1$ is $2MK$. Similarly, the update of $\theta_2$ is also $2MK$. In addition, the update of the projected error vector $e(i + 1)$ requires $MK$
multiplications. Therefore, the total number of multiplications required for the pixel update is $5MKN$. This results in a total computational cost for the ICD/Born method of $5(M + K)LN + (7MK + 2)N$ complex multiplications per iteration.

To compare this result with the computational cost for DBIM, let us first briefly explain the DBIM algorithm. In the DBIM, a new parameter estimate is computed from the perturbation equation using a trust region constraint, where the Fréchet derivative $f'(x^n)$ is again defined as in (21) and (23). Each iteration of the DBIM also requires the computation of the Fréchet derivative and a regularized inverse. If we use the same linear solver for the computation of the Green’s function and flux, and if we also use the reciprocity relation (31), the number of multiplications required for the Fréchet derivative update is again $5(M + K)LN$, where the iteration number $L$ must typically be chosen to be larger due to the greater change in $x$ for each DBIM iteration. To fill in the Fréchet derivative (21) a total of $(2MK + 2)N$ multiplications is again required. The computational cost for the regularized inverse using QR-decomposition is at least $2(MK)^2N - (MK)^3/3$ [33]. Therefore, the total number of multiplications for a complete update of the DBIM is $5(M + K)LN + 2((MK)^2 + MK + 1)N - (MK)^3/3$.

Table 1 summarizes the computational complexity results. The bottom row of the table lists the dominant (i.e. highest order) terms for each method. Notice, that the essential difference is that the DBIM contains a $(MK)^2N$ term, whereas ICD only contains a $MKN$ term. This difference is particularly significant as the number of sources and detectors grows.

The computational advantage of ICD/Born algorithm over the DBIM becomes clear when we use actual numbers. Table 2 shows the number of complex multiplications required when we use $M = 12$ detectors, $K = 12$ sources, and $N = 1089$ pixels (e.g. a $33 \times 33$ discretization domain). As discussed earlier, the number of the forward linear solver iterations for the two algorithms will vary. We used the MUDPACK (multigrid software for elliptic partial differential equations) libraries [34] as our forward solver, and controlled the number of iterations by setting the relative error tolerance in MUDPACK [34]. Therefore, the number of forward solver iterations varies dynamically with respect to $x$. For the results shown in Table 2, we pick typical iteration numbers $L_{DBIM}$ and $L_{ICD}$
that we believe to be reasonable. These choices are justified later by Table 3 which shows the actual
CPU time per iteration.

Two cases are considered, one which uses the same number of forward iterations for both
methods, and one which uses different numbers of forward iterations. The results listed in Table 2
show that for these two cases the computation using the ICD/Born method is reduced by a factor
of 13 and 19 as compared to the DBIM. The bottom row of the table lists the estimated complex
multiplications for a three dimensional reconstruction scenario with similar geometry. We can see
that for this case the computational speedup of the ICD/Born method is dramatic.

VIII Numerical Results

Simulation results are presented here to assess the performance of the new algorithm. The entire
region, \( \Omega \) (including the homogeneous background) is considered unknown. A total of 12 sources
and 12 detectors are located uniformly over the boundary of a 8 cm \( \times \) 8 cm domain, as shown in
Figure 1.

To illustrate the effect of noise on the stability of the algorithm, random noise with a complex
Gaussian distribution is added to the measured flux data. Referring to (9), the signal-to-noise ratio
(SNR) of the current measurement for the \( m \)-th detector, and the \( k \)-th source can be represented
by the magnitude of the flux and is given by

\[
\text{SNR}_{mk} = \frac{|E[i]|^2}{\sigma_o^2} \approx \frac{1}{2\alpha} |\phi_k(d_m, \omega)|, \tag{32}
\]

where the constant \( \alpha \) is defined in Appendix I. (Notice that the SNR increases proportionally to
the magnitude of the flux at the detector.) To determine the value of \( \alpha \) in (32), we consider a
1 MHz bandwidth detection system \((B = 1 \text{ MHz})\) and a 200 MHz modulation frequency with a
modulation depth \( \beta = 1.0 \). For detector parameters, a common laboratory PMT, the R928 from
Hamamatsu, is considered. Here the cathode responsivity \( \kappa \) is 68 mA/W [35] and \( \gamma \) is set to 1 mm\(^2\).
The noise is generated by a Gaussian random number generator and added to the signal as

\[
\text{Re}[y_{mk}] = \text{Re}[\phi_k(d_m, \omega)] + \sqrt{\alpha|\phi_k(d_m, \omega)|} \times N(0, 1)
\]
\[ Im[y_{mk}] = Im[\phi_k(d_m, \omega)] + \sqrt{\alpha|\phi_k(d_m, \omega)|} \times N(0, 1), \]  

(33)

where \( N(0, 1) \) is a zero mean Gaussian random variable with unit variance. To compare the overall accuracy of the reconstructions, we introduce the normalized root-mean-square error (NRMSE) of the reconstructed image as a function of iteration, defined as

\[
NRMSE_n = \sqrt{\frac{\sum_{i=1}^{N} (\mu_a^n(r_i) - \mu_a(r_i))^2}{\sum_{i=1}^{N} (\mu_a(r_i))^2}},
\]  

(34)

where \( \mu_a^n(r_i) \) is the reconstructed value of the absorption coefficient at mesh location \( r_i \) at the \( n \)-th iteration, and \( \mu_a(r_i) \) is the correct value. For the simulations, the image is discretized on a \( 33 \times 33 \) grid. In all the simulations, the scattering coefficient is set to a constant value \( \mu_s = 10.0 \text{ cm}^{-1} \). The simulation was performed on a Sun Ultra Sparc 30.

Figure 3(a) shows the original \( \mu_a \) image used for the first numerical experiment. Figure 3(b) shows the DBIM reconstruction result [8]. The reconstruction shows a high noise level and incorrect peak heights. This is due to the \( L_2 \) norm used in each update as a trust region constraint, which does not have a noise smoothing effect. For the GGMRF prior model, we used an eight-point neighborhood system, with \( b_{i,j} = (2\sqrt{2} + 4)^{-1} \) for nearest neighbors and \( b_{i,j} = (4\sqrt{2} + 4)^{-1} \) for the diagonal neighbors. The MAP reconstructions were computed by running the ICD/Born algorithm with a positivity constraint. The stopping criterion was fixed CPU time (60 seconds).

Figure 3(c) shows the MAP reconstruction for a Gaussian prior \( (p = 2, \sigma = 1.0 \times 10^{-3}) \), while Figure 3(d) gives the reconstruction for \( p = 1.1 \) and \( \sigma = 2.3 \times 10^{-4} \). The reconstruction using the Gaussian prior \( (p = 2) \) reduces the background noise compared to the reconstruction by the DBIM, but suffers from some smoothing of the edges as the price for noise suppression. The boundaries can be sharpened by using a larger value of \( \sigma \), but at the expense of larger noise artifacts. Sharper edges and good noise suppression are obtained for \( p = 1.1 \) (Figure 3(d)). Here, due to the edge preserving nature of the GGMRF prior, the edges are more noticeable in the reconstruction while suppressing the noise.

Figure 4 shows a plot of the NRMSE versus CPU time for this simulation. With a single user environment, the CPU times were about 80% of the wall clock time. The average CPU time
for a single iteration of each algorithm is also given in Table 3. The difference in computational speed between DBIM and ICD/Born was predicted by the computational complexity analysis of Section VII (see Table 1 and Table 2). Notice that more CPU time is required for ICD/Born with $p = 1.1$ than for $p = 2.0$. This is due to the increased number of the half-interval searches for $p = 1.1$. In Figure 4, we can also observe that the DBIM results eventually diverge. This is again due to the insufficient regularization of the DBIM [22].

Figure 5 shows a plot of the cost function in (18) versus iteration for the ICD/Born method with $p = 2.0$ and $p = 1.1$ for the above simulation. Note that the ICD/Born is very stable and has reasonably fast convergence. The convergence behavior of ICD/Born is also consistent with the previous single site update algorithm [18].

The problem of choosing $p$ and $\sigma$ is complex, and a maximum likelihood (ML) estimation technique for these parameters can be used [28]. However, in this paper we choose values empirically that give the best results. For $\sigma$, the range of $0 < \sigma < 0.02 cm^{-1}$ was considered (note the units of $\sigma$). In Figure 6, the reconstructions obtained using the GGMRF prior with $p = 1.1$ for different values of $\sigma$ are shown. A small value of $\sigma = 2.84 \times 10^{-5} cm^{-1}$ results in an over-smoothed reconstruction, as observed in Figure 6(b). Note that in this case the amplitude of the inhomogeneity at the center is under-estimated. A larger value of $\sigma = 1.52 \times 10^{-2} cm^{-1}$ produces background noise in the reconstruction, even though the edges are improved and the value of the inhomogeneity at the center is improved (Figure 6(d)). Figure 6(c) with $\sigma = 2.31 \times 10^{-4} cm^{-1}$ shows the best trade-off between the smoothness of the image and the edge improvement.

Figure 7 shows a variety of more complicated absorption images. Reconstructions using the ICD/Born method for these images are shown in Figure 8. The parameters used for the reconstructions and the final NRMSE are given in the caption to Figure 8. In each case, the reconstructions are accurate both quantitatively and qualitatively, demonstrating that our new algorithm performs well on more complex images.
IX Conclusion

We have formulated the optical diffusion inverse problem in a Bayesian framework and implemented a maximum a posteriori (MAP) reconstruction of the absorption coefficient. The Bayesian framework enables incorporation of prior knowledge of the unknown parameters as well as detection statistics. The shot noise detection statistics we described here provide a natural weighting for the measurement data. As a prior model of the unknown parameter, we use the generalized Gaussian random field (GGMRF). This has resulted in stable and edge preserving regularization for this inverse problem. As an optimization technique for the Bayesian framework, we developed a new method which combines the iterative coordinate descent (ICD) and the Born approximation which we called the ICD/Born method. This method significantly reduces the computational complexity as compared to methods such as the conventional distorted Born iterative method (DBIM). Numerical simulations show that the Bayesian framework together with the ICD/Born method significantly improves the quality of reconstructions.

APPENDIX I: Detector Noise Model

We develop in this appendix a noise model for the measured data \( y_{km} \). A measurement will sense the optical power leaving the scattering domain, the photon current \( J \), which is given by Fick’s law [4] as \( J = -D \nabla \phi \). The average (DC) optical power \( P \) at a detector is therefore

\[
P = |J^+(d_m)|A_e
\]

\[
= D(d_m)|\nabla \phi_k(d_m, \omega)|A_e,
\]

where \( J^+(d_m) \) is the outward photon current from the domain at the detector, \( A_e \) is the effective detector aperture, and \( D(d_m) \) is the diffusion constant at the receiver location \( d_m \).

We consider an absorbing boundary condition as follows (Figure 9) [4]. The detectors are located on a boundary, denoted by \( \partial \Omega_d \), inside the computational boundary \( \partial \Omega \), such that all the incident photons are absorbed at the detector boundary. This means that the inward photon current at the detectors \( J^-(d_m) = 0 \). This boundary condition is implemented by setting the flux \( \phi = 0 \) on the
computational boundary ∂Ω at a distance $2.131D(d_m)$ from $\partial \Omega$ [4]. Referring to Figure 9, we can therefore approximate $|\nabla \phi_k(d_m, 0)|$ by $\phi_k(d_m, 0)/(2.131D(d_m))$ and the optical power by

$$P = \gamma |\phi_k(d_m, 0)|$$

(36)

where $\gamma = A_e/2.131$.

We assume an ideal photo detector in which thermal noise is negligible so that the noise is dominated by shot noise. Shot noise for the photo detector current is described by Gaussian statistics in which the variance is proportional to the mean [19]. The probability density function for the complex envelope $i$ of the detector current is then

$$p(i) = \frac{1}{2\pi \sigma_o^2} \exp \left( -\frac{|i - E[i]|^2}{2\sigma_o^2} \right),$$

(37)

where the noise current variance for shot noise is given by

$$\sigma_o^2 = 2eB_i_o,$$

(38)

with $e$ the electron charge, $B$ the bandwidth and $i_o = E[i]$ the DC current [19]. The current $i_o$ is proportional to the detector DC optical power, i.e., $i_o = \kappa P$, where $\kappa$ is the detector responsivity [35], so that using (36)-(38) allows the probability density function for the data $y_{km} = \phi_k(d_m, \omega)$ to be derived as

$$p(y_{km}|x) = \frac{1}{2\pi \eta \phi_k(d_m, 0)} \exp \left[ -\frac{|y_{km} - \phi_k(d_m, \omega)|^2}{2\eta \phi_k(d_m, 0)} \right],$$

(39)

where $\eta = 2eB/(\kappa \gamma)$ and $p(y_{km}|x)$ is the probability density function for each measured datum given a particular image $x$. Furthermore, we show in Appendix II that for typical tissue optical parameter and typical (low) modulation frequencies, $\phi_k(d_m, 0) \simeq |\phi_k(d_m, \omega)|/\beta$, so that (39) can be replaced by

$$p(y_{km}|x) = \frac{1}{2\pi \alpha |\phi_k(d_m, \omega)|} \exp \left[ -\frac{|y_{km} - \phi_k(d_m, \omega)|^2}{2\alpha |\phi_k(d_m, \omega)|} \right],$$

(40)

where $\alpha = \eta/\beta$. 

22
APPENDIX II: Approximation of DC Photon Flux Density

Consider the time domain diffusion equation (1) in an infinite homogeneous region with \( S(t) = 1 + \beta \exp(j\omega t) \) located at \( s_k \), which yields the closed form solution [36]

\[
\psi_k(r, t) = \text{Re}\left\{ \frac{1}{4\pi D z} \exp\left[ -z \left( \frac{\mu_a}{D} \right)^{1/2} \right] \right. \\
+ \frac{\beta}{4\pi D z} \exp\left[ -z \left( \frac{c^2 \mu_a^2 + \omega^2}{c^2 D^2} \right)^{1/4} \cos \left\{ \frac{1}{2} \arctan \left( \frac{\omega}{c \mu_a} \right) \right\} \right] \\
\times \exp\left[ -j z \left( \frac{c^2 \mu_a^2 + \omega^2}{c^2 D^2} \right)^{1/4} \sin \left\{ \frac{1}{2} \arctan \left( \frac{\omega}{c \mu_a} \right) \right\} + j \omega t \right\} \right\}. \quad (41)
\]

where \( z = |s_k - d_m| \). The first term corresponds to the DC intensity at the detector location \( d_m \), while the second term corresponds to the AC component. Using reference signals \( 2 \cos(\omega t) \) and \( 2 \sin(\omega t) \) for the in-phase and quadrature phase of the measured current, respectively, the magnitude of the modulation envelope of the current becomes \( |i| = \kappa \gamma |\phi_k(d_m, \omega)| \), where for \( \omega \neq 0 \),

\[
|\phi_k(d_m, \omega)| = \frac{\beta}{4\pi D z} \exp\left[ -z \left( \frac{c^2 \mu_a^2 + \omega^2}{c^2 D^2} \right)^{1/4} \cos \left\{ \frac{1}{2} \arctan \left( \frac{\omega}{c \mu_a} \right) \right\} \right]. \quad (42)
\]

For typical optical parameters in tissue and modulation frequencies \( (\frac{\omega}{\pi} \leq 2 \times 10^8 \text{Hz}) \), simulations (Figure 10) show that (42) is nearly independent of \( \omega \) and equal to the DC value times \( \beta \). This is due to \( [(c^2 \mu_a^2 + \omega^2)/(c^2 D^2)]^{1/4} \) being an increasing function of \( \omega \), while \( \cos\left\{ \frac{1}{2} \arctan\left( \frac{\omega}{c \mu_a} \right) \right\} \) is a decreasing function, for small \( \omega \). Therefore, \( \phi_k(d_m, 0) \simeq \frac{1}{\beta} |\phi_k(d_m, \omega)| \). We do not know the exact value of \( \phi_k(d_m, \omega) \) because of measurement noise, so we use the estimate \( \phi_k(d_m, \omega) \simeq y_{mk} \).

X Acknowledgments

The authors thank Thomas J. Downar of the School of Nuclear Engineering at Purdue University for numerous discussions of diffusion equation solver related to this work. This work was supported by a Purdue Research Foundation fellowship and NSF grant MIP 97-07763.
References


Table 1: Computational complexity of the DBIM and the ICD/Born method in terms of number of complex multiplications per iteration.

<table>
<thead>
<tr>
<th></th>
<th>DBIM (cflops)</th>
<th>ICD/Born (cflops)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green’s function and $\phi_k$ update</td>
<td>$5(M + K)LN$</td>
<td>$5(M + K)LN$</td>
</tr>
<tr>
<td>Fréchet derivative</td>
<td>$(2MK + 2)N$</td>
<td>$(2MK + 2)N$</td>
</tr>
<tr>
<td>Pixel update</td>
<td>$2(MK)^2N - (MK)^3/3$</td>
<td>$5MKN$</td>
</tr>
<tr>
<td>Total order of computation</td>
<td>$2(MK)^2N + 5(M + K)LN$</td>
<td>$7MKN + 5(M + K)LN$</td>
</tr>
</tbody>
</table>
Table 2: Comparison of the computation required (complex multiplications) for one iteration of the DBIM and the ICD/Born method.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DBIM (cflops)</th>
<th>ICD/Born (cflops)</th>
<th>Speed up</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{DBIM} = 10, L_{ICD} = 10, K = 12, K = 12, N = 33^2$</td>
<td>45,790,290</td>
<td>2,406,690</td>
<td>19 to 1</td>
</tr>
<tr>
<td>$L_{DBIM} = 30, L_{ICD} = 20, K = 12, K = 12, N = 33^2$</td>
<td>48,403,890</td>
<td>3,713,490</td>
<td>13 to 1</td>
</tr>
<tr>
<td>$L_{DBIM} = 30, L_{ICD} = 30, K = 54, K = 54, N = 33^3$</td>
<td>$6.034 \times 10^{11}$</td>
<td>$1.316 \times 10^9$</td>
<td>459 to 1</td>
</tr>
</tbody>
</table>
Table 3: Average CPU time per iteration for the first simulation (Figure 3).

<table>
<thead>
<tr>
<th></th>
<th>DBIM</th>
<th>ICD/Born (p=2.0)</th>
<th>ICD/Born (p=1.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU time per iteration</td>
<td>4 sec</td>
<td>0.228 sec</td>
<td>0.318 sec</td>
</tr>
</tbody>
</table>
Figure 1: Simulation geometry with the locations of sources and detectors for inversion of synthetic data. The sources and detectors are uniformly spaced along the edges.

Figure 2: Pseudo-code specification of the ICD/Born algorithm.

Figure 3: Reconstruction results for $\mu_a$: (a) original absorption image; (b) reconstruction by the DBIM; (c) reconstruction by the new algorithm with a Gaussian prior ($p = 2.0, \sigma = 1.00 \times 10^{-3}$); (d) reconstruction by the new algorithm with a GGMRF prior ($p = 1.1, \sigma = 2.31 \times 10^{-4}$).

Figure 4: Normalized root mean square error (NRMSE) plots for the ICD/Born (with $p = 1.1$ and $p = 2$) and the DBIM.

Figure 5: Cost function (log scale) versus iteration for the ICD/Born algorithm with $p = 1.1$ and $p = 2$.

Figure 6: Reconstruction for $\mu_a$ showing the effect of $\sigma$ for a GGMRF prior and with $p = 1.1$: (a) original absorption image; (b) $\sigma = 2.85 \times 10^{-5}$; (c) $\sigma = 2.31 \times 10^{-4}$; (d) $\sigma = 1.52 \times 10^{-2}$.

Figure 7: A variety of true absorption images used for simulations.

Figure 8: Reconstructions of the absorption images shown in Figure 7 using the ICD/Born algorithm with the GGMRF prior and $p = 1.1$ and $\sigma = 2.31 \times 10^{-4}$. The stopping criterion is a CPU time of 60 seconds on a Sun Ultra Sparc 30. Normalized RMSE values for the final reconstructions are: (a) $5.83 \times 10^{-2}$, (b) $5.56 \times 10^{-2}$, (c) $1.92 \times 10^{-1}$, (d) $1.25 \times 10^{-1}$, (e) $7.70 \times 10^{-2}$, (f) $2.18 \times 10^{-1}$, (g) $8.34 \times 10^{-2}$, (h) $1.26 \times 10^{-1}$, (i) $2.08 \times 10^{-1}$.

Figure 9: Illustration of the zero input photon current or absorbing boundary condition for the diffusion equation, where all incident light from within the scattering boundary is lost to free space. Setting $\phi = 0$ on an extrapolated boundary at 0.7104(3D), where 3D is the mean free path, is equivalent to the zero input current condition on the physical boundary. In this example, $z$ is
the distance variable perpendicular to the interface and $\mathbf{a}_z$ is the unit vector.

**Figure 10:** Normalized magnitude (with respect to the source strength) of the complex optical signal versus modulation frequency for several absorption coefficients. The scattering coefficient is set to be $\mu_s = 10.0 \text{cm}^{-1}$ and the distance is 4 cm. The plots are computed using the analytic form in (42).
Figure 1:
For \( n := 1 \) until stopping criterion; \(/^* \) for each full iteration through \( x \) */

1. Solve the frequency domain diffusion equation for each source and detector combination using an absorption coefficient estimate \( x^n = \{\mu_a^n(r_i)\}_{i=1}^N \).

2. Fill in the Fréchet derivative matrix entries (21) by computing (23).

3. Compute the initial error vector, \( e(1) = y - f(x^n) \).

4. For \( i := 1 \) until \( N \); \(/^* \) for each pixel in \( x \) */

   (a) \( \theta_1 = -2Re \left[ \left[ f'(x^n)_{si} \right]^H \Lambda e(i) \right] \).

   (b) \( \theta_2 = 2 \left| f'(x^n)_{si} \right|^H \Lambda f'(x^n)_{si} \).

   (c) \( \hat{x}_i = \arg \min_{\tilde{x}_i \geq 0} \left[ \theta_1 (\tilde{x}_i - x^n_i) + \frac{\theta_2}{2} (\tilde{x}_i - x^n_i)^2 + \frac{1}{\rho \sigma^p} \sum_{j \in N_i} b_{i-j} |\tilde{x}_i - x^n_j|^p \right] \).

   (d) \( e(i + 1) = e(i) - f'(x^n)_{si} (\hat{x}_i - x^n_i) \);

   (e) \( x^n_i := \hat{x}_i \).

end

end;

Figure 2:
Figure 3:
Figure 4:
Figure 5:
Figure 6:
Figure 7:
Figure 8:
\( J^* = 0 \)

\( J^* = -D \nabla \phi \cdot a_z \)

Figure 9:
Figure 10: